An Intelligent Ambulatory Fall Risk Assessment Method Based on the Detection of Compensatory Balance Reactions and Environmental Factors

by

Mina Nouredanesh

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Mechanical and Mechatronics Engineering

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Dr. Stephen Robinovitch Professor, Dept. of Biomedical Physiology and Kinesiology, Simon Fraser University
Supervisors:	Dr. James Tung Associate Professor, Dept. of Mechanical and Mechatronics Engineering, University of Waterloo
	Dr. William Melek Professor, Dept. of Mechanical and Mechatronics Engineering, University of Waterloo
Internal-External Member:	Dr. Jennifer Howcroft Lecturer, Dept. of Systems Design Engineering, University of Waterloo
Internal Members:	Dr. Jan Huissoon Professor, Dept. of Mechanical and Mechatronics Engineering, University of Waterloo
	Dr. Arash Arami Assistant Professor, Dept. of Mechanical and Mechatronics Engineering, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This dissertation includes first-authored material that has (or will be) appeared in journals/conference proceedings:

- Nouredanesh, M., Godfrey, A., Howcroft, J., Lemaire, E. D., & Tung, J. (2021). Fall risk assessment in the wild: A critical examination of wearable sensor use in free-living conditions. Gait & Posture, 85, 178-190. https://doi.org/10.1016/j.gaitpost.2020.04.010

- Nouredanesh, M., Gordt, K., Schwenk, M., & Tung, J. (2019). Automated detection of multidirectional compensatory balance reactions: A step towards tracking naturally occurring near falls. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(2), 478-487. https://doi.org/10.1109/TNSRE.2019.2956487

- Nouredanesh, M., Li, A. W., Godfrey, A., Hoey, J., & Tung, J. (2018, September). Chasing Feet in the Wild: A Proposed Egocentric Motion-Aware Gait Assessment Tool. In European Conference on Computer Vision (pp. 176-192). Springer, Cham. https://doi.org/10.1007/978-3-030-11024-6_12

- Nouredanesh, M., Ojeda, L., Alexander, N. B., Godfrey, A., Schwenk, S., Melek, W., & Tung, J. "Automated Detection of Older Adults' Naturally-Occurring Compensatory Balance Reactions: Translation from Laboratory to Free-living Conditions" *(Submitted as a journal manuscript, Currently Under Review).*

- Nouredanesh, M., Godfrey, A., Powell, D., & Tung, J. "Egocentric vision-based detection of surfaces: Towards context-aware free-living digital biomarkers for gait and fall risk assessment" *(Submitted as a journal manuscript, Currently in the Review Process).*

Abstract

Falls in older adults are a critical public health problem worldwide and impact one in three older adults at least once each year. In addition to physical consequences, e.g., hip fracture and traumatic brain injury, falls can lead to negative mental health outcomes, such as fear of falling and depression.

Fall risk assessment (FRA) is the initial step for fall prevention programs and interventions. In particular, clinicians aim to understand what factors put older adults at high risk of falling to inform the selection and timing of fall prevention interventions (e.g., strengthening programs). These risk factors are generally categorized as intrinsic or biological (e.g., gait and balance disorders) and extrinsic or environmental (e.g., slippery surfaces). While *supervised* FRAs, including performance-based (e.g., Timed up and Go) and instrumented methods (e.g., motion capture systems), capable of quantifying intrinsic risks have advanced significantly, falls still remain a major priority in geriatric medicine and public health. This can be due to the Hawthorne effect, the heterogeneous nature of older adults' health, lifestyle, and behaviors, and the complex, multifactorial etiology of falls.

To address the limitation of supervised FRAs, a growing body of literature has focused on wearable sensor-based methods for *free-living* (or ambulatory) FRA. These studies, reviewed in Chapter 2, investigated the relationships between free-living digital biomarkers (FLDBs) extracted from wearable sensors data (generally, inertial data) and the frequency of prospective/retrospective falls in older adults. However, many FLDBs exhibited inconsistent fall predictive powers across studies, indicating 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 and the frequency of naturally-occurring compensatory balance reactions (CBRs), have yet to be investigated in depth. Considering controlled studies reported balance impairment as one of the strongest risk factors for falls, the investigation of balance-related FLDBs may lead to more stable risk assessments and provide new insights into fall prevention in older adults.

Although gait-related FLDBs extracted from inertial data can be impacted by both intrinsic and environmental factors, their respective impacts have not been differentiated by the majority of free-living FRA methods. This may lead to the ambiguous interpretation of the subsequent FLDBs, and less precise intervention strategies to prevent falls. A *context-aware* free-living FRA would elucidate the interplay between intrinsic and environmental risk factors and clarifies their respective impacts on fall predictive powers of FLDBs. This may subsequently enable clinicians to target *more specific* intervention strategies including environmental modification (e.g., eliminating tripping hazards) and/or rehabilitation interventions (e.g., training to negotiate stairs/transitions).

This doctoral thesis aims to address the aforementioned research gaps by proposing multiple machine learning frameworks and incorporating an egocentric camera along with wearable inertial measurement units (IMUs). Chapter 3 discusses the development of random forest models to differentiate between normal gait episodes and multidirectinoal CBRs (e.g, slip-like, trip-like, sidestep) elicited by a perturbation treadmill in controlled conditions in healthy young adults, where the CBR detection model achieved the overall accuracy of $\approx$ 96%. This chapter established the infrastructure for Chapter 4, where a validation study was performed to detect older adults' CBRs under free-living conditions. Random forest models were trained on independent/unseen datasets curated from multiple sources, including perturbation treadmill CBRs. 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. A subset of models differentiated between naturally-occurring CBRs and free-living activities with high sensitivity (100%) and specificity ($\geq$99%) suggesting that accurate detection of naturally-occurring CBRs is feasible. Moreover, to address the limitations of IMUs in terms of the estimation of step width in free-living conditions, Chapter 5 presents a novel markerless deep learning-based model to obtain gait patterns by localizing feet in the egocentric vision data captured by a waist-mounted camera.

With the aim of improving the interpretability of gait-related FLDBs and investigating the impact of environment on older adults' gait, Chapter 6 proposes a vision-based framework to automatically detect the most common level walking surfaces. Using a belt-mounted camera and IMUs worn by fallers and non-fallers (mean age 73.6 yrs), a unique dataset was acquired (a subset of Multimodal Ambulatory Gait and Fall Risk Assessment in the Wild (MAGFRA-W) dataset). A series of ConvNets were developed: *EgoPlaceNet* categorizes frames into indoor and outdoor; and *EgoTerrainNet* (with outdoor and indoor versions) detects the enclosed terrain type in patches. *EgoPlaceNet* detected outdoor and indoor scenes in MAGFRA-W with 97.36% and 95.59% (leave-one-subject-out) accuracies, respectively. *EgoTerrainNet*-Indoor and -Outdoor achieved high detection accuracies for pavement (87.63%), foliage (91.24%), gravel (95.12%), and high-friction materials (95.02%), which indicate the models' high generalizabiliy.

Overall, promising results encourage the integration of wearable cameras and machine learning approaches to complement IMU-based free-living FRAs, towards stable context-aware FLDBs for fall prevention in older adults. Implications for further research to examine the relationships between naturally-occurring CBRs and fall risk, and clinical applications are discussed.

Acknowledgements

I would like to express my sincere gratitude to my supervisors for supporting my doctoral research. I am immensely grateful to Dr. James Tung for his invaluable advice, guidance, and support at every stage of my graduate studies. I would also like to thank Dr. William Melek for all his unwavering support and encouragement during my PhD journey.

I am thankful to my thesis examining committee members: Dr. Stephen Robinovitch, Dr. Jan Huissoon, Dr. Jennifer Howcroft, and Dr. Arash Arami for their time, feedback, and insightful comments.

I would like to extend my sincere thanks to research collaborators: Dr. Alan Godfrey (Northumbria University), Dr. Michael Schwenk (Heidelberg University), Dr. Neil B. Alexander (University of Michigan) and Lauro Ojeda (University of Michigan), for supporting my research on fall risk assessment in older adults by contributing to some of the datasets used in this thesis.

I would like to thank AGE-WELL NCE Inc. (Canada's Technology and Aging Network) for supporting my doctoral research (AGE-WELL Doctoral Student Award in Technology and Aging and AGE-WELL ACCESS Award). Additionally, I would like to thank the Vector Institute for AI (Toronto) and Waterloo AI Institute for their support.

Thanks should also go to my friends and colleagues at the University of Waterloo, especially the current and former members of: the Computational Health Informatics Laboratory and Neural and Rehabilitation Engineering Laboratory.

I am extremely grateful to my parents and brother, for their support and encouragement throughout my doctoral studies. The completion of my dissertation would not have been possible without their support.

Dedication

This doctoral thesis is dedicated to my mother, *Mozhgan*, who has been a constant source of support, inspiration, and encouragement, to my father, who has supported me all the way, and to my amazing brother.

This thesis is also dedicated to all older adults and people with gait and/or balance disorders.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

ACC: Accelerometer or acceleration signal (depending on the context)
ADLs: Activities of daily living
AP: Antero-posterior
BoS: Base of support
CBR: Compensatory balance reaction
CO: Crossover (a CBR strategy)
CoM: Center of mass
ER: Elimination rate
FIVR: Free-living inertial and voice recorder (dataset)
FLDB: Free-living digital biomarker
FPV: First-person vision
FRA: Fall risk assessment
GT: Ground truth
Gyro: Gyroscope or Angular velocity signal (depending on the context)
HFM: High-friction materials
IMU: Inertial measurement unit
IMUFD: Inertial Measurement Unit Fall Detection (dataset)
IoU: Intersection-over-union
LHS: Left heel strike
LOSO: Leave-one-subject-out
LOTO: Leave-one-trial-out
LSTM: Long short term memory
MAE: Mean absolute error
MAGFRA-C: Multimodal Ambulatory Gait and Fall Risk Assessment in Clinic (dataset)
MAGFRA-W: Multimodal Ambulatory Gait and Fall Risk Assessment in the Wild (dataset)
MAGFRA: Multimodal Ambulatory Gait and Fall Risk Assessment (dataset)
MBB: Micro ambulatory bout-based
MGE: Micro gait event-based

ML: Mediolateral
NG: Normal gait
OA: Older adult
PD: Parkinson's disease
PFR: Proposed foot region
PT: Perturbation treadmill
RF: Random forest
RHS: Right heel strike
RMS: Root mean square
ROI: Region of interest
S-ADL: Simulated activity of daily living
S-CBR: Simulated compensatory balance reaction
S-NG: Simulated normal gait
sEMG: Surface electromyography
SS: Side step (a CBR strategy)
SVA: Signal vector amplitude
TPV: Third-person vision
TUG: Timed Up and Go
VT: Vertical

Chapter 1

Introduction

[1]

It is estimated that 1 in 3 older adults ($\geq$ 65 yrs) fall at least once each year [139, 60]. Falls are a major public health problem worldwide that can lead to serious physical (e.g., hip fractures, traumatic brain injuries) [188, 112] and/or psychological (depression and fear of falling) [57, 143] consequences, and incur heavy expenses [141, 45]. Although falls are considered in the class of preventable injuries [88], to date, there has been no method that can precisely predict older adults' falls.

Falls among older adults and people with neurodegenerative diseases have a complex etiology, often resulting from multiple *interacting factors* [88, 9], which can be unique to an individual's characteristics, lifestyle, and the surrounding environment (see Fig. 1.1). Intrinsic or biological risk factors are specific to the individual's capabilities and health status, and include chronic conditions such as Parkinson's disease [107], dementia [168], muscle weakness, and/or balance impairment [97]. Environmental (or extrinsic) risks include low-friction terrains, obstacles, uneven surfaces, and poor lighting conditions [20, 50]. Generally, gait and balance disorders and environmental hazards have been reported to be the most important risk factors contributing to $\approx 17\%$ and $\approx 31\%$ of falls in older adults, respectively [162]. These risk factors can be aggravated by social and behavioural risk factors such as loneliness and social isolation [18], anxiety [54], and depression [76]. Moreover, exposure to high-risk activities, such as hurried activity beyond the limits of stability [107], influences individual risk for falls. Considering the large number of potential risk factors,

[1] The content of this chapter is partly obtained from [122], which was co-authored by the present thesis' author.

identifying individual risks in a multifactorial assessment approach is an essential first step for best practice fall prevention interventions (Fig. 1.1).

By identifying the various risks specific to an individual, fall risk assessment (FRA) can inform clinical decisions on the most appropriate preventive interventions to reduce the risk for future fall events (Fig. 1.1). To date, commonly used FRA methods involve easy-to-implement movement-based tasks with minimal equipment requirements, such as total time to complete a timed-up-and-go (TUG) [145, 172] or Tinetti Test [189]. Based on a meta-analysis, the diagnostic accuracy of TUG was poor to moderate for fall prediction in healthy high-functioning older adults and the cut-off thresholds for TUG-based identification of fallers were highly inconsistent within the included studies [167]. These limitations have led to a methodological shift towards the use of more detailed assessments using instrumented approaches.

As the adopted gold standard, electronic-based tools, such as three-dimensional motion capture and instrumented walkways, can be used to offer detailed quantitative assessments. Yet, these tools remain resource-intensive and fixed to specialized clinics/locations, offering snapshots during scripted functional tasks. Additionally, extrinsic risk factors for falls can be self-reported using diaries or voice-recorders [56], however, this often lacks accuracy and adequately detailed descriptions. To systematically investigate the impact of environmental conditions on older adults' tendency to fall, researchers have designed paradigms to mimic challenging natural conditions in a laboratory setting. For example, minimum foot clearance was measured in different lighting conditions in [55] to understand the nature of trips on stairs in older adults, where, in contrast to young adults, the lack of precautionary increase in older adults' foot clearance under reduced lighting contributed to falls on stairs. However, due to the Hawthorne effect, i.e., alteration of performance associated with awareness of being observed, during controlled gait and balance tests [159], supervised FRA measures may not necessarily reflect naturalistic and multitasking behaviour [32, 63, 155]. For instance, a weak association ($r = 0.333$, $p < 0.001$) between natural gait speed and in-laboratory gait speed was reported [184]. Similarly, free-living gait speed and step regularity measures were significantly lower compared with in-lab usual walking and tended to be more similar to in-lab dual-task walking [63]. Thus, novel free-living FRAs to identify fallers based on their free-living behaviour in their natural daily living environments could provide complementary information to supervised FRAs.

There have been a wide range of methods proposed to measure mobility behaviour in out-of-lab conditions. Ambient sensors, such as radar [38], passive infrared [91], third-person video [158], and depth cameras [46, 26, 144, 4] have been investigated as a means to extract gait parameters, detect falls, and track longitudinal changes in a person's mobility patterns. Although ambient sensors can provide valuable information on an older adults'

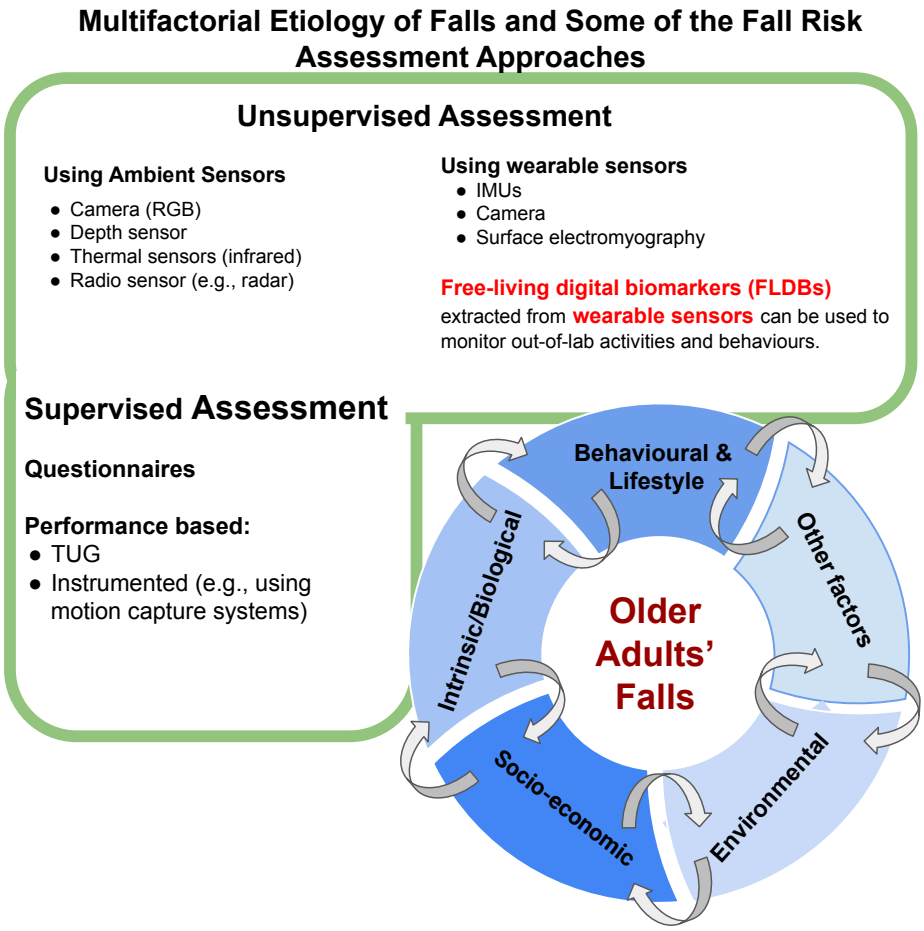


Figure 1.1: Considering the multifactorial etiology of falls, and complex interactions between risk/protective factors (shown by grey arrows) at the individual level, different fall risk assessment approaches (and instruments) have been proposed to examine exposure to different risk factors.

risk of falls in an unobtrusive manner, especially in inpatient units [114], they have limitations due to visual occlusions (e.g., furniture), inability to extract spatiotemporal data when full-body view is unavailable, and tracking the same person in spaces with multiple residents with similar body characteristics [180]. Moreover, they are restricted to the environments the sensors are installed in. In contrast, wearable sensors and their data have greater utility beyond the living space where ambient sensor data is recorded, at the expense of additional burden in donning and maintaining devices. These technologies include

wearable inertial sensors (e.g., tri-axial accelerometers, gyroscopes and magnetometers), in-shoe plantar pressure sensors [34, 118, 40] and wearable cameras [127, 92, 185].

Recent attention has focused on the examination of fall predictive powers of wearable sensor-based free-living digital biomarkers (FLDBs), such as total time walking/lying, frequency-based (e.g., the amplitude of dominant frequency) and temporal (e.g., step time) measures extracted from detected activity bouts (e.g., gait) and events (e.g., turns). Several studies have shown that wearable-based FLDBs can either outperform or complement clinical (supervised) FRA tests [83, 156, 201]. For instance, a machine learning-based model developed on transition-based FLDBs outperformed its counterpart developed on clinical test scores (e.g., TUG) in discriminating between older fallers and non-fallers [83]. However, there are no clear solutions for transparent deployment of wearables for free-living FRA at this time due to the ongoing novel developments within the field. Furthermore, the utility of existing free-living FRA methods to inform interventions remains limited.

Chapter 2 (peer-reviewed and published in Gait & Posture [122]) provides a detailed literature review, highlighting the research gaps in the proposed free-living FRA approaches. This chapter explores journal papers investigating free-living data collected by wearable sensors for a duration of at least 24 hours to identify fall-prone older adults. The search yielded twenty-four studies, in which inertial measurement units (IMUs)[2] were the only wearable system employed for FRA in the wild. This review highlighted that while early free-living FRA approaches have demonstrated promise, many of the explored FLDBs exhibited inconsistent fall predictive powers across studies, indicating that they may not be *stable* in distinguishing fall-prone individuals. Moreover, inter-study inconsistencies were observed in experimental design (e.g., sensor placement, duration of free-living data collection), sample demographics (e.g., number of participants, neurological status), use of processing hyperparameters (e.g., cut-off thresholds to define an ambulatory bout), and definition/measurement of FLDBs (e.g., different central tendency measures), and domains. Chapter 2 further provides recommendations towards the harmonisation and standardization of outcomes for free-living FRA.

The literature review in Chapter 2 further highlighted that the relationships between falls and free-living dynamic postural control measures, such as *step width* [103] and the *frequency of naturally-occurring compensatory balance reactions*, have yet to be investigated in depth [122]. Considering balance impairment as one of the strongest risk factors for falls [190], 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

[2]While IMUs refer to sensor systems that include tri-axial accelerometers, gyroscopes and magnetometers, for simplicity, the present thesis may refer to an inertial sensor as IMU.

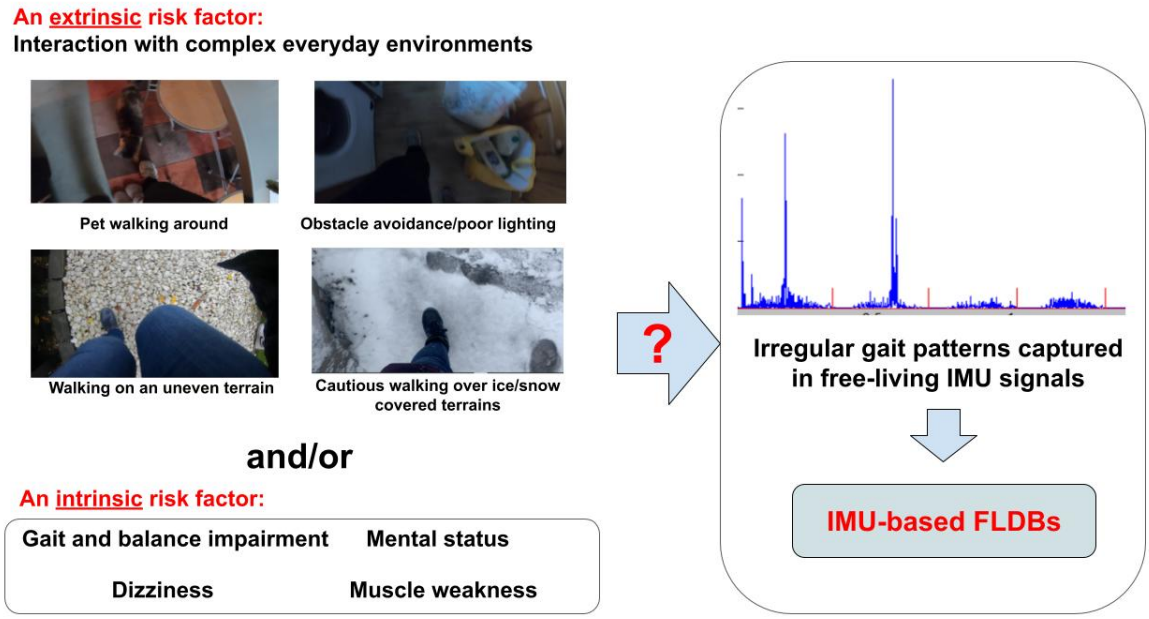


Figure 1.2: Without detailed information of the mobility context, such as terrain characteristics and obstacles, the ability to interpret free-living digital biomarkers (FLDBs) is constrained. Therefore, understanding the possible interplay between intrinsic and extrinsic features (i.e., Environment × Intrinsic), and their impact on FLDBs requires further investigation. This necessitates the development of *context-aware* fall risk assessment approaches.

the development of robust models to identify balance impairment in older adults under free-living conditions. To address these research gaps, Chapter 3 describes a study investigating methods of detecting multidirectional CBRs using young adults' in-lab data captured over a perturbation treadmill (peer-reviewed and published in IEEE Transactions on Neural Systems and Rehabilitation Engineering [125]). Based on the findings of the aforementioned study, Chapter 4 proposes another machine learning framework, which was further validated using older adults' free-living data encompassing naturally-occurring CBRs. Moreover, to address the IMUs' limitation to estimate step width [149], Chapter 5 presents a deep egocentric vision-based approach to estimate this gait parameter in pixels (peer-reviewed and published in Proceedings of the European Conference on Computer Vision (ECCV) Workshops 2018 [127]).

Chapter 2 discussed another literature gap originating from the inability of IMUs in

terms of the provision of detailed contextual information on human-environment interaction. While IMU-based digital biomarkers have shown to be sensitive to intrinsic risk factors (e.g., chronic motor impairment), they can be also impacted by different environmental features (e.g., surfaces, slopes, stairs) [193, 95, 202] (see Fig. 1.2). However, the majority of free-living FRA methods have not differentiated between the impacts of environmental and intrinsic factors. This may lead to the ambiguous interpretation of the subsequent IMU-based FLDBs (see Fig. 1.2), and therefore, less precise intervention strategies to prevent falls. For instance, depending on the mobility context, higher variability in acceleration signal (e.g., a lower amplitude of the dominant frequency as a FLDB) in the mediolateral direction captured by a lower back-mounted accelerometer during gait could indicate a higher adaptability to the environment [201] (and potentially a lower risk of falls), and/or could represent an abnormality in gait due to an intrinsic factor (and potentially a higher risk of falls) [79].

With the aim of improving the interpretability of IMU-based FLDBs and investigating the impact of environment on older adults' gait, towards *context-aware FLDBs*, Chapter 6 proposes a deep vision-based framework to automatically detect the most common level walking surfaces in indoor and outdoor environments. The framework was further validated using older non-fallers and fallers' free-living data. Chapter 7 summarizes the key contributions of the thesis and discusses overarching themes arising from the aforementioned studies.

Chapter 2

Fall risk assessment in the wild: A critical Review of wearable sensor use in free-living conditions

[1]

To address the imitations of supervised fall risk assessment (FRA) approaches, a growing body of literature has focused on wearable sensor-based methods for free-living FRA, to ultimately reduce fall occurrence. These studies investigated the relationships between wearable-based free-living digital biomarkers (FLDBs) and falls. This chapter reviews journal papers investigating natural data collected by wearable sensors for a duration of at least 24 hours to identify fall-prone older adults. After summarizing the key aspects of the experimental protocols used to collect free-living data (e.g., sensor placement, duration of free-living data collection, demographics), this chapter reviews sources of inconsistencies between the proposed free-living FRA approaches. At the end, gaps in the literature were highlighted and recommendations were provided to inform future work towards achieving a harmonized free-living FRA model.

The next chapters of this thesis will address some of the research gaps highlighted in the present chapter.

[1]The content of this chapter is partly obtained from the following peer-reviewed research paper: **Nouredanesh, M.**, Godfrey, A., Howcroft, J., Lemaire, E. D., & Tung, J. (2021). Fall risk assessment in the wild: A critical examination of wearable sensor use in free-living conditions. Gait & Posture, 85, 178-190

2.1 Search criteria

Three databases, Scopus, PubMed, and Google Scholar, were searched up to and including September 2019 (2010 to 2019). Search terms were ["home" or "unsupervised" or "real-world" or "community" or "ambulatory"] and ["fall" or "fall risk assessment"] and ["elderly" or "senior" or "aged"] and ["wearable sensor" or "accelerometers" or "inertial" or "wearable camera"]. Journal articles were included if they: 1) assessed the relationships between falls and features extracted from free-living data, 2) collected data from wearables used by older adults (or people with neurodegenerative diseases if compared with older adults (≥65 yrs) as controls) for a duration of at least 24 h per participant, and 3) primarily considered categorization of participants into faller and non-faller cohorts for data analysis [2]. After the initial title screen, abstracts were reviewed. Twenty six (n=26) papers from databases met the inclusion criteria (Fig. 2.1). Multiple papers from two research groups investigated the same/overlapping datasets or very similar sets of FLDBs: (a) Hausdorff and colleagues examined datasets from healthy older adults [201, 81, 80, 78] and PD older adults [203, 82], and (b) Pijnappels and colleagues investigated: b-1. fall risk assessment in older adults (FARAO) dataset that was collected from >300 older adults [195, 120, 77], and b-2. overlapping (but different) subsets of FARAO dataset to address different research questions [155, 156, 197, 196]. Considering the high degree of overlap in b-2, the most relevant, largest sample, and/or highly cited paper examining fall risk was included for the purposes of this review [156] and [196] (see Figure 2.1). Therefore, the key methodological/demographic information from n=24 papers was extracted and provided in Tables 1-3 (the key aspects of [155] and [197] were highlighted in section 2.3) [3].

2.2 Results

2.2.1 Study characteristics

Tables 2.1 to 2.3 show the study designs for capturing free-living data, demographic information, key methodological aspects, types of wearables, sensor anatomical location, general

[2] There were a group of papers [177, 86, 93, 19] that examined the associations between falls and intensity of physical activities, such as falls per 100 hours walked or falls per individual physical activity exposure time. For data analysis, these papers did not explicitly considered the categorization of participants into faller and non-faller cohorts.

[3] This literature search was repeated in May 2021, and no further study was found to meet the inclusion criteria.

description of outcome measures, and the length of free-living recording for all included papers. The general procedure for wearable-based ambulatory FRA is shown in Fig. 2.2.

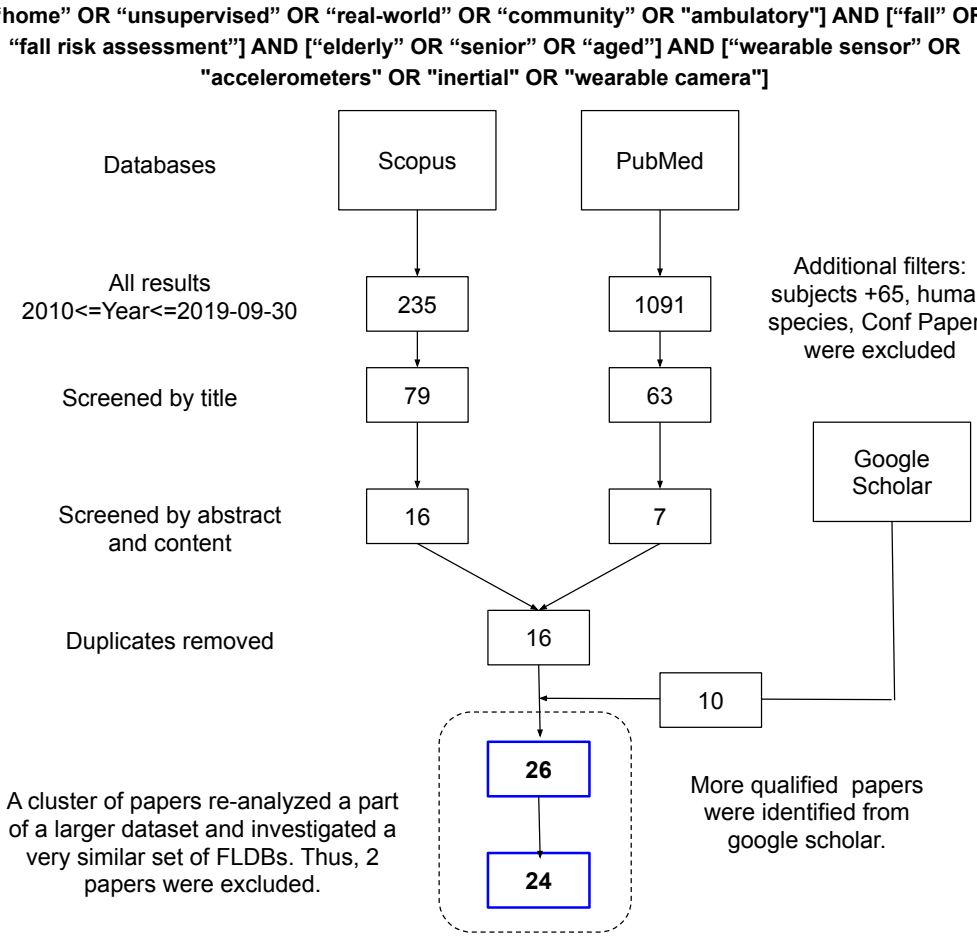


Figure 2.1: PRISMA flow chart of study design, illustrating search strategy results and filters at each stage of the study selection process.

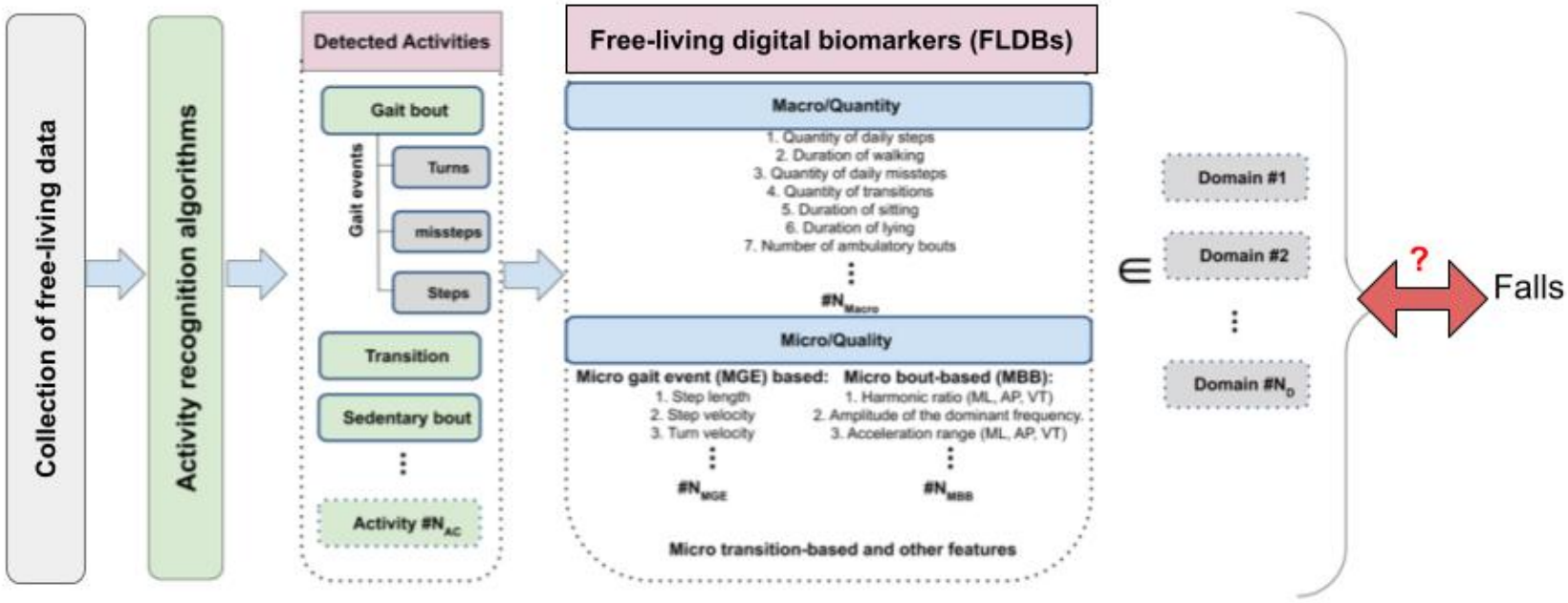


Figure 2.2: General process flow to acquire, process, and extract predictors for free-living fall risk assessment. Free-living data are collected using wearable sensors, segmented into bouts of activity types (e.g., gait, sitting), then predictors are extracted from activity bouts. There is a high degree of inconsistency in the literature in the categorization of extracted predictors into free-living FRA domains (dashed boxes)

In the reviewed studies, a range of inertial-based wearables were employed as described in Table 2.3. Most studies used a single tri-axial accelerometer-based wearable, with a minority using a uniaxial [107, 65] or combined tri-axial accelerometer and tri-axial gyroscope (e.g., [83, 99]). Typically, one inertial-based wearable was worn on the lower back, including the pelvis, sacrum, and L3 to L5 vertebrae [83, 201, 196, 31, 157, 81, 80, 78, 203, 82, 195, 120, 77], and midsagittal plane of the lower back [99]. Other wearable locations included chest/sternum [17, 116, 168], middle of the thigh [65], upper-thigh [107] dominant and nondominant hand/wrist [148]. In one study, multiple wearables were attached, two on shoes and one at L5 [110]. Free-living data were recorded from 24 hours [168] to 58 days (average over participants) in [17]. Most studies monitored community-dwelling older adults without neurological disorders (i.e., OA in Table 2.2). However, a number of studies also investigated differences between fallers and non-fallers in other populations: PD [203, 82, 65, 31, 107], dementia [168, 49], and varying frailty levels [116]. Additionally, fallers were further categorized as single fallers and recurrent fallers in two papers [107, 110].

Studies performed analysis to understand the relationships between wearable-based FLDBs and prospective falls [195, 120, 107, 116, 168, 148, 49], retrospective falls [83, 201, 81, 80, 203, 82, 65, 31, 157, 16] and both [196, 99, 110]. In one study [107], falls were also categorized with respect to the associated pre-fall events and allocated to one of three categories: 1) transitions during changes of posture (e.g., turning, rising from chair); 2) ambulation (e.g., everyday walking activities, including stair climbing) and 3) advanced activities including complex high-risk motor tasks (e.g., skiing, hill walking). Multivariate analysis or deep learning techniques were also applied to discriminate between fallers and non-fallers [120, 77, 49] without corresponding univariate analysis to investigate the individual FLDBs with respect to their predictive ability for falls.

2.2.2 Free-living activity/event detection

Detecting bouts of activity were critical step in extracting FLDBs. The detection of ambulatory bouts were most common [83, 201, 65, 99, 31, 157, 17, 16, 116, 168, 110, 81, 80, 78, 195, 120, 77, 196, 107] followed by, sitting [195, 196, 116, 168], lying [195, 196, 116, 168], sedentary (sitting and lying together) [65, 99] and standing [195, 196, 65, 116, 168]. Bout of activity detection was the initial methodological step required for the extraction of FLDBs from inertial sensor data (Fig. 2.2). Transitions between consecutive ambulatory and sedentary bouts (i.e., walk-to-sit and sit-to-walk) were also quantified for insights to FRA [83, 195, 196, 116, 148]. Ambulatory bouts were further examined for detection of discrete gait events (Fig. 2.2) such as initial and final contact within the gait cycle (e.g.,

in [201, 195, 196, 157], turns [99, 110], and missteps [82]. Although the term 'misstep' may mostly imply the involvement of *intrinsic* risk factors in the episodes of imbalance, in this thesis, missteps, compensatory balance reactions, and near-falls were interchangeably used, and were broadly defined as reactions to regain stability following a loss of balance.

Cut-off thresholds for identification of ambulatory bouts and turns

To identify an ambulatory bout, different minimum/maximum cut-off thresholds were defined according to steps, time or a combination of both [31, 17, 16] (Table 2.3). For example, Del Din et al. [31] referred to an ambulatory bout between 3 steps to 60s, 60-120s, and longer than 120s as 'short', 'medium', and 'long' walks, respectively. Alternatively, Brodie et al. considered short walks as those <7 s and <8 s and longer than three steps in two different studies [17, 16]. The minimum cut-off thresholds ranged from 1 step [107] to 120s [31] and a minimum of three steps was the most frequent cut-off threshold used within the reviewed studies from distinct datasets [31, 17, 16, 116, 168]. Discrete angular thresholds were also used to identify turns. For instance, one study [110] examined those greater than 45^o but elsewhere different turn resolutions, e.g. small (50-100^o), medium (100-150^o), and large (150-200^o) were taken into account [99], Table 2.3. The detected activity bouts and events were later used independently for the extraction of FLDBs, which were statistically analyzed with respect to falls (Fig. 2.2).

2.2.3 Conceptual models

As ambulatory bouts were the most investigated free-living activity for FRA, research groups defined different conceptual FRA models to classify and interpret a range of gait-based FLDBs. Each model consists of several domains, including a homogeneous group of FLDBs usually in terms of their mathematical description (Fig. 2.3, models a [31], b [17], c [16], d [196], and e). Model e represents the merged domains from a set of research papers [201, 81, 80, 203] as discussed in section 2.1. In Fig. 2.3-model e, complexity and local dynamic stability measures reported in [81, 80] were categorized into the same class because of their mathematical similarities (e.g., Lyapunov components [89]).

Broadly speaking, the reviewed literature examined the following features:

1. the 'quantity' of gait events or ambulatory bouts and their duration over days/weeks [201, 203, 17, 16] also termed 'macro' (as discussed in 2.2.3) [107], [65], [31] and 'amount of gait' [195, 196],

2. FLDBs that are obtained by performing a higher resolution analysis of the inertial signals or gait events, which include spatial (e.g. step length), temporal (step time), and frequency-based (e.g. harmonic ratio) features. These features termed as 'micro' (referring to more detailed micro-structural characteristics of gait as discussed in 2.2.3) [31, 65] [107] (Fig. 2.3 model a) or 'quality' of gait (e.g. in [201, 203, 195, 196, 17, 16] (Fig. 2.3),

3. models of quantity/quality also extended to categorize turns [99, 110] and transition features [83, 148].

As depicted in Fig. 2.3-model d [196], gait quality was represented by six domains, each with its own set of FLDBs, such as: intensity (e.g., standard deviation, range); variability (e.g., autocorrelation, slope, and magnitude of the dominant frequency); smoothness (e.g. index of harmonicity) and complexity (sample entropy). Alternatively, quality was presented within quantity-intensity-exposure-quality models (Fig. 2.3, model b) and quantity-intensity-quality (Fig. 2.3-model c) comprising of different predictors within three domains (e.g., between-walk adaptability). Due to these inconsistencies associated with the use of 'quality', here the general categorization of FLDBs into macro and micro is used to describe quantity and quality of activities, respectively.

Macro/quantity FLDBs

Macro outcomes were generally described by duration or volume of an activity or the quantity of daily occurrences. Commonly used macro FLDBs include the number of ambulatory bouts [195, 196, 107, 65, 17] total steps within each bout [201, 203, 195, 196, 17], number of daily turns [99, 110], number of daily compensatory balance reactions [82], and number of transfers/transitions [83, 195, 196, 116, 148]. In addition to the aforementioned linear features, macro outcomes were utilized for non-linear analyses [107, 65, 31] (Fig. 2.3-model a), including: 1. alpha (α), which is a unit-less FLDB derived from the power distribution of ambulatory bouts with respect to the cut-off thresholds and 2. within subject variability of bout length (S2) obtained from a maximum likelihood technique as the distribution of bout length.

For sedentary (lying and sitting) and standing bouts, only macro features were investigated within studies [195, 196, 107, 65, 99, 116, 168], which includes: total standing time [195, 196, 116, 168], total sedentary time [195, 196, 99, 116, 168], lying, sitting, and standing bout duration (mean, maximum, and 90th percentile) [116, 168], standing and sedentary bout duration variability [107], [65], number of sedentary and standing bouts [65], and alpha measures for sedentary and standing bouts [65].

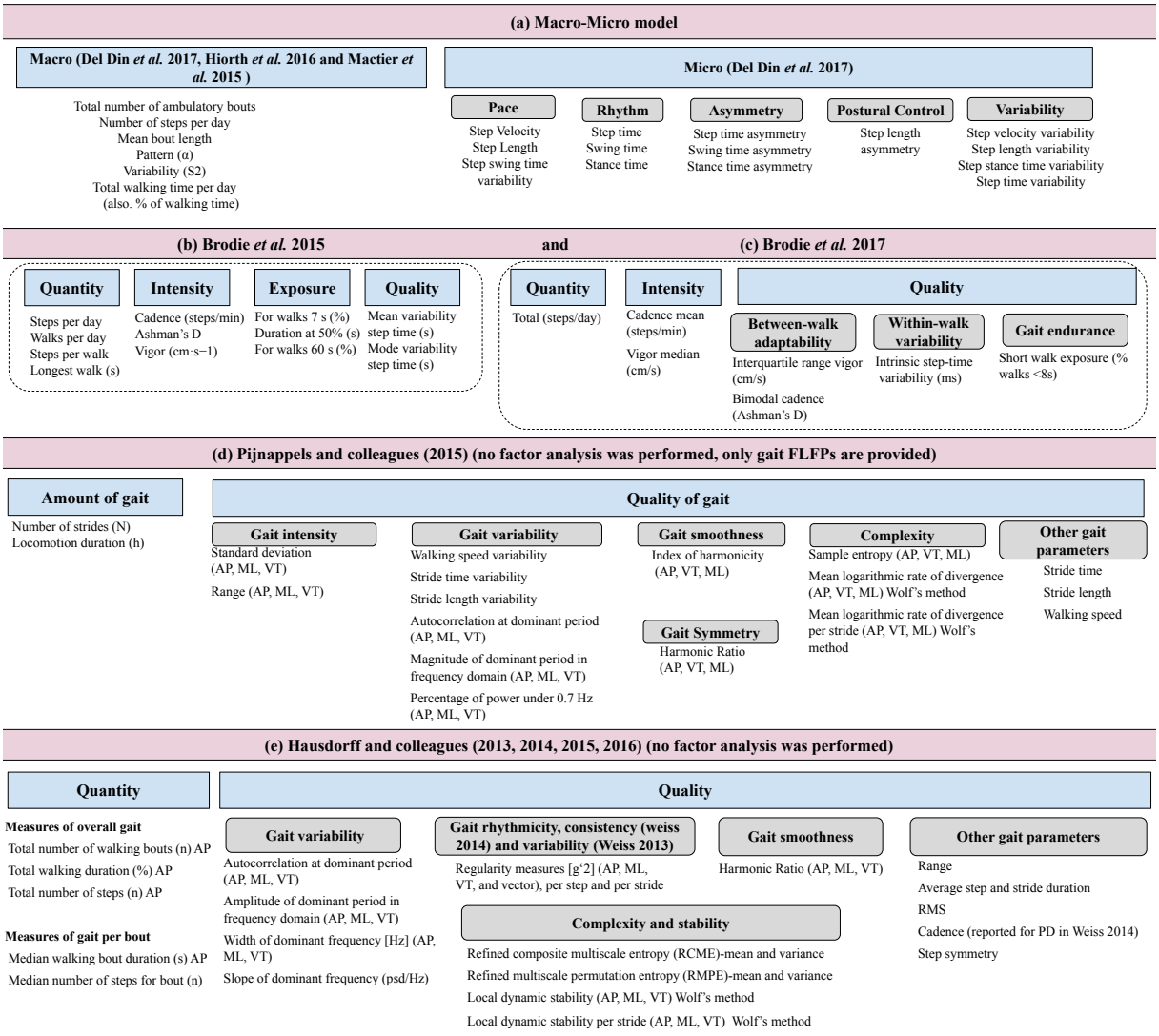


Figure 2.3: Conceptual models proposed by researchers to categorize gait-related features for fall risk assessment

Micro/quality FLDBs

Considering micro FLDBs that were investigated in studies, these features were categorized into three main classes:

1. Micro gait event-based (MGE) FLDBs represent features requiring detection of gait

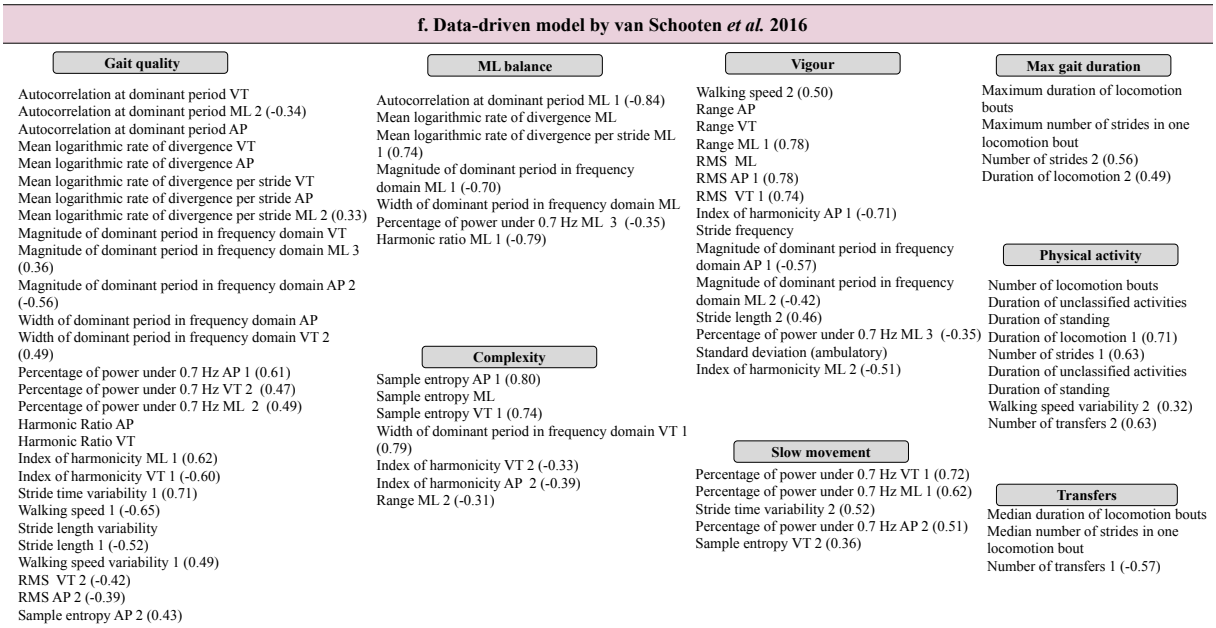


Figure 2.4: Data-driven model proposed by van Schooten et al, 2016 to categorize FLDBs. Only 8 relevant factors were taken into account (e.g., none of the FLDBs were loaded into domains such as history of falls, so those domains were not taken into account). If the FLDB is loaded into more than one domain (i.e., loading$>$ 0.3 in the paper), top three domains to which they indicate highest associations are indicated with integers (1 to 3). The numbers in parenthesis indicate the loading of the variable on varimax-rotated principal components.

events from ambulatory bouts in order to be quantified. For example, turn duration [99, 110] and step/stride length (e.g., examined in [195, 196, 31]) require detection of a gait event, such as turn and foot contacts, respectively. Consequently, established spatiotemporal gait parameters are generally considered as the MGE features, including stance time, double support, step length. Turn-based MGEs were investigated in [99, 110], and they included mean turn velocity, peak turn velocity, turn duration, variability of turn duration, mean turn angle, turn angle variability and the logarithm of normalized jerk.

2. Micro ambulatory bout-based (MBB) outcomes include high-level temporal (e.g., root mean square) and frequency-based (e.g., mean logarithmic rate of divergence) FLDBs [201, 195, 196, 157] extracted from either the detected ambulatory bouts (based

on the cut-off thresholds discussed in 2.2.2) or the subsequent epochs (discussed later), regardless of the enclosed gait events. MBB predictors were either direction-dependent features or based on signal vector magnitude. Directionally-dependent FLDBs were extracted from antero-posterior (AP), mediolateral (ML), and vertical (VT) accelerations (e.g. regularity measures in [201, 203]). A number of FLDBs were extracted from the signal vector magnitude associated with the bout or epoch, including regularity measure [157], phase-dependent generalized multiscale entropy [77], and phase-dependent local dynamic stability [81]). To avoid possible sample size-related bias, each axial/signal vector segment attributed to a macro bout was split into fixed-size epochs for some studies. For example, bouts longer than 10 s and 60 s were split into the fixed 10 s [157] and 50 s [80] epochs, respectively, and each epoch was used separately for the extraction of MBB outcomes. MBB features are less intuitive compared to spatiotemporal gait FLDBs; and are assumed to be indicative of different aspects of gait based on their mathematical description. For instance, slope, width, and amplitude of the dominant frequency in acceleration signals were linked to variability of gait domain [201, 203, 195, 196, 157] and entropy measures (e.g., sample entropy [196], multiscale and phase-dependant entropy [80], [77]), were commonly extracted as potential markers of complexity domain (in [80] a lower entropy extracted from acceleration signals was linked to loss of complexity and an increased regularity).

3. Micro-transitions: similar to MBB predictors, these FLDBs consist of high-level temporal (e.g., peak velocity, range) or frequency-based (e.g. entropy) outcomes, which were either direction dependent (i.e., roll, pitch, yaw [83, 148]), or extracted from the signal vector magnitude (e.g. jerk [148]) attributed to sit-to-walk (or stand) and walk (or stand)-to-sit transitions.

2.2.4 Data-driven models

One study presented a data-driven model by performing factor analysis to systematically categorize FLDBs into independent domains. In [195] principal component analysis was performed on 75 outcomes including: macro (e.g., median number of strides in one ambulatory bout); MGE (e.g., stride time, stride length variability); MBB (e.g., mean logarithmic rate of divergence), and several questionnaire-based and test outcomes (e.g., inability to use public transportation, ability to ascend/descend stairs). Although principal component analysis revealed 18 domains, wearable-based FLDBs were only loaded into 8 classes (considering their 3 highest weights that exceeded 0.3) as shown in Fig. 2.4-model f.

2.3 Discussion

To date, inertial sensors using primarily acceleration signals, have been the preferred approach used to identify fallers based on their natural free-living behaviour over prolonged periods. These systems have demonstrated adequate capabilities to monitor and detect free-living activities, e.g., gait [83, 201, 65, 157, 17, 16, 116, 168, 65, 31, 81, 80, 78, 195, 120, 77, 196, 107], lying [195, 196, 116, 168] and gait events such as turns [99, 110]. However, it was observed that similar FLDBs that were examined by different studies indicated different levels of fall predictive ability, which can be due to the different experimental protocols used to collect free-living data (e.g. sensor placement, duration of free-living data collection, demographics), different mathematical/statistical methods, and algorithms used to define/detect activities (e.g., different cut-off thresholds). Due to these inconsistencies, developing conclusive interpretations of existing evidence remains limited. In the next subsections, the potential sources of inconsistency in methodology and categorization of FLDBs into domains are discussed. Following the sources of inconsistency, recommendations are provided towards harmonization of free-living FRA methods to advance the field.

2.3.1 Inconsistencies in free-living FRA models

Similar FLDBs, different predictive power for falls

Inconsistency in ambulatory bout and turn cut-off thresholds Considering the initial step in processing free-living inertial-based signals is detecting bouts, the observed variability in defining ambulatory bout thresholds is a large source of inconsistency, potentially affecting the fall predictive ability of the extracted bout-based FLDBs. For instance, walking duration and the number of ambulatory bouts (two FLDBs) obtained from bouts longer than 3 s (i.e., 3 s as the minimum cut-off threshold) showed no associations with falls [31]. However, by changing the minimum cut-off threshold to 120 s, these same FLDBs (i.e., walking duration, number of bouts) showed a statistically significant predictive ability for falls [31]. Another example of inconsistent results arising from bout definition differences was variability in bout length. Using the definition of bouts >3 s yielded significant association with falls; whereas variability in bouts > 120 s was not significantly associated [31]. Similarly, while exposure to short duration walks < 7 s [17] and < 8 s [16]) was significantly associated with falls, exposure from walks shorter than 60 s was not discriminative ($p \approx 0.1$) [17].

It was also observed that discretizing angular cut-off thresholds can impact fall predictive power of turn-related FLDBs. For instance, although no relationship between the total number of daily turns (considering turns with different resolutions) and falls was reported [99, 110], after dividing them into three separate angular levels, the quantity of turns in each resolution turned out to be significantly lower for prospective fallers [99]. As only two studies were concerned with turns [99, 110]), the effects of varying cut-off threshold to determine bouts and/or events and subsequent impact on fall predictive ability remains underexamined.

Central tendency measures to estimate FLDBs To extract FLDBs from free-living data, measures of central tendency used to calculate predictors were inconsistent between studies. The different statistical methods resulted in inter- and intra-study inconsistencies in terms of fall predictive values for similar FLDBs. For instance, mode of step time variability in [17] was significantly associated with falls; while the mean estimation did not indicate any relationship. In addition to medians, in [156] extremes of FLDBs were estimated (i.e., the 10th and 90th percentiles of gait characteristics) over 10s epochs/bouts; whereas in [195] only the medians of MBB FLDBs (e.g., entropy, amplitude of the dominant frequency) were reported. For instance, compared to median values, a stronger association was reported for some of the extreme estimations and falls [156]. Similarly, macro gait features such as ambulatory bout duration, mean [31, 116, 168], maximum [201, 195, 196, 17, 168], 90th percentiles [116], and medians [201, 203, 195, 196, 17] were reported. Overall, the lack of consistency limits the capacity to compare across studies and synthesize the evidence.

Free-living data collection protocols Across the reviewed studies, inconsistent data collection protocols may play a key role in fall predictive ability of FLDBs. Specifically, length of data collection, sampling frequency, and location of wearable sensor were inconsistent across studies. Although eight days of free-living data was reported to be sufficient for the identification of fall-prone individuals [17], only one day of free-living data per participant was investigated in [116, 168]. In [197] up to five days was reported to be required for the estimation of median duration of locomotion bouts; while a minimum of two days of free-living data resulted an inter class correlation greater than 0.7 for most activities (sitting, standing and shuffling, except for lying). Sampling frequency was inconsistently used within the studies, ranging from 10 Hz in [107] to 100 Hz in [148] (see Table 2.3). Considering the impact of sampling frequency on the wearable unit battery life [191], the identification of the optimal sampling frequency requires further investigation. Moreover,

estimates of gait characteristics may suffer from errors due to discrepancies in accelerometer location [154]. Although lab-based data suggest that inertial-based wearables mounted on shins can outperform other anatomical locations [117] for detection of fallers, no study has considered this location for the collection of free-living data with lower back as the most frequently used location for sensor placement. To date there is no consensus about the most robust location for free-living FRA, which requires further exploration.

Inconsistencies in the proposed free-living FRA models

Efforts to develop an FRA model based on predictors generated from wearables is needed to interpret FLDBs related to fall risk. These interpretations are critical to understanding the underlying causes or factors indicating risks and informing interventions for clinicians. In contrast, black-box models (e.g., deep models) that estimate risk without interpretive value are less useful. The reviewed studies with models demonstrated considerable inconsistency, likely reflective of the on-going advancement in the field.

As discussed in 2.2.3, quality was found to be the most inconsistently used term with discrepancies in definition and application. There were many inconsistent terminologies observed across a range of domains and FLDBs within the examined models. For instance, harmonic ratio was an indicator of gait smoothness in conceptual model e and an indicator of symmetry in model d (instead, index of harmonicity was the measure of smoothness in model d). In some cases, different terminologies were used to describe the same FLDB. For example, endurance and exposure were defined by the same calculation in [17, 16], albeit described in different domains in models b and c, respectively (Fig. 2.3). In another case, the term 'vigour' was used to describe a domain (Fig. 2.4) [195] and defined as a FLDB (root mean square of vertical angular velocity) in models b and c [17, 16].

In contrast to conceptual models (e.g., Fig. 2.3, models d and e), where all of the AP, ML, VT, and vector-based components of the same MBB FLDB where considered under the same domain (e.g., harmonic ratio in ML, AP, and VT were considered under gait smoothness), different anatomical directions of some of the MBB FLDBs were loaded onto independent domains in data-driven model f (Fig. 2.4). Many of the FLDBs extracted from ML acceleration were loaded into a distinct domain called 'ML balance' (e.g., harmonic ratio in ML direction was loaded into ML balance, while the other two components were considered under quality). Similarly, index of harmonicity in AP direction was loaded into vigour; although the other two directions were loaded into quality. It is clear the proposed models are a work in progress as the field is continually advancing. The existing inconsistencies highlighted in the current review suggest the need for deeper discussion to harmonize interpretive models.

2.3.2 Less-investigated FLDBs and domains

There are a number of FLDBs and domains that have not been well-investigated in the context of free-living FRA. For instance, none of the free-living FRA studies included in this chapter investigated foot clearance measures (e.g. maximal heel clearance, minimum toe clearance, and their variability) or Micro FLDBs attributed to postural control [103] (or base of support [68]) domain. In addition to step width and its variability (investigated in factor analyses by [103, 68]), other features such as the frequency of compensatory balance reactions [82], and the number of steps to recover balance in a misstep, are likely to be linked to this domain. Controlled laboratory evidence supports the view that impaired ability to execute compensatory balance reactions is associated to a higher risk of falling [109]. However, only a few studies [178, 48, 82] have examined the links between falls and the frequency of naturally-occurring CBRs. Among them, only one was instrumented [82], and therefore, included in this literature review. The latter reported a strong association between suspected missteps and retrospective falls in PD populations. Moreover, step width variability has shown to be linked to a fall history in older adults [13]. However, no study was found exploring step width (and its variability) in fallers and non-fallers in free-living condition. This requires to be investigated in future studies. Other domains and FLDBs that have not been investigated in the free-living context, include, but not limited to: arm symmetry measures during gait [69], foot strike angle measures [182], micro transition and micro standing measures [27], which are recommended to be investigated in future studies.

2.3.3 Recommendations for harmonization and clinical implications

Precise activity recognition algorithms

One possible reason for inconsistencies in fall predictive power of FLDBs is the use of black-box thresholds in activity recognition algorithms [52, 7] in some of the studies, resulting in a low specificity and clarity to analyze free-living activities. For instance, studies included here were only concerned with general types of activities, such as ambulation, standing, sitting, and lying. A subset of studies differentiated between patterns of gait, such as ascending/descending stairs, or fast walking. For instance, [156] excluded locomotion bouts suspected to contain running episodes as they caused severe outliers in FLDB estimations for some of the participants. Additionally, turns and compensatory balance reactions were not excluded from ambulatory bouts before the extraction of MBB features. For

instance, the misstep detection method in [82] revealed a number of suspected missteps from 5-second ambulatory bouts. However, in a subsequent study [203], MBB features were extracted from the same dataset without excluding those suspected missteps, which may have affected the fall predictive values of the MBB features.

Multimodal approaches, such as a combination of IMUs, surface electromyography [130], electroencephalogram [30], heart rate variability and ECG [21] and pressure-sensitive shoe-insoles [34, 118, 40] may increase the specificity of activity recognition algorithms, and therefore, to improve the sensitivity of FLDBs to identify fallers. As an example, using a force sensing insole equipped with an IMU and a barometric sensor, a framework to discriminate between level walking and stairs patterns of gait was developed and evaluated [41]. State-of-the-art algorithms, such as long-short term memory and deep convolutional neural networks have achieved near-human accuracy levels in detection of a broad range of activities from multimodal public datasets [138], which can be employed as a replacement to threshold-based methods to detect a broader range of natural activities. Only one study [120] investigated the use of deep learning models to identify fallers based on their preprocessed 10-second walking patterns; where deep models slightly outperformed the baseline approach based on the biomechanical FLDBs as discussed in [195].

Interpreting FLDBs by acquiring contextual information

In [193] it was shown that the mobility measures are affected by the environmental features (e.g., sidewalk slopes and crossings) and it was hypothesized that individuals would adapt to challenging environments by decreasing gait speed, increasing cadence, and shortening stride length. Moreover, a higher variability in ML direction, e.g., a lower amplitude of the dominant frequency, could indicate a higher adaptability to the environment [201]. However, the intrinsic meaning of these measures and terminologies (e.g., adaptability) in different anatomical directions and contexts remains unclear, inertial sensors do not provide sufficient information on human-environment interaction. Although applying cutting-edge algorithms (as discussed in 2.3.3) can boost the interpretation of context (e.g. stair climbing, walking downhill), the validity of these algorithms in complicated free-living scenarios needs to be carefully examined (also refer to the last column in Table 2.3, which shows the algorithms used in included studies were not mostly validated in free-living conditions).

In [133] time-stamped self-reporting (voice recordings) was along with IMU data to increase the interpretability of IMU data and to locate compensatory balance reactions (slips, trips, stumbles) collected in 4 weeks. However, such a tool may also suffer from subjectivity and under-reporting. On the other hand, egocentric first-person video, acquired

via body-worn cameras have been used as a gold standard [52] for the purpose of validating IMU-based algorithms or identification of specific events or environmental features [32, 62, 186]. First person video captures more contextually relevant information on the properties of the environment compared to IMU alone. This includes, but is not limited to, varying slope and surface navigation as well as static and dynamic objects (e.g. obstacles, pedestrians) that influence mobility behaviour. As an alternative to frame-by-frame investigation (visual inspection) of first person video data, algorithms are in development to automatically detect extrinsic risk factors from first person video data. For instance, image processing (e.g., Gabor Barcodes) techniques have shown promising performance in automatic detection of environmental fall-related hazards, including slope changes (e.g., stairs, curbs, ramps) and different surfaces (e.g., gravel, grass, concrete) [128]. By augmenting IMU approaches with egocentric videos, more insight can be readily gained from specific motoric activity. For example, gait data pertaining to micro/quality gait (from an IMU) within a new residential environment under low-level lighting conditions (video) or within crowded open spaces during daylight offer different challenges for fallers. Combined IMU and video approaches may allow healthcare professionals to target individualised approaches for rehabilitation strategies, ensuring safer navigation and reduced falls.

An all-inclusive data-driven model

As discussed in 2.3.1, different hyperparameters, such as ambulatory bout length, central tendency measures, and data collection protocols, such as the length of free-living data and sampling frequency, can impact fall predictive power of FLDBs. However, there is not sufficient evidence indicating the optimal values for these hyperparameters to achieve the highest predictive values specific to each FLDB. Moreover, many FLDBs were not investigated in any of the previous data-driven models (e.g., Fig. 2.4-model f and controlled models such as [103]). Therefore, debate continues about their real identity in terms of their allocation to free-living domains. The aforementioned gaps in the literature indicate a need to obtain a standardized model to define discrete independent domains by performing factor analysis on a comprehensive range of wearable-based FLDBs derived from a broad range of video-validated activity bouts. This comprehensive set may also include similar FLDBs but different in terms of hyperparameters (e.g., ambulatory bout lengths, turn cut-off thresholds, and central tendency measures). It is also feasible to simulate shorter/longer collection periods, and sampling rates (by up- and down-sampling signals) and examine FLDBs' sensitivity with respect to these factors. Performing factor analysis on the aforementioned comprehensive set of features (including those discussed in 2.3.2) altered based on different hyperparameters would permit deeper insights on developing more structured

free-living models and provides more information on the differences between laboratory and free-living features in prediction of falls.

2.4 Conclusion

Overall, free-living FRA using wearables is a promising avenue for fall prevention; however, due to the high level of heterogeneity in the use of wearables; e.g., sensor location, diverse cohorts, stratified employment (e.g., 1 vs 7 days), definition of free-living domains, and the selection of free-living bout resolutions, the evidence for the relationships between FLDBs and falls remains inconclusive. Moreover, many FLDBs were specific to research groups and were not systematically investigated in an all-inclusive factor analysis. Therefore, achieving a data-driven model is necessary to systematically identify the FLDBs, bout resolutions, and domains with the highest predictive values for falls to eventually address intervention programs and prevent older adults from falling. Publishing well-annotated video-validated free-living datasets to support harmonization efforts is strongly recommended.

Table 2.1: Study designs for capturing free-living data. ACC: accelerometer, Gyro: gyroscope, FLDB: free-living digital biomarker.

Study	Free-living data-log duration	Type of activities	Modality(ies)	Sensor placement
Weiss (2013) [201]; Ihlen (2015, 2016a, 2016b) [81, 80, 78]	3 days	Gait	DynaPort MoveMonitor	At the L5 level
Iluz (2014) [82]	3 days	Missteps during gait	DynaPort Hybrid (ACC+Gyro)	Lower back(L4-5)
Weiss (2014) [203]	3 days	Gait		
Iluz (2015) [83]	3 days	Sit-to-walk and walk-to-sit transitions	ACC+Gyro (one unit)	Lower back
Rispens (2015a) [157]	2 weeks	Amount of physical (in-)activity and quality of daily-life gait	DynaPort MoveMonitor	Around the waist and set along the lumbar spine
van Schooten (2015a) [196]	8 days	Amount of physical (in-)activity	DynaPort MoveMonitor	Dorsally on the trunk at L5
Rispens (2015b) [156]	1 week	Amount of physical (in-)activity and quality of daily-life gait		
van Schooten (2016) [195]	1 week	Bouts of locomotion, sitting, lying, standing		
Nait Aicha (2018) [120] and Ihlen (2018) [77]		only gait in Nait Aicha (2018) [120] and Ihlen (2018) [77]	DynaPort MoveMonitor	At the L5 level
Brodie (2015) [17]	Mean 58 days/participant: Fallers: 44.0(29.0) days, NFs: 67.0(29.0) days	Activities during walking hours were monitored, only gait was investigated	Philips (Senior Mobility Monitor) pendant (ACC+Barometer)	Sternum Level
Brodie (2017) [16]	7 days (≥6hr data per day)	Activities during walking hours were monitored, only gait was investigated	Philips (Senior Mobility Monitor) pendant (ACC+Barometer)	Sternum Level
Mancini (2016) [110]	≈10 hours/day for 7 days	Turning mobility	3 IMUs(APDM Opal)	One on belt (L5), 2 on shoes
Leach (2018) [99]	5-9 days	Turns and gait	Android smartphone (ACC+Gyro) and uFALL app	Midsagittal plane of the lower back
Mohler (2016) [116]	48 hours	Walking, sitting, standing, lying and postural transitions (all time except while showering)	PAMSys BioSensic	Into a T-shirt with a device pocket at sternum level
Pozaic (2016) [148]	7 days	Sit-to-stand transitions (focus of the study)	Bosch sensor tech Gmbh (ACC, Gyro, magnetometer)	One attached to wrist
Mactier (2015) [107]	7 days	Gait	activPAL (uniaxial ACC)	Upper thigh
Hiorth (2016) [65]	7 days (at least 4 days)	1) Volume, 2) pattern, 3) accumulation, 4) variability of sedentary behavior (sitting, lying), standing, and ambulatory bouts	activPAL	Middle of thigh

Del Din (2017) [31]	7 days	Macro- and micro-gait parameters (spatiotemporal)	Axivity AX3	Lower back
Schwenk (2014) [168]	24hours	Walking, sitting, standing, lying	Physilog (2ACC+1Gyro)	Attached to the chest (pocket)
Gietzelt (2014) [49]	7 days after each visit (4 visits in 8 months)		One SHIMMER sensor	

Table 2.2: Demographic data, Fs: fallers, NFs: non-fallers, OA: community-dwelling older adults, PD: people with Parkinson's disease, f: female, m: male, FLDB: free-living digital biomarker.

Study	Specific disease (Yes/No)	Number of participants, Age, female:male		Categorization of fallers and non-fallers based on	
				Prospective falls	Retrospective falls
Weiss (2013) [201]; Ihlen (2015, 2016a, 2016b) [81, 80, 78]	No	NFs: 39, 78.77(4.39)y, fs: 64.10%	Fs:32, 77.86(5.09)y, fs: 65.62%	12 participants reported $\geq$ 2 falls 6 months following the experiment [4]	$\geq$ 2 falls one year prior to measurements
Weiss (2014) [203]	All PD	NFs: 67(40 to 85y)	Fs: 40(40 to 85y)	1 year follow-up(each month returned) [5], 14 NFs turned to Fs	$\geq$1 fall prior year
Iluz (2014) [82]	All PD	NFs+Fs:40, 62.16(10.02)y, Fs:9, f:m=8:32		N/A	$\geq$ 2 falls in the past 6 months
Iluz (2015) [83]	No	NFs: 38, 78.65(4.35), f: 63.15%	Fs: 33, 77.89(4.99), f: 66.66%	No	$\geq$2 falls in the previous year
Rispens (2015a) [157]	No	110 (Fs+NFs),78.4(7.8)y, f:m= 77:33		No	$\geq$2 falls in the previous year; 1.2(1.6) falls in 12 months prior the measurements
van Schooten (2015a) [196]	No	Retrospective-NFs: 109, Prospective-NFs: 110	Retrospective-Fs: 60, Prospective-Fs: 59	$\geq$ 1 Falls in 6 months follow-up	Falls in the preceding 6 months
		NFs+Fs: 169, 75.4(6.8)y, f: 52.1%			
Rispens (2015b) [156]	No	NFs: 132, 75.1(6.6)y, f: 50%	Fs: 70, 75.6(6.1)y, f: 53%	$\geq$ 1 falls 6 months follow-up	Used in models but FLDBs were analyzed with respect to prospective falls

[4]Statistical analysis was done with respect to retrospective falls only

[5]Statistical analysis for FLDBs was only done with respect to retrospective falls; but separate survival analysis/Cox regression was perfomed for prospective falls

van Schooten (2016)[6] [195]	No	Complete data for 294/319 participants were analyzed. Total 294 participants: 75.3(6.8)y, f: 50.8%, 48.8% had $\geq$ 1 falls in past year, 25.2% had $\geq$ 2 falls in past year		Initially 6 months, extended to 12 months if willing to continue	$\geq$ 1 and $\geq$ 2 falls in past year
Nait Aicha (2018) [120]		101/296 participants had $\geq$ 1 fall (34.1%), m: 74.1%			
Ihlen (2018) [77]		Total NFs: 199; Matched NFs for Single-Fs: 75.9(6.7)y, f: 51%; Matched NFs for Recurrent-Fs: 75.2(6.4)y, f: 48.8%	Single-Fs: 58, 76.1(6.8)y, f: 51%; Recurrent-Fs (2< falls): 46, 75.5(6.7)y, f: 48.8%	6 months follow-up	
Brodie (2015) [17]	No	NFs: 11, 84.0(7.9)y, f:m=7:0	Fs: 7, 82.2(5.9)y, f:m=7:4	No	$\geq$1 falls in the past 12 months
Brodie (2017) [16]	No	NFs: 63, 75.8(7.3)y, sex: 0.48(0.50) (considering f=1 and m=0)	Fs: 33, 74.9(8.5)y, sex: 0.81(0.39) (considering f=1 and m=0)	No	$\geq$ 1 fall in prior 12 months
Mancini (2016) [110]	No	NFs: 16, 83.9(7.0)y, f:m=3:13	Recurrent-Fs: 7, 88.4(8.8)y, f:m=2:5; Single-Fs=12, f:m=8:4, 86.0(7.0)y	6 months following the 7 day recording, 7/35 experienced one or more fall	N$\geq$2 falls in the previous year
Leach (2018) [99]	No	Retrospective-NFs: 154 Prospective-NFs = 153	Retrospective-Fs = 6 Prospective-Fs = 7	$\geq$ 2 falls in 12 months	$\geq$ 2 falls in 12 months
Mohler (2016) [116]	Non-frail, pre-frail and frail	Non-frail NFs: 23, 74.7 (6.7)y, fs: 19 Pre-frail NFs: 38, 79.7(8.5)y, fs:28 Frail NFs: 10, 86.6(5.9)y, fs:7	Non-frail Fs: 20, 74.4(6.6)y, fs: 17 Pre-frail Fs: 19, 79.4(8.8)y, fs:15 Frail Fs: 9, 80.9(9.8)y, fs:9	Falls in the 6 months after the initial baseline study visit	N/A
Pozaic (2016) [148]	No	NFs: 123, 72.4(5.6)y, fs in NFs:57.7%	Fs: 13, 74.2(5.3)y, fs in Fs:69.2%	$\geq$ 1 fall in one month follow-up (overall 21 falls, 4 participants $\geq$2 falls)	N/A
Mactier (2015) [107]	All PD (no healthy control)	NFs: 70, 67.5(60.6-75.0)y, ms: 65.7%	Single-Fs: 17, 72.2 (64.1–75.7)y, ms: 76.5% Recurrent-Fs: 24, 69.3 (61.5–74.5)y, ms: 75%; Based on pre-fall events: (a) Ambulation-Fs: 17, 64.6 (57.7–72.1)y, ms: 70.6% (b) Transition-Fs: 14, 73.0 (67.7–78.2)y, ms:85.7%	12 months	N/A

[6]Fall risk assessment in older adults (FARAO) dataset

Hiorth et al. (2016) [65]	All PD	NFs: 28	Fs: 20	N/A	6 months fall history
Del Din (2017) [31]	OA and PD	OA-NFs: 50, PD-NFs:15; For all NFs: 69.05(7.67)y f: 56%	OA-Fs: 122, PD-Fs:155; For all Fs: 73.33(6.78)y, f: 42%	N/A	≥2falls past 6 months, NFs: not fallen in at least 18 months
Schwenk (2014) [168]	People with dementia	NFs: 49, 81.8(5.9)y, m: 34.7%	Fs: 28, 82.0(7.1)y, m: 21.4%	3-months follow-up	
Gietzelt (2014) [49]	People with dementia	At the beginning 40 Fs+NFs (fs: 20), 76.0(8.3)y, Fs=13 (n=8 fell once, n=2 fell 2 times, n=1 fell 4 times, n=2 fell 5 times, 12 drop-outs		A prospective cohort study with 3 phases: short-term (start to month 2): n=38(2 drop-outs (DOs) , 6 falls, 18 missing dataset, (month 2-4): n=33 (5 DOs), 11 falls, 2 missing dataset, (month4-6): 30 (3 DOs): 8 falls, 5 missing dataset, (month 6-8):n=28 (2 DOs), 1 fall, 11 missing dataset.	

Table 2.3: Bout of walking and activity definitions. ACC: accelerometer, Gyro: gyroscope, 3D: three-dimensional. FLDB: free-living digital biomarker, OA: older adults, VT: vertical, AP: antero-posterior, F: faller, NF: non-faller.

Study	Bout of activities/walking definition and Specific constraints	signal identifying bout	Sampling rate	Validation
Weiss (2013) [201]; Weiss (2014) [203], Ihlen(2015, 2016a,2016b) [81, 80, 78]	Ambulatory bouts≥ 60s were taken into account for FLDB extraction (dissected to 60-second intervals, e.g., in [201]). The bouts were identified based on two filters: 1. a signal magnitude area (SMA) threshold-based activity detection monitor, 2. a threshold of the energy in the frequency domain (windows with energy <0.05 were excluded). In Ihlen (2016) [80] ambulatory bouts ≥60s were divided into 50-second epochs for the extraction of entropy/complexity features.	ACC(3D)	100Hz	In Weiss (2013) [201] it was mentioned that the validated methods discussed in papers by Weiss (2011), Moe-Nolssen (2004) and Yack (1993) were used to quantify different aspects of gait. Validation in free-living conditions: N/A.
Iluz (2014) [82]	5-second windows (running window of 5s) with 2-15 steps were detected and considered for FLDB extraction	ACC(3D), Gyro(3D)	100Hz	To develop the algorithm 29 missteps (negotiating obstacles while walking) were captured in a laboratory setting and more than 60 hours of data were recorded. Their rule-based algorithm achieved a 93.1% hit ratio and 98.6% specificity on this dataset.

Iluz (2015) [83]	Detection of the subject's gross activity (e.g., gait, sitting) was performed based on the local mean of the acceleration signals (e.g., negative values for local mean of the vertical acceleration signal indicate lying; while positive values may correspond to: gait, standing, or sitting), with additional criteria applied to increase the robustness of the detection. The transitions were identified after detection of sitting and gait episodes.	ACC (3D), Gyro (3D)	100 Hz	Validation in free-living conditions: N/A.
Rispens (2015a) [157]	Locomotion episodes $\geq$ 10s, each episode was split into 10-second epochs (to avoid sample-size related bias)	ACC(3D)	100Hz	Validation in free-living conditions: N/A.
van Schooten (2015a) [196] and van Schooten (2016) [195] and Rispens (2015b) [156], Nait Aicha (2018) [39]	Locomotion episodes $\geq$ 10s, each episode was split into 10-second epochs (to avoid sample-size related bias), sufficient quantity of gait bouts ($\geq$ 50s) per participants were included in analysis. In Rispens(2015b) [156] locomotion bouts with suspected running events were discarded.	ACC(3D)	100Hz	Bouts of non-wearing, locomotion, sitting, lying and standing were identified using manufacturer's algorithm (Dijkstra et al.) Validation in free-living conditions: N/A.
Ihlen (2018) [77]	First, ambulatory bouts$\geq$3s were identified by a commercially available activity detection algorithm. Then, 3D ACC signals attributed to ambulatory bouts$\geq$30s were included for the analysis. The included ambulatory bouts were split into equal-sized 30-second epochs to provide a consistent sample size to estimate entropy measures. The epochs were further checked (visually) and non-walking activity bouts were excluded. Inclusion criteria for walking epochs were: (a) distinct impact peaks in VT and/or AP component of the ACC signal, (b) distinct cyclical ACC pattern in VT and/or AP component(s), and (c) criteria (a) and (b) were satisfied for at least 80% of the epochs, where max 20% was considered for gait initiation, turning or transitional micro-breaks.			The aforementioned inclusion criteria were based on the visual inspection of fast, normal and slow walking pattern discussed in a 'validation' study for activity detection in OA by Bourke et al., (2017).
Brodie (2015) [17]	Ambulatory bouts were defined by consecutive heel strikes identified by vertical acceleration peaks in the level 4 and 5 Daubechies 5th-order wavelet decomposition. Daily-life gait quantity was quantified by: (a) Steps per day from ambulatory bouts$\geq$3 steps), (b) Walks per day (of ambulatory bout$\geq$8 steps), (c) Steps per walk (mean of walks$\geq$8 steps). For intensity and quality analyses: ambulatory bout$\geq$7 steps and for exposure analysis ambulatory bout $<$7s and $<$60s were taken into account.	ACC(3D)	50Hz (25Hz barometer-not used)	Validation in free-living conditions: N/A.
Brodie (2017) [16]	Walks were defined by 3 consecutive heel strike peaks less than 3s apart. Short walk exposure was calculated by the percentage of walking duration $<$8s.	ACC(3D)	50Hz (25Hz barometer-not used)	Validation in free-living conditions: N/A.

Mancini (2016) [110]	Ambulatory bouts longer or equal to 10s were first detected using 3D angular velocity signals and then were further investigated for the detection of turns, 0.5s<turnDuration<10s and 45^o $\leq$turn angles. Turn angles were obtained by integrating the angular rate of the lumbar sensor about the VT axis.	3D angular velocity signals	128 Hz	In El-Gohary (2013), compared to Motion Analysis and video, the algorithm maintained a sensitivity of 0.90 and 0.76 and a specificity of 0.75 and 0.65, respectively. The turn detection algorithm was further applied to data collected in the home from 12 PD and 18 control participants and the algorithm successfully detects turn characteristics.
Leach (2018) [99]	Turn type and angular range of turn include 50-100^o, 100-150^o and 150-200^o (analyzed only if happened during gait with 0.5-5s duration). The turn detection algorithm was based on the angular rotational rate of the pelvis about the VT axis.	ACC(3D) + Gyro(3D)	N/A	The turn detection algorithm was validated in El-Gohary (2013) as above.
Mohler (2016) [116]	Walking bouts >5s and > 3 steps were taken into account (based on Tosizadeh(2015))	ACC(3D)	N/A	Validation in free-living conditions: N/A.
Pozaic (2016) [148]	N/A	ACC(3D)	100Hz	Particular trigger events (such as rotation of the wrist above a predefined threshold), as well as periodical or motionless/dormancy situations after these events. An ACC-based arm swing detector was used for the detection of the walking phase. Methods was validated in a pilot study with 28 OA (65-90y), who performed eight different types of the sit-to-stand transitions in a controlled environment (i.e. camera-supervised lab) as part of the protocol that simulated activities of daily living. The algorithm showed 71.4% precision for the non-dominant hand and 67.9% precision for the dominant hand.
Mactier (2015) [107]	Windows of 15s were used and walking episodes of $\geq$one step were taken into account for FLDB extraction. Total number of steps in ambulatory bouts of<20 steps, 20-100 steps and > 100 steps were also taken into account.	ACC(1D)	10Hz	Previous work in OA against other accelerometer and video recordings in people with rheumatoid arthritis during simulation of ADL in the laboratory. It was mentioned that the validated activePAL can identify postures (e.g. sitting, lying, standing).
Hiorth (2016) [65]	Ambularory bouts<10 strides, 10-50 strides and >50 strides were taken into account.	ACC(1D)	20Hz	Validation in free-living conditions: N/A.
Del Din (2017) [31]	All ambulatory bouts >3 steps (minimum bout length) were taken into account for the analysis. For for micro-gait FLDBs, ambulatory bout> 10s were taken into account. For macro-gait FLDBs, ambulatory bouts> 3steps (short), Ambulatory bout> 60s (medium) and Ambulatory bout>120s (large) were considered. A threshold of 2.5s was set for the maximum resting period between consecutive ambulatory bouts	ACC(3D)	100Hz	The ambulatory bout detection algorithm was validated in a study with wearable cameras (Hickey et al., 2017). Characteristics of gait were selected based upon a model of gait validated both in OA and in people with PD in two distinct studies.
Gietzelt (2014) [49]	Ambulatory bouts of $\geq$ 20 s were taken into account.	ACC(3D)	N/A	Validation in free-living conditions: N/A.

Schwenk (2014) [168]	A walking period was defined as an interval with at least 3 successive steps as described in the validation study of the Physilog. Activities with < 3 steps were considered as standing (e.g. working in the kitchen and moving < 3 steps).	Two ACCs and one Gyro	N/A	It was mentioned that the algorithm was sensitive (87-99%) and specific (87-99.7%) for detection of the physical activity pattern in different samples of OA and patients. Validation in free-living conditions: N/A.

Chapter 3

Automated detection of multidirectional compensatory balance reactions: Insights from controlled studies

[1]

As discussed in Chapter 2, earlier studies 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 [116] and turns [110]) and micro (e.g., spatiotempral measures such as step time [31], or frequency measures [195] including index of harmonicity). 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, Chapter 2 highlighted that the relationships between falls and free-living dynamic postural control measures have yet to be investigated in depth [122]. Considering balance impairment as one of the strongest risk factors for falls [190], the investigation of balance-related FLDBs may lead to more stable risk assessments and provide new insights into fall prevention in older adults. As an alternative to the step-dependent

[1]The content of this chapter is mainly obtained from the following peer-reviewed research paper: **Nouredanesh, M.**, Gordt, K., Schwenk, M., & Tung, J. (2019). Automated detection of multidirectional compensatory balance reactions: A step towards tracking naturally occurring near falls. IEEE transactions on Neural Systems and Rehabilitation Engineering, 28(2), 478-487

approaches, monitoring the frequency and characteristics of naturally-occurring compensatory balance reactions (CBRs) has been suggested as a promising direction. This necessitates the development of *robust models* to detect older adults' CBRs under free-living conditions.

The present chapter discuses proof-of-concept testing for CBR detection in 9 young, healthy adults, and establishes the infrastructure for Chapter 4, where methods for the detection of older adults' naturally-occurring CBRs are investigated.

3.0.1 Compensatory Balance Reactions (CBRs)

CBRs, also known as missteps, near-falls, or reactive balance responses, are reactions to recover stability following a loss of balance. In anterior-posterior (AP) direction, CBRs are often elicited from slips and trips. During slips, which occur mainly due to the loss of traction on a low-friction surface or because of wearing an inappropriate footwear [205, 94], a person's base of support (BoS) is displaced in the forward direction relative to the center of mass (CoM) [100]. A trip also occurs when the CoM moves beyond the BoS, and is typically caused by a collision between the swing foot and an obstacle (e.g., curbs, power cords) [165]. In [48], tripping was reported to be the most common cause of near-falls in Parkinson's disease (PD) fallers. There are two general strategies as responses to trips [43]: 1) elevating, which is caused by a perturbation (impact from an obstacle) during early swing phase (i.e., 5-25% of stride duration in normal walking [166]) and 2) lowering, which is caused by a perturbation during late swing (i.e, 55-75% of stride duration in normal walking [166]). For the first strategy, the leading (swing) leg is lifted over the obstacle while the support leg creates the push-off reaction after the passive reaction [170, 43]. During the lowering strategy, the swing leg rapidly touches the ground in front of the obstacle. The trailing leg is swung forward after clearing the obstacle and positioned in front of the body to recover stability [170]. In mid swing (30-50% of stride duration in normal walking), either strategy may occur [166, 43]. In Fig. 3.1, C2, C5, C6, and C9 represent the aforementioned reactions in AP direction.

Moreover, frontal plane instability and lateral balance impairments have shown to be an important risk factor for falls in older adults [64]. Comparing reactive responses to perturbations in frontal and sagittal plane, McIntosh et al. showed that lateral support surface perturbations are the most challenging to postural control in older adults [113], as accumulating evidence indicates impaired mediolateral (ML) stability requires active control in contrast to more passively stable antero-posterior (AP) stability [67, 131]. Additionally, physical consequences of falls (e.g., hip fractures) are more likely to be associated

with falls in lateral direction [115]. These CBRs generally include: 1) the sidestep (SS) or lateral ankle strategy (Fig. 3.1-C4 and C7), which comprises of a medial or lateral movement of the foot [67], and 2) crossover step (CO) strategy, in which the stepping foot crosses in front of the stance foot [67] (Fig. 3.1, C3 and C8).

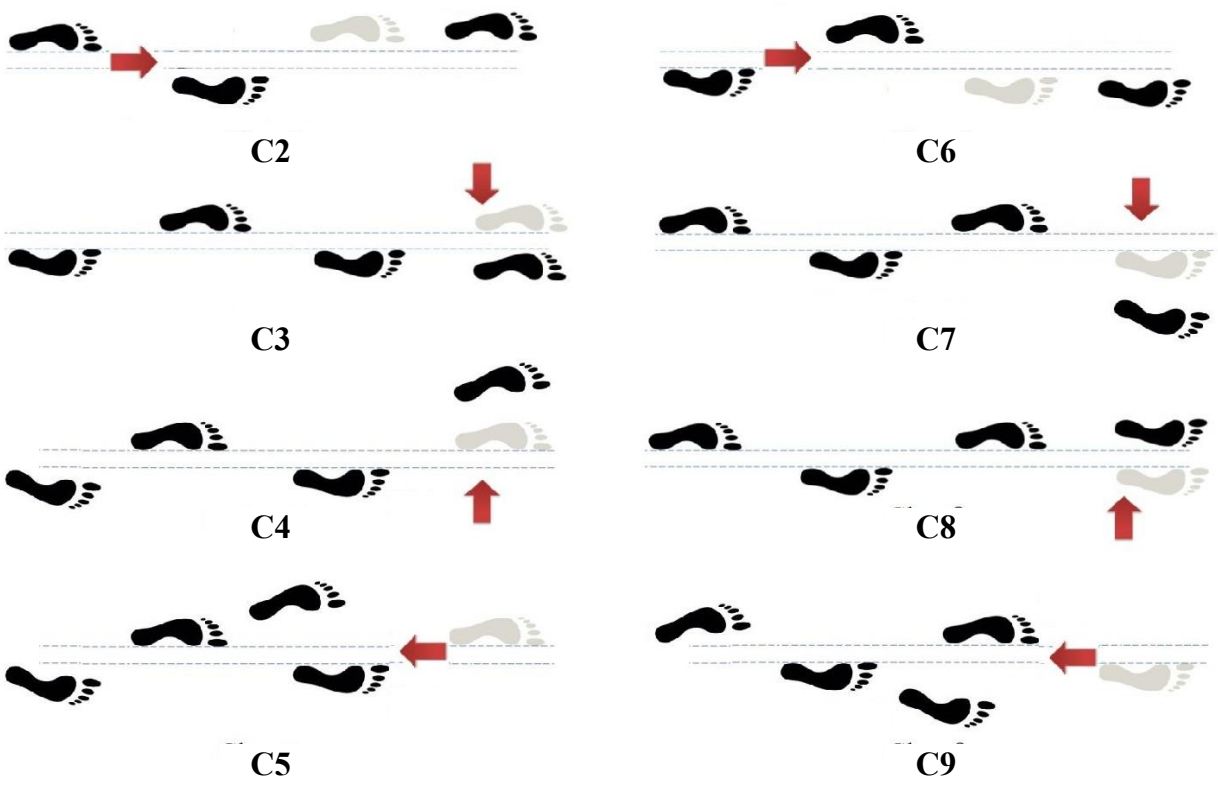


Figure 3.1: Types of compensatory balance responses (CBR), black and grey areas represent the actual and expected foot contacts, respectively Row 1 depicts reactions elicited by forward perturbation (slip-like reaction), rows 2 and 3 show reactions elicited by lateral perturbation: right foot loaded at perturbation (arrow), eliciting a left foot side step (SS), left foot loaded at perturbation (arrow), eliciting right foot crossover (CO) step, forward CBR. Row 4 presents the trip-like reaction induced by backward perturbation. Note that C1 includes normal gait segments and not presented in this figure. C: Class.

3.0.2 Prior CBR detection studies

While controlled laboratory evidence supports the view that impaired ability to execute CBRs is linked to a higher risk of falling [109], there are only a few studies focusing on the detection of CBRs [130, 129, 126, 6, 204, 82, 24, 90]. In [6], a support vector machine model was developed based on the features (e.g., mean, variance of signals) extracted from seven IMUs mounted on the ankles, thighs, sternum, waist, and head. This approach used a dataset (e.g., 150 simulated CBRs including recovery from tripping, missteps, collisions with pedestrians, and incorrect transfers) collected from healthy, young participants in controlled conditions. Although the model distinguished the simulated near-falls from activities of daily living (ADLs) with 100% sensitivity and specificity, as naturally-occurring CBRs are likely to exhibit faster coordinated responses [109] than volitional stepping efforts, the generalizability of the proposed method to detect CBRs occurring involuntarily with no attention needs to be examined (addressed in Chapter 4).

In another controlled study [24], a total of 100 stumbles (in 45 minutes of walking) were induced in 9 participants by placing wooden barriers at random times during gait. During this process, the participants were blindfolded and listened to music through headphones. The anomaly detection algorithm developed using the chest-mounted accelerometer data resulted in 0.2% false alarm rate. In another study, trip-like stumbles were elicited on a treadmill by winding/unwinding a stiff robe attached to the ankles at different phases of gait [90]. The subsequent wavelet-based detection algorithm further identified stumbles from treadmill walking with a specificity of 99.9% and a sensitivity of 98.4% [90].

A threshold-based algorithm for the detection of tripping events, induced by placing obstacles on a treadmill, was proposed by Weiss et al. [204], which achieved $\approx 85\%$ sensitivity and specificity. Another study from the same research group investigated a rule-based algorithm to detect missteps in 40 individuals with PD while negotiating obstacles [82]. The method achieved 93.1% hit ratio and 98.6% specificity when applied to controlled/in-lab data [82]. When applied to three days of free-living IMU data, the quantity of 'suspected'[2] missteps was reported to be positively ($\rho = 0.010$) associated with retrospective falls in people with PD. This supports the findings of a paper by Srygley et al. [178], where the quantity of self-reported missteps was shown to be positively associated with the frequency of prospective falls in older adults (relative risk 3.89). In contrast, Gazibara et al. [48] showed that self-reported near-falls in the first 6 months were not associated with falling in the latter 6 months of follow-up in people with PD. Considering the aforementioned find-

[2]The term 'suspected' for this FLDB reflects the lack of criterion (gold) standard data to reliably validate the employed threshold-based CBR detection approach in free-living conditions (this will be further discussed in Chapter 4).

ings, further research on the validity of CBR detection methods needs to be undertaken to reliably examine the associations between the frequency of naturally-occurring CBRs and falls.

To address the lack of existing methods for the *automated* detection of lateral CBRs, Nouredanesh et al. developed machine learning-based models to detect CO and SS reactions using signals captured by three wearable IMUs (on the sternum, right shank and thigh), and surface electromyography (sEMG) sensors (mounted on biceps femoris, gastrocnemius, tibialis anterior, rectus femoris) [130]. The models detected CBRs (CO+SS) elicited by lateral perturbations with the leave-one-subject-out (LOSO) cross-validation accuracies of 99.21% and 83.95% using IMU and sEMG sensor streams, respectively. While lateral CBR detection was strong using IMU signals, observed errors indicated difficulty distinguishing between SS and CO stepping reactions. These findings indicate wearable sEMG sensors may provide complementary information, particularly when CBR movements are small or attenuated [130]. However, considering the additional setup and variability associated with skin preparation and electrode placement, coupled with marginal improvement in detection accuracy, the present thesis considers IMU-based signals only.

3.0.3 The proposed CBR detection models

To develop a robust machine learning-based CBR detection framework, it is important to train the model on a relatively large dataset. However, acquiring large sets of naturally-occurring CBRs is logistically challenging as they are relatively rare events [133, 24]. For instance, only 46 self-reported CBRs were reported by three older non-fallers in 107 person-day of data in [133]. As an alternative, in the present study, a perturbation treadmill (PT) was hypothesized to be a reproducible and safe option for eliciting *multidirectional* CBRs (PT-CBRs here). Therefore, the present chapter discusses the development of machine learning models, i.e., random forest (RF) to automatically distinguish between:

1. PT-CBRs and normal gait (NG) episodes captured on the treadmill (i.e., PT-NGs),
2. different types of PT-CBRs.

Eight classes of PT-CBRs were considered and induced by delivering mechanical perturbations in the following 4 directions: 1,2) left and right (inducing CO and SS strategies), 3) forward (inducing trip-like reactions), and 4) backward (evoking slip-like responses); at right and left stance conditions (4 directions $\times$ 2 stance conditions = 8 classes), see Fig. 3.1. Five wearable IMUs were placed on participants' waist (pelvis) and bilaterally on

their ankles and thighs. Therefore, the classification results using a multi-IMU approach was discussed in sections 3.1.4 and 3.2.1). Moreover, the impact of individual sensor location on the CBR detection results was investigated in sections 3.1.4 and 3.2.2. For the development and assessment of the aforementioned models, the leave-one-trial-out (LOTO) cross-validation approach was considered.

To further assess the generalizability of the proposed CBR detection framework:

1. the LOSO cross-validation approach was used to assess the framework's *robustness against inter-participant differences* (discussed in 3.1.4 and 3.2.3), and

2. an independently captured dataset, i.e., the "Inertial Measurement Unit Fall Detection" (IMUFD) [6, 5], was employed as the test dataset, to assess the framework's *generalizability to never-seen data*.

The IMUFD dataset, which was the only relevant publicly available dataset at the time of the experiment, represents common real-life fall scenarios observed in long-term care facilities [5] and contains different types of simulated CBRs (S-CBRs), as discussed in 3.0.2. The procedure for this generalizability assessment as well as the test results are provided in sections 3.1.4 and 3.2.4, respectively.

3.1 Materials and methods

3.1.1 Participants and protocol

Nine healthy young participants (mean age = 26 years) wore five IMUs (Opal model, APDM Inc., Portland). Two APDM IMUs were mounted bilaterally on the ankles below treadmill IMUs (see Fig. 3.2), two at the center of anterior side of the thighs, and one at the pelvis. Six channels were collected at $f_{s,APDM} = 128$ Hz by each IMU unit: 3 channels of acceleration (vertical: X^+ points downward, medial-lateral: Y^+ points left, anterior posterior: Z^+ points forward, see Fig. 3.2, and angular velocity in three directions: yaw, pitch and roll) resulting in overall 30 (=5 (sensors)×6 (channels)) signal streams for each participant. Accelerometers and gyroscopes were set to operating ranges of ±16g and ±2000 deg/s, respectively. To trigger perturbations at the correct phase of gait, two treadmill IMUs comprising of 1D accelerometer and 1D gyroscope were placed bilaterally on the participants' shanks (≈1 cm above the APDM units, see Fig. 3.2). These sensors were used to record the timing and type of the mechanical impulses induced by the treadmill

over the course of the experiment. These treadmill IMUs captured data at the sampling frequencies of 50 Hz for two participants (ID1, ID2) and 37 Hz for the rest (ID3-9). In order to synchronize the treadmill and APDM IMUs, three knocks were exerted simultaneously on all sensors prior to stepping onto the running treadmill.

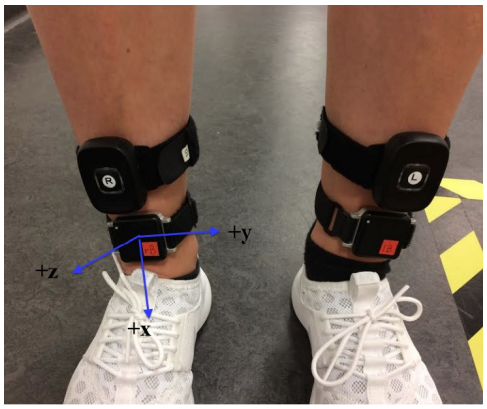


Figure 3.2: Axes and placement of treadmill and APDM IMUs on the shanks. Treadmill sensors were placed above the APDMs and used to annotate data. In addition to these units, 3 additional APDM IMUs were mounted on pelvis, right and left thigh.

While wearing a full-body harness to prevent falls, participants walked on a perturbation treadmill (BalanceTutor, Meditouch) at the constant speed of 1.11 m/s [125]. Left and right perturbations were induced by automatically moving the treadmill surface in ML-direction (12.8 cm and 1.5 $ms^{-}2$). Forward and backward perturbations were induced by acceleration and deceleration of the belt. To present forward perturbations, the belt speed accelerated to 2.5 ms^{-1} and subsequently decelerated to 1.1 ms^{-1}. The backward perturbation was presented by deceleration of the belt speed to 0 ms^{-1} and subsequent acceleration to 1.1 ms^{-1}.

Perturbations were delivered in two separate phases, during single support of the: 1) right, and 2) left leg. For each phase, 80 perturbations were delivered with 20 in each direction. Overall, participants received 160 perturbations resulted in eight different classes of PT-CBRs (see Fig. 3.1).

Treadmill IMU signals were upsampled to 128 Hz after data collection to allow for synchronization. All data were processed using MATLAB (R2016b, MathWorks Inc). The study has received ethics clearance and was reviewed and approved by the Medical Faculty, Tübingen University, Germany (No: 266/2016MP2).

3.1.2 Signal segmentation for feature extraction

This section describes the process and rationale for selecting time windows for feature extraction. Defining the time window of $W_{Label,S,M,a,j}$ corresponding to the jth perturbed or normal gait trial ($Label = \{PT-CBR, PT-NG\}$) from all 30 $Signal_{M,a}$ streams ($S = \{Shank_{Right}, Thigh_{Right}, Pelvis, Shank_{Left}, Thigh_{Left}\}$, $M = \{ACC, Gyro\}$, and $a = \{ML, AP, VT\}$) (ML: mediolateral, AP: antero-posterior, VT: vertical), plays a critical role in the quality of the features to be extracted from, and consequently, the performance of the proposed CBR detection models. Specifically, the segment onset time and width are parameters to be selected.

Based on the findings reported in [129, 126, 130], the peaks observed in signal vector amplitude (SVA) of acceleration at the sternum is a reliable signal-based indicator of CBR onset in response to lateral perturbations (see Appendix A). In the current study, the pelvis (waist) sensor is close to the CoM and was hypothesized to be comparable to the sternum sensor (e.g., considering the findings of [33]) for the task of CBR detection. A custom graphic user interface (GUI) was designed to: 1) synchronize treadmill and APDM IMU signals on different clock and sampling rates, 2) segment and 3) annotate the segments by plotting the $SVA_{Pelvis,ACC}$ signal along with other 30 signals. To identify the onset of the j^{th} PT-CBR, i.e. $ind_{PT-CBR,j}$, the samples close to the peak SVA of the pelvis (i.e. close to $argmax(SVA_{Pelvis,ACC,j})$) were manually selected if the type was medial-lateral or forward. For backward PT-CBRs, a clear plateau in $SVA_{Pelvis,ACC}$ signal was observed, and a point in that plateau was selected to further identify the backward PT-CBR.

The segments attributed to a PT-CBR event needs to be sufficiently wide to encompass important transitional information attributed to the mechanical and postural adjustments evoked after a perturbation. In previous studies, segment widths of 1 s [129], 2.5 s [6], and 5 s [82] were utilized to detect different types of CBRs. Considering older adults require more steps to regain balance following perturbation [67], a relatively large window size may be necessary to capture multiple steps. On the other hand, an overly wide window may suppress prominent changes encoded by the statistical features (e.g., mean, variance). To find the optimal width, published timings of EMG reflex activities during corrective reactions and normal gait in proximal and distal muscles for different types of CBRs were examined in Appendix A. In the current study, the average duration from the onset of a mechanical impulse to treadmill's steady state for all perturbations was ≈4.5 s.

Considering all of the aforementioned criteria, a segment width of 300 samples (≈ 2.34 s) before and after of each recorded perturbation onset were selected to create the j^{th} PT-CBR epoch: $W_{PT-CBR,S,M,a,j} = [ind_{PT-CBR,j} - 300\ ,\ ind_{PT-CBR,j} + 300]$ ($width(W_{PT-CBR})$= 601 samples or ≈ 4.69 s). This window was also hypothesized to be sufficiently wide to

compensate for possible errors in annotations (i.e., wide enough to contain the specified PT-CBR with a high probability). PT-NG epochs were extracted from the 'steady state' intervals between the two consecutive PT-CBRs. The j^{th} steady state interval includes the samples $\in [ind_{PT-CBR,j} + 1000\ ,\ ind_{PT-CBR,j+1} - 500]$. The PT-NGs were then identified based on the $argmax(SVA_{steady-interval,Pelvis,j})$ in that interval and similar to the PT-CBR epochs, the same width of 601 samples was taken into account.

The first set of data (perturbations during right leg stance) of participant number 4 and normal gait epochs from participant number 7 (perturbations during left leg stance), were excluded due to the poor synchronization of treadmill and APDM IMU signals and shorter width of the steady state intervals, respectively. Moreover, the data recorded before the second perturbation were excluded for all participants to include more steady patterns. Overall, 1,335 PT-CBR and 1,222 PT-NG segments were annotated (711 and 624 PT-CBRs were elicited during single support of the left and right legs, respectively). The segments were detrended to compensate for the changes in orientation or placement of the sensor as they may be tilted or shifted slightly during trials.

3.1.3 Feature extraction

Extraction of discriminative features from acquired IMU segments is an important step in our machine learning-based approach for recognition of CBR patterns. Overall, 295 linear acceleration- and angular velocity-based features were extracted from $W_{Label,S,ACC,a,j}$ and $W_{Label,S,Gyro,a,i}$, respectively, forming a $X_{IMU,2557\times295}$ matrix. While filtering the input signals may impact the ability of the subsequent features to detect CBRs[3], assessing the impact of different filtering options requires further investigation and is out of the scope of the present study. In accordance with our previous research works [129, 130], the present study considered raw inertial signals for feature extraction.

Acceleration based features The intensity, range, dynamics, and shape of acceleration signals were measured using the following features extracted for each time window of $W_{Label,S,ACC,a,j}$: 1) maximum peak $max(|W_{Label,S,ACC,a,j}|)$ (absolute value), 2) root mean square (RMS), 3) $mean(|W_{Label,S,ACC,a,j}|)$ (absolute value), 4) variance $var(W_{Label,S,ACC,a,j})$, 5) skewness, 6) kurtosis, 7) the maximum of the acceleration derivative (jerk) $max(|dW_{Label,S,ACC,a,j}/dt|)$ (absolute value), 8) mean jerk $mean(|dW_{Label,S,ACC,a,j}/dt|)$ (absolute value), 9) variance of jerk $var(dW_{Label,S,ACC,a,j}/dt)$, 10) maximum

[3]As an example, in [126] the impacts of different filtering hyperparameters on the performance of sEMG-based CBR detection models were examined.

amplitude of the total acceleration $max(SVA_{Label,S,ACC,j})$, $SVA_{Label,S,ACC,j} = \sqrt[2]{W^2_{Label,S,ACC,AP,j} + W^2_{Label,S,ACC,ML,j} + W^2_{Label,S,ACC,VT,j}}$), 11) Shannon entropy, 12) Segment's integral, 13) amplitude of the dominant frequency (periodogram PSD) and 14) the dominant frequency in the segment ($\pi\omega$).

Angular velocity-based features The following features for each time window segment $W_{Label,S,Gyro,a,j}$ were extracted: 1) maximum peak $max(|W_{Label,S,Gyro,a,j}|)$ (absolute value), 2) $mean(|W_{Label,S,Gyro,a,j}|)$ (absolute value), 3) variance $var(W_{Label,S,Gyro,a,j})$, 4) maximum of the angular velocity derivative $max(|dW_{Label,S,Gyro,a,j}/dt|)$ (absolute value), 5) mean of the angular velocity derivative $mean(|dW_{Label,S,Gyro,a,j}/dt|)$ (absolute value), 6) variance of the derivative $var(dW_{Label,S,Gyro,a,j}/dt)$, 7) maximum of total rotational rate $max(SVA_{Label,S,Gyro,j})$, $SVA_{Label,S,Gyro,j} = \sqrt[2]{W^2_{Label,S,Gyro,AP,j} + W^2_{Label,S,Gyro,ML,j} + W^2_{Label,S,Gyro,VT,j}}$ [111], and 8) Shannon entropy.

3.1.4 Machine learning models

In previous work on a similar classification problem examining multiple machine learning techniques [126][130], RF method (bootstrap-aggregated decision trees) [15] outperformed other classic methods (e.g. support vector machines, neural networks) for the task of CBR detection. Considering RFs permit parallel processing and demonstrates good robustness against nonlinear relationships and imbalanced datasets, this approach was selected for the current study. Based on the initial tests on $X_{IMU,2557\times295}$, the classification error plateaus when the parameters were set to 100 (RF_{100}) grown trees. For consistency, the same parameters were used for all models, although optimal values could be obtained for each model separately. 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).

CBR detection and type identification models

Here, the CBR detection model is a binary classifier which aims to discriminate PT-CBRs from PT-NGs (*Model 1*). *Model 2* further specifies the direction of the PT-CBRs (i.e. 5-class classification) regardless of their timing with respect to the phase of gait (e.g. swing or stance phases of the leading leg). Additionally, considering the possibility of

developing a two-level sequential classifier (following the CBR detection model), *Model 3* was only trained on the PT-CBR data to identify their type (i.e. 8-class classification problem). *Model 4* is a 9-class classifier (i.e., 8 classes of PT-CBRs and PT-NGs) which detects the CBRs during gait and identifies their type. The training dataset for *Model 4* was imbalanced, with more PT-NG samples (majority class) than the minority PT-CBR type classes. Although imbalanced, the data used in this model resemble every-day real-life data in which the frequency of naturally-occurring CBRs is expected to be lower than ambulatory bouts. Under this model, using accuracy as a performance metric can be misleading as performance may be skewed by correctly classifying the majority class even if it wrongly classifies the minority class. Therefore, F_1 $(= 2\sum_{c=1}^{\xi}(\frac{n_c}{N})\frac{Recall_c \times Precision_c}{Recall_c + Precision_c})$, score would be more meaningful to assess the performance of this model, where $Precision_c = \frac{True-Positives}{True-Positives+False-Negatives}$ and $Recall_c = \frac{True-Positives}{True-Positives+False-Positives}$, $\xi = 9$ (number of classes), n_c refers to the number of samples in class $c = \{NG, C_2, ..., C_9\}$, and $N = 2,557$. To compare the performance of model trained using a (smaller) balanced dataset, *Model 5* was trained by randomly selected exemplars of PT-NG and all PT-CBR types. For all of the models, a LOTO cross-validation (i.e. 2557-fold cross-validation: 2,557 times of training using 2,556 trials for training, one trial for testing) was used.

Sensor location evaluation

To identify the individual sensor location with the highest performance for CBR detection (binary classification) problem, 5 different RF_{100} models were implemented. In each procedure, the model was developed based on the 59 features attributed to an individual IMU, i.e. $X_{S,2557\times59}$ (pelvis (*Model 6*), Shank_{Right} (*Model 7*), Thigh_{Right} (*Model 8*), Shank_{Left} (*Model 9*), Thigh_{Left} (*Model 10*), see table 3.2). *Models 11* and *12* were also developed based on the combination of features obtained from both shanks (118 features) and thighs (118 features) senors (see Table 3.2). LOTO cross-validation was taken into account to assess the performance of these models.

Assessment of inter-participant generalizability: leave-one-subject-out (LOSO) cross-validation

To assess the model generalizability across different individuals, a LOSO cross-validation evaluation approach was performed. For each RF_{100} model, the data matrix belonging to one participant is considered as a test dataset, while the data belonging to the remaining participants is used to train the model. The training/testing process was repeated 10 times and the mean classification error is reported. N_{ID} in $295\times N_{ID}$ matrix of D_{Test} (Table 3.2)

denotes the total number of epochs obtained for each participant separately. Results for participant No. 4 are not included due to missing data.

Detection generalizability on IMUFD dataset

To evaluate the generalizability of the Model developed on all PT-CBR and PT-NG data (≈*Model 1*), the IMUFD dataset acquired and published by Aziz et al was taken into account [5]. In the IMUFD dataset, 10 actors simulated 5 types of CBRs (i.e. S-CBRs): 1) slips, 2) trips, 3) incorrect transfer while rising from sitting to standing, 4) misstep while walking, and 5) hit and bump by another person, which were shown to be common CBR incidents by analyzing videos captured in longterm care facilities (3 trials per actor per category, overall 150 S-CBR segments). The simulated ADLs include eight categories (3 trials per actor per category, overall 240 epochs): 1) walking (30 epochs overall), 2) standing quietly, 3) rising from sitting, 4) descending from standing to sitting, 5) descending from standing to lying, 6) picking up an object from the ground, 7) ascending and 8) descending stairs. The 30 walking trials included in the dataset were treated separately (i.e S-NG subset), along with the remaining 210 ADLs (i.e. S-ADL subset).

Since the employed sensor type, locations (pelvis≈waist, [33]) and the sampling rate (f_s =128 Hz) were comparable to the experimental design discussed in the present chapter, similar procedures for signal preprocessing, segmentation and feature extraction were employed using the data of all 5 sensor modalities. Overall, 195 S-ADL, 135 S-CBR, and 30 S-NG segments were included, which satisfied the following criterion: $0 < [argmax(SVA_{Pelvis,ACC}) - 300]$ or $[argmax(SVA_{Pelvis,ACC}) + 300] < width(W_{S-ADL})$. Note that the sequences in IMUFD dataset had an approximate width of 1,900 samples. The S-CBR and/or some of the S-ADL events attributed to one participant was not considered for the analysis due to processing issues at the time of the experiment. The segments were detrended to compensate for the changes in orientation.

3.2 Results

3.2.1 Overall classification accuracy

Confusion matrices and accuracies for all trained models are tabulated in Table 3.1. In general, *Model 1* (CBR detection) demonstrated high overall accuracy (96.60%, LOTO) in distinguishing multidirectional PT-CBRs from PT-NG patterns. Representing 5- and 8-

class problems, *Models 2* and *3* demonstrated diminishing accuracies (88.15% and 74.46%, LOTO). While *Model 4* displayed a satisfactory accuracy (80.64%, LOTO) for the 9-class classification problem (all CBR classes and NG), *Model 5* trained on a more balanced dataset demonstrated 74.59% overall accuracy for the same classification problem.

3.2.2 Sensor location evaluation

Table 3.2 reports the effect of sensor location on CBR detection accuracy (binary classification). A single IMU placed on participants' pelvis outperformed all other sensor locations (94.70% vs $\approx$ 91% mean accuracy) for the purpose of CBR detection. No significant difference was found between the performance of *Model 6* (pelvis sensor only) and multi-sensor models of: 11 (93.78%, right and left shank sensors) and 12 (94.02%, right and left thigh sensors).

3.2.3 Inter-participant generalizability

Results of the LOSO cross-validation are shown for each participant in Table 3.2. Note that the model for participant number 4 was not developed due to an incomplete data set (missing perturbed gait during right leg stance). The models developed for different participants showed various performance (accuracy range: 67.7-97.7%). This resulted in a lower mean LOSO detection accuracy of 83.18% compared with *Model 1* (96.60%).

3.2.4 Detection generalizability evaluation (IMUFD dataset)

To assess different aspects of the framework's generalizability to unseen data Models i-iii were developed (see Table 3.3). *Model (i)*, for which the training and test datasets were completely independent, the sensitivity of 100% was achieved (i.e., all S-CBRs were detected correctly); however, the S-NGs were poorly identified. The differences between the PT-NG and S-NG segments was hypothesized to be the underlying reason (discussed in 3.3).

To address the aforementioned hypothesis, 50% of S-NGs were added to the PT dataset to form the training dataset for *Model ii* (see Table 3.3). This resulted in 100% specificity and sensitivity, supporting the aforementioned hypothesis.

Moreover, *Model iii* was developed to examine the framework's capability to differentiate between CBRs from all other activities. For this model, the sensitivity dropped from 100% to 84.15%. These results are further discussed in section 3.3.

For each RF_{100} model (in *Models i-iii*, Table 3.3), the training procedure was repeated 10 times. MATLAB randperm was used to randomly select a proportion of the IMUFD data (e.g., 50%) that was subsequently added to the training set for *Models ii* and *iii*.

Table 3.1: Confusion matrices for different trained models. To obtain these models, leave-one-trial-out cross-validation (Training data set: 295×2556, Testing dataset: 295×1) and RF_{100} method were used.

Model 1: CBR detection

Accuracy: 96.60% $F_1 score = 0.9660$

	NG	CBRs	Total
NG	0.9714	0.0286	1222
CBRs	0.0390	0.9610	1,335

Model 2: Detection and type identification of specific recovery strategies to perturbations in left, right, forward (trip-like), and backward(slip-like) during gait, 5-class classification.

Accuracy: 88.15%, $F_1 score = 0.8761$

	NG	FWD	Left	Right	Reverse	Total
NG	0.9885	0.0033	0.0033	0.0025	0.0025	1222
FWD	0.0533	0.8669	0.0444	0.0325	0.0030	338
Left	0.1416	0.0324	0.7227	0.0885	0.0147	339
Right	0.2087	0.0654	0.1277	0.5888	0.0093	321
Reverse	0.0267	0	0.0089	0.0178	0.9466	337

Model 3: identification of eight specific recovery strategies to perturbations in left, right, forward (trip-like), and backward(slip-like)

Accuracy: 74.46%, $F_1 score = 0.7441$

		C2	C3	C4	C5	C6	C7	C8	C9	Total
C2	-	0.8291	0.0380	0.0633	0	0.0506	0.0063	0.0063	0.0063	158
C3	-	0.0881	0.6981	0.1069	0	0.0126	0.0755	0.0189	0	159
C4	-	0.1000	0.1067	0.6867	0.0067	0.0067	0.0133	0.0733	0.0067	150
C5	-	0.0064	0.0127	0.0191	0.7197	0	0	0	0.2420	157
C6	-	0.0444	0	0	0	0.8111	0.0722	0.0556	0.0167	180
C7	-	0	0.0722	0.0056	0.0056	0.0722	0.7111	0.1111	0.0222	180
C8	-	0.0058	0.0292	0.0702	0	0.0468	0.1170	0.6959	0.0351	171
C9	-	0	0	0	0.1611	0	0.0167	0.0278	0.7944	180

Model 4: CBR detection + type identification (9-class classification)

Accuracy: 80.64%, F_1 score: 0.7934

	NG	C2	C3	C4	C5	C6	C7	C8	C9	Total
NG	0.9935	0.0016	0.0016	0.0016	0	0.0008	0	0	0.0008	1222
C2	0.1013	0.7595	0	0.0506	0	0.0759	0.0127	0	0	158
C3	0.2013	0.0377	0.6164	0.0566	0	0.0063	0.0629	0.0189	0	159
C4	0.3467	0.0667	0.1133	0.3733	0	0.0133	0.0133	0.0733	0	150
C5	0.0382	0	0.0064	0	0.7006	0	0	0	0.2548	157

C6	0.0611	0.0444	0	0	0	0.8111	0.0389	0.0333	0.0111	180
C7	0.1833	0	0.0944	0	0.0056	0.0667	0.5556	0.0722	0.0222	180
C8	0.2807	0.0117	0.0292	0.0643	0	0.0526	0.0526	0.4854	0.0234	171
C9	0.0167	0	0	0.0056	0.1778	0	0.0167	0.0333	0.7500	180

Model 5: CBR detection + type identification with more balanced dataset (9-class classification), only 200 epochs of NG were randomly selected out of the total 1,222 NGs

Accuracy: 74.59%, $F_1 score = 0.7431$

	NG	C2	C3	C4	C5	C6	C7	C8	C9	Total
NG	0.9300	0.0100	0.0050	0.0250	0.0050	0.0050	0.0050	0.0050	0.0100	200
C2	0.0380	0.7911	0.0127	0.0443	0	0.0949	0	0.0127	0.0063	158
C3	0.0692	0.0440	0.7107	0.0629	0	0	0.0755	0.0377	0	159
C4	0.0733	0.0667	0.1400	0.6467	0.0067	0	0.0067	0.0533	0.0067	150
C5	0.0318	0	0	0.0064	0.6943	0	0	0	0.2675	157
C6	0.0111	0.0611	0	0	0	0.8111	0.0667	0.0389	0.0111	180
C7	0.0889	0	0.0722	0	0.0056	0.0722	0.6333	0.1000	0.0278	180
C8	0.1053	0.0058	0.0351	0.0468	0	0.0585	0.0936	0.6257	0.0292	171
C9	0.0056	0	0	0	0.1333	0	0.0056	0.0333	0.8222	180

Table 3.2: The effect of sensor location and between-participant generalizability evaluation on overall accuracy. Results for binary classification using the RF_{100} method. Note: results for participant No. 4 are not included due to missing data.

	Data	**CBR detection during gait** Binary classification
5 IMUs (LOTO)	295×2557	96.60
Modality location evaluation (LOTO)		
Pelvis (Model 6)	59×2557	Model 6: 94.70
Right Shank (Model 7)	59×2557	91.06
Right Thigh (Model 8)	59×2557	91.04
Left Shank (Model 9)	59×2557	91.00
Left Thigh (Model 10)	59×2557	91.12
Right + Left Shank (Model 11)	118×2557	93.78
Right + Left Thigh (Model 12)	118×2557	94.02
13: LOSO cross-validation: mean accuracy 83.18%		
ID 1	295×311: D_{Test}	95.51
ID 2	295×310: D_{Test}	76.10
ID 3	295×311: D_{Test}	91.42
ID 5	295×312: D_{Test}	67.70
ID 6	295×313: D_{Test}	81.00
ID 7	295×219: D_{Test}	85.27
ID 8	295×314: D_{Test}	70.72
ID 9	295×310: D_{Test}	97.74

Table 3.3: Testing the gernarlizability of the PT-CBR detection model to detect other types of CBRs during activities of daily living (using IMUFD dataset [5]). All models are binary classifiers and the results reflect the average of 10 times of training/testing. In models ii and iii, for each of the 10 steps, a percentage of IMUFD data was randomly added to PT data by performing random permutation.

Training set	**Test set**	**Accuracy**	**Sensitivity**	**Specificity**
Model (i): PT-NGs + PT-CBRs	all S-CBRs + all S-NGs	82.00	100	10.00
Model (ii): PT-NGs + PT-CBRs + 50% of S-NGs	all S-CBRs + 50% of S-NGs	100	100	100
Model (iii): PT-NGs + PT-CBRs + 50% of S-ADLs + 50% of S-NGs	all S-CBRs + 50% of S-ADLs+ 50% of S-NGs	90.97	84.15	99.20

3.3 Conclusion and future work

This chapter discussed the development of a machine learning based framework, to distinguish between multidirectional CBRs and normal walking episodes. In general, testing demonstrated high accuracy (96.60%) in distinguishing CBRs from normal walking patterns (a binary classification approach) and satisfactory accuracy (80.64%, F_1 score =0.79) in distinguishing between multiple classes of CBRs and normal gait (a 9-class classification approach).

Overall, the detection of slip- and trip-like corrective reactions (C2, C5, C6, C9) was more accurate than the detection of CBRs in ML direction (C3, C4, C7, C8). As shown in the confusion matrix for model 2 (Table 3.1), the PT-CBRs elicited in left (C4, C8) or right (C3, C7) directions were either confused with each other or with normal gait episodes. For instance, 8.8% and 14.1% of the left PT-CBRs categorized as right and PT-NGs, respectively. In a more specific analysis, where the phase of gait is considered, 18.33-34.67% of ML PT-CBRs (C3, C4, C7, C8) were confused with PT-NGs, while 1.67-10.13% of AP PT-CBRs (C2, C5, C6, C9) were misclassified as PT-NGs (confusion matrix for *Model 4*, Table 3.1).

Applying a more balanced dataset partially mitigated these errors. For instance, in *Model 4*, C8 showed 28.07% misclassification rate as PT-NG; which dropped to 10.53% in *Model 5*. By visual inspection of IMU signals, ML PT-CBRs were less distinguishable compared to their AP counterparts, and more similar to PT-NG signals. These are in

line with our previously reported findings [130], where SS and CO CBRs were frequently confused with each other. The inferior CBR detection performance in the ML direction in the present study may be also attributable to the lower intensities of the ML perturbations, as well as the possible adaptation gained during the data collection protocol.

Examining the effect of sensor location yielded a clear benefit of using the pelvis sensor (94.7%), which outperformed other sensor locations ($\approx$ 91% for other locations) for the task of CBR detection (this finding informed Chapter 4). Considering the sternum and pelvis are both close to the CoM, this supports previous findings of [130], where the sternum sensor outperformed shank and thigh sensors in CBR detection. Moreover, the CBR detection accuracy attributed to the pelvis sensor was slightly higher than the accuracies obtained from both shanks (93.78% accuracy) and both thighs (94.02% accuracy) (see Table 3.2), while slightly lower compared to the 5-IMU model (*Model 1*: 96.60% accuracy).

Table 3.2 shows that for some participants (e.g. ID2, ID5, ID8), the LOSO accuracies were considerably lower. This can be attributed to the differences between these individuals' characteristics compared to the other 6 participants ($>$ 80% accuracy), which requires further investigation. To improve the average LOSO accuracy, and thus, the generalizability of the CBR detection framework, a larger dataset captured from people with varying intrinsic characteristics (e.g., body mass index, leg length, age, and height) and gait pattern is suggested.

By assessing the generalizability of the binary CBR detection model (trained on PT-CBR data) to an independent/unseen dataset, it was observed that while unseen CBRs in the IMUFD datasets were all detectable, the normal gait segments in the same dataset were not accurately detected resulting in a low specificity. By comparing the results of *Models i* and *ii*, it can be concluded that the inclusion of less constrained walking episodes, e.g., overground, in the training set can improve the overall framework's performance. All PT-NG segments include steady state gait with constant speed. On the other hand, S-NGs episodes in the IMUFD dataset were collected under fewer constraints, and are likely to include a broader range of walking behaviors and patterns (e.g., gait initiation and/or termination). The possibility of differentiating between CBRs from all other ADLs was further tested in *Model iii*, for which the accuracy regressed to 85.9%. This poor result can be attributed to the similarities between the kinematic characteristics of PT-CBRs and S-ADLs, which increased the false alarm generation rate. To improve the robustness of the CBR detection framework against the generation of false alarms, future datasets are recommended to include a broad spectrum of gait patterns as well as other ADLs. The latter is an important consideration in Chapter 4.

A clear limitation of the present chapter is the use of controlled data acquired from

healthy, young adults to develop CBR detection models. Previous research conducted in controlled conditions [67] showed that older adults correct perturbations less accurately, with a higher variability, and more steps to regain balance. Similarly, in another controlled study, older participants slipped longer and faster [102]. Supporting the long-term vision for ambulatory assessment of fall risk by tracking naturally-occurring CBRs in older adults, this study aimed to inform the optimization of sensor location as well as hyperparameters (e.g., segment length); and assess the utility of a perturbation treadmill paradigm in terms of the provision of appropriate and sufficiently large training datasets to develop generalizable and robust CBR detection models. The findings of this chapter form the infrastructure for Chapter 4, where older adults' free-living data were employed.

Chapter 4

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

[1]

Among the studies that met the inclusion criteria for the review process in Chapter 2, only one study explored missteps or compensatory balance reactions (CBRs) [82], 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 Parkinson's disease (PD). The thresholds used in the aforementioned CBR detection approach were mostly determined based on trial and error [82]. Surprisingly, the highest number of suspected missteps was 1,007 within 4,148 gait windows (window length: 5 s, ≈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 rates of false positives was attributed to the presence of high amplitudes in the vertical (VT) acceleration signal and more inconsistent gait patterns compared to

[1]The content of this chapter is mainly obtained from the following unpublished research work (*journal submission and currently under review*): **Nouredanesh, M.**, Ojeda, L., Alexander, N. B., Godfrey, A., Schwenk, S., Melek, W., & Tung, J.. *"Automated Detection of Older Adults' Naturally-Occurring Compensatory Balance Reactions: Translation from Laboratory to Free-living Conditions"*

controlled conditions [82]. As discussed in Chapter 3, the term 'suspected' for this free-lving digital biomarker (FLDB) highlights the lack of criterion (gold) standard data to reliably validate the employed *threshold-based* CBR detection approach in free-living conditions.

Moreover, as discussed in Chapter 3, there have been *machine learning-based* CBR detection methods, developed based on surface electromyography (sEMG) [126, 130] or IMU [129, 6, 125] features, where the IMU-based models presented a more satisfactory performance compared to the sEMG-based ones [130]. The majority of the proposed 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 [134]. 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 precise identification/localization of CBR onsets in the corresponding IMU data, and may provide information on the circumstances leading to false alarms. This information can subsequently be used to assess the performance of the IMU-based CBR detection models.

This chapter proposes one of the first 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 4.0.1.

4.0.1 Key considerations for models' training and validation

Previous studies conducted in controlled conditions have suggested that a single IMU placed on participants' trunk (e.g., sternum [130], waist [6]) or pelvis [125] (discussed in Chapter 3) outperforms all other sensor locations, including ankles and thighs, for the purpose of CBR detection. Moreover, as discussed in Chapter 3, the use of IMUs mounted on 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%) [125]. 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 [59], a CBR detection model dependent on a gait detection algorithm (e.g., [59, 82]) may exhibit a limited performance in some scenarios (please refer to section 4.3 for a more detailed description). Thus, differentiation between CBRs from all other activities of daily living was hypothesized to be a superior approach and considered in the present chapter.

Alternate methods of acquiring and applying training and test datasets were considered in the study design. Previously, hold-out, k-fold, and leave-one-subject-out [59, 125, 130] cross-validation approaches were considered for CBR model development and performance assessment. Similar to our previously published research works [127, 125] and as briefly discussed in Chapter 3, in the present study it was hypothesized that machine learning models trained on a dataset independent of the test dataset would exhibit more realistic results in terms of generalization to unseen data (although lower accuracies are expected to be obtained [125], see Table 3.3 and 3.2.3 in the previous chapter). Moreover, this approach facilitates managing the distribution of samples in the training dataset, which is otherwise very challenging. In addition to their rarity, a major challenge in examining naturally-occurring CBRs is the varying occurrence frequencies for different types of CBRs. For instance, trips were reported to be the most common CBR type in PD fallers [48]. Therefore, compared to the other CBR types, they are potentially easier to be captured in order to form a training dataset. While the investigation of CBRs in sagittal plane has attracted more attention from the researchers (e.g., in [59, 134]), the ability to detect different CBR types, including those in the frontal plane, may provide a more comprehensive insight into older adults' balance impairment [125]. To address this, datasets from multiple sources, including from controlled conditions, can be curated to form a sufficiently large training dataset with adequate samples over different CBR classes. The latter is a necessary step enabling the subsequent machine learning models to learn specific inertial patterns attributed to different CBR types (e.g., crossover, sidestep, slip-like). The findings of Chapter 3 [125] indicated that a perturbation treadmill (PT) is a safe and reproducible option to elicit multidirectional CBRs (PT-CBRs). Additionally, it was hypothesized that the incorporation of PT and over-ground walking data as well as other activities in the training dataset can augment the performance of the subsequent CBR detection models [125] (see 3.3). 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 [5], which includes

young adults' simulated CBR and non-CBR events (simulated activities of daily living), as discussed in Chapter 3,

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) [125] 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 rates of false positives when applied to unseen/free-living data [125, 82]. 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 [125]. 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 4.1.4).

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., egocentric vision, Fig. 4.1). 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 sections in section 4.2. The future directions and clinical implications of the framework are further highlighted in section 4.3.

4.1 Methods and Procedures

4.1.1 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 [33]. 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.

IMUFD

The IMUFD dataset includes 150 CBRs and 240 non-CBR epochs simulated by 10 healthy young participants between 22 and 32 yrs [5]. 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).

PT dataset

In [125], nine healthy young participants (mean age = 26 yrs) wore five IMUs (Opal model, APDM Inc., accelerometers and gyroscopes were set to operating ranges of ±16g and ±2000 deg/s, respectively) at the pelvis, and bilaterally on thighs and ankles. While wearing a full-body harness to prevent falls, participants walked on a perturbation treadmill (BalanceTutor, Meditouch) at the constant speed of 1.11m/s. Perturbations in 4 directions (right, left, backward, forward) were induced during the right or left leg stance phases (in 2 separate 20-minutes 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).

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 [56]. As detailed elsewhere [56], 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 [132]. Overall, 7 CBRs (all trips), with ≈ 10 minutes before and after of each event (overall 140 minutes of data) were taken into account for model validation (see FIVR D1 to D7 in Figures 4.3 to 4.6, D: dataset).

The study reviewed and approved by the University of Michigan Institutional Review Board (HUM00073568).

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 ≈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 participant's age was below 65 yrs, however, as the person was a recurrent faller, the associated data were considered for further analysis. One older adult's data (female, 80 yrs, non-faller) were used for model training (see 4.1.3) and 6 participants' data were considered for models' validation (see 4.1.4, Fig. 4.7 and 4.8). MAGFRA-W D3 and MAGFRA-W D5 include inertial data collected in two participants' homes and neighbourhoods. Two participants (e.g., in MAGFRA-W D6) used walking aids during the data collection.

The project received ethics approval (reference number: 17589) from Northumbria University Research Ethics Committee, Newcastle upon Tyne, UK. All participants gave written informed consent before participating in the study.

4.1.2 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 dataseets.

For each of the simulated CBR and non-CBR trials in the IMUD dataset (with an approximate width of $15s$/trial), each of the 6 acceleration and angular velocity signals was detrended (removing the DC offset) separately. Similarly, for the PT dataset, as the orientation of IMUs did not significantly change over the course of data collection, all six acceleration (ACC) and angular velocity (Gyro) signals corresponding to each set were detrended separately. However, for FIVR and MAGFRA-W datasets, as the orientation of the trunk-mounted IMU during free-living activities changes significantly, instead of detrending the full-length signals acquired 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 ACC and Gyro signals, and the overlapping data were detrended separately. All data were processed using MATLAB (R2019a, MathWorks Inc).

4.1.3 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 4.1.3) and preparation of the training datasets (i.e., feature matrices X) for Models 1 and 2.

Data segmentation

Based on the findings reported in [126, 130, 129] and Chapter 3 [125], 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 Chapter 3 [125], a segment width of ≈ 4.69 ($= 601$ samples at $f_s = 128$ Hz) created by cropping $\approx 2.34s$ (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).

IMUFD signal segmentation After calculation of the SVA signal and detection of $argmax(SVA_{ACC})$ for each trial in IMUFD (section 4.1.1), 227 non-CBR and 148 CBR segments were considered for feature extraction to form the training datasets for models 1 and 2.

PT signal segmentation Considering the possible adaptation happening over the coarse of data collection (80 CBRs/set for each participant as discussed in 4.1.1) [108], 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 ith 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 [125]). These segments were further considered for feature extraction (discussed in 4.1.3) to form the training dataset for Model 2.

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 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 of 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 4.1.3) to form the training dataset for Model 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 [125], 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 (see section 4.3). 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, 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) skewnes of the signal derivative, 17) kurtosis of the signal derivative, 18) RMS of the signal derivative, 19) integral of the signal derivative, and 20) the Shannon entropy of signal derivative. In addition to the aforementioned features, $argmax(SVA_{Gyro})$ 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 [125, 6, 130].

Training procedure

In our previous work on a similar classification problem to detect CBRs, multiple machine learning techniques were examined [126][130][125], where the random forest (RF) method (bootstrap-aggregated decision trees) [15] outperformed other classic approaches (e.g., support vector machines, artificial neural networks). 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 $X_{375\times41}$ matrix), and 2) IMUFD, PT, and MAGFRA-W segments (a $X_{615\times41}$ matrix). Based on the initial tests, an RF model with 19 trees (RF_{19}) showed satisfactory results on all validation datasets, while more number of 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 4.1.4). 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).

4.1.4 Models validation based on free-living data

Validation/test dataset

FIVR dataset Data discussed in 4.1.1 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 Figures 4.3 to Fig. 4.6.

MAGFRA-W dataset By visual inspection of the recorded egocentric vision data in the MAGFRA-W dataset, only 1 naturally-occurring (hit/bump) CBR was identified (see Fig. 4.1 and Fig. 4.6-MAGFRA-W D1, in which the CBR event happened at $\approx t = 631$ s). The participant (older non-faller) hit a light pole and lifted 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 Fig. 4.2).

Data from 5 more participants (Fig. 4.7 and 4.8) were examined to asses the models' robustness across varying contexts during which the models could generate false alarms.

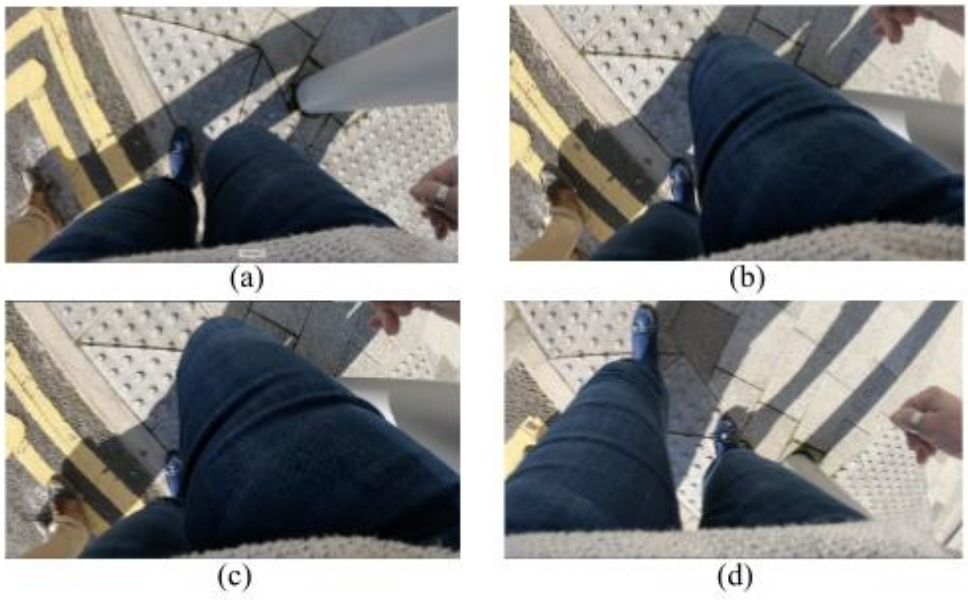


Figure 4.1: One CBR was observed in an older adult's egocentric vision data (MAGFRA-W D1). The participant hit a light pole and lifted right leg forward.

Regions of interest

The validation dataset was segmented similar to the method described in section 4.1.3. 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, $ind_{ROI_j} = argmax(SW_{SVA_{ACC,5s,j}})$, 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.

Possibly-noisy ROIs Preliminary results (Appendix B) indicated that all CBRs were detectable either by Model 1 or 2; with the exception of 1 CBR event depicted in FIVR

D2 (Fig. 4.3). 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. 4.3 to Fig. 4.6). Moreover, high rates of false positives were initially observed in FIVR D2 (Fig. 4.3) and FIVR D6 (Fig. 4.5) as shown in Appendix B.

As mentioned in subsection 4.0.1, 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. [67] reported that older adults corrected perturbations with higher variability, less accurately, regained balance with more steps and 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 [11], 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 [32, 63, 47, 155] as well as discrepancies between older and young adults' gait [83]. Compared to controlled gait, acceleration signals attributed to free-living gait represent lower regularity [63]), more diverse range, e.g., in the antero-posterior (AP) [155] and VT [82, 155] 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 [98], and some of the digital biomarkers extracted from older adults' treadmill and free-living gait data were significantly different [155]. 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, it was hypothesized that by detecting and filtering/smoothing 'possibly noisy' ROIs in free-living data, inconsistencies between the training and validation datasets can be compensated, and subsequently, the overall performance of the CBR detection models can be improved.

As mentioned earlier, gait speed impacts compensatory stepping characteristics [11]. Based on knowledge that walking speed was strongly correlated with range in ACC_{VT} and ACC_{AP} signals [195], and considering that both ACC_{VT} and ACC_{AP} demonstrated significantly higher ranges during free-living gait compared to their in-lab counterparts [155] (with little-to-no difference was reported in mediolateral direction), it was 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 a deeper investigation of controlled and free-living data, here, a 'possibly-noisy' condition is met, if the range in ACC_{AP} or ACC_{VT} in a window (with the length of $\phi = 2.32s$) before or after of the ROI is above a certain threshold (θ_{AP} =8.55m/s^2 and θ_{VT} =11.36 m/s^2). These hyperparameters/thresholds were obtained based on the results reported in [195], 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 (ϕ =2×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_{VT} signals during gait, respectively [195]. 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.

Feature extraction from ROIs For each ROI, either low-pass filtered or unchanged, the 41 features discussed in 4.1.3 were extracted.

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).

4.2 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 4.1). 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 4.1). This model generated 13 false positives yielding the overall specificity of 99.80%.

By comparing the preliminary CBR detection results in Appendix B (obtained for 8 datasets: FIVR D1-D7 and MAGFRA-W D1) with the corresponding results in Table 4.1, 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. 4.3) 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 4.1) due to the

integration of the ROI elimination approach. This resulted in ≈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, older non-faller), who walked on different indoor and outdoor surfaces (e.g., stairs, gravel, grass, transitions), are included in Fig. 4.7. The only context leading to the false positive (shown in Fig. 4.7-upper panel) was a sudden change in walking direction on carpet (indoor environment). Moreover, Fig. 4.2 represents an anticipatory pattern, obstacle avoidance, which could generate false positives according to previous studies' reports [82]. 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. 4.6). There were several peaks with higher SVA_{ACC} amplitudes than the spotted CBRs', e.g., in Fig. 4.4-FIVR D4 and Fig. 4.5-FIVR D5, for which the models did not generate false positives, indicating their robustness against such signal features. Several data points with high amplitudes can be seen in MAGFRA-W D4 and D6 (Fig. 4.8), for which Models 2 and 2' did not generate false positives.

Figure 4.2: Obstacle avoidance, anticipatory reactions with similar patterns to sidestep and crossover CBRs. The detection models (correctly) rejected this event as a CBR.

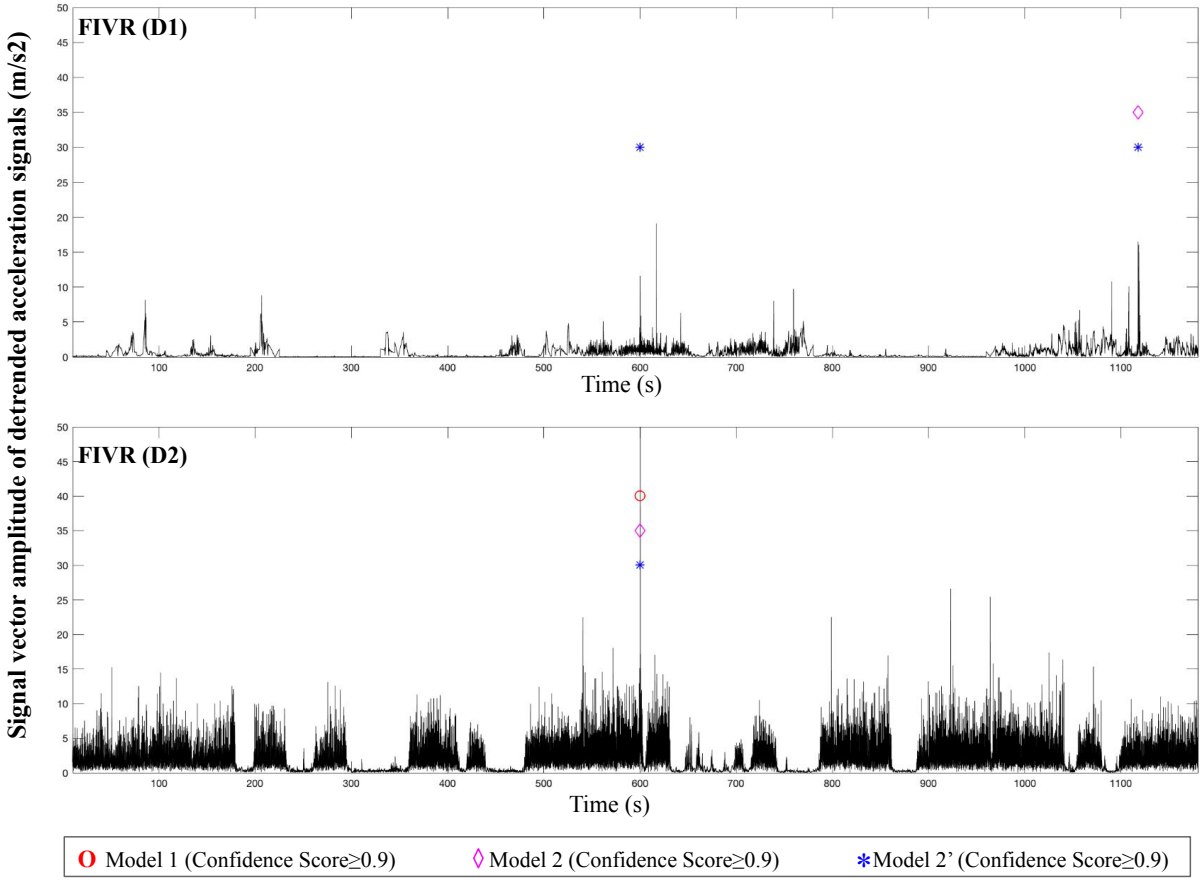


Figure 4.3: CBR detection Models 1, 2, and 2' were applied to the FIVR datasets (the CBR events are located at 600±3 s).

4.3 Discussion

This study presented 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). Considering our previous research [127, 125], we hypothesized that models trained on a dataset independent of the validation/test dataset

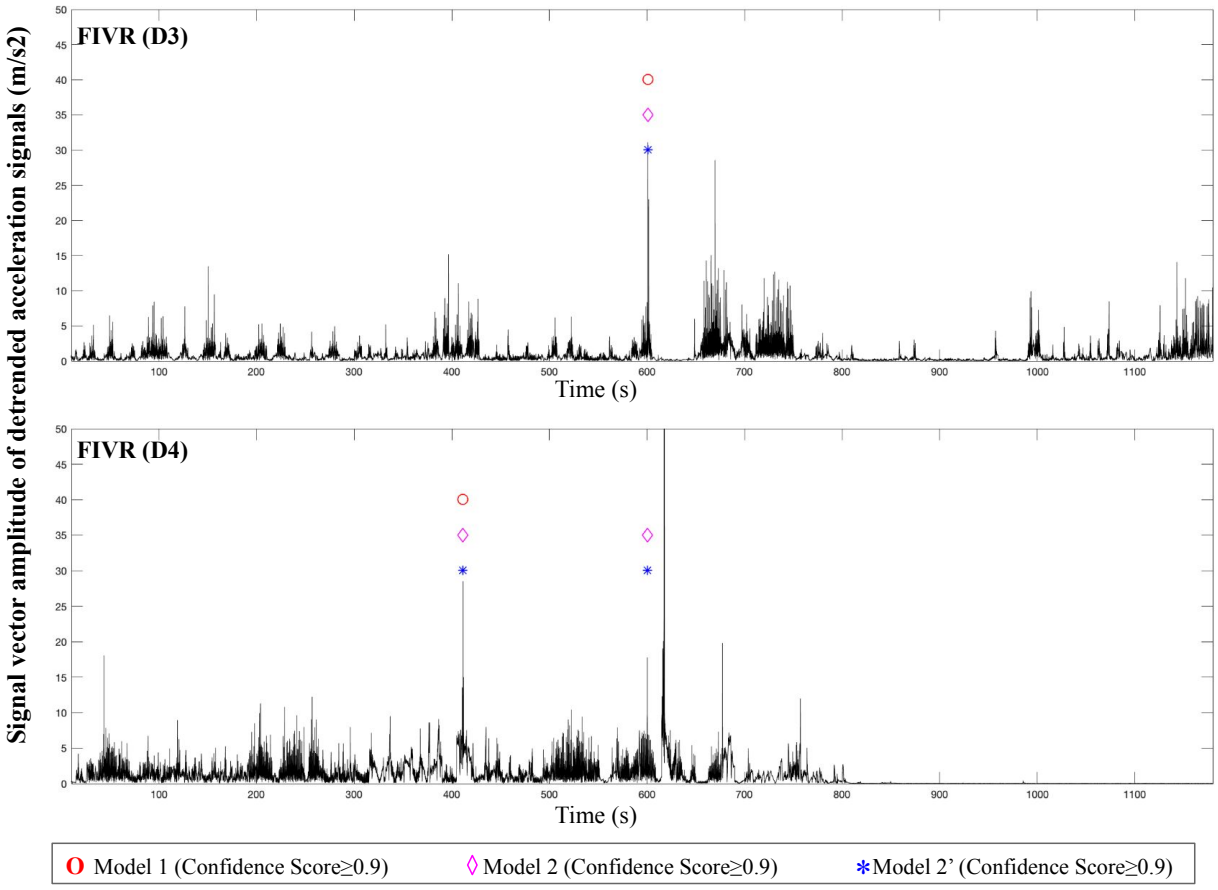


Figure 4.4: CBR detection Models 1, 2, and 2' were applied to the FIVR datasets (the CBR events are located at 600±3 s).

could represent more realistic performance in terms of generalization to complex free-living data captured from individuals with different characteristics. 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-

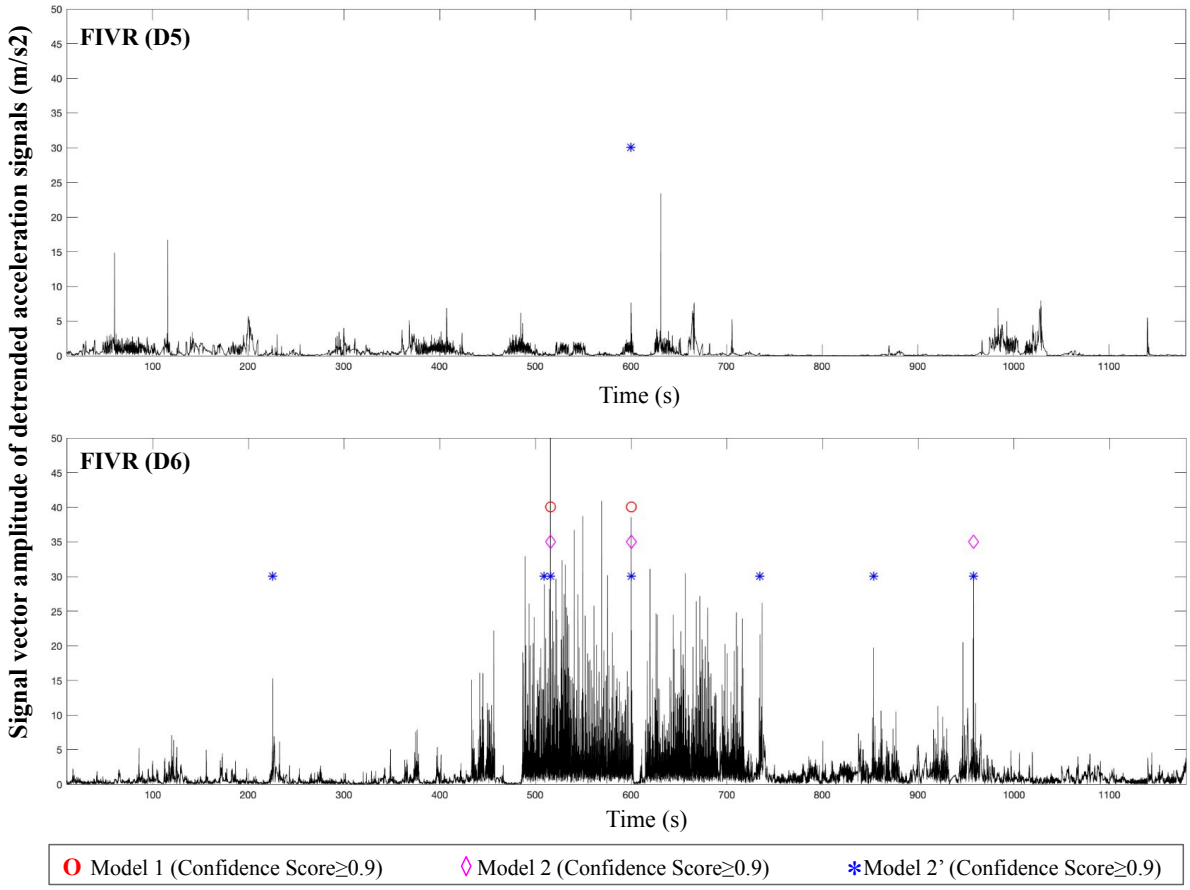


Figure 4.5: CBR detection Models 1, 2, and 2' were applied to three FIVR datasets (the CBR events are located at 600±3 s) and one older adult's data from the MAGFRA-W dataset (the CBR event is located at t=631 s).

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)

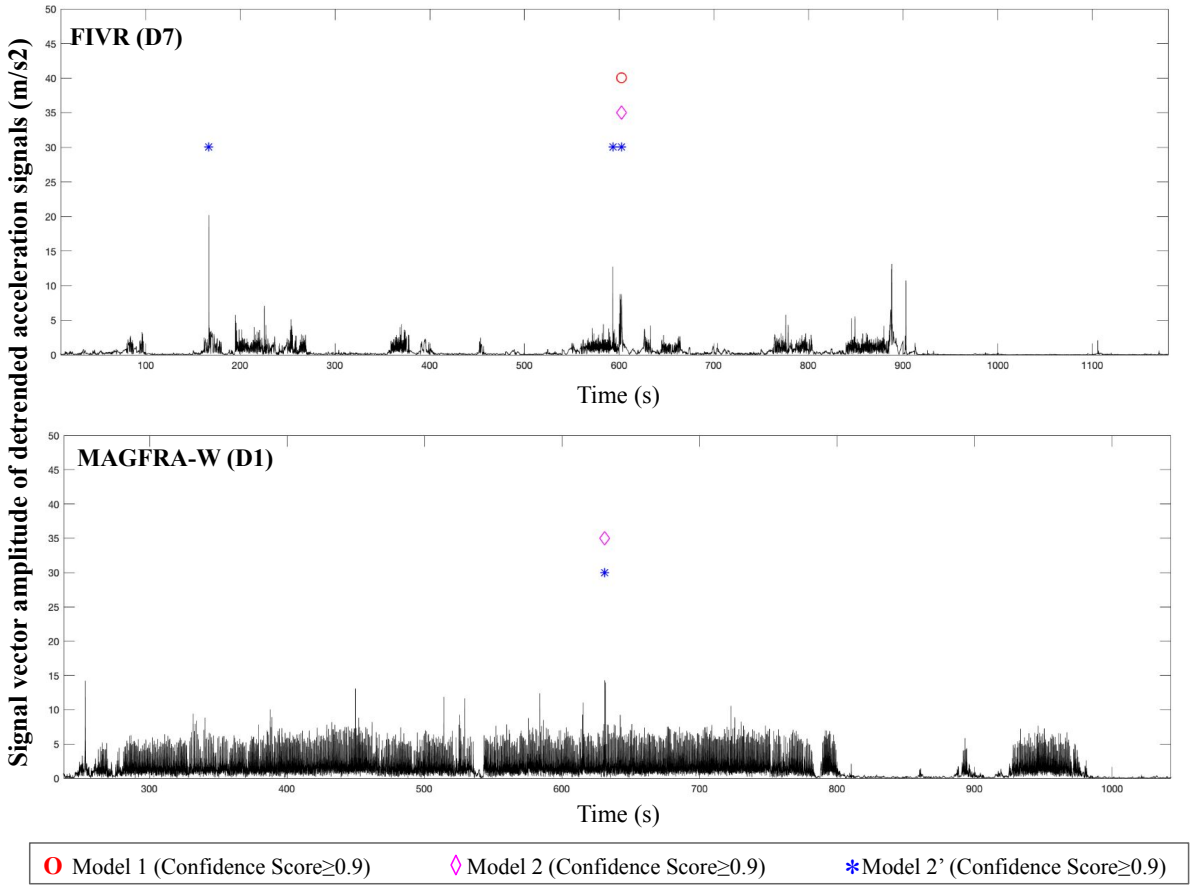


Figure 4.6: CBR detection Models 1, 2, and 2' were applied to three FIVR datasets (the CBR events are located at 600±3 s) and one older adult's data from the MAGFRA-W dataset (the CBR event is located at t=631 s).

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. The simulated CBRs performed by participants in the IMUFD [6] mostly include anticipatory adjustments preceding voluntary movements [125]. 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 of environmen-

Table 4.1: Test results obtained after applying the models on the extracted regions of interest (ROIs) from the validation datasets. FP: false positive, SN: Sensitivity, SP: specificity.

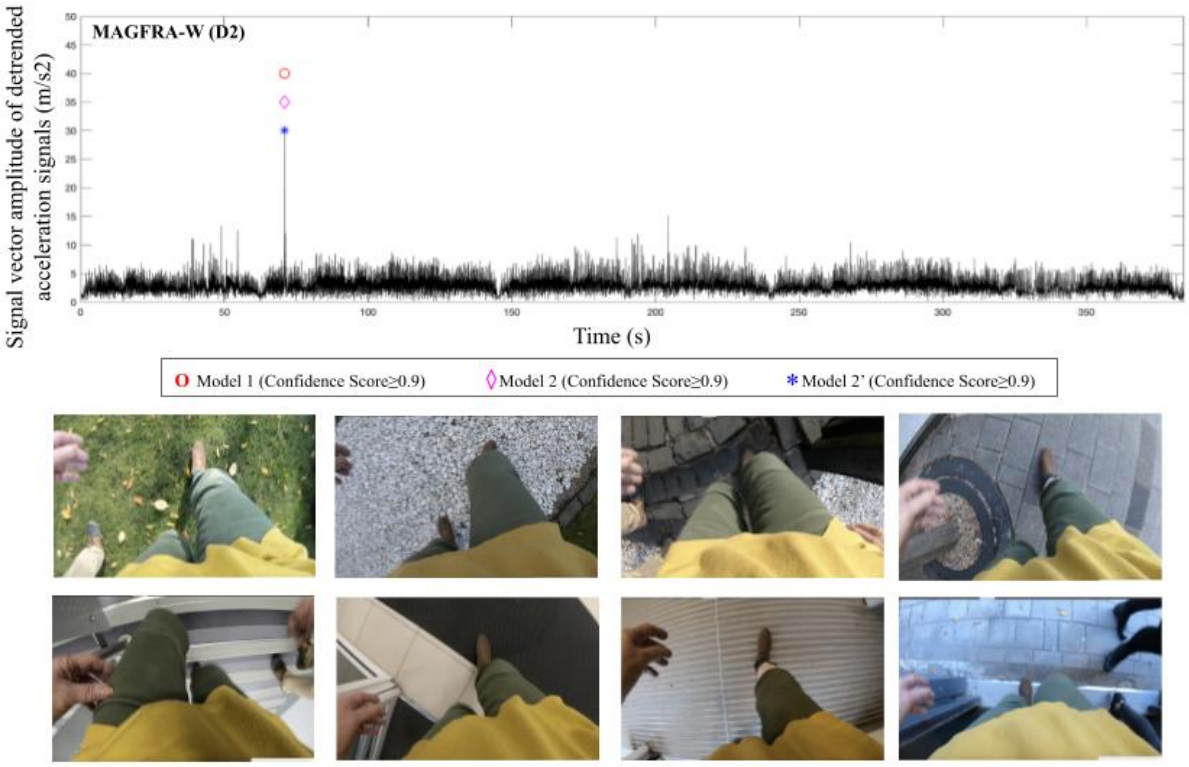
		Model 1			Model 2			Model 2'		
	$\#ROIs$	SN	SP	#FP	SN	SP	#FP	SN	SP	#FP
FIVR D1	190	0.00	100	0	0.00	99.47	1	100	99.47	1
FIVR D2	203	100	100	0	100	100	0	100	100	0
FIVR D3	197	100	100	0	100	100	0	100	100	0
FIVR D4	198	0.00	99.49	1	100	99.49	1	100	99.49	1
FIVR D5	200	0.00	100	0	0.00	100	0	100	100	0
FIVR 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

tal challenges, and are performed automatically with no attention [67, 125]. Considering the reproduciblity 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 deep learning models, which may outperform the random forest models and the engineered features discussed in the present study [125].

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 that 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

Figure 4.7: Multimodal data collected form 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: grass, gravel, ascending stairs, pavement, descending stairs, and transitions between different surfaces.



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 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.

Due to the challenges associated with the collection of naturally-occurring CBRs, only 8 CBRs were verified and investigated in the present study. Further research is required to deeper assess the performance of the proposed framework (machine learning models, possibly-noisy condition, ROI elimination approach) for the detection of naturally-occurring CBRs in larger scale studies. 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 [179], the proposed framework exhibited a satisfactory performance.

No requirement for gait detection Previous work suggested a two-step approach for CBR detection, which require gait detection as the first step [82, 59]. 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 [17], they can be missed/disregarded by commonly used gait detection algorithms [122]. 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 [62], due to reduced gait speed [175] and different variations of gait patterns (e.g., scuffling, dragging of the feet [174]) happening frequently during activities such as household cleaning [62]. 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.

Robustness against sensor orientation misconfigurations Previous research showed that the protocol for free-living data collection, in terms of sensor placement (location or orientation) may not be followed by the participants. For example, in [181], 15.6% of participants who wore accelerometers for seven days did not follow the protocol for at least one day, resulting in prospective miscalculaions of physical activity by more than 20%. Considering the aforementioned point, and since the data in the present study were acquired from multiple studies with different data collection protocols, while detection of noisy context is dependent on the acceleration range in V and AP directions, the consid-

eration of SVA_{ACC} and SVA_{Gyro}, rather than all 6 axes, could generally decrease the risk of misclassification that can possibly be caused by sensor misalignment in the long run.

Clinical applications The clinical applications of the proposed approach are further discussed in Chapter 7.

4.4 Conclusion and future work

The findings of the present chapter suggest that accurate detection of older adults' naturally-occurring CBRs is feasible. Future research will focus on applying the proposed CBR detection framework to large-scale free-living IMU datasets collected from older fallers and non-fallers in a longitudinal manner to examine and understand the associations between falls and CBR-related FLDBs (e.g., direction, duration, number of steps to recover balance).

In the present study, the *contextual information* acquired either by the voice-recorder or wearable camera played a critical role in improving the interpretability of the results. Without detailed information of the mobility context, such as terrain characteristics and obstacles, interpreting the outputs of the proposed CBR detection framework would have remained constrained. For instance, for FIVR-D6, the high rates of false positives was attributed to the irregular inertial patterns induced by walking in a construction site. Moreover, the egocentric vision data captured in MAGFRA-W provided rich contextual information about the factors leading to CBRs (e.g., a light pole, Fig. 4.1) as well as the contexts associated with generating false positives (e.g., a sudden change in walking direction, Fig. 4.7). By identifying contexts associated with verified CBRs, risky features of the environment can be detected and targeted for intervention. Therefore, the development of *automated* egocentric vision-based methods to detect walking surfaces (e.g., irregular gravel-covered) is discussed in Chapter 6. Moreover, Chapter 5 examines the feasibility of detecting CBRs, by tracking feet in the video data acquired by a wearable camera alone.

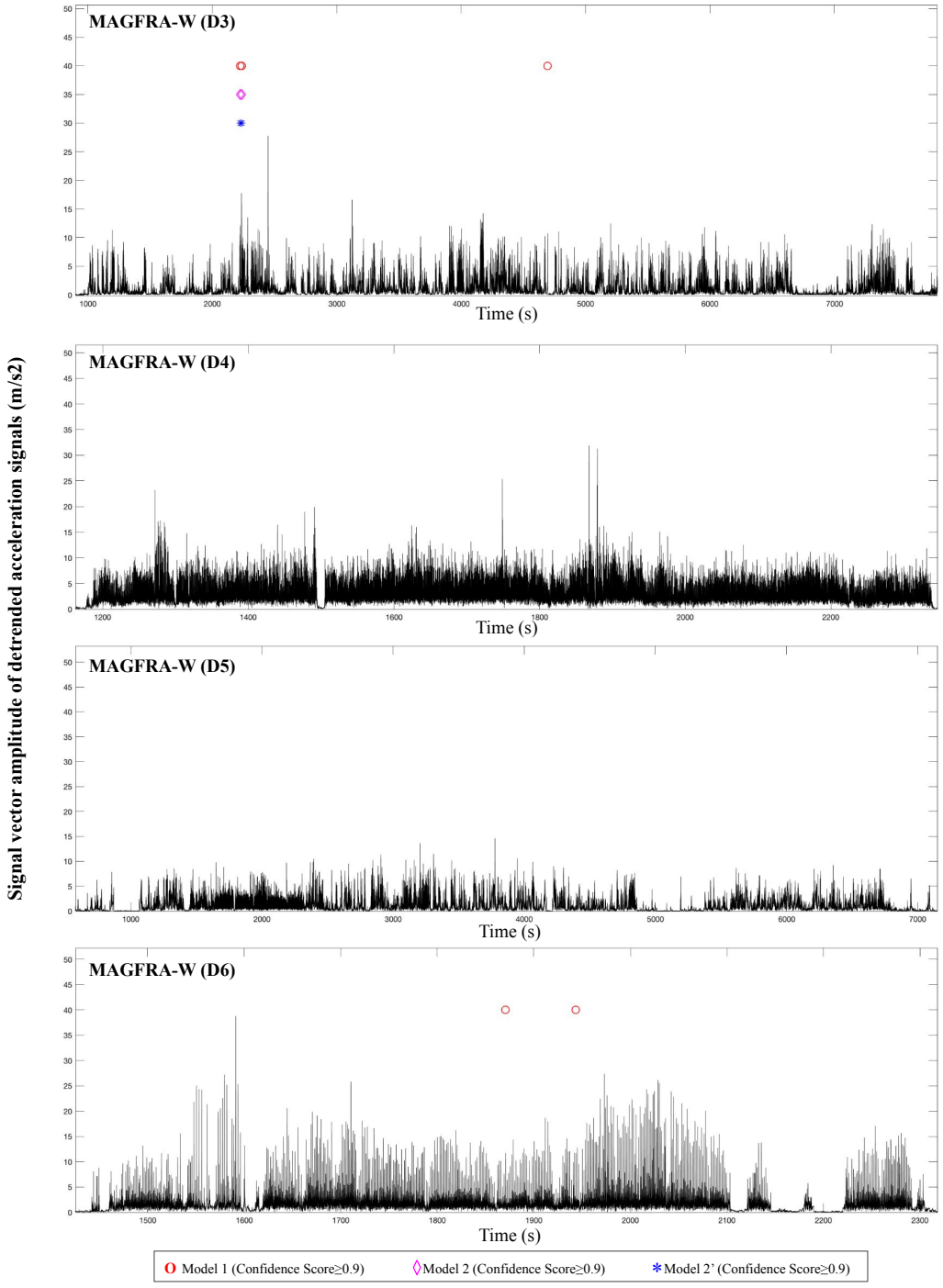


Figure 4.8: CBR detection Models 1, 2, and 2' were applied to 4 older adults' inertial data from MAGFRA-W dataset. The investigation of egocentric vision data confirmed that there was no CBR event in these datasets.

Chapter 5

FootChaser: A proposed egocentric motion-aware gait assessment tool

[1]

Chapter 2 explored free-living fall risk assessment (FRA) approaches that employed wearable IMUs to identify activity bouts and extract gait-related free-living digital biomarkers (FLDBs). Among the explored FLDBs, only 3 types of spatial FLDBs, i.e. step/stride length, step/stride length variability [195, 196, 31], and step length asymmetry [31], were investigated. Although controlled studies have shown that step width variability has a high predictive power for falls in older adults [13], this measure has not been investigated in any of the free-living FRA studies. While body-worn IMUs have demonstrated excellent capabilities to measure temporal gait parameters, a critical drawback associated with the use of IMUs is inaccurate estimation of key spatial parameters, in particular, *step width* [149]. In addition to drift effects, this measurement limitation is largely attributed to a relative lack of motion in the frontal plane during gait, resulting in small IMU excitation and low signal-to-noise ratio.

As discussed in Chapters 2 (2.3.3) and 4, egocentric or first-person vision (FPV) data, acquired via body-worn cameras, can be used as criterion (gold) standard data, outperforming IMUs in terms of the provision of rich contextual information[2]. Bearing in mind

[1] The content of this chapter is mainly obtained from the following published and peer-reviewed research paper: **Nouredanesh, Mina**, Aaron W. Li, Alan Godfrey, Jesse Hoey, and James Tung. "Chasing Feet in the Wild: A Proposed Egocentric Motion-Aware Gait Assessment Tool." In European Conference on Computer Vision, pp. 176-192. Springer, Cham, 2018.

[2] As discussed in Chapter 6 and Appendix C

a waist-worn camera pointed down and ahead of the user, FPV data may also outperform IMUs by offering strong spatial signals (e.g., in the frontal plane) through capturing the wearer's feet locations in the camera's field of view. To address this, previous research examined the utility of a smartphone-based camera mounted on the waist and foot markers to quantify gait characteristics [92]. However, the requirement for placing additional markers on the feet can be considered as a limitation of the aforementioned method, suggesting the development of markerless approaches.

In the present chapter, an automated and markerless framework, i.e., *FootChaser*, is proposed to localize feet in 2D coordinates of video frames captured by a belt-mounted camera. In comparison to head- and chest-mounted camera views, we proposed that a waist-level view would offer the best view for three reasons. First, waist-level FPV data offer a consistent view of the legs and feet even when turning. In contrast, head- or chest-mounted views tend to rotate in anticipation of turns or changes in attention, which reduces the available views of the feet. Second, a waist-level view affords greater resolution of the feet than views higher on the body. Finally, camera egomotion is hypothesized to provide a rich source of temporal information to segment body parts [105].

The *FootChaser* framework (see Fig. 5.3) comprises of two deep models:

1. the *FootRegionProposer*, a ConvNet that proposes regions (or bounding boxes) in RGB frames with high probability of containing feet (based on feet/shoes global appearance), and

2. the *LocomoNet*, a ConvNet sensitive to the periodic gait patterns, which further examines the temporal content in the stacks of optical flow corresponding to the proposed regions by the *FootRegionProposer*. This model is primarily used to filter out false positives generated by the *FootRegionProposer* to further locate the accurate feet regions among the bounding boxes.

Please note that the proposed *FootChaser* framework aims to generate *pixel-wise* foot placement outputs towards the eventual goal of estimating spatial parameters (e.g., step width). The transformation between pixel outputs to distances, likely using 2D metrology approaches, is beyond the scope of the current study and will be examined in subsequent works.

5.1 Related work

While there have been third-person vision (TPV) based research efforts utilizing smartphone or ambient camera video to assess gait (e.g., [144, 46, 26]) and estimate pose (e.g., [53, 23, 75, 200, 42]), the challenges and signal features associated with FPV are distinct. There are several factors that impact the performance and development of the proposed framework: 1) occlusion or extreme illumination conditions, 2) similar objects/terrain patterns to the feet (e.g., other people's feet), and 3) motion blur from fast movements. To inform the chosen methodologies, i.e. camera type and location, this section reviews previous studies that used FPV and addressed the aforementioned challenges.

There are relatively few previous works aiming to extract spatial gait parameters using FPV. An interesting and novel approach was using a walker-mounted depth and/or color camera to estimate 3D pose of lower limbs, mainly in frontal plane [140, 73, 121]. To achieve this, Ng et al. [121] used general appearance model (texture and colour cues) within a Bayesian probabilistic framework. In [73], a Kinect (depth) sensor along with two RGB cameras were placed on a moving walker, and the 3D pose was formulated as a particle filtering problem with a hidden Markov model. The key limitation of these works is the dependency on a stable platform (i.e., walker) to afford consistent views of the lower limbs and monitor pose over time, which is not generalizable to individuals that do not require a walking aid for ambulations.

The possibility of using one or several body-mounted cameras is investigated for 3D full body [171, 87, 206] and upper limb (arms and hands) [160, 119] pose estimation. In [171, 87], outward-looking body-mounted cameras along with optimization approaches were used to estimate 3D body pose. In [171] more than ten cameras were attached to all the person's joints, and structure from motion approach was used to localize the cameras, estimate the joint angles and reconstruct human motion. The main limitation of the proposed method is the obtrusive multi-camera setup and intensive computational load required to infer pose in a video sequence. To alleviate the main weaknesses of [171], Jiang et al. [87] developed a model based on synchronized egocentric videos captured by a chest-mounted camera and a Kinect sensor. The 3D body pose model employs camera egomotion and contextual cues to infer body pose, without direct views of the key body parts (i.e., legs, feet) desired for gait assessment. Moreover, the videos were restricted to relatively static activities (i.e., sitting, standing). Such restrictions and the failure to examine more complex (i.e., dynamic) scenarios limits the applicability is the important limitation of of their approach to the gait assessment problem.

In contrast to the previous studies, [152] and [206] utilized body-related visual cues (outside-in/top-down view) provided by fisheye cameras attached to a bike helmet and

baseball cap, respectively. In [206], a ConvNet for 3D body pose estimation was developed to address limitations in its former version [152], including dependency on 3D actor model initialization and inability to run in real-time. Although the authors compensated for the distortion imposed by the fisheye lens, estimation of the lower body 2D heatmaps (ankles, knees, hip, and toes) was less accurate due to the strong perspective distortion (i.e., a large upper body and small lower body).

The closest approach in spirit to the proposed approach is a hybrid method which combines both global object appearance (spatial network) and motion patterns (temporal network) in a two-stream ConvNets structure. This approach was inspired by Simonyan and Zisserman [173], in which a ConvNet was trained by stacks of optical flow for the task of TPV-based activity recognition. Similar architecture is also employed in FPV-based methods to recognize different activities [105, 176]. To capture long-term sequential information from FPV data, recurrent neural network/long-short term memory (LSTM) was used by Abebe et al. [2, 1] where stacked spectrograms generated over temporal windows from mean grid-optical-flow vectors were used to represent motion [176].

Modeling temporal information in the regions enclosed by bounding boxes in consecutive frames is investigated in previous TPV-based studies [14, 187]. In [85] an object-centric motion compensation scheme was implemented by training CNNs as regressors to estimate the shift of the person from the center of the bounding box. These shifts were further applied to the image stack (a rectified spatiotemporal volume) so that the subject remains centered. More related to the proposed *LocomoNet* approach is the work by Brattoli et al. [14], in which a fully connected network was trained to analyze the grasping behavior of rats over time. Based on optical flow (temporal) data of both initial positives (paw regions) and random negatives cropped from other regions, temporal representation was learned to detect paws.

5.2 The *FootChaser* framework

As an alternative to inferring gait parameters from 3D pose estimates, we hypothesized that tracking the centers of the person's feet in 2D plane of walking over time could provide accurate spatial estimates. The scope of this study is to localize feet by incorporating both temporal and spatial data achieved from FPV data.

Let I_i be the i^{th} frame in a video sequence with the length N, captured by a belt-mounted camera with an outside-in top-down view ($i = \{1, 2 \cdots N\}$). The manually annotated ground truth (GT) data is in the form of bounding boxes $GT_{f,i} = [x_{f,i}^{GT}, y_{f,i}^{GT}, w_{f,i}^{GT}, h_{f,i}^{GT}]$

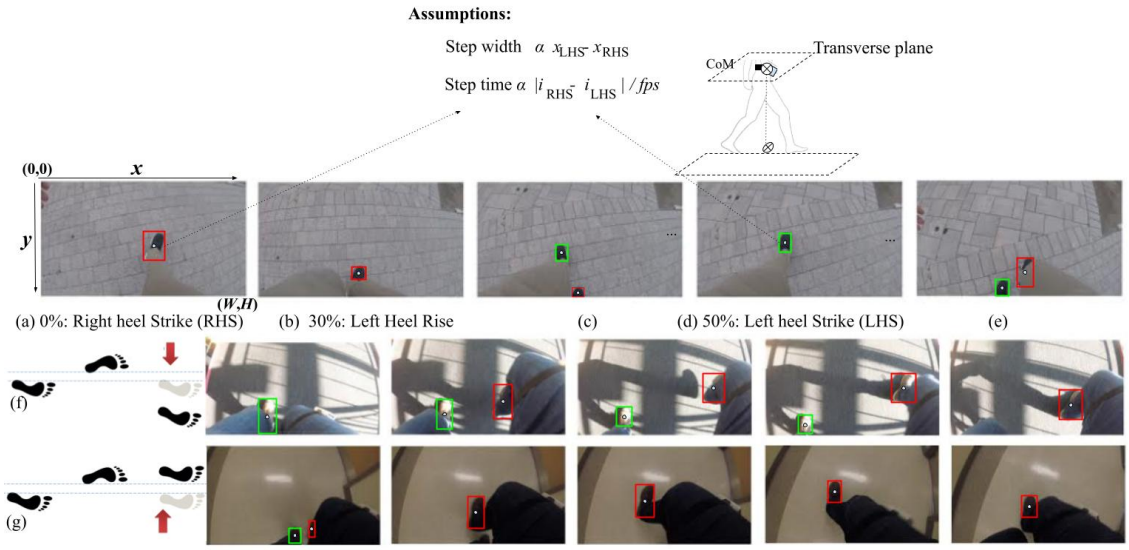


Figure 5.1: Egocentric camera-based gait assessment overview. Panels a,b,c,d,e represent different phases of gait captured by a belt-mounted camera. The x and y location of the right foot (red bounding boxes) and left foot (green boxes) over consecutive frames (CoM: center of mass). Rows f and g depict lateral sidestep and lateral crossover compensatory balance reactions, respectively. These reactions are important behaviours related to fall risk. Note the transformation between pixel-wise box coordinates to distances is not covered in the current study.

indicate the camera wearer's feet ($f = \{left, right\}$) in 2D 1080×1920 coordinate system of each frame (see Fig. 5.1), where x and y denote the center ($C^{GT}_{f,i}$) (Fig. 5.3), and w and h represent the width and height of the bounding box(es), respectively. The goal of the *FootChaser* framework is to detect and localize the center of each foot (if presents in the frame) in the form $P_{f,i} = [x^P_{f,i}, y^P_{f,i}, w^P_{f,i}, h^P_{f,i}]$ during the gait. In an ideal case, the error measure (E) will be minimized for x ($E(x^{GT}_{f,i}, x^P_{f,i})$) and y ($E(y^{GT}_{f,i}, y^P_{f,i})$) trajectories and the underlying area should be the same for the Ps and GTs, i.e, the intersection over union (IoU) measure will be maximized ($IoU = 1$). The predicted x ($\approx$ frontal axis) and y ($\approx$ sagittal axis) trajectories can be used to estimate pixel-wise step width and step length gait parameters, respectively.

To investigate the feasibility of pixel-wise step-by-step gait parameter extraction, the x^{GT}_{left}, x^{GT}_{right} data are plotted in Fig. 5.2. While y^{GT}_{left} and y^{GT}_{right} were examined for measurement of step length, the main focus of this study is on step width estimation. We observed that (a) the trajectories roughly resemble the center of pressure (CoP) data captured by

forceplates, (b) the local maxima and minimuma seem to be correlated with right heel strike (RHSs) and left heel strike (LHSs), respectively (further investigation is required using gold-standard gait analysis methods, e.g., Vicon), and (c) GT data can be divided into frames with one foot ($GT-One$), and both feet ($GT-Two$).

In most of the $GT-Two$ frames, a small portion of the trailing foot is observable (see Fig. 5.1), and is irrelevant for extraction of gait parameters. Considering shape distortions affect detection results, it was hypothesized that the ConvNet is more likely to detect the other foot rather than the less-visible one similar to the findings of Huang et al. [74] and Rozamtsev et al. [161]. In other words, in the frames with two GT, the network often locates the center of the foot that is required for the extraction of gait parameters.

Considering these cues, it was surmised that tracking each foot separately is unnecessary and frames with only one predicted foot center can be used to extract step width. Specifically, (C_i^{P-one}) obtained from the *FootChaser* ($P-One = [x_i^{P-one}, y_i^{P-one}, w_i^{P-one}, h_i^{P-one}]$), regardless of the foot type f. As the key signals for the calculation of spatiotemporal gait parameters (e.g., LHS and RHS points), these can be observed from the x^{P-one} and y^{P-one} trajectories.

To achieve feet localization, the present proposes a two-stage *FootChaser* framework comprised of two ConvNets: 1) *FootRegionProposer* and 2) *LocomoNet*. The *FootRegionProposer* proposes $n \in \mathbb{N}$ bounding boxes as 'proposed foot regions', or $PFR_{j,i}$, $j = \{1,...,n\}$ in the i^{th} frame. As there may be several false positives in the proposed regions, it was hypothesized that the *FootRegionProposer* results may be boosted by applying another ConvNet, called *LocomoNet*, trained to be sensitive to the periodic/specific movement patterns embedded in the user's feet regions during gait. In other words, the *LocomoNet* is expected to filter out false positives by selecting the most confident regions. After applying the *LocomoNet* on $PFR_{j,i}$, only the frames with a single PFR are used for step width estimation (see Fig. 5.2).

5.2.1 FootRegionProposer

The *FootRegionProposer* is a ConvNet fine-tuned to propose PFRs in a frame. The j^{th} proposed region is in the form of a bounding box $PFR_{j,i} = [x_{j,i}, y_{j,i}, w_{j,i}, h_{j,i}]$, where $x_{j,i}$, $y_{j,i}$, $w_{j,i}$, and $h_{j,i}$ denote the center coordinates, and width and height of the box, respectively (see sample PFRs marked by red rectangles in Fig. 5.3). The training procedure for the *LocomoNet* is discussed in subsection 5.3.2. As noted above, there are several factors that may challenge the performance of the *FootRegionProposer*: 1) occlusion or extreme illumination conditions can increase the number of false negatives, 2) objects or terrain

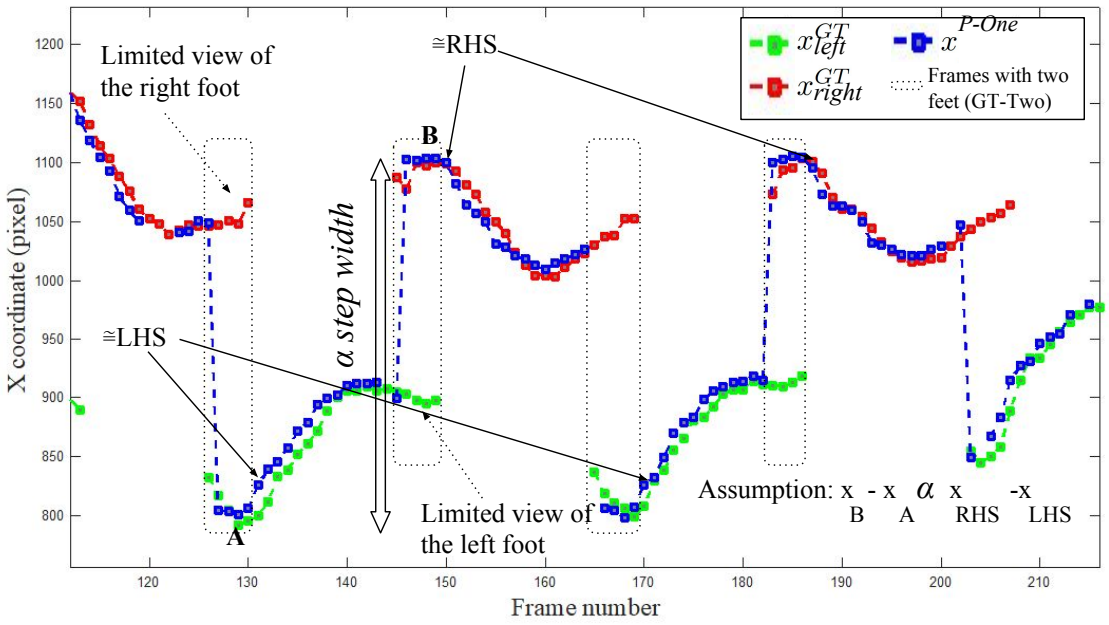


Figure 5.2: Sample bounding box X-coordinate time series data from dataset 2. Ground truth (GT) data for left (green) and right (red) feet, and *FootChaser* predictions with 1 identified region (blue). The expected x location of left heel strike (LHS) and right heel strike (RHS) are marked (further investigation is required using gold-standard gait analysis methods, e.g., Vicon). Periods with 2 identified feet (GT-Two) are indicated by dotted boxes.

similar to the feet (i.e., noise, see Fig. 5.4-c), and 3) motion blur from fast movements. In addition to incorporating a fast and precise object localization/detection ConvNet (e.g., faster R-CNN [151], or YOLO [150]), a second ConvNet was applied to the *FootRegion-Proposer* output to filter false PFRs (subsection 5.2.2).

5.2.2 *LocomoNet*: Learning from gait patterns

To reduce the number of proposed false positives (i.e., false PFRs) by *FootRegionProposer* Network (towards the goal of 'one' true PFR), the dynamic temporal structure of the $PFR_{j,i}$ will be further examined by the proposed *LocomoNet* ConvNet. Inspired by Simonyan and Zisserman's work [173], the present study considered examining optical flow features to deliver bounding boxes with higher confidence of representing feet.

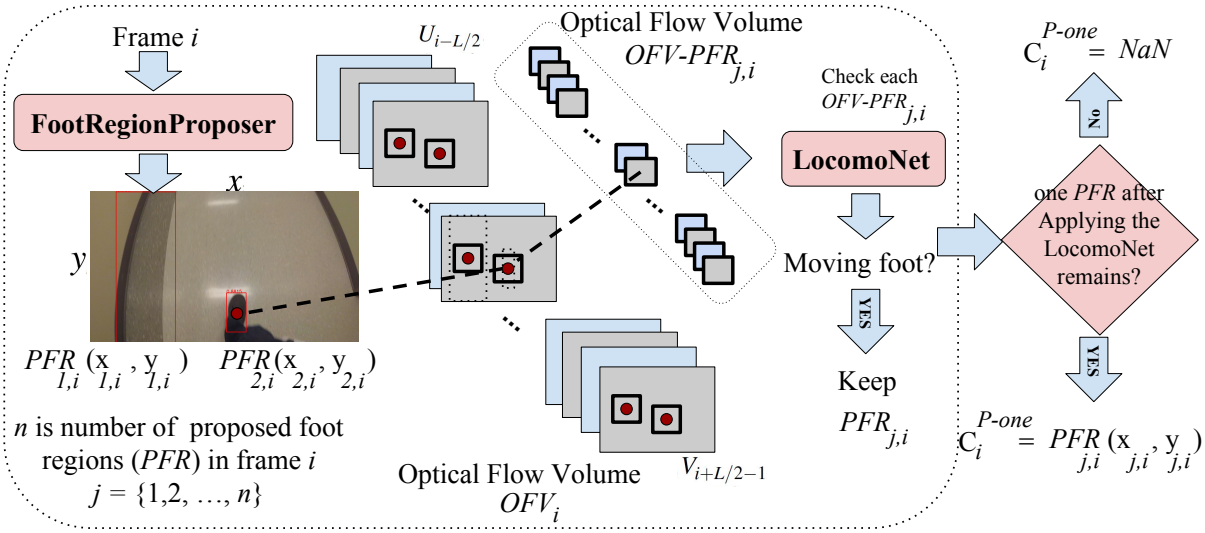


Figure 5.3: The *FootChaser* framework. First, the *FootRegionProposer* proposes $n \in \mathbb{N}$ $PFR_{j,i}$ bounding boxes (red boxes), $j = \{1,2,...,n\}$) in the i^{th} frame. Multiple regions proposed are examined by *LocomoNet* to filter out false positives. After obtaining the stacks of optical flow volume OFV_i (V and U are vertical and horizontal 2D flow components) from the $[i - L/2, i + L/2 - 1]$ frames (L denotes the depth/length of stack), *LocomoNet* inputs are obtained by cropping fixed size regions centered at the center of each $PFR_{j,i}$, i.e., $(x_{j,i}, y_{j,i})$, which creates the optical flow volumes from PFRs ($OFV - PFR_{j.i}$). Final *FootChaser* outputs reflect frames with a single proposed region (C_i^{P-one}).

Figure 5.4: Sample frames reflecting high inter- and intra-class variability in terms of: 1) intense illuminations conditions and shadows (row 1-a,b), 2) different phases of gait, 3) different walking surfaces, e.g., color, texture (each column corresponds to a specific environment and walking surface), and 4) motion blur during crossover and side-step compensatory reactions (row 3-a,b).

The horizontal $U = \{U_1, U_2, ..., U_{N-1}\}$ and vertical optical flow $V = \{V_1, V_2, ..., V_{N-1}\}$ can be calculated separately for each two consecutive frames in the video sequence (the height and width of the U and V components are equal to the frame's 2D dimension, i.e., 1080×1920). Considering a fixed length of L consecutive frames, the optical flow volume $OFV_i = \{U_{i-L/2}, V_{i-L/2}, ..., U_{i+L/2-1}, V_{i+L/2-1}\}$ is obtained for the i^{th} frame. In order to represent the temporal information of $PFR_{j,i}$, a fixed $(W_c \times H_c)$ region centered at $(x_{j,i}, y_{j,i})$ is cropped from OFV_i, which ends up to a $(2L \times W_c \times H_c)$ volume of interest $(OFV - PFR_{j,i})$ corresponding to that proposal (see Fig. 5.3). Each of these volumes are fed into the *LocomoNet* for filtering. The training procedure for *LocomoNet* is discussed in subsection 5.3.3. After applying the *LocomoNet*, if the output frame has only one remaining PFR, the center of that $PFR_{j,i}$ will be saved in the center vector (C_i^{P-One}). Otherwise, the corresponding component will be replaced by NaN and will not be considered in the evaluation.

5.3 Experiments

5.3.1 Dataset

Sufficiently large datasets are challenging to collect, often the primary bottleneck for deep learning. However, there are no publicly available datasets specific to our needs, i.e., large dataset captured by a belt-mounted camera including the images/videos of feet from different people with a considerable diversity in appearance (e.g., shoes with different colors, shape, barefoot, socks) and movement (i.e., gait). To facilitate training, the ConvNet was fine-tuned [137] based on real images with normal optics from large scale datasets, which also was hypothesized to boost the generalizability of the network. Therfore, the ConvNet was fine-tuned on Footwear (footgear) subcategory images ($\approx$ 1300 images with bounding boxes, and 446 images of shoes from top-down view with and without bounding boxes, and we added the bounding boxes manually) from the ImageNet 2011 [163] dataset. Such images resemble more realistic appearance of one's footwear from different views (compared to alternatives such as UT-Zap50K [209]).

Three healthy young adults participated and the FPV data were collected using a GoPro Hero 5 Session camera centered on participants' belt (30 fps, 1080×1920), with no specific calibration and setup. Overall, 5 datasets (including 2 separate datasets from 2 participants in different environments) were captured in five different indoor (tiles, carpet) and outdoor environments (bricks, grass/muddy), resulting in 4505 ($= 5 \times N, N = 901$) total frames (Fig. 5.4 shows samples from the dataset). Frames were annotated by drawing bounding boxes around the right and left shoes (in PASCAL VOC format), using the LabelImg tool [194].

In addition to the normal walking sequences, in two datasets, simulated compensatory balance reactions (CBRs: lateral sidestep, crossover stepping) during gait were also collected (see Fig. 5.4-row 3 columns a,b for sets 1 and 2, and the GT plot for dataset 2 in Fig. 5.6). CBRs (near falls) are reactions to recover stability following a loss of balance (see Fig. 5.1-panels f and g), characterized by rapid step movements (or reaching) to widen the base of support. CBRs also introduce more challenge to our dataset as the corresponding FPV data is usually blurry (i.e., fast foot displacement) (see Fig. 5.4) and the field of view may be occluded.

5.3.2 FootRegionProposer Training

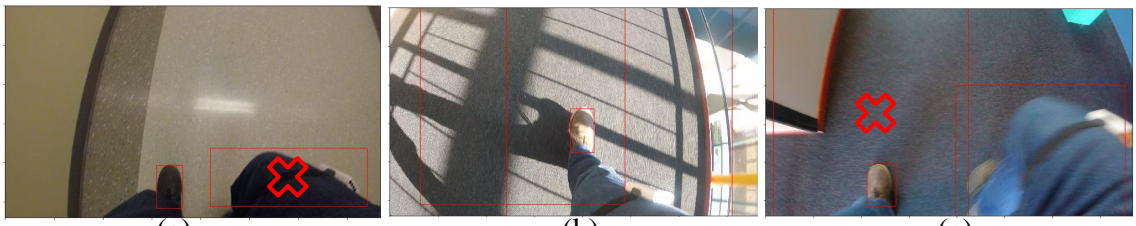
There are several models that can be taken into account for *FootRegionProposer*, including Single Shot MultiBox Detector (SSD) [101], faster R-CNN [151], R-FCN [28]. In [74], it is shown that SSD models typically have (very) poor performance on small objects, e.g. the relatively small feet regions in our experimenrs. Among related approaches, YOLO [150] shows state-of-the-art results in terms of speed and accuracy.

To implement the *FootRegionProposer*, the original YOLO version 2 in [150] was employed. The pre-trained weights on the large-scale ImageNet dataset were used for network initialization, which was then fine-tuned on ImageNet shoe sub-category. The ConvNet was further fine-tuned on images of shoes that are captured in realistic scenes from a top-down view. All of the network inputs were resized to $K \times 3 \times 832 \times 832$, where $K = 64$ was the batch size (mini-batch size: 32). Moreover, the stochastic gradient descent with momentum was used as optimization method, with an initial learning rate of $\gamma = 0.001$, momentum: 0.9, and decay rate of 0.0005 (at steps 100 and 25,000) selected using a Nvidia Titan X GPU. To further address the problem of limited data, the data was augmented (i.e., random crops and rotation) to improve the generalization of the network.

5.3.3 *LocomoNet* training

Although YOLO is very fast, it often suffers from a high number of false positives. The goal of the *LocomoNet* is to improve *FootChaser* performance by reducing the number of false proposals. The *LocomoNet* output maps each OFV to one of the two possible classes. Similar to [176, 105, 199], the TVL1 optical flow algorithm [210] is chosen, here with OpenCV GPU implementation. Moreover, similar to [173, 199, 105], the stack length of $L = 10$ (i.e., 20 input modality channels for *LocomoNet*) is selected, and crop size is set to $W_c = H_c = 224$.

Based on our experiments, a 224×224 region and the stack length of $L = 10$ provided sufficient temporal information for foot regions during gait. Moreover, off-the-frame crops were handled by shifting the 224×224 box in the opposite direction in place of resizing to retain the aspect ratio. To train the *LocomoNet*, 300 positive (shoe/foot regions) volumes were extracted for left and right feet in each of the 5 datasets, resulting in a total of 3000 ($=2 \times 300 \times 5$) true positive volumes. An equal number of negative volumes (i.e. 3000) were also randomly cropped from the non-shoe regions from the consecutive frames, with a constraint of $IoU \approx 0$ with the shoe regions at the i^{th} frame, the past and next frames in the volume were not constrained to allow for a more realistic evaluation.



(a) (b) (c)

Figure 5.5: Example *FootRegionProposer* results (PFRs) for three frames marked by red boxes. Correct foot regions were identified by the *FootRegionProposer*; however, false positives were also proposed. After applying the *LocomoNet*, some false positives were filtered out (marked with (×)). In (a) and (c) false positive(s) are successfully removed, (b) shows a case of intense illumination and shadows challenging *LocomoNet*, resulting two false positives that were not filtered out.

The approach proposed in [199], where the authors demonstrated the possibility of pre-training temporal nets with ImageNet model, was applied in the current study. After extracting optical flow fields and discretizing the fields into [0, 255], the authors averaged the ImageNet model filters of first layer across the channel to account for the difference in input channel number for temporal and spatial nets (20 vs. 3), then copied the average results 20 times as the initialization of temporal nets. Considering such an approach, a motion stream ConvNet (ResNet-101 [61] architecture) pre-trained on video information in UCF101 dataset was used, with stochastic gradient descent and cross entropy loss. Batch size, initial learning rate, and momentum were set to $K = 64$, 0.01, and 0.9, respectively.

5.4 Results

1) Model generalizability. To evaluate the extent to which *subject-related movement patterns* in *different environments* can be handled by *LocomoNet*, a leave-one-dataset-out (LODO) cross-validation was performed. To achieve this, a $LocomoNet_{N_D}$ ($N_D = \{1, 2,...5\}$) model was trained using the whole dataset except N_D dataset (i.e., 4800 volumes for training) and tested on the dataset N_D (i.e., 1200 volumes for testing), and repeated 5 times. The following LODO accuracies were obtained for our 5 datasets: 1: 92.41%, 2: 91.16%, 3: 98.33%, 4: 83.83%, and 5: 96.25%. The high accuracies indicate the generalizability of *LocomoNet* to discriminate foot-related $OFV-PFR$ in unseen datasets. The following average IoU scores were obtained for each set: 1: 0.7626, 2: 0.7304, 3: 0.3794, 4: 0.7155, and 5: 0.5235. Considering an IoU threshold of 0.5 is typically used in object detection evaluation to determine whether detection is positive (IoU of true positive$>$ 0.5)

Table 5.1: Number of proposed foot regions (N_{PFR,N_D}) and elimination rate (ER) in different intersection-over-union (IoU) intervals indicating *LocomoNet* ability to remove false positives by dataset. N_{PFR,N_D} dramatically reduced after applying the *LocomoNet*. ER_T is the weighted average of elimination rate, $IoU > 0.5$ and < 0.5, representing the true and false positives, respectively [44].)

	IoU									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$N_{PFR,1}$	1219	36	7	4	11	22	114	218	312	110
$N_{PFR,2}$	654	10	2	3	10	26	122	282	277	76
$N_{PFR,3}$	781	0	4	12	13	35	89	156	116	15
$N_{PFR,4}$	1225	2	2	1	6	31	119	293	294	36
$N_{PFR,5}$	229	18	17	27	55	106	188	195	83	10
$N_{PFR,T-}$	4108	66	32	47	95	220	632	1144	1082	247
ER1	73.83	55.55	42.85	0.00	0.00	4.54	4.38	8.25	7.05	1.81
ER2	92.20	100.00	0.00	0.00	10	11.53	13.11	17.37	13.35	10.52
ER3	97.18	100.00	0.00	8.33	7.69	5.71	0.00	1.28	3.44	6.66
ER4	83.91	50.00	100	100.00	16.66	35.48	31.93	27.30	26.87	19.44
ER5	83.40	77.77	0.00	0.00	0.00	0.00	3.72	4.61	8.43	20.00
ER_T	83.25	68.18	15.62	2.14	3.15	7.72	9.82	13.81	13.77	8.09

[44], we interpret that the generalizability of the model except for $N_D = 3$, is satisfactory. The lower performance of the network on dataset 3 was attributed to the patterns of walking surface (tiles with different sizes, see Fig. 5.4-c).

2) The number of proposed regions with $IoU < 0.2$ (false positives) dramatically reduced after applying the *LocomoNet* on PFRs. To assess the false positive removal performance of the $LocomoNet_{N_D}$, in this study, elimination rate metric was defined as $ER_{N_D} = \frac{\text{Number of filtered PFRs in a specific IoU interval}}{\text{Total number of PFRs in a specific IoU interval}} \times 100$, ($IoU$=$Area(GT \cap P)/Area(GT \cup P)$). As shown in Table 5.1, the PFRs in a low IoU score range ($\in [0, 0.2)$), representing false positives, were removed with a high rate (e.g., in $IoU_{[0,0.1)}$ with 83.25% reduction). The relatively low true positive removal score (i.e., in $IoU_{[0.9,1)}$ with 8.09% reduction) reflects satisfactory performance of *LocomoNet* in retaining the true positives (refer to Fig. 5.2 for some failure and success cases).

3) *FootChaser* prediction trajectories closely match ground truth trajectories. The performance of the *FootChaser* in tracing the GT data can be assessed by measuring 1) the individual IoU scores, and 2) the pixel-wise distance (error, E) between the predicted foot center and its corresponding point in GT data, i.e. as discussed in section 5.2, by comparing the predicted $P-One$ bounding boxes with $GT-one$

Table 5.2: Mean absolute error (MAE) results for the $GT - One$ region in absolute pixels and as a fraction of image resolution. MAE $= 1/N \sum |GT - One_{a,f,i} - P - One_{a,i}|$, where $a = \{x, y\}$, $f = \{left, right\}$, $N = length(GT - One)$. MAE/R as a fraction of image resolution (R), where: R_x=1920, R_y = 1080.

	MAE (pixel)				MAE/R			
Dataset	x_{Left}	x_{Right}	y_{Left}	y_{Right}	x_{Left}	x_{Right}	y_{Left}	y_{Right}
D_1	41.68	87.50	55.66	54.81	0.021	0.045	0.051	0.050
D_2	32.90	44.00	54.29	55.94	0.017	0.022	0.050	0.051
D_3	125.74	194.85	75.19	154.46	0.065	0.101	0.069	0.143
D_4	64.40	62.57	76.11	74.11	0.059	0.070	0.057	0.068
D_5	99.31	37.68	101.52	92.04	0.051	0.019	0.094	0.085

Table 5.3: Mean absolute error (MAE) for $GT - Two$ regions in absolute pixels and as a fraction of resolution (MAE/R), where (R:) R_x=1920, R_y = 1080..

	MAE (pixel)		MAE/R	
Dataset	x	y	x	y
D_1	58.11	84.00	0.030	0.077
D_2	36.12	80.44	0.018	0.074
D_3	121.47	117.78	0.063	0.109
D_4	103.55	94.90	0.053	0.087
D_5	25.28	101.52	0.013	0.094

$(E(a^{P-One}, a^{GT-One}), a = \{x, y\})$, where mean absolute error (MAE) is taken into account as the error metric E (see Table 5.2). For $GT - Two$ (e.g., the black dotted parts in Fig. 5.2), the performance was evaluated by comparing the a_i^{P-One} with the nearest GT point regardless of the foot type (Table 5.3 displays the results). At first glance, this may appear to be a weak metric. However, as discussed in section 5.2 and depicted in Fig. 5.6 and 5.2, in $GT-Two$ data the *FootChaser* is biased toward proposing regions corresponding to the nearly-full-view feet (rather than partially-observable ones). In this application, the observed bias to larger objects is a strength as it predicts the center of the foot required for the extraction of spatiotemporal gait parameters. This can be attributed to the fact that the *FootRegionProposer* is trained on ImageNet dataset that mainly includes the full-view images of feet. Moreover, this is in line with the findings of [161, 74], where a higher performance was reported for the detection of bigger objects in videos. Considering these points, the error criteria for $GT - Two$ regions seem to be a satisfactory representation of performance.

In addition to the relatively low error rates ($< 10\%$ for the x trajectories), as presented in Fig. 5.6, the framework also predicted many of the points at the timings of CBRs (spikes). Therefore, these trajectories can be a promising avenue for the detection of CBRs. High E values for D_3 (Tables 5.2 and 5.3) also support the low IoU rate achieved for that dataset (due to the patterns of the walking surface).

5.5 Conclusion and future work

As the main contribution, this study investigated the feasibility of incorporating a body-mounted camera data to develop automated markerless models, towards assessing gait in natural environments. The *FootChaser* prediction trajectories closely match the ground truth trajectories in x and y directions, suggesting that the pixel-wise estimation of step width as well as the detection of abnormal events such as CBRs using FPV data alone are feasible. Although the usability of the proposed framework requires a deeper investigation, the findings of the present study advance our long-term objective to extract meaningful and stable FLDBs, to complement existing IMU-based methods.

In order to validate the performance of the proposed *FootChaser* framework for the estimation of step width in older adults, Muldimodal Gait and Fall Risk Assessment in Clinic (MAGFRA-C) dataset [124] was prepared (not presented in this thesis). This dataset includes criterion (gold) standard measures recorded by Vicon motion capture system. Older adults walked over a treadmill while wearing a waist-mounted egocentric camera (GoPro), and multiple IMU devices (AX6, Axivity, Newcastle upon-Tyne, UK), affixed

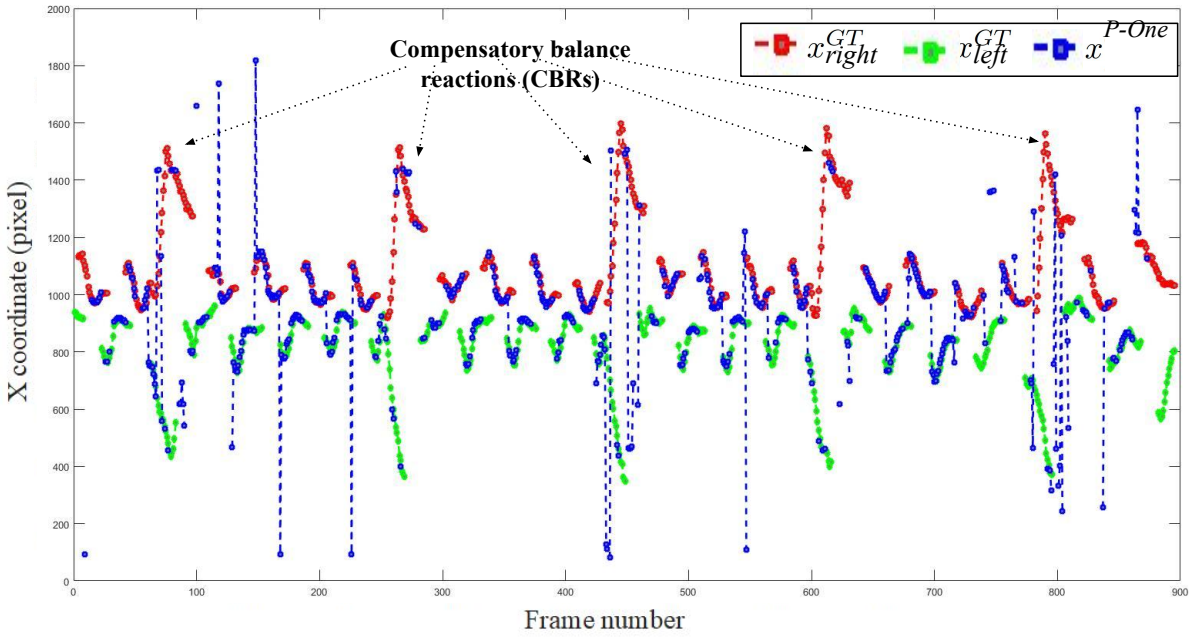


Figure 5.6: Time series plot of X coordinate center of the most confident proposed foot regions (PFR, blue) predicted by the *FootChaser* framework for dataset 2. Ground truth (GT) for the left and right feet are plotted in green and red, respectively. Spikes represent compensatory balance reactions (CBRs) performed by the participant.

to both wrists, lower-back, both legs, head, pelvis, and the camera). Vicon markers were placed on the feet (center) and on IMUs. Outside of the scope of the current thesis, future research will compare the pixel-wise outputs of FootChaser with the x and y coordinates of feet centers recorded by the Vicon system. Therefore, using these data, another model can be developed to directly convert the pixel-wise results of the *FootChaser* into the commonly-used distance units (e.g., m or cm).

Given massive amounts of unlabeled FPV data that aim to be collected during longer-term studies, approaches need to be developed that can robustly handle significant diversity in movement patterns (e.g., rhythm, speed), different populations (e.g., older fallers), and varying clothing and footwear appearance. To address these aspects, similar to [23], we aim to *personalize* both of the *FootRegionProposer* and *LocomoNet* ConvNets and introduce an adaptive pipeline (i.e., *AdaFootChaser*) in our future work.

Chapter 6

Egocentric vision-based detection of surfaces: Towards context-aware free-living digital biomarkers for gait and fall risk assessment

[1]

As discussed in Chapters 1 and 2, although IMU-based free-living digital biomarkers (FLDBs) can be impacted by both intrinsic and environmental features [193, 32, 202], their respective impacts on FLDBs' fall predictive powers have not been differentiated [122]. For instance, higher variability in acceleration signal (measured by the amplitude of the dominant frequency in the mediolateral direction, as a FLDB) during gait could indicate appropriate adaptation to the environment [201] (and potentially a lower risk of falls) and/or exhibit gait impairment (and potentially a higher risk of falls) [80]. Similarly, frequent missteps detected in free-living IMU data can be an indicator of impaired dynamic balance control (and a higher risk for falls [82]) and/or false alarms generated by anticipatory locomotion adjustment while walking on an irregular terrain (e.g., construction site as discussed in the previous chapter, see section 4.4). This ambiguity in interpretation leads to less precise intervention strategies to prevent falls.

[1] The content of this chapter is mainly obtained from the following unpublished research work (*journal submission and currently in the review process*): **Nouredanesh, M.**, Godfrey, A., Powell, D., & Tung, J. "Egocentric vision-based detection of surfaces: Towards context-aware free-living digital biomarkers for gait and fall risk assessment"

A *context-aware* free-living FRA would elucidate the interplay between intrinsic and environmental risk factors and clarifies their respective impacts on fall predictive powers of FLDBs. This would subsequently enable clinicians to target *more specific* intervention strategies including environmental modification (e.g., securing carpets and eliminating tripping hazards) and/or rehabilitation interventions (e.g., training to negotiate stairs and transitions). Ideally, a context-aware free-living FRA method would be capable of examining the relationships between the frequency of falls, FLDBs, and different environmental fall-related features such as presence of dynamic obstacles (e.g., pedestrians, pets), unstable furniture, lighting condition, and terrain types. As a step towards this longer-term goal, the focus of the present chapter is to develop an automated method to differentiate between different walking surfaces commonly observed in everyday environments.

6.1 Related research

A wrist-mounted voice recorder was previously utilized to capture contextual information following misstep events (trips) [56], which could be limited to observations made by the user and may lack spatial and temporal resolution. To objectively identify terrain types, several studies examined the feasibility of using wearable IMU data recorded during gait [58, 71, 72]. For instance, machine learning models achieved 89% accuracy (10-fold cross-validation) to detect six different terrains including soil and concrete using two IMUs in [58]. These studies investigated datasets mostly sampled from young participants in controlled conditions (i.e., walking repetitively over a few surface types with constant properties), and primarily reported machine learning models' holdout or k-fold cross-validation measures. However, cross-validation approaches such as leave-one-subject-out (LOSO) or models' assessment using independent test and training datasets represent a more reliable picture of models' robustness against inter-participant differences and generalizability to unseen data [125, 127]. Appendix C reports the drastic difference between the k-fold and LOSO results of machine learning models implemented using the same IMU data (open access dataset [104]) to differentiate between the walking patterns over stairs, gravel, grass, and flat/even surfaces.

Egocentric or first person vision (FPV) data recorded by wearable cameras affords the ability to provide rich contextual information more readily than IMU-based data alone [127]. In [185], seven days of data were collected from fallers and controls during daily activities using ankle-mounted IMUs and a neck-mounted camera. Subsequently, the frames attributed to walking bouts were investigated and annotated manually. The most frequent terrain type manually identified for all participants were outdoors on pavement, indoors

on carpet and polished or hardwood flooring. Other terrain observations included grass, gravel, and multiple environments. However, the manual identification of walking surfaces, especially in large-scale free-living studies, is a laborious and inefficient process. To advance the field of free-living FRA and gait assessment, there exists a need to develop automated vision-based methods for terrain type specification.

Automated vision-based methods for terrain type identification have been investigated in other fields of assistive technology and robotics. For instance, in [3] head-mounted camera data were used for adaptive control of legged (humanoid) robot's posture and dynamic stability on different terrains. Engineered features such as intensity level distribution, complex wavelet transform, and local binary pattern were extracted and a support vector machine model was developed to categorize 1,000 training images to three classes: (a) hard (e.g., tarmac, bricks, rough metal); (b) soft (e.g., grass, soil, gravels, snow, mud); and (c) unwalkable (static and moving obstructions). Although useful, this approach may not provide sufficient descriptive information to inform FRA. For instance, while snow, gravel and grass were considered into the same class, they would be expected to induce different patterns of gait. A relatively high accuracy of 82% was achieved when the model was applied to a 40-second video. However, this approach's high computational cost was considered a limitation. Elsewhere, to control a powered prosthetic leg, a camera and IMU were mounted on the prosthetic and the relationship between image sharpness and acceleration was considered to trigger the camera [36]. Twenty minutes of data were collected from 5 able-bodied participants walking over 6 different types of terrain (asphalt, carpet, cobblestone, grass, mulch, and tile). Using a bag of word approach (SURF), an average classification accuracy of 86% was achieved based on 5-fold cross-validation. Deep learning approaches have shown strong potential to outperform engineered and bag-of-word-based approaches from many aspects, particularly inference time and accuracy [169, 135]. By integrating both order-less texture details and local spatial information, a Deep Encoding Pooling Network model was developed [207]. The model was trained on the images in Ground Terrain in Outdoor Scenes (GTOS) dataset [208], and tested on GTOS-mobile dataset. The former contains 30,000 images across 40 outdoor terrain classes captured by a camera mounted on a mobile exploration robot with a fixed distance between the camera and the ground. GTOS-mobile data was captured by a mobile phone and with more flexible viewpoint, still relatively close to the ground. Although promising results were achieved, due to low intra-class diversity, limited viewpoint, and restriction to outdoor terrains, the GTOS(-mobile) models may not be generalizable to address the problem of terrain identification in complex everyday environments. More relevant to the context of FRA, data of a chest-mounted camera and Gabor Barcodes [128] were used to automatically detect 17 environmental fall-related features such as slope changes (e.g., ramps) and surfaces (e.g.,

gravel, grass). Although high (88.5%) accuracy was achieved, the incorporated dataset was restricted to young adults, limited to public environments lacking at-home data. Moreover, only k-fold cross-validation results were reported.

To address the previous research works' limitations, this paper employs a unique dataset, i.e., Multimodal Ambulatory Gait and Fall Risk Assessment in the Wild (MAGFRA-W), collected from older non-fallers and fallers in out-of-lab conditions and presents a vision-based deep framework to classify level walking surfaces (see Fig. 6.1). To *maximize* the framework's generalizability and *minimize* its dependence on sample size, an independent training dataset with *high* intra-class variance was formed by curating data from relevant datasets, such as GTOS (section 6.2.3). The curated dataset includes the following 8 classes (a) outdoor: pavement, foliage/grass, gravel, soil, and snow/slush and (b) indoor: high-friction materials, tiles, hardwood flooring. Subsequently, the framework's generalizability to older adults' data and its robustness against inter-participant differences were assessed (e.g., using LOSO cross-validation). The proposed framework provides one of the first investigations into the contextualization of free-living gait and fall risk assessment in older adults.

6.2 Materials and Methods

6.2.1 Recruitment and data collection

The project received ethics approval (reference number: 17589) from Northumbria University Research Ethics Committee, Newcastle upon Tyne, UK. All participants gave written informed consent before participating in the study.

Using wearable IMUs, cameras, and a motion capture system, a unique dataset, Multimodal Ambulatory Gait and Fall Risk Assessment (MAGFRA), was collected from fallers and older non-fallers in laboratory/clinic (MAGFRA-C) and/or *in the wild* (MAGFRA-W) [124]. In the present study, FPV data from nine participants (2 males, 7 females, mean age ≈73.6 yrs, 3 fallers based on the number of self-reported falls in the previous 12 months) from MAGFRA-W dataset were used. Participants are referred to as 'P' (P1 to P9) in this chapter (see Table 6.1). One participant's age was below 65 yrs, but as the person was a recurrent faller, the associated data were considered for further analysis.

Considering our previous findings [128, 127], it was hypothesized that a waist-level camera would offer a greater resolution of the feet and texture of surfaces than views higher on the body (e.g., a chest-level camera) for the purpose of informing free-living

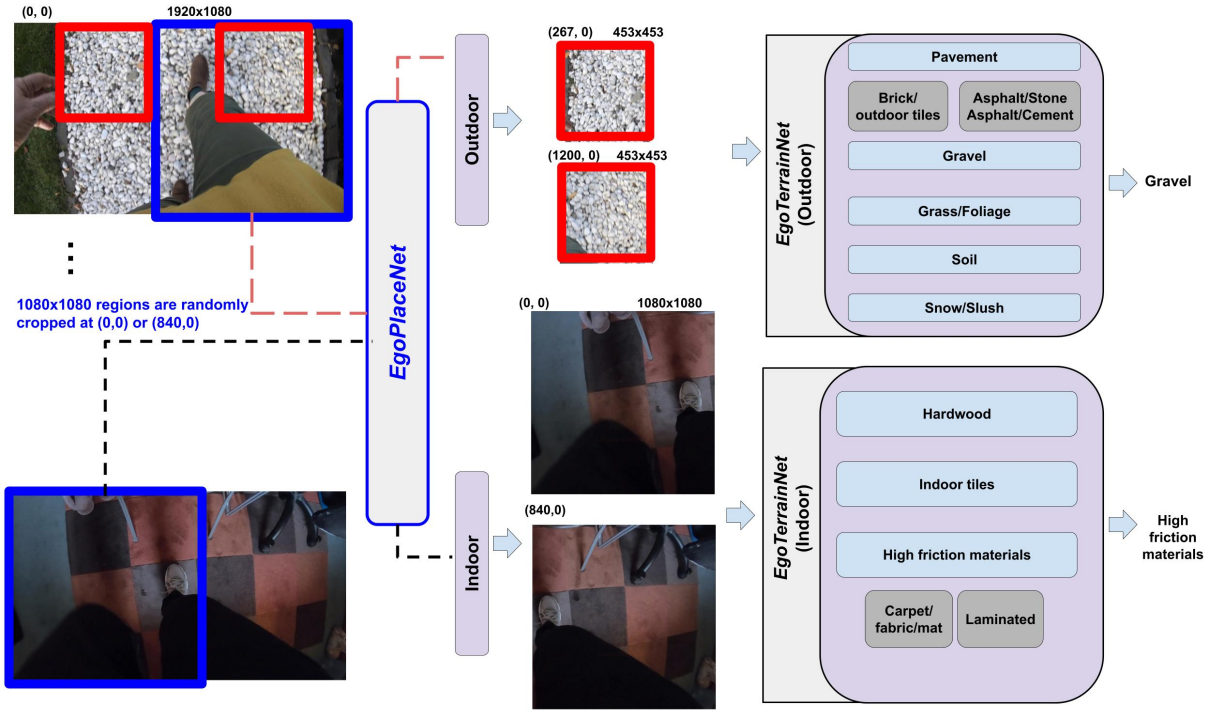


Figure 6.1: The proposed framework consists of two models: (a) *EgoPlaceNet*, which classifies scenes (one 1080×1080 region for each frame cropped randomly either from right or left corner, the blue square) into indoor and outdoor, and (b) *EgoTerrainNet*, with Indoor and Outdoor versions, which classifies two 453×453 (red squares) and 1080×1080 patches based on the enclosed terrain type.

FRA. Moreover, as discussed in Chapter 5), waist level views offer a consistent view of the feet even during sharp turns [127]. In contrast, head- and leg-mounted views tend to rotate in anticipation of turns or shift in attention, which reduces views of the feet and the terrain underneath and increases risk of motion blur [3, 36]. Thus, video data were collected using a GoPro Hero 5 Session or Hero 6 Black camera (30 fps, 1920×1080, wide view, except for P2 and P3 as marked by $\star$ in tables, see section 6.2.2), centered at each participant's waist by means of a belt attachment. The camera 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.

Data collection was performed in (a) public environments within Northumbria University, during which participants had to navigate through different indoor and/or outdoor environments while walking alongside a researcher (the walking paths were not predefined for participants to allow capturing different environmental features), or (b) participants'

homes or their neighbourhood (for P2, P3, P6 and P8) for between 1-2 hours with no researcher in attendance. Data collection in outdoor environments was performed during daylight hours. Two participants used a cane/stick during the data collection. P3 and P6 were living in the same home as marked by † in tables.

Table 6.1: The distribution of annotated crops/frames over different classes. $\star$: camera was unintentionally mounted upside-down by the participants or was set to take photos (not videos) resulted in smaller sample size, † : Participants living in the same home, HFM: high-friction materials.

	Outdoor Terrain Patches				Indoor Terrains Patches			In/Out frames	
	Pavement	Foliage	Gravel	Soil	HFM	Tile	Wood	Outdoor	Indoor
P_1	5101	153	0	39	2078	444	0	2884	1284
$P_2^\star$	-	-	-	-	106	0	0	0	106
$P_3^\star$†	51	8	0	3	60	16	0	51	71
P_4	3351	45	0	0	1674	611	0	1724	1197
P_5	1945	107	258	0	1523	343	0	1250	960
$P_6^\dagger$	78	29	0	0	261	241	0	232	319
P_7	2053	159	351	0	1079	226	47	1283	714
P_8	6239	151	0	0	427	0	312	3451	767
P_9	0	0	0	0	987	351	0	0	681
Total	18818	652	609	42	8195	2232	359	10875	6099

6.2.2 Preprocessing

Gait/ambulatory bout definition is highly inconsistent in the literature, but is often defined as any walking ≥3 steps [122]. In the MAGFRA-W dataset, FPV data attributed to level walking bouts ≥3 steps (stairs ascending/descending episodes were excluded) were taken into account for annotation. FPV data collected during short pauses/stances between longer walks were not necessarily excluded. Frames attributed to the identified gait bouts were sampled at 6 Hz using MATLAB R2019b. Compared to 1/15 Hz in previous work [185], this sampling rate was appropriate to capture changes in environment during gait. FPV data for P2 and P3 (marked by $\star$ in tables) were accidentally collected with a lower sampling rate (resulting in a smaller quantity of annotated images, Table 6.1) and a higher resolution. Therefore, the subsequent frames were resized to align with the rest of data. Additionally, P2, P3 and P6 wore the camera upside-down (marked by $\star$ in tables). Subsequently, a rotation was applied to permit comparisons with other data.

All sampled frames and image patches used for models development are in the RGB color space (e.g., 3×1920×1080), however, for simplicity, '3×' is removed here when describing the dimensions.

6.2.3 Considerations for the framework's structure and annotation of MAGFRA-W data

Two-layer framework vs end-to-end approach

Depending on the phase of gait and camera angle with respect to the frontal plane, a portion of the frames captured by a waist-mounted camera can be obscured by lower extremities and/or hands (see Fig. 6.2, Fig. 6.4, and Fig. 6.5). For instance, in Fig. 6.5-narrow *outdoor* path, a considerable portion of the frame is covered by participants' blue jeans. In this case, it can be hypothesized that the color of pants/clothing may impact the prediction of an end-to-end model when the full frame, rather than its specific regions, is fed as input.

To address this, frames were investigated to identify robust regions in terms of the provision of terrain-related visual features. Two 453×453 patches cropped at $(267, 0)$ and $(1200, 0)$ in 1920×1080 frames (Fig. 6.1) showed minimal overlap with upper and lower extremities during walking, which were initially considered as representatives of surfaces underneath the participants' left and right feet, respectively. These two patches were primarily cropped from all frames attributed to gait bouts.

From visual inspection of cropped patches, it was observed that 453×453 regions attributed to different indoor and outdoor surfaces can resemble each other closely in terms of colour and texture (Fig. 6.2), which may lead to a low classification accuracy. Moreover, it was noticed that due to the higher complexity of indoor scenes (compared to outdoor scenes), there could be a higher likelihood of overlap between the two 453×453 indoor patches with objects occluding views of the terrain such as walls or furniture. Thus, two larger 1080×1080 regions cropped at $(0, 0)$ and $(840, 0)$ (Fig. 6.1 and Fig. 6.4) were considered as better representatives of indoor terrains. The smaller outdoor and larger indoor patches were also more similar to the images in GTOS and Material in Context (MINC) [8]) datasets, respectively, which were further considered to form an independent training dataset (dicsussed in subsection 6.2.3).

The aforementioned points necessitated the development of a two-layer framework, rather than an end-to-end approach (8-class classification considereing all terrain types), to first categorize frames based on their location into indoor and outdoor classes. The first-layer's (i.e., *EgoPlaceNet* model) prediction further determines the frames' regions

that need to be cropped and fed into the second layer (i.e., *EgoTerrainNet*-Outdoor or -Indoor models) for terrain type identification (see Fig. 6.1).

Annotation of MAGFRA-W FPV data

Using MATLAB R2019b, 453×453 and 1080×1080 regions were cropped from outdoor (at (267,0) and (1200,0)) and indoor frames (at (0,0), (840,0))), respectively. Outdoor crops were annotated as: (a) pavement (e.g., outdoor tiles, bricks, asphalt and cement), (b) gravel/stone (including pebble, shale), (c) soil, (d) grass/foliage; and the indoor patches were grouped as (a) high friction materials (including carpet, fabric, laminate flooring, gym surfaces), (b) indoor tiles, and (c) wood (Fig. 6.4 and Table 6.1). Less than two patches were annotated for frames that either considerably overlapped with non-terrain materials (e.g., walls), had *fully* occluded field of view, or with unknown terrain type due to poor lighting ($\approx 7\%$ and 11% of outdoor and indoor patches remained unlabeled, respectively).

Assessing and augmenting models' generalizability

As discussed in section 6.1, high accuracies obtained from holdout and k-fold approaches may not necessarily indicate model's generalizability and robustness against environmental and/or inter-individual differences. Moreover, while MAGFRA-W data possesses high intra-class variance (see Fig. 6.4), surfaces may differ significantly from one older adult's home to another (e.g., carpet comes in a wide range of colours, patterns, and textures). As collecting a sufficient sample to capture this heterogeneity across older adults' everyday environments may not be feasible, similar to Chapter 5 [127], it was hypothesized that curating a training dataset from other (sufficiently similar) databases could increase intra-class variance, in terms of textures, colours, geometry, lighting conditions and clutter, and reduce the possible propensity to sample size bias. Subsequently, the framework's generalizability to unseen datasets could be improved. The procedure for curating the training dataset is discussed in subsection 6.2.3.

The LOSO cross-validation approach (which was used in Chapter 4) was considered as the next best to evaluate the framework's generalizability and robustness against inter-participant variance, for which the MAGFRA-W dataset alone was used for training.

Figure 6.2: Patches cropped from right or left parts of sample frames: (a) laminate flooring (high-friction material), (b) asphalt, (c) carpet (high-friction material), (d) partial view of furniture. Although the type of the walking surfaces are different, the 453×453 patches are very similar in terms of color and texture. *EgoPlaceNet* was adopted to classify frames into outdoor and indoor before terrain type identification to improve the framework's performance.

Independent training dataset A separate dataset was curated from other resources including public datasets: MINC-2500 (or MINC here) [8], GTOS datasets [208, 207][2], and HUJI EgoSeg (or 'EgoSeg' here) [146, 147][3], [4]. These datasets complement each other to address identification of various terrain types observed under free-living conditions[5]. For instance, while the MINC-2500 dataset does not contain images of asphalt, there are asphalt and stone asphalt classes in GTOS (which includes outdoor terrain patches only). Moreover, although there are 2,500 images of carpet, wood and tiles in MINC-2500, only a small proportion resemble the images that could be taken from a top-down view. Considering a large proportion of images in MINC-2500 are irrelevant to MAGFRA-W (e.g. furniture, or cabinet in class 'wood'), only relevant images from MINC-2500 were selected (e.g., 445/2500 from wood as hardwood flooring, see Table 6.2).

[2]Xue, J., Zhang, H., Dana, K., & Nishino, K. (2016). Differential Angular Imaging for Material Recognition. arXiv e-prints, arXiv-1612.

[3]Poleg, Y., Ephrat, A., Peleg, S., & Arora, C. (2016, March). Compact cnn for indexing egocentric videos. In 2016 IEEE winter conference on applications of computer vision (WACV) (pp. 1-9). IEEE.

[4]Poleg, Y., Arora, C., & Peleg, S. (2014). Temporal segmentation of egocentric videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2537-2544).

[5]Appendix D shows the preliminary test results of *EgoTerrainNet*-Outdoor and -Indoor fine-tuned on (a) a subset of GTOS and (b) all 2,500 images of 'wood', 'carpet', and 'tiles' in MINC-2500, when applied to MAGFRA-W. The results suggest that one dataset alone may not result in generalizable models.

As mentioned in section 6.1, images in GTOS were collected while the camera-ground distance is much smaller than the height of the waist-mounted camera. This field of view resulted in very low complexity and intra-class variance in GTOS (e.g., pedestrian's feet were observed in the image) compared to the higher view in MAGFRA-W and may reduce the prospects for generalizability to everyday terrains. Although there are 40 different classes of outdoor terrains in GTOS, differentiating between each may not provide relevant additional information for gait assessment and free-living FRA. For example, separate GTOS classes of asphalt, cement, or pavement bricks may not result in substantially different walking patterns. Thus for the purpose of this study, images from the relevant classes were combined.

To further address the limitations of MINC-2500 and GTOS, the suitability of several FPV-based datasets (e.g. EPIC-Kitchens 2018 [29]) was examined. Among public FPV datasets, EgoSeg was considered a suitable candidate, as the camera wearers walked in diverse outdoor environments. EgoSeg video data were collected from a head-mounted GoPro Hero3+ camera during a range of activities (e.g., walking, riding bus, driving). After resizing 720p frames to 1920×1080, patches of 453×453 were cropped from the lower-central, right, and left parts of the resized frames. Considering head-mounted cameras may not provide a consistent view of terrain, only a handful were annotated and included.

As GTOS and MINC-2500 datasets contain no images of snow/slush-covered terrains, a smartphone at waist level was used to capture videos of slush- or snow-covered terrains by the authors. Patches (453×453) were cropped from the right and left corners of the frames and added to the training dataset. Although this snow/slush class may not have representatives in the test dataset (MAGFRA-W), snow-covered terrains are frequently observed in regions with low average yearly temperature, impact gait patterns and are a potential risk factor for falls. Therefore, adding this class would improve the framework's relevance and generalizability.

Overall, 3,651 and 5,773 image patches were extracted from the aforementioned datasets to form training datasets for indoor and outdoor surfaces, respectively. The distribution of patches extracted from different datasets as well as sample patches for snow/slush has been shown in Table 6.2. The open access image/FPV datasets discussed here can be accessed and viewed from their corresponding data repositories.

Table 6.2: Training dataset, the relevant images, mostly from top-down view, were either extracted from available datasets, MINC-2500, GTOS and EgoSeg, or collected by the authors.

	Total	MINC-2500	GTOS(-Mobile)	EgoSeg	Author
Outdoor					
Pavement (Asphalt /Brick/Cement)	1503	250 (brick)	591 (brick+ asphalt+ cement+ stone asphalt + stone cement)	662	
Foliage/Grass	1382	0	1227 (turf+leaf+grass)	155	0
Gravel/Stone	1266	0	1266 (pebble+shale)	0	0
Snow	323	0	0	0	323
Soil	1299		1230	69	0
Indoor					
High-friction materials	2184 (e.g., fabric and carpet (top-down view))	2184	-	-	-
Wood (top-down view)	445	445	-	-	-
Indoor tiles (top-down view)	1018	810	-	208	-

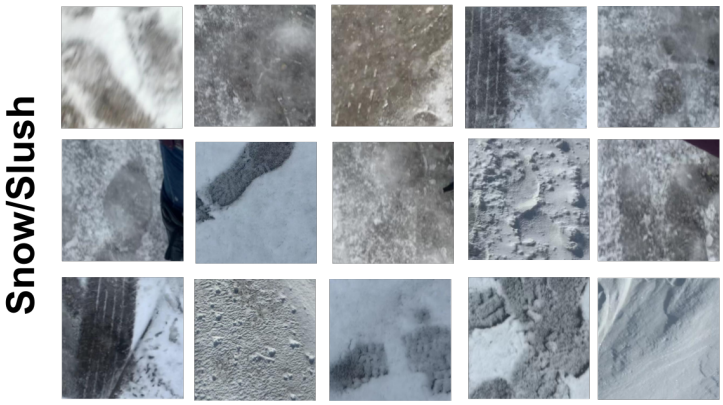


Figure 6.3: Sample patches representing class 'slippery/snow', the data was captured from a smartphone from waist level.

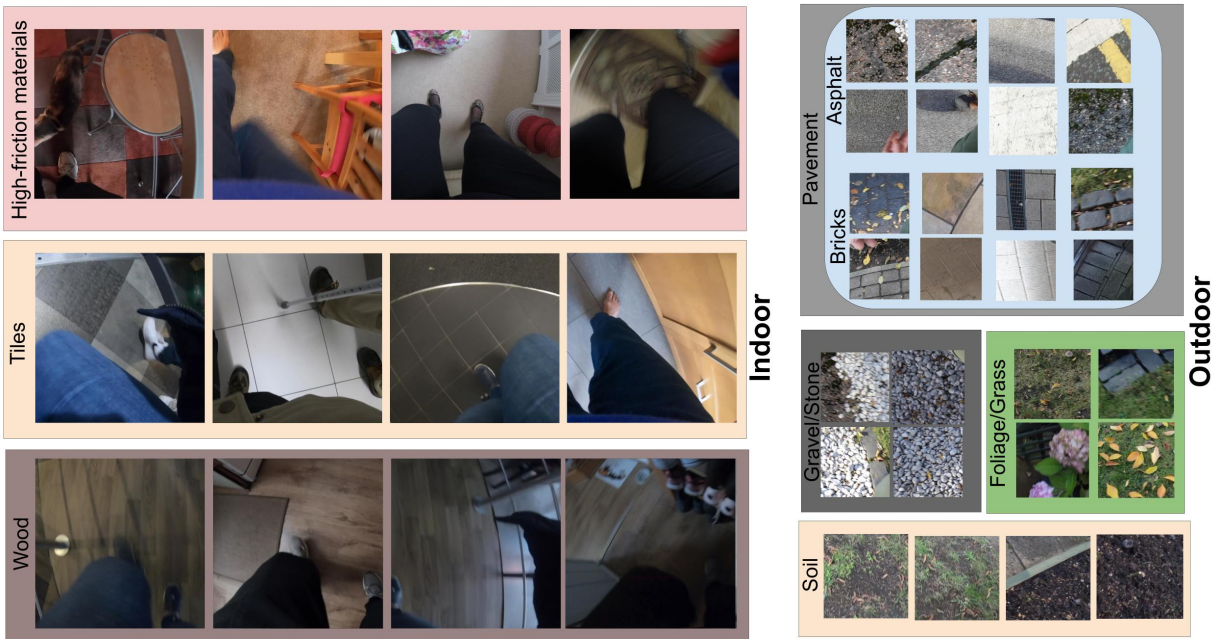


Figure 6.4: Sample patches from MAGFRA-W dataset. Outdoor patches were cropped at (267,0) and (1200,0) from the 1920 × 1080 outdoor frames during gait. 1080 × 1080 regions were cropped from upper left and right corners for indoor scenes. These dimensions were carefully selected to be compatible with the datasets used to train *EgoTerrainNet*-Outdoor and -Indoor.

6.2.4 Pre-trained ConvNets

Considering the size of the curated training dataset and MAGFRA-W (also used for training in the LOSO approach), training a deep ConvNet from scratch was not feasible. Therefore, the transfer learning approach [137] was considered. This subsection discusses the criteria for selecting the backbone models (ConvNet pre-trained on a large-scale dataset) for *EgoPlaceNet* and *EgoTerrainNet*.

For applications in prosthestics and exoskeleton, the real-time detection of environmental features is a critical part of the control loop. While on-device detection of environmental features is not necessary for the purpose of FRA, this allows processing of frames without the need for storing videos, and may subsequently mitigate privacy and ethical issues associated with FPV data use. By benchmark analysis of state-of-the-art deep neural network architectures (in terms of accuracy, size of the learnable parameters, memory usage, computational complexity using the floating-point operations, and inference time), SqueezeNets, MobileNets, ResNet-18, GoogLeNet, and AlexNet achieved optimal real-time performance, while no significant relationship between model complexity and recognition

accuracy was reported [12]. Building upon the idea of depth-wise separable convolution from MobileNetV1 [70], MobileNetV2 pushed the state of the art for mobile image classification [164] using the inverted residual with linear bottleneck as a novel layer module. This resulted in faster and more accurate performance while using $\approx 30\%$ fewer parameters compared to MobileNetV1. Therefore, MobileNetV2 pretrained on ImageNet [163] was considered as the initial candidate for backbone models in our study.

In contrast to ImageNet categories, where indoor and outdoor scenes were not separated, images in Places365 dataset [212] were categorized into indoor and outdoor macro-classes (e.g., indoor and outdoor categories for ice skating rink) and the models were trained on millions of scene photographs. Therefore, deep networks trained on this dataset have learned different feature representations for a wide range of indoor and outdoor images compared to ImageNet, and hypothesized to be a better candidate for *EgoPlaceNet*. Among the available pre-trained deep models on Places365 dataset[6,7], e.g., AlexNet (over 60 million parameters for 227×227 images, 8 layers, [96]), GoogLeNet ($\approx$12 times fewer parameters compared to AlexNet, 22 layers [183]) model was considered as the backbone model for *EgoPlaceNet*. Fine-tuning procedures for all models are discussed in section 6.2.5.

6.2.5 Experiments

The first classifier, *EgoPlaceNet* considers full FPV frames to determine their location (indoors vs outdoors) in a binary classification problem. As discussed in subsection 6.2.4, GoogLeNet pre-trained on Places365 was selected as the backbone for *EgoPlaceNet*. To maintain 1:1 aspect ratio for each of the indoor (6,099) and outdoor (10,875) frame (Table 6.1), a 1080×1080 region was randomly cropped either from the top right or top left corners of the down sampled 1920×1080 frames (1 crop for each frame), Fig. 6.1, blue squares .

To assess the generalizability of *EgoPlaceNet* at different levels, two different training procedures were considered. In *EgoPlaceNet*.v1, the GoogLeNet pre-trained on Places365, was fine-tuned on indoor and outdoor images in the curated training dataset (described in 6.2.3). Secondly, LOSO cross-validation (*EgoPlaceNet.LOSO*$_n$, $n = \{1, \ldots, 9\}$) was performed to investigate robustness against inter-participant variations, where the GoogLeNet-Places365 was fine-tuned based on the dataset acquired from 8 participants, and tested on the remaining data from one participant. The GoogLeNet-Places365 models were fine-tuned by freezing the weights of 10 earlier layers in the network.

[6]https://github.com/CSAILVision/places365

[7]https://www.mathworks.com/help/deeplearning/ref/googlenet.html

For *EgoTerrainNet*-Outdoor and -Indoor versions, MobileNetV2's were fine-tuned using the curated training dataset discussed in 6.2.3, by replacing the last fully connected layer and the final classification layer of the network.

Depending on the model (i.e., *EgoPlaceNet.LOSO*$_n$, *EgoPlaceNet.v1*, *EgoTerrainNet*-Outdoor, *EgoTerrainNet*-Indoor) the relevant training dataset (e.g., indoor or outdoor, as discussed in section 6.2.3, or remaining 8 Ps for *EgoPlaceNet.LOSO*$_n$) was randomly divided into training (70%) and validation (30%) with images resized to 224×224. Experiments were performed on a workstation (Intel(R) Core (TM)i7-6700, 3.4GHz with Nvidia GeForce GTX 750 Ti), with MATLAB R2019b. The mini-batch sizes of $K = 10$ and $K = 64$ were used (due to the limited memory), for *EgoPlaceNet* and *EgoTerrainNet*, respectively. Epochs were set to 8; however, the training procedure was terminated manually in some cases if no change was observed in validation loss/accuracy (validation patience was set to 20).

The initial learning rate of $\gamma = 0.01$ for *EgoTerrainNet*-Outdoor and -Indoor and $\gamma = 3e\text{-}4$ for *EgoPlaceNet* models resulted in the best validation accuracies. Stochastic gradient descent with momentum was considered as the optimization method. Moreover, the following hyperparameters were employed: momentum: 0.9, L2 Regularization: $1e\text{-}4$, gradient threshold method: L2 norm, and decay rate of 0.0005. To further address the problem of a small dataset, improve the generalization of the network, data were augmented by including random crops, translation, rotation $\in [-20 + 20]$ deg (accounting for changes in camera orientation during gait) and vertical reflection/flip over y axis. Considering the viewpoints of images in the training dataset as well as data captured by a belt-mounted camera, horizontal reflection was not considered for augmentation.

6.3 Results

During the training process, the validation accuracies of 93.97%, 98.19% (mean over participants), 99.23%, and 85.26% were achieved for *EgoPlaceNet*.v1, *EgoPlaceNet.LOSO*$_n$, *EgoTerrainNet*-Outdoor and *EgoTerrainNet*-Indoor, respectively.

The *EgoPlaceNet*.v1 resulted in (mean-over-participant) test accuracies of 91.14% and 78.04% (Table 6.3) for the detection of outdoor and indoor scenes in MAGFRA-W, respectively. However, these rose to 97.36% and 95.59% when LOSO cross-validation was performed (9 models, *EgoPlaceNet.LOSO*$_n$).

Confusion matrices and per-class detection accuracies for each participant were examined separately to better assess the impact of inter-participant differences, colour of cloth-

ing, camera placement, and environmental features on models' performance (Table 6.4). Therefore, although the distribution of image patches over classes were different for participants, due to the aforementioned points, the mean-over-participant detection accuracy for each class was considered as a more suitable metric compared to overall detection accuracy for each class (e.g., total true positives from all participants for class 'pavement'/total number of patches labeled as 'pavement' from all participants).

Table 6.3: Results for *EgoPlaceNet*: 1. validation accuracy of 93.97 was obtained for *EgoPlaceNet.v1*, which was fine-tuned on the selected training dataset from MINC-2500+EgoSeg+GTOS, tested on MAGFRA-W. 2. for participant n *EgoPlaceNet-LOSO*$_n$ was trained on the data from the other 8 participants and tested on the data from n.

	EgoPlaceNet.v1		***EgoPlaceNet-LOSO***$_n$		
	Out	In	LOSO$_{val}$	Out	In
P_1	91.26	78.12	98.32	99.10	99.92
$P_2^\star$	-	90.57	98.35	-	100
$P_3^\star\dagger$	84.31	88.73	98.70	94.74	98.59
P_4	99.54	72.60	98.33	99.83	98.16
P_5	77.28	77.81	98.07	98.16	99.69
$P_6^\dagger$	93.53	77.12	99.08	95.69	83.39
P_7	99.22	53.64	97.68	99.84	94.12
P_8	92.90	84.22	99.61	94.20	99.74
P_9	-	79.59	95.58	-	86.78
mean	91.14	78.04	98.19	97.36	95.59

EgoTerrainNet-Outdoor exhibited satisfactory performance (mean-over-participant accuracies) for the identification of pavement (87.63%), grass/foliage (91.24%), and stone/gravel (95.12%). However, it failed to detect soil (Table 6.4). *EgoTerrainNet*-Indoor detected high-friction materials including carpet and laminate flooring with a high accuracy (mean over participants: 95.02%). However, the mean-over-participant accuracies drastically decreased to 71.15% and 64.76% for tiles and wood, respectively. While tile identification accuracy was high in most participants (P1: 88.51%, P4: 94.93%, P5: 90.96%, P7: 88.05%, P9: 92.88%), the results of P3 (0%) and P6 (42.74%) decreased the mean accuracy for tile detection. It was interesting as the color of tiles in P3' and P6's home was grey, and similar to some sample patches from other participants' data (see Fig. 6.4 and 6.5). Similarly, wood identification achieved a high and low accuracies for P7 (93.62%) and P8 (35.90%), respectively.

6.3.1 Deeper analysis of lower accuracies

Results marked by △ in Table 6.4 are further discussed in this subsection.

First, pavement is mainly confused with soil (e.g., in P8) and snow (e.g., P1). In P1's data, the 453×453 patches overlapped with regions of pavement with white paintings/street signs (Fig. 6.5), which could be confused with snow. Moreover, many images in asphalt and soil classes in the curated training dataset *EgoTerrainNet*-Outdoor (section 6.2.3) share similar visual features such as colour and texture, which partially explains the aforementioned confusion.

Samples in soil were mainly confused with grass/foliage in P1. In P1's data, soil was frequently mixed with or covered by grass/foliage (Fig. 6.5). Considering the lack of a standard definition for the annotation of patches, either grass/foliage or soil may have been assigned to those patches, which may explain the subsequent results.

The camera's field of view during free-living data collection in P3's and P6's home was heavily occluded by participants' clothing (e.g., blue pants/jeans). In addition to covering the tiles texture, the participants' clothing likely confused *EgoTerrainNet*-Indoor to classify tiles as high-friction material (e.g., fabric/carpet) (see Fig. 6.5). The images were also blurry in many cases (Fig. 6.5), concealing the texture of tiles in the subsequent images.

The lower wood detection accuracy for P8, compared to P7, could be attributed to poor lighting conditions in the former's home (Fig. 6.5). The image patches attributed to 'wood' for this participant were mainly categorized as high-friction materials, likely because the texture of wood was not differentiable in the subsequent 1080×1080 patches due to poor lighting condition.

Table 6.4: Confusion Matrices at participant level: for *EgoTerrainNet*-Outdoor and -Indoor, MobileNetV2's pre-trained on ImageNet dataset were fine-tuned. The validation accuracies for -Outdoor and -Indoor versions were 99.23 and 85.26, respectively. ⋆ : camera was unintentionally mounted upside-down by the participants or was set to take photos (not videos), † : Participants living in the same home, HFM: high-friction materials, △: cases that are discussed in 6.3.1

	Outdoor							Indoor				
		Pavement	Foliage	Gravel	Soil	Snow	Accuracy		HFM	Tile	Wood	Accuracy
P_1	Pavement	4446	4	0	3	648$^{\triangle}$	87.16	HFM	1962	52	64	94.42
	Foliage	34	113	0	0	6	73.86$^{\triangle}$	Tile	31	393	20	88.51
	Gravel	0	0	0	0	0	-	Wood	0	0	0	-

	Soil	16	21	0	2	0	$5.13^{\triangle}$	-				
$P_2^{\star}$	Pavement	-	-	-	-	-	-	HFM	103	0	3	97.17
	Foliage	-	-	-	-	-	-	Tile	0	0	0	-
	Gravel	-	-	-	-	-	-	Wood	0	0	0	-
	Soil	-	-	-	-	-	-					
$P_3^{\star\dagger}$	Pavement	47	3	1	0	0	92.16	HFM	58	0	2	96.67
	Foliage	1	7	0	0	0	87.50	Tile	15	0	1	$0^{\triangle}$
	Gravel	0	0	0	0	0	-	Wood	0	0	0	-
	Soil	3	0	0	0	0	$0^{\triangle}$	-				
P_4	Pavement	3010	5	5	42	289	89.82	HFM	1607	34	33	96.00
	Foliage	0	45	0	0	0	100	Tile	29	580	2	94.93
	Gravel	0	0	0	0	0	-	Wood	0	0	0	-
	Soil	0	0	0	0	0	-					
P_5	Pavement	1769	0	17	2	157	90.95	HFM	1500	16	7	98.49
	Foliage	0	107	0	0	0	100	Tile	30	312	1	90.96
	Gravel	8	0	249	0	1	96.51	Wood	0	0	0	-
	Soil	0	0	0	0	0	-					
$P_6^{\dagger\star}$	Pavement	63	0	0	0	15	80.77	HFM	256	0	5	98.08
	Foliage	1	26	0	2	0	89.66	Tile	131	103	7	$42.74^{\triangle}$
	Gravel	0	0	0	0	0	-	Wood	0	0	0	-
	Soil	0	0	0	0	0	-					
P_7	Pavement	1731	11	48	2	261	84.32	HFM	1023	12	44	94.81
	Foliage	9	150	0	0	0	94.34	Tile	19	199	8	88.05
	Gravel	21	1	329	0	0	93.73	Wood	3	0	44	93.62
	Soil	0	0	0	0	0	-					
P_8	Pavement	5506	228	5	329	171	88.25	HFM	357	5	65	83.61
	Foliage	4	141	6	0	0	93.38	Tile	0	0	0	-
	Gravel	0	0	0	0	0	-	Wood	183	17	112	$35.90^{\triangle}$
	Soil	0	0	0	0	0	-					
P_9	Pavement	-	-	-	-	-	-	HFM	947	31	9	95.95
	Foliage	-	-	-	-	-	-	Tile	25	326	0	92.88
	Gravel	-	-	-	-	-	-	Wood	0	0	0	-
	Soil	-	-	-	-	-	-					

6.4 Discussion/Conclusion

This paper proposes an egocentric vision-based framework to automatically detect indoor and outdoor level walking surfaces. To the best of the authors' knowledge, this work is the first to present a deep learning-based model tested on older adults' everyday FPV data towards the development of a context-aware free-living FRA.

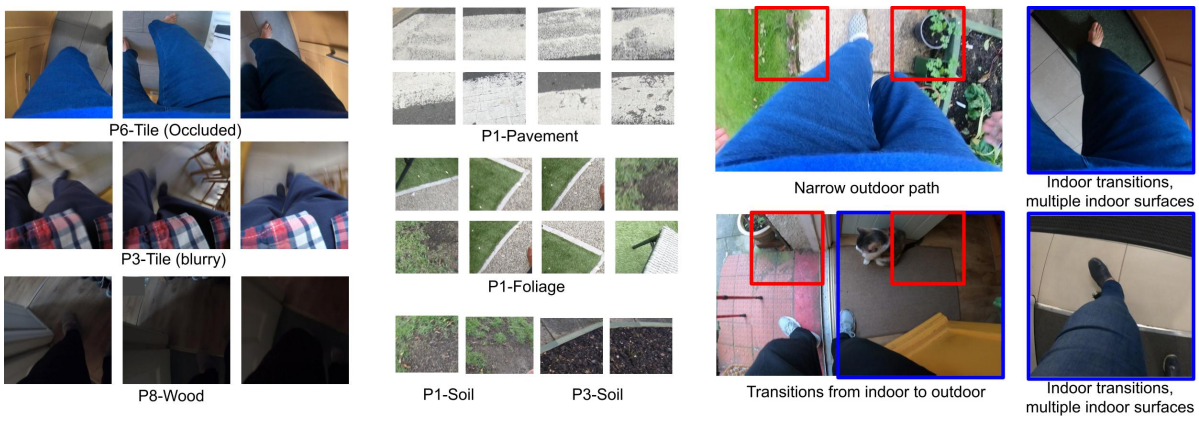


Figure 6.5: Sample frames/patches illustrating conditions challenging the performance of the proposed framework.

The MAGFRA-W dataset offers a considerable diversity in terms of terrain types, feet appearance (e.g., shoes with different colours, barefoot, socks, slippers), and gait patterns (e.g., using walking aids) leading to more ecologically valid classification results compared with data collected in controlled conditions. After investigating the participants' FPV data and other relevant public datasets, a two-layer structure was considered superior to an end-to-end approach for terrain type identification. Subsequently, the training and test datasets were prepared according to this hypothesis. Overall, it can be concluded that aggregating *EgoPlaceNet* trained on outdoor-indoor images captured by a belt-mounted camera, followed by *EgoTerrainNet*s trained on an independent dataset leads to the best terrain identification performance in terms of accuracy and generalizability.

To train and test *EgoPlaceNet* and *EgoTerrainNet*s, several approaches could have been considered: (a) holdout and k-fold cross-validation, (b) LOSO using MAGFRA-W, or (c) using MAGFRA-W as the test dataset and incorporating an independent (but sufficiently similar) training dataset for fine-tuning deep models. Considering that the discrepancy between the distributions of training and test datasets in approach (c) avoids the generation of unrealistically high accuracies, we considered this option to be superior. Furthermore, option (c) is aligned with previous research work presented in Chapter 5 [127] and represents a pragmatic picture of the proposed framework's generalizability. Option (b) was considered as the next best to evaluate the framework's robustness against inter-participant differences.

To form an independent training set for approach (c), relevant images (or frames) from different datasets (e.g., MINC-2500, EgoSeg, and GTOS) were selected. The sub-

sequent fine-tuned ConvNets on this dataset, i.e., *EgoTerrainNet*-Outdoor and -Indoor, were applied to specific regions of outdoor (453×453) and indoor (1080×1080) frames. Promising results exhibited the models' generalizability to detecting a broad range of terrains. Although the sample size for the curated dataset was relatively small (9,424 images overall), the results indicate that this dataset captured a high variations of texture, colour, and shape in everyday scenes, which bypasses the requirement for prolonged data collection from a large cohort of older adults to form a heterogeneous training dataset. This approach also outperformed the models that were solely fine-tuned on one dataset (e.g., GTOS or MINC-2500, as shown in Appendix D).

EgoPlaceNet.v1 achieved 91.14% and 78.04% detection accuracies for outdoor and indoor scenes, respectively. The relatively poor performance of this binary classifier supports the hypothesis that an end-to-end approach, i.e., an 8-class classification problem (more complex compared to the binary classification) may not exhibit a robust performance if option (c) is considered. On the other hand, high *EgoPlaceNet.LOSO*$_n$ accuracies ($\geq$95% for both indoor and outdoor scenes, Table 6.3) confirm the models' robustness against variations in participants' characteristics, camera view and partial occlusions (e.g., lower extremities, walking aids).

High detection accuracies were consistently observed for pavement, gravel, foliage/grass, and high-friction materials for all participants. Among the outdoor terrain types, soil had the lowest detection accuracy as well as a low per-class quantity in MAGFRA-W (only 42 samples, see Table 6.1). Additionally, no sample of snow was found in the MAGFRA-W dataset. These points necessitate further investigation of *EgoTerrainNet*-Outdoor's performance using a more inclusive test dataset in future studies. Moreover, while tiles (in different patterns and colors such as grey, white, see Fig. 6.4) in public environments were detected with high accuracies (P1, P4-5, P7, P9; ranging from $\approx$ 88% to $\approx$ 95%), in-home tiles (mostly grey) captured in P3's and P6's home were mainly confused with high-friction materials. The same trend was observed for 'wood', which was detected with 93.62% and 35.90% accuracies for P7 (public environment) and P8 (in-home), respectively. As detection of wood and indoor tiles require capturing fine details of terrain textures, partially-obscured views as well as blurry and/or dark images due to dim lighting conditions in in-home settings were considered as the primary reasons for this inferior performance. A similar phenomenon was observed in other studies [39, 198], where image blur/noise led to a considerable drop in classification accuracies. Methods have been proposed to exclude or skip blurry images [36, 3], at the expense of heavier computational demand. Other works have suggested that classification performance of deep architectures could be improved by fine-tuning the models on blurry images [198]. In [35], authors jointly trained a deblurrer combined with a high-level computer vision network. Therefore, the integration of similar

pipelines into the proposed framework in the present study may augment the performance of *EgoTerrainNet*-Indoor.

The backbone deep models considered here (i.e., MobileNetV2 and GoogLeNet pre-trained on ImageNet and Places365 datasets, respectively) were selected based on multiple criteria and previous comparison studies (discussed in subsection 6.2.4), and exhibited satisfactory performance in terms of detection accuracy. Further investigation using larger-scale datasets is required to identify the optimal deep architecture addressing terrain type identification in the wild.

While the collection of FPV data in controlled conditions facilitates the process of image annotation by providing high quality and consistent data, the complex nature of everyday scenes captured in the MAGFRA-W dataset challenged the process of image patch annotation. First, a subset of image patches (7% and 11% of the extracted outdoor and indoor patches, respectively) remained unlabelled due to their significant overlap with non-terrain materials such as walls, dim lighting conditions, or obscured views. Therefore, although the accuracies for *EgoPlaceNet* and *EgoTerrainNet* were calculated separately, the overall framework's accuracy (the sequential approach) could not be reported. The addition of class 'others' [8] in the training dataset could have been considered to address this limitation, however, the preparation of relevant samples collected from the top-down view to form this class was out of the scope of the present study. Secondly, in addition to mix surfaces (e.g., soil and grass in Fig. 6.5), transitions between different locations and surfaces (see Fig. 6.5) challenged the annotation of ground truth data. For example, while only one label was attributed to each 1080×1080 patch, in Fig. 6.5 (right panel) each foot is placed on a different surface. Subsequently, both tiles and high-friction materials could be considered as valid labels for the patch. Such a discrepancy in the annotations could introduce errors to the reported results. This issue occurred less frequently during the annotation of outdoor patches, as due to their smaller size (compared to indoor patches) the enclosed outdoor terrain type was generally more consistent. Considering a belt-mounted camera's field of view, a separate 453×453 region was expected to represent the terrain type around each foot in outdoor scenes. However, there were exceptions. For instance in Fig. 6.5, a participant is walking on a narrow 'brick-covered' ($\in$pavement) surface, and the right and left patches partially overlap with foliage, which is irrelevant to the walking surface type. In Chapter 5 [127], the *FootChaser* framework was proposed to localize feet in the video data captured from a belt-mounted camera for the purpose of gait assessment. Therefore, rather than cropping frames' fixed regions (considered in the present study), the integration of *FootChaser* model into the proposed framework is expected to allow cropping more specific regions (with varying sizes) of frames in the proximity of each localized foot. This may permit a more accurate identification of walking surfaces.

Considering the preliminary results of [202] and Appendix C (discussed LOSO results) regarding the feasibility of stair walking detection in older adults using IMU data, the present study focused on vision-based detection of level walking surfaces. In our future work, other details of the walking surfaces will be considered. For instance in P4's multimodal data (IMU and FPV), 1 naturally-occurring (hit/bump) misstep was identified, during which a light pole was considered as the environmental fall risk. Therefore, algorithms to detect such a static obstacle, as well as other tripping hazards including dynamic obstacles (pedestrians and pets, Fig. 6.4) and cracks in pavement [211] will be considered to provide complementary information on the properties of environment, towards a comprehensive context-aware free-living gait and fall risk assessment method. Moreover, the performance of a terrain identification model developed on a multimodal dataset will be examined.

Overall, encouraging results suggest that the integration of wearable cameras as well as deep learning approaches can provide objective information on the properties of walking surfaces, towards context-aware FLDBs for gait and fall risk assessment in the wild. Considering IMU data were collected along with FPV data in MAGFRA-W, the impact of environmental features on the fall predictive power of IMU-based gait-related FLDBs will be investigated in our future works.

Chapter 7

Conclusion

Older adults' falls are highly complicated phenomena with a multifactorial etiology. To address the limitations of laboratory-based or supervised fall risk assessment (FRA) approaches, a growing body of literature has focused on the development of wearable sensor-based methods for free-living FRA to assess older adults' activities in their natural environments. These studies investigated the relationships between wearable sensor-based free-living digital biomarkers (FLDBs) and falls. However, there are no clear solutions for transparent deployment of wearables for free-living FRA at this time due to the ongoing novel developments within the field. To highlight the gaps in the literature, Chapter 2 provided a detailed review of investigated free-living FRA approaches. The subsequent chapters of this thesis addressed priority research gaps by proposing multiple machine learning-based frameworks towards a more *stable* and *context-aware* free-living FRA approach advancing fall prevention in older adults. In this final chapter, key themes investigated in this thesis are discussed.

7.1 Validation studies in older adults

Performing validation studies using older adults' free-living data to assess the performance of the proposed machine learning-based frameworks can be considered as one of the key strengths of the present research work. Specially, the majority of machine learning models were either developed by integrating a training dataset independently collected from the test/validation dataset (i.e., curated training dataset from multiple sources) or were tested using leave-one-subject-out cross-validation approach. As discussed in multiple chapters,

these approaches outperform other cross-validation/testing methods such as k-fold, and exhibit more realistic results in terms of generalization to unseen data/users.

Chapter 4 presented one of the first validation studies of machine learning models to detect older adults' naturally-occurring compensatory balance reactions (CBRs). Compared to steps, CBRs are rare events. Therefore, prolonged collection of criterion standard data (along with IMU data) was required to validate model's performance in free-living conditions. By investigating 11 fallers' and older non-fallers' free-living criterion standard data, 8 naturally-occurring CBRs (i.e., 7 trips and 1 hit/bump) 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 by a perturbation treadmill (discussed in Chapter 3). Subsequently, the models' translation/generalization to older adults' out-of-lab data were assessed. A subset of models differentiated between naturally-occurring CBRs and free-living activities with high sensitivity (100%) and specificity ($\geq$99%). The findings suggest that accurate detection of naturally-occurring CBRs is feasible.

With the aim of improving the interpretability of gait-related FLDBs and investigating the impact of environment on older adults' gait, Chapter 6 discussed a deep vision-based framework proposed to automatically detect the most common level walking surfaces. The proposed framework provides one of the first investigations into the contextualization of free-living gait and FRA in older adults. Using a belt-mounted camera and IMUs worn by fallers and older non-fallers (mean age 73.6 yrs), a unique dataset (i.e., Multimodal Gait and Fall Risk Assessment in the Wild (MAGFRA-W)) was acquired. The frames and image patches attributed to nine participants' gait were annotated: (a) outdoor terrains: pavement (asphalt, cement, bricks, outdoor tiles), gravel, foliage/grass, soil, snow/slush; and (b) indoor terrains: high-friction materials (e.g., carpet, laminated floor), wood, and tiles. A series of ConvNets were developed: *EgoPlaceNet* categorizes frames into indoor and outdoor; and *EgoTerrainNet* (with outdoor and indoor versions) detects the enclosed terrain type in patches. *EgoPlaceNet* detected outdoor and indoor scenes in MAGFRA-W with 97.36% and 95.59% (leave-one-subject-out) accuracies, respectively. *EgoTerrainNet*-Indoor and -Outdoor (MobileNetV2's fine-tuned on the curated dataset) achieved high detection accuracies for pavement (87.63%), foliage (91.24%), gravel (95.12%), and high-friction materials (95.02%), which indicate the models' high generalizabiliy. The combination of applying data acquired directly from older adults (i.e., MAGFRA-W) and utilizing independent training and test datasets (e.g., leave-one-subject-out) directly supports the feasibility and reliability evaluations for use with the target population.

In order to validate the performance of *FootChaser* framework discussed in Chapter 5 for the estimation of step width in older adults, Muldimodal Gait and Fall Risk Assessment

in Clinic (MAGFRA-C) dataset [124] was prepared (not presented in this thesis). This dataset includes criterion (gold) standard measures recorded by Vicon motion capture system. Older adults walked over a treadmill while wearing a waist-mounted egocentric camera (GoPro), and multiple IMU devices (Axivity, Newcastle upon-Tyne, UK), affixed to both wrists, lower-back, both legs, head, pelvis, and the camera. Vicon markers were also placed on the feet (center) and on IMUs. Outside of the scope of the current thesis, future research will compare the pixel-wise outputs of FootChaser with the x and y coordinates of feet centers recorded by the Vicon system, and other spatial gait measures that can be estimated from the IMUs. Considering that the initial version of the *FootChaser* model yielded pixel-wise outputs of feet centers, to develop spatial estimates with clinical values to be interpreted by physicians, our future research will focus on the development of *FootChaser.v2*, which will output standard distance units (e.g. m or cm). Moreover, by comparing the IMU-based estimations with criterion standard data, the feasibility of developing machine learning models to estimate step width using IMU data alone can be investigated.

7.2 Clinical and translational implications

Considering the promising results of the machine learning based frameworks proposed in this thesis, subsequent research will focus on applying the described models to large-scale multimodal datasets collected from older fallers and non-fallers in a longitudinal manner, towards a deeper understanding of falls etiology. This would allow further investigation of: 1. the associations between falls and CBR-related FLDBs (e.g., direction, duration, number of steps to recover balance) and 2. the associations between falls, different features of environment, and FLDBs (i.e., context-aware FLDBs). The proposed multimodal approach can be used to elucidate the interplay between intrinsic and environmental risk factors and clarifies their respective impacts on fall predictive powers of FLDBs, which would subsequently inform the on-going development of clinical use of wearables for FRA. The egocentric vision data captured in MAGFRA-W dataset provided rich contextual information about the factors leading to CBRs (e.g., a light pole, Fig. 4.1) and false alarms (e.g., a sudden change in walking direction, Fig. 4.7). 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 hypothesized to be prevented (see Fig. 7.1).

Moreover, there are a number of perturbation training interventions currently being

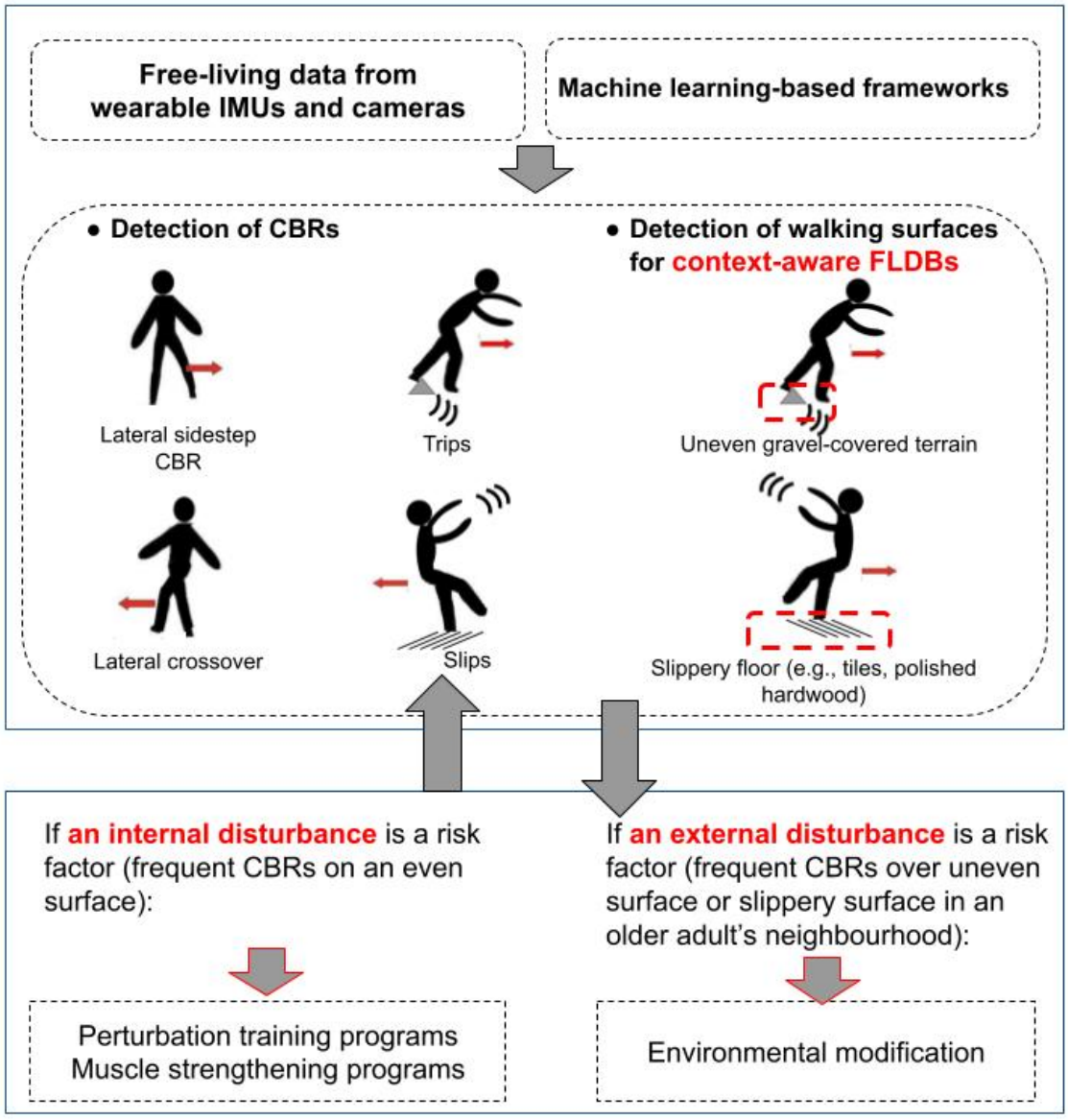


Figure 7.1: The proposed multimodal approach can be used to elucidate the interplay between intrinsic and environmental risk factors and clarifies their respective impacts on fall predictive powers of free-living digital biomarkers (FLDBs), which may subsequently enable clinicians to target *more specific* intervention strategies.

developed and tested in clinical settings [153, 136, 142]. However, the transfer of balance recovery skills gained during these in-clinic programs to everyday scenarios has not been well-investigated. The methods proposed in Chapter 4 may be applied to track responsiveness to these programs by providing objective information on the timing and frequency of naturally-occurring reactive responses induced by real-life perturbations.

7.3 Systematic investigation of FLDBs' and domains' fall sensitivities

7.3.1 The integration of different sensor modalities for fall risk assessment

Due to the complex and interconnected etiology of older adults' falls, any single model developed based on any single type of data can touch only a small part of the entire knowledge regarding this important public health problem (see Fig. 1.1). To date, social & behavioural health scientists, movement disorder researchers, and imaging researchers investigate the fall predictive power of biomarkers that are specific to their domains, often in isolation from other potential risk/protective factors. In addition to all FLDBs discussed in this thesis, social and behavioural factors such as level of education [51], loneliness and social isolation [18], anxiety [54], depression [76] and diet [106] have been reported to be linked to falls. More recently, a genome-wide association study examined genetic basis of falling risk susceptibility, where three novel fall-associated loci were identified [192]. Another 2020 study indicated that falls in people with Parkinson's disease may arise from altered cortical processing of body spatial orientation, can possibly be predicted by abnormal cortical metabolism [84]. Thus, integrating meaningful biomarkers from different data types, including clinical measures, free-living (using both ambient and wearable sensors) and other data types could be a promising avenue for the early identification of individuals with high propensity for falls. Addressing this research problem requires interdisciplinary collaboration between researchers from different disciplines.

At the time of performing the literature review on free-living FRA approaches [122], inertial sensors were the only wearable system employed for collecting older adults' data for the duration of at least 24 hours to address solutions for fall risk assessment. This thesis aimed to target some of the limitations of IMUs, e.g., in the estimation of step width [149], by integrating a body-mounted camera. While the results of Chapter 6 and Appendix C indicate that egocentric vision data potentially outperform IMUs for the provision of con-

textual data, incorporating an additional sensor modality along with a trunk-mounted IMU may inhibit older adults from following the protocol in larger-scale field studies, and negatively impact the usability of the framework, due to ethical concerns. To mitigate privacy and ethical issues associated with egocentric vision data use, although on-device detection of environmental features is not necessary for the purpose of FRA, Chapter 6 considered methods to allow processing of frames without the need for storing videos. In addition to improving the specificity of activity recognition algorithms (discussed in 2.3.3), integrating other sensor types and multimodal approaches may provide solutions with minimal obtrusiveness to identify different properties of environment.

7.3.2 An all-inclusive data-driven model, towards standardization for fall risk assessment in older adults

[1]

Chapter 2 provides a broad range of FLDBs that have been investigated by different research groups to identify measures with highest predictive powers for falls. Moreover, this thesis aimed to investigate novel FLDBs (e.g., pixel-wise step width, CBR-related measures) that can be linked to falls. Considering free-living FRA is an active research area, including generation of new markers, the critical question that remains unanswered is which FLDBs and free-living domains have the highest predictive powers for future falls. As indicated in Chapter 2, synthesizing evolving evidence is challenged by high inter-study inconsistencies. Furthermore, the utility of existing free-living FRA methods to inform interventions remains limited, largely due to challenges interpreting unconventional metrics of free-living behaviour (e.g., entropy, index of harmonicity) [123]. Addressing the aforementioned points is important, as continuous monitoring of the key FLDBs and domains (with high predictive powers for falls) during the targeted intervention programs could track an individual's responsiveness to the intervention, and ultimately reduce fall frequency.

The aforementioned gaps in the literature indicate a need to obtain a standardized model to systematically reveal the independent domains of free-living behavior. This can be investigated by performing factor analysis on a comprehensive range of wearable-based FLDBs (including the CBR-related measures) derived from a broad range of video-validated activity bouts or *context-aware* FLDBs. This comprehensive set may also include similar FLDBs but different in terms of hyperparameters (e.g., ambulatory bout lengths, turn cut-off thresholds, and central tendency measures) as discussed in Chapter 2. The forthcoming

[1] This subsection is obtained from [122]

standardized model can improve the capabilities of free-living FRAs by identifying the most important FLDBs and domains that are linked to falls, and eventually address intervention programs targeting behaviours that correspond to identified domains and ultimately prevent older adults from falling.

While a deeper discussion on developing better models by including established lab-based models of fall risk is beyond the scope of the current thesis, efforts towards a more comprehensive fall risk model leveraging both laboratory and free-living sources of evidence are on-going. Such efforts will better inform pragmatic efforts for which gait and other functional movements may be useful to identify surrogate markers of incipient pathology, inform diagnostic algorithms, track disease progression, and measure the efficacy of interventions [122].

References

[1] Girmaw Abebe and Andrea Cavallaro. Inertial-vision: cross-domain knowledge transfer for wearable sensors. In *Proc. of International Conference on Computer Vision Workshops (ICCVW), Venice, Italy*, volume 7, 2017.

[2] Girmaw Abebe and Andrea Cavallaro. A long short-term memory convolutional neural network for first-person vision activity recognition. In *Proc. of International Conference on Computer Vision Workshops (ICCVW)*, 2017.

[3] Nantheera Anantrasirichai, Jeremy Burn, and David Bull. Terrain classification from body-mounted cameras during human locomotion. *IEEE transactions on cybernetics*, 45(10):2249–2260, 2014.

[4] Edouard Auvinet, Franck Multon, Victoria Manning, Jean Meunier, and JP Cobb. Validity and sensitivity of the longitudinal asymmetry index to detect gait asymmetry using microsoft kinect data. *Gait & posture*, 51:162–168, 2017.

[5] Omar Aziz, Magnus Musngi, Edward J Park, Greg Mori, and Stephen 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. *Medical & biological engineering & computing*, 55(1):45–55, 2017.

[6] Omar Aziz, Edward J Park, Greg Mori, and Stephen N Robinovitch. Distinguishing near-falls from daily activities with wearable accelerometers and gyroscopes using support vector machines. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5837–5840. IEEE, 2012.

[7] Gillian Barry, Brook Galna, Sue Lord, Lynn Rochester, and Alan Godfrey. Defining ambulatory bouts in free-living activity: Impact of brief stationary periods on bout metrics. *Gait & Posture*, 42(4):594–597, 2015.

[8] Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala. Material recognition in the wild with the materials in context database. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3479–3487, 2015.

[9] Robert L Berg, Joseph S Cassells, et al. Falls in older persons: risk factors and prevention. In *The second fifty years: Promoting health and preventing disability*. National Academies Press (US), 1992.

[10] W Berger, V Dietz, and J Quintern. Corrective reactions to stumbling in man: neuronal co-ordination of bilateral leg muscle activity during gait. *The Journal of physiology*, 357(1):109–125, 1984.

[11] Tanvi Bhatt, JD Wening, and Y-C Pai. Influence of gait speed on stability: recovery from anterior slips and compensatory stepping. *Gait & posture*, 21(2):146–156, 2005.

[12] Simone Bianco, Remi Cadene, Luigi Celona, and Paolo Napoletano. Benchmark analysis of representative deep neural network architectures. *IEEE Access*, 6:64270–64277, 2018.

[13] Jennifer S Brach, Jaime E Berlin, Jessie M VanSwearingen, Anne B Newman, and Stephanie A Studenski. Too much or too little step width variability is associated with a fall history in older persons who walk at or near normal gait speed. *Journal of neuroengineering and rehabilitation*, 2(1):21, 2005.

[14] Biagio Brattoli, Uta Büchler, Anna-Sophia Wahl, Martin E Schwab, and Björn Ommer. Lstm self-supervision for detailed behavior analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, 2017.

[15] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[16] Matthew A Brodie, Milou J Coppens, Andreas Ejupi, Yves J Gschwind, Janneke Annegarn, Daniel Schoene, Rainer Wieching, Stephen R Lord, and Kim Delbaere. Comparison between clinical gait and daily-life gait assessments of fall risk in older people. *Geriatrics & gerontology international*, 17(11):2274–2282, 2017.

[17] Matthew A Brodie, Stephen R Lord, Milou J Coppens, Janneke Annegarn, and Kim Delbaere. Eight-week remote monitoring using a freely worn device reveals unstable gait patterns in older fallers. *IEEE Transactions on Biomedical Engineering*, 62(11):2588–2594, 2015.

[18] Feifei Bu, Jessica Abell, Paola Zaninotto, and Daisy Fancourt. A longitudinal analysis of loneliness, social isolation and falls amongst older people in england. *Scientific reports*, 10(1):1–8, 2020.

[19] David M Buchner, Eileen Rillamas-Sun, Chongzhi Di, Michael J LaMonte, Stephen W Marshall, Julie Hunt, Yuzheng Zhang, Dori E Rosenberg, I-Min Lee, Kelly R Evenson, et al. Accelerometer-measured moderate to vigorous physical activity and incidence rates of falls in older women. *Journal of the American Geriatrics Society*, 65(11):2480–2487, 2017.

[20] Susan E Carter, Elizabeth M Campbell, Rob W Sanson-Fisher, Selina Redman, and William J Gillespie. Environmental hazards in the homes of older people. *Age and ageing*, 26(3):195–202, 1997.

[21] Rossana Castaldo, Paolo Melillo, Raffaele Izzo, Nicola De Luca, and Leandro Pecchia. Fall prediction in hypertensive patients via short-term hrv analysis. *IEEE journal of biomedical and health informatics*, 21(2):399–406, 2017.

[22] Rakié Cham and Mark S Redfern. Lower extremity corrective reactions to slip events. *Journal of biomechanics*, 34(11):1439–1445, 2001.

[23] James Charles, Tomas Pfister, Derek Magee, David Hogg, and Andrew Zisserman. Personalizing human video pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3063–3072, 2016.

[24] Nabil Hajj Chehade, Pinar Ozisik, James Gomez, Fabio Ramos, and Greg Pottie. Detecting stumbles with a single accelerometer. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6681–6686. IEEE, 2012.

[25] Stacie A Chvatal and Lena H Ting. Common muscle synergies for balance and walking. *Frontiers in computational neuroscience*, 7:48, 2013.

[26] Enea Cippitelli, Samuele Gasparrini, Susanna Spinsante, and Ennio Gambi. Kinect as a tool for gait analysis: validation of a real-time joint extraction algorithm working in side view. *Sensors*, 15(1):1417–1434, 2015.

[27] Alice Coni, Sabato Mellone, Marco Colpo, Jack M Guralnik, Kushang V Patel, Stefania Bandinelli, and Lorenzo Chiari. An exploratory factor analysis of sensor-based physical capability assessment. *Sensors*, 19(10):2227, 2019.

[28] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. In *Advances in neural information processing systems*, pages 379–387, 2016.

[29] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, and Toby Perrett. Will price, and michael wray. 2018. scaling egocentric vision: The epic-kitchens dataset. In *Proceedings of the European Conference on Computer Vision*, pages 753–771.

[30] D De Venuto, VF Annese, M de Tommaso, E Vecchio, and AL Sangiovanni Vincentelli. Combining eeg and emg signals in a wireless system for preventing fall in neurodegenerative diseases. In *Ambient Assisted Living*, pages 317–327. Springer, 2015.

[31] Din Del, Brook Galna, Alan Godfrey, Esther MJ Bekkers, Elisa Pelosin, Freek Nieuwhof, Anat Mirelman, Jeffrey M Hausdorff, Lynn Rochester, 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. *Journals of Gerontology Series A: Medical Sciences*, 2017.

[32] Silvia Del Din, Alan Godfrey, Brook Galna, Sue Lord, and Lynn Rochester. Free-living gait characteristics in ageing and parkinson's disease: impact of environment and ambulatory bout length. *Journal of neuroengineering and rehabilitation*, 13(1):1–12, 2016.

[33] Silvia Del Din, Aodhán Hickey, Naomi Hurwitz, John C Mathers, Lynn Rochester, and Alan Godfrey. Measuring gait with an accelerometer-based wearable: influence of device location, testing protocol and age. *Physiological measurement*, 37(10):1785, 2016.

[34] Mirko Di Rosa, Jeff M Hausdorff, Vera Stara, Lorena Rossi, Liam Glynn, Monica Casey, Stefan Burkard, and Antonio Cherubini. Concurrent validation of an index to estimate fall risk in community dwelling seniors through a wireless sensor insole system: A pilot study. *Gait & posture*, 55:6–11, 2017.

[35] Steven Diamond, Vincent Sitzmann, Stephen Boyd, Gordon Wetzstein, and Felix Heide. Dirty pixels: Optimizing image classification architectures for raw sensor data. *arXiv preprint arXiv:1701.06487*, 2017.

[36] Jean P Diaz, Rafael L da Silva, Boxuan Zhong, He Helen Huang, and Edgar Lobaton. Visual terrain identification and surface inclination estimation for improving human locomotion with a lower-limb prosthetic. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1817–1820. IEEE, 2018.

[37] V Dietz, J Quintern, and W Berger. Corrective reactions to stumbling in man: functional significance of spinal and transcortical reflexes. *Neuroscience letters*, 44(2):131–135, 1984.

[38] Giovanni Diraco, Alessandro Leone, and Pietro Siciliano. A radar-based smart sensor for unobtrusive elderly monitoring in ambient assisted living applications. *Biosensors*, 7(4):55, 2017.

[39] Samuel Dodge and Lina Karam. Understanding how image quality affects deep neural networks. In *2016 eighth international conference on quality of multimedia experience (QoMEX)*, pages 1–6. IEEE, 2016.

[40] Christopher Moufawad El Achkar, Constanze Lenbole-Hoskovec, Anisoara Paraschiv-Ionescu, Kristof Major, Christophe Büla, and Kamiar Aminian. Classification and characterization of postural transitions using instrumented shoes. *Medical & biological engineering & computing*, 56(8):1403–1412, 2018.

[41] Christopher Moufawad el Achkar, Constanze Lenoble-Hoskovec, Anisoara Paraschiv-Ionescu, Kristof Major, Christophe Büla, and Kamiar Aminian. Instrumented shoes for activity classification in the elderly. *Gait & posture*, 44:12–17, 2016.

[42] Ahmed Elhayek, Edilson de Aguiar, Arjun Jain, J Thompson, Leonid Pishchulin, Mykhaylo Andriluka, Christoph Bregler, Bernt Schiele, and Christian Theobalt. Marconi—convnet-based marker-less motion capture in outdoor and indoor scenes. *IEEE transactions on pattern analysis and machine intelligence*, 39(3):501–514, 2017.

[43] Janice J Eng, David A Winter, and Aftab E Patla. Strategies for recovery from a trip in early and late swing during human walking. *Experimental Brain Research*, 102(2):339–349, 1994.

[44] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[45] Curtis S Florence, Gwen Bergen, Adam Atherly, Elizabeth Burns, Judy Stevens, and Cynthia Drake. Medical costs of fatal and nonfatal falls in older adults. *Journal of the American Geriatrics Society*, 66(4):693–698, 2018.

[46] Moshe Gabel, Ran Gilad-Bachrach, Erin Renshaw, and Assaf Schuster. Full body gait analysis with kinect. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1964–1967. IEEE, 2012.

[47] Irina Galperin, Inbar Hillel, Silvia Del Din, Esther MJ Bekkers, Alice Nieuwboer, Giovanni Abbruzzese, Laura Avanzino, Freek Nieuwhof, Bastiaan R Bloem, Lynn Rochester, et al. Associations between daily-living physical activity and laboratory-based assessments of motor severity in patients with falls and parkinson's disease. *Parkinsonism & related disorders*, 62:85–90, 2019.

[48] Tatjana Gazibara, Darija Kisic Tepavcevic, Marina Svetel, Aleksandra Tomic, Iva Stankovic, Vladimir S Kostic, and Tatjana Pekmezovic. Near-falls in people with parkinson's disease: Circumstances, contributing factors and association with falling. *Clinical neurology and neurosurgery*, 161:51–55, 2017.

[49] Matthias Gietzelt, Florian Feldwieser, Mehmet Gövercin, Elisabeth Steinhagen-Thiessen, and Michael Marschollek. A prospective field study for sensor-based identification of fall risk in older people with dementia. *Informatics for health and social care*, 39(3-4):249–261, 2014.

[50] Thomas M Gill, Christianna S Williams, Julie T Robison, and Mary E Tinetti. A population-based study of environmental hazards in the homes of older persons. *American Journal of Public Health*, 89(4):553–556, 1999.

[51] Tiffany Gill, Anne W Taylor, and Ann Pengelly. A population-based survey of factors relating to the prevalence of falls in older people. *Gerontology*, 51(5):340–345, 2005.

[52] Alan Godfrey. Wearables for independent living in older adults: Gait and falls. *Maturitas*, 100:16–26, 2017.

[53] Rıza Alp Güler, Natalia Neverova, and Iasonas Kokkinos. Densepose: Dense human pose estimation in the wild. *arXiv preprint arXiv:1802.00434*, 2018.

[54] David John Hallford, Geoff Nicholson, Kerrie Sanders, and Marita P McCabe. The association between anxiety and falls: A meta-analysis. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 72(5):729–741, 2017.

[55] Kathryn A Hamel, Noriaki Okita, Jill S Higginson, and Peter R Cavanagh. Foot clearance during stair descent: effects of age and illumination. *Gait & posture*, 21(2):135–140, 2005.

[56] Shirley Handelzalts, Neil B Alexander, Nicholas Mastruserio, Linda V Nyquist, Debra M Strasburg, and Lauro V Ojeda. Detection of real-world trips in at-fall risk community dwelling older adults using wearable sensors. *Frontiers in medicine*, 7:514, 2020.

[57] Klaas A Hartholt, Ed F van Beeck, Suzanne Polinder, Nathalie van der Velde, Esther MM van Lieshout, Martien JM Panneman, Tischa JM van der Cammen, and Peter Patka. Societal consequences of falls in the older population: injuries, healthcare costs, and long-term reduced quality of life. *Journal of Trauma and Acute Care Surgery*, 71(3):748–753, 2011.

[58] Muhammad Zeeshan Ul Hasnain Hashmi, Qaiser Riaz, Mehdi Hussain, and Muhammad Shahzad. What lies beneath one's feet? terrain classification using inertial data of human walk. *Applied Sciences*, 9(15):3099, 2019.

[59] Jeremiah Hauth, Safa Jabri, Fahad Kamran, Eyoel W Feleke, Kaleab Nigusie, Lauro V Ojeda, Shirley Handelzalts, Linda Nyquist, Neil B Alexander, Xun Huan, 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*, 21(14):4661, 2021.

[60] Helen Hawley-Hague, Elisabeth Boulton, Alex Hall, Klaus Pfeiffer, and Chris Todd. Older adults' perceptions of technologies aimed at falls prevention, detection or monitoring: a systematic review. *International journal of medical informatics*, 83(6):416–426, 2014.

[61] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[62] Aodhán Hickey, Silvia Del Din, Lynn Rochester, and Alan Godfrey. Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis. *Physiological measurement*, 38(1):N1–N15, 2017.

[63] Inbar Hillel, Eran Gazit, Alice Nieuwboer, Laura Avanzino, Lynn Rochester, Andrea Cereatti, Ugo Della Croce, Marcel Olde Rikkert, Bastiaan R Bloem, Elisa Pelosin,

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. *European review of aging and physical activity*, 16(1):1–12, 2019.

[64] Marjorie Johnson Hilliard, Katherine M Martinez, Imke Janssen, Beatrice Edwards, Marie-Laure Mille, Yunhui Zhang, and Mark W Rogers. Lateral balance factors predict future falls in community-living older adults. *Archives of physical medicine and rehabilitation*, 89(9):1708–1713, 2008.

[65] Ylva Hivand Hiorth, Jan Petter Larsen, Kirsten Lode, Ole-Bjørn Tysnes, Alan Godfrey, Sue Lord, Lynn Rochester, and Kenn Freddy Pedersen. Impact of falls on physical activity in people with parkinson's disease. *Journal of Parkinson's disease*, 6(1):175–182, 2016.

[66] AL Hof and Jaak Duysens. Responses of human hip abductor muscles to lateral balance perturbations during walking. *Experimental brain research*, 230(3):301–310, 2013.

[67] At L Hof, SM Vermerris, and WA Gjaltema. Balance responses to lateral perturbations in human treadmill walking. *Journal of Experimental Biology*, 213(15):2655–2664, 2010.

[68] John H Hollman, Eric M McDade, and Ronald C Petersen. Normative spatiotemporal gait parameters in older adults. *Gait & posture*, 34(1):111–118, 2011.

[69] Fay B Horak, Martina Mancini, Patricia Carlson-Kuhta, John G Nutt, and Arash Salarian. Balance and gait represent independent domains of mobility in parkinson disease. *Physical therapy*, 96(9):1364–1371, 2016.

[70] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[71] B Hu, PC Dixon, JV Jacobs, JT Dennerlein, and JM Schiffman. Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface- and age-related differences in walking. *Journal of biomechanics*, 71:37–42, 2018.

[72] B Hu, S Li, Y Chen, R Kavi, and S Coppola. Applying deep neural networks and inertial measurement unit in recognizing irregular walking differences in the real world. *Applied Ergonomics*, 96:103414, 2021.

[73] Richard Zhi-Ling Hu, Adam Hartfiel, James Tung, Adel Fakih, Jesse Hoey, and Pascal Poupart. 3d pose tracking of walker users' lower limb with a structured-light camera on a moving platform. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pages 29–36. IEEE, 2011.

[74] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al. Speed/accuracy trade-offs for modern convolutional object detectors. In *IEEE CVPR*, volume 4, 2017.

[75] Yinghao Huang. Towards accurate marker-less human shape and pose estimation over time. In *3D Vision (3DV), 2017 International Conference on*, pages 421–430. IEEE, 2017.

[76] Andrea Iaboni and Alastair J Flint. The complex interplay of depression and falls in older adults: a clinical review. *The American Journal of Geriatric Psychiatry*, 21(5):484–492, 2013.

[77] Espen AF Ihlen, Kimberley S Van Schooten, Sjoerd M Bruijn, Jaap H Van Dieen, Beatrix Vereijken, Jorunn L Helbostad, and Mirjam Pijnappels. Improved prediction of falls in community-dwelling older adults through phase-dependent entropy of daily-life walking. *Frontiers in aging neuroscience*, 10:44, 2018.

[78] Espen AF Ihlen, Aner Weiss, Yoav Beck, Jorunn L Helbostad, and Jeffrey M Hausdorff. A comparison study of local dynamic stability measures of daily life walking in older adult community-dwelling fallers and non-fallers. *Journal of biomechanics*, 49(9):1498–1503, 2016.

[79] Espen AF Ihlen, Aner Weiss, Alan Bourke, Jorunn L Helbostad, and Jeffrey M Hausdorff. The complexity of daily life walking in older adult community-dwelling fallers and non-fallers. *Journal of biomechanics*, 49(9):1420–1428, 2016.

[80] Espen AF Ihlen, Aner Weiss, Alan Bourke, Jorunn L Helbostad, and Jeffrey M Hausdorff. The complexity of daily life walking in older adult community-dwelling fallers and non-fallers. *Journal of biomechanics*, 49(9):1420–1428, 2016.

[81] Espen AF Ihlen, Aner Weiss, Jorunn L Helbostad, and Jeffrey M Hausdorff. The discriminant value of phase-dependent local dynamic stability of daily life walking in older adult community-dwelling fallers and nonfallers. *BioMed research international*, 2015, 2015.

[82] Tal Iluz, Eran Gazit, Talia Herman, Eliot Sprecher, Marina Brozgol, Nir Giladi, Anat Mirelman, and Jeffrey M Hausdorff. Automated detection of missteps during community ambulation in patients with parkinson's disease: a new approach for quantifying fall risk in the community setting. *Journal of neuroengineering and rehabilitation*, 11(1):1–9, 2014.

[83] Tal Iluz, Aner Weiss, Eran Gazit, Ariel Tankus, Marina Brozgol, Moran Dorfman, Anat Mirelman, Nir Giladi, and Jeffrey M Hausdorff. 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. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 71(11):1459–1465, 2016.

[84] Ioannis U Isaias, Joachim Brumberg, Nicoló G Pozzi, Chiara Palmisano, Andrea Canessa, Giorgio Marotta, Jens Volkmann, and Gianni Pezzoli. Brain metabolic alterations herald falls in patients with parkinson's disease. *Annals of clinical and translational neurology*, 7(4):579–583, 2020.

[85] Arjun Jain, Jonathan Tompson, Yann LeCun, and Christoph Bregler. Modeep: A deep learning framework using motion features for human pose estimation. In *Asian conference on computer vision*, pages 302–315. Springer, 2014.

[86] Barbara J Jefferis, Dafna Merom, Claudio Sartini, S Goya Wannamethee, Sarah Ash, Lucy T Lennon, Steve Iliffe, Denise Kendrick, and Peter H Whincup. Physical activity and falls in older men: the critical role of mobility limitations. *Medicine and science in sports and exercise*, 47(10):2119, 2015.

[87] Hao Jiang and Kristen Grauman. Seeing invisible poses: Estimating 3d body pose from egocentric video. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3501–3509. IEEE, 2017.

[88] Pekka Kannus, Harri Sievänen, Mika Palvanen, Teppo Järvinen, and Jari Parkkari. Prevention of falls and consequent injuries in elderly people. *The Lancet*, 366(9500):1885–1893, 2005.

[89] Holger Kantz and Thomas Schreiber. *Nonlinear time series analysis*, volume 7. Cambridge university press, 2004.

[90] Joël MH Karel, Rachel Senden, Joep EM Janssen, HHCM Savelberg, B Grimm, IC Heyligers, Ralf Peeters, and Kenneth Meijer. Towards unobtrusive in vivo monitoring of patients prone to falling. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 5018–5021. IEEE, 2010.

[91] Jeffrey Kaye, Nora Mattek, Hiroko Dodge, Teresa Buracchio, Daniel Austin, Stuart Hagler, Michael Pavel, and Tamara Hayes. One walk a year to 1000 within a year: continuous in-home unobtrusive gait assessment of older adults. *Gait & posture*, 35(2):197–202, 2012.

[92] Albert Kim, Junyoung Kim, Shirley Rietdyk, and Babak Ziaie. A wearable smartphone-enabled camera-based system for gait assessment. *Gait & posture*, 42(2):138–144, 2015.

[93] Jochen Klenk, Ngaire Kerse, Kilian Rapp, Thorsten Nikolaus, Clemens Becker, Dietrich Rothenbacher, Richard Peter, Michael Dieter Denkinger, and ActiFE Study Group. Physical activity and different concepts of fall risk estimation in older people–results of the actife-ulm study. *PloS one*, 10(6):e0129098, 2015.

[94] Thomas D Koepsell, Marsha E Wolf, David M Buchner, Walter A Kukull, Andrea Z LaCroix, Allan F Tencer, Cara L Frankenfeld, Milda Tautvydas, and Eric B Larson. Footwear style and risk of falls in older adults. *Journal of the american geriatrics society*, 52(9):1495–1501, 2004.

[95] Daniel B Kowalsky, John R Rebula, Lauro V Ojeda, Peter G Adamczyk, and Arthur D Kuo. Human walking in the real world: Interactions between terrain type, gait parameters, and energy expenditure. *PLoS one*, 16(1):e0228682, 2021.

[96] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[97] Nabil Kronfol. Biological, medical and behavioral risk factors on falls. *World Health Organisation http://www who int/ageing/project/falls_prevention_older_age/en/inded html*, 2012.

[98] Brandi S Row Lazzarini and Theodore J Kataras. Treadmill walking is not equivalent to overground walking for the study of walking smoothness and rhythmicity in older adults. *Gait & posture*, 46:42–46, 2016.

[99] Julia M Leach, Sabato Mellone, Pierpaolo Palumbo, Stefania Bandinelli, and Lorenzo Chiari. Natural turn measures predict recurrent falls in community-dwelling older adults: a longitudinal cohort study. *Scientific reports*, 8(1):4316, 2018.

[100] Anna Lee, Tanvi Bhatt, and Yi-Chung Pai. Generalization of treadmill perturbation to overground slip during gait: effect of different perturbation distances on slip recovery. *Journal of biomechanics*, 49(2):149–154, 2016.

[101] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[102] Thurmon E Lockhart, Jeffrey C Woldstad, and James L Smith. Effects of age-related gait changes on the biomechanics of slips and falls. *Ergonomics*, 46(12):1136–1160, 2003.

[103] Sue Lord, Brook Galna, Joe Verghese, Shirley Coleman, David Burn, and Lynn Rochester. Independent domains of gait in older adults and associated motor and nonmotor attributes: validation of a factor analysis approach. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 68(7):820–827, 2012.

[104] Yue Luo, Sarah M Coppola, Philippe C Dixon, Song Li, Jack T Dennerlein, and Boyi Hu. A database of human gait performance on irregular and uneven surfaces collected by wearable sensors. *Scientific data*, 7(1):1–9, 2020.

[105] Minghuang Ma, Haoqi Fan, and Kris M Kitani. Going deeper into first-person activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1894–1903, 2016.

[106] Marcos D Machado-Fragua, Ellen A Struijk, Juan-Manuel Ballesteros, Rosario Ortolá, Fernando Rodriguez-Artalejo, and Esther Lopez-Garcia. Habitual coffee consumption and risk of falls in 2 european cohorts of older adults. *The American journal of clinical nutrition*, 109(5):1431–1438, 2019.

[107] Karen Mactier, Sue Lord, Alan Godfrey, David Burn, and Lynn Rochester. The relationship between real world ambulatory activity and falls in incident parkinson's disease: influence of classification scheme. *Parkinsonism & related disorders*, 21(3):236–242, 2015.

[108] Forough Madehkhaksar, Jochen Klenk, Kim Sczuka, Katharina Gordt, Itshak Melzer, and Michael 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*, 13(4):e0195902, 2018.

[109] Brian E Maki and William E Mcilroy. Control of compensatory stepping reactions: age-related impairment and the potential for remedial intervention. *Physiotherapy theory and practice*, 15(2):69–90, 1999.

[110] Martina Mancini, Heather Schlueter, Mahmoud El-Gohary, Nora Mattek, Colette Duncan, Jeffrey Kaye, and Fay B Horak. Continuous monitoring of turning mobility and its association to falls and cognitive function: a pilot study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 71(8):1102–1108, 2016.

[111] Martina Mancini, Heather Schlueter, Mahmoud El-Gohary, Nora Mattek, Colette Duncan, Jeffrey Kaye, and Fay B Horak. Continuous monitoring of turning mobility and its association to falls and cognitive function: a pilot study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 71(8):1102–1108, 2016.

[112] Tahir Masud and Robert O Morris. Epidemiology of falls. *Age and ageing*, 30(suppl_4):3–7, 2001.

[113] Emily I McIntosh, John L Zettel, and Lori Ann Vallis. Stepping responses in young and older adults following a perturbation to the support surface during gait. *Journal of motor behavior*, 49(3):288–298, 2017.

[114] Sina Mehdizadeh, Andrea Sabo, Kimberley-Dale Ng, Avril Mansfield, Alastair J Flint, Babak Taati, and Andrea Iaboni. Predicting short-term risk of falls in a high-risk group with dementia. *Journal of the American Medical Directors Association*, 22(3):689–695, 2021.

[115] Marie-Laure Mille, Marjorie E Johnson, Katherine M Martinez, and Mark W Rogers. Age-dependent differences in lateral balance recovery through protective stepping. *Clinical Biomechanics*, 20(6):607–616, 2005.

[116] M Jane Mohler, Christopher S Wendel, Ruth E Taylor-Piliae, Nima Toosizadeh, and Bijan Najafi. Motor performance and physical activity as predictors of prospective falls in community-dwelling older adults by frailty level: application of wearable technology. *Gerontology*, 62(6):654–664, 2016.

[117] Luis Montesinos, Rossana Castaldo, and Leandro Pecchia. Wearable inertial sensors for fall risk assessment and prediction in older adults: A systematic review and meta-analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(3):573–582, 2018.

[118] Christopher Moufawad El Achkar, Constanze Lenoble-Hoskovec, Anisoara Paraschiv-Ionescu, Kristof Major, Christophe Büla, and Kamiar Aminian. Physical behavior in

older persons during daily life: Insights from instrumented shoes. *Sensors*, 16(8):1225, 2016.

[119] Franziska Mueller, Dushyant Mehta, Oleksandr Sotnychenko, Srinath Sridhar, Dan Casas, and Christian Theobalt. Real-time hand tracking under occlusion from an egocentric rgb-d sensor. In *Proceedings of International Conference on Computer Vision (ICCV)*, volume 10, 2017.

[120] Ahmed Nait Aicha, Gwenn Englebienne, Kimberley van Schooten, Mirjam Pijnappels, and Ben Kröse. Deep learning to predict falls in older adults based on daily-life trunk accelerometry. *Sensors*, 18(5):1654, 2018.

[121] Samantha Ng, Adel Fakih, Adam Fourney, Pascal Poupart, and John Zelek. Towards a mobility diagnostic tool: tracking rollator users' leg pose with a monocular vision system. In *International Conference of IEEE Engineering in Medicine and Biology Society (EMBC)*, volume 1, pages 662–666, 2009.

[122] Mina Nouredanesh, Alan Godfrey, Jennifer Howcroft, Edward D Lemaire, and James Tung. Fall risk assessment in the wild: A critical examination of wearable sensor use in free-living conditions. *Gait & Posture*, 85:178–190, 2021.

[123] Mina Nouredanesh, Alan Godfrey, and James Tung. Fall risk assessment with wearables in the wild: Towards recommended free-living outcomes. *International Society of Posture & Gait Research 2019 World Congress*, pages 52–52, 2019.

[124] Mina Nouredanesh, Alan Godfrey, and James Tung. First-person vision-based assessment of fall risks in the wild, towards fall prevention in older adults. *Journal of Computational Vision and Imaging Systems*, 5(1):1–1, 2019.

[125] Mina Nouredanesh, Katharina Gordt, Michael Schwenk, and James Tung. Automated detection of multidirectional compensatory balance reactions: A step towards tracking naturally occurring near falls. *IEEE transactions on neural systems and rehabilitation engineering*, 28(2):478–487, 2019.

[126] Mina Nouredanesh, Sunil L Kukreja, and James Tung. Detection of compensatory balance responses using wearable electromyography sensors for fall-risk assessment. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1680–1683. IEEE, 2016.

[127] Mina Nouredanesh, Aaron W. Li, Alan Godfrey, Jesse Hoey, and James Tung. Chasing feet in the wild: A proposed egocentric motion-aware gait assessment tool. In

Laura Leal-Taixé and Stefan Roth, editors, *Computer Vision – ECCV 2018 Workshops*, pages 176–192, Cham, 2019. Springer International Publishing.

[128] Mina Nouredanesh, Andrew McCormick, Sunil L Kukreja, and James Tung. Wearable vision detection of environmental fall risk using gabor barcodes. In *Biomedical Robotics and Biomechatronics (BioRob), 2016 6th IEEE International Conference on*, pages 956–956. IEEE, 2016.

[129] Mina Nouredanesh and James Tung. Machine learning based detection of compensatory balance responses to lateral perturbation using wearable sensors. In *Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE*, pages 1–4. IEEE, 2015.

[130] Mina Nouredanesh and James Tung. Imu, semg, or their cross-correlation and temporal similarities: Which signal features detect lateral compensatory balance reactions more accurately? *Computer methods and programs in biomedicine*, 182:105003, 2019.

[131] Shawn M O'Connor and Arthur D Kuo. Direction-dependent control of balance during walking and standing. *Journal of neurophysiology*, 102(3):1411–1419, 2009.

[132] Lauro Ojeda and Johann Borenstein. Non-gps navigation for security personnel and first responders. *The Journal of Navigation*, 60(3):391–407, 2007.

[133] Lauro V Ojeda, Peter G Adamczyk, John R Rebula, Linda V Nyquist, Debra M Strasburg, and Neil B Alexander. Reconstruction of body motion during self-reported losses of balance in community-dwelling older adults. *Medical engineering & physics*, 64:86–92, 2019.

[134] Lauro V Ojeda, Peter G Adamczyk, John R Rebula, Linda V Nyquist, Debra M Strasburg, and Neil B Alexander. Reconstruction of body motion during self-reported losses of balance in community-dwelling older adults. *Medical engineering & physics*, 64:86–92, 2019.

[135] Emmanuel Okafor, Pornntiwa Pawara, Faik Karaaba, Olarik Surinta, Valeriu Codreanu, Lambert Schomaker, and Marco Wiering. Comparative study between deep learning and bag of visual words for wild-animal recognition. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8. IEEE, 2016.

[136] Yoshiro Okubo, Daina L Sturnieks, Matthew A Brodie, Lionne Duran, and Stephen R Lord. Effect of reactive balance training involving repeated slips and trips on balance recovery among older adults: a blinded randomized controlled trial. *The Journals of Gerontology: Series A*, 74(9):1489–1496, 2019.

[137] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1717–1724, 2014.

[138] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.

[139] World Health Organization, World Health Organization. Ageing, and Life Course Unit. *WHO global report on falls prevention in older age*. World Health Organization, 2008.

[140] Solenne Page, Maria M Martins, Ludovic Saint-Bauzel, Cristina P Santos, and Viviane Pasqui. Fast embedded feet pose estimation based on a depth camera for smart walker. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 4224–4229. IEEE, 2015.

[141] Marco Pahor. Falls in older adults: prevention, mortality, and costs. *JAMA*, 321(21):2080–2081, 2019.

[142] Yi-Chung Pai, Tanvi Bhatt, Feng Yang, Edward Wang, and Stephen Kritchevsky. Perturbation training can reduce community-dwelling older adults' annual fall risk: a randomized controlled trial. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 69(12):1586–1594, 2014.

[143] Jane A Painter, Leslie Allison, Puneet Dhingra, Justin Daughtery, Kira Cogdill, and Leonard G Trujillo. Fear of falling and its relationship with anxiety, depression, and activity engagement among community-dwelling older adults. *American Journal of Occupational Therapy*, 66(2):169–176, 2012.

[144] Lorraine J Phillips, Chelsea B DeRoche, Marilyn Rantz, Gregory L Alexander, Marjorie Skubic, Laurel Despins, Carmen Abbott, Bradford H Harris, Colleen Galambos, and Richelle J Koopman. Using embedded sensors in independent living to predict gait changes and falls. *Western journal of nursing research*, 39(1):78–94, 2017.

[145] Diane Podsiadlo and Sandra Richardson. The timed "up & go": a test of basic functional mobility for frail elderly persons. *Journal of the American geriatrics Society*, 39(2):142–148, 1991.

[146] Yair Poleg, Chetan Arora, and Shmuel Peleg. Temporal segmentation of egocentric videos. In *CVPR*, 2014.

[147] Yair Poleg, Ariel Ephrat, Shmuel Peleg, and Chetan Arora. Compact cnn for indexing egocentric videos. In *WACV*, 2016.

[148] Tomislav Pozaic, Ulrich Lindemann, Anna-Karina Grebe, and Wilhelm Stork. Sit-to-stand transition reveals acute fall risk in activities of daily living. *IEEE journal of translational engineering in health and medicine*, 4:1–11, 2016.

[149] John R Rebula, Lauro V Ojeda, Peter G Adamczyk, and Arthur D Kuo. Measurement of foot placement and its variability with inertial sensors. *Gait & posture*, 38(4):974–980, 2013.

[150] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[151] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[152] Helge Rhodin, Christian Richardt, Dan Casas, Eldar Insafutdinov, Mohammad Shafiei, Hans-Peter Seidel, Bernt Schiele, and Christian Theobalt. Egocap: egocentric marker-less motion capture with two fisheye cameras. *ACM Transactions on Graphics (TOG)*, 35(6):162, 2016.

[153] Markus M Rieger, Selma Papegaaij, Frans Steenbrink, Jaap H Van Dieën, and Mirjam 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*, 20(1):1–12, 2020.

[154] Sietse M Rispens, Mirjam Pijnappels, Kimberley S van Schooten, Peter J Beek, Andreas Daffertshofer, and Jaap H van Dieën. Consistency of gait characteristics as determined from acceleration data collected at different trunk locations. *Gait & posture*, 40(1):187–192, 2014.

[155] Sietse M Rispens, Jaap H Van Dieën, Kimberley S Van Schooten, L Eduardo Cofré Lizama, Andreas Daffertshofer, Peter J Beek, and Mirjam Pijnappels. Fall-related gait characteristics on the treadmill and in daily life. *Journal of neuroengineering and rehabilitation*, 13(1):1–9, 2016.

[156] Sietse M Rispens, Kimberley S van Schooten, Mirjam Pijnappels, Andreas Daffertshofer, Peter J Beek, and Jaap H van Dieën. Do extreme values of daily-life gait characteristics provide more information about fall risk than median values? *JMIR research protocols*, 4(1):e4, 2015.

[157] Sietse M Rispens, Kimberley S van Schooten, Mirjam Pijnappels, Andreas Daffertshofer, Peter J Beek, and Jaap H van Dieen. Identification of fall risk predictors in daily life measurements: gait characteristics' reliability and association with self-reported fall history. *Neurorehabilitation and neural repair*, 29(1):54–61, 2015.

[158] Stephen N Robinovitch, Fabio Feldman, Yijian Yang, Rebecca Schonnop, Pet Ming Leung, Thiago Sarraf, Joanie Sims-Gould, and Marie Loughin. Video capture of the circumstances of falls in elderly people residing in long-term care: an observational study. *The Lancet*, 381(9860):47–54, 2013.

[159] Verónica Robles-García, Yoanna Corral-Bergantiños, Nelson Espinosa, María Amalia Jácome, Carlos García-Sancho, Javier Cudeiro, and Pablo Arias. Spatiotemporal gait patterns during overt and covert evaluation in patients with parkinson's disease and healthy subjects: is there a hawthorne effect? *Journal of applied biomechanics*, 31(3):189–194, 2015.

[160] Grégory Rogez, James S Supancic, and Deva Ramanan. First-person pose recognition using egocentric workspaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4325–4333, 2015.

[161] Artem Rozantsev, Vincent Lepetit, and Pascal Fua. Flying objects detection from a single moving camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4128–4136, 2015.

[162] Laurence Z Rubenstein, Christopher M Powers, and Catherine H MacLean. Quality indicators for the management and prevention of falls and mobility problems in vulnerable elders. *Annals of internal Medicine*, 135(8_Part_2):686–693, 2001.

[163] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[164] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[165] AM Schillings, BMH Van Wezel, and J Duysens. Mechanically induced stumbling during human treadmill walking. *Journal of neuroscience methods*, 67(1):11–17, 1996.

[166] AM Schillings, BMH Van Wezel, TH Mulder, and Jaak Duysens. Muscular responses and movement strategies during stumbling over obstacles. *Journal of Neurophysiology*, 83(4):2093–2102, 2000.

[167] Daniel Schoene, Sandy M-S Wu, A Stefanie Mikolaizak, Jasmine C Menant, Stuart T Smith, Kim Delbaere, and Stephen R Lord. Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. *Journal of the American Geriatrics Society*, 61(2):202–208, 2013.

[168] Michael Schwenk, Klaus Hauer, Tania Zieschang, Stefan Englert, Jane Mohler, and Bijan Najafi. Sensor-derived physical activity parameters can predict future falls in people with dementia. *Gerontology*, 60(6):483–492, 2014.

[169] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 806–813, 2014.

[170] Takaaki Shiratori, Brooke Coley, Rakié Cham, and Jessica K Hodgins. Simulating balance recovery responses to trips based on biomechanical principles. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 37–46, 2009.

[171] Takaaki Shiratori, Hyun Soo Park, Yaser Sheikh, Jessica K Hodgins, et al. Motion capture from body mounted cameras, July 22 2014. US Patent 8,786,680.

[172] Anne Shumway-Cook, Sandy Brauer, and Marjorie Woollacott. Predicting the probability for falls in community-dwelling older adults using the timed up & go test. *Physical therapy*, 80(9):896–903, 2000.

[173] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pages 568–576, 2014.

[174] Anke H Snijders, Bart P Van De Warrenburg, Nir Giladi, and Bastiaan R Bloem. Neurological gait disorders in elderly people: clinical approach and classification. *The Lancet Neurology*, 6(1):63–74, 2007.

[175] Cristina Soaz and Klaus Diepold. Step detection and parameterization for gait assessment using a single waist-worn accelerometer. *IEEE Transactions on Biomedical Engineering*, 63(5):933–942, 2015.

[176] Sibo Song, Vijay Chandrasekhar, Bappaditya Mandal, Liyuan Li, Joo-Hwee Lim, Giduthuri Sateesh Babu, Phyo Phyo San, and Ngai-Man Cheung. Multimodal multi-stream deep learning for egocentric activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 24–31, 2016.

[177] Karin Srulijes, Jochen Klenk, Michael Schwenk, Cornelia Schatton, Lars Schwickert, Kristin Teubner-Liepert, Miriam Meyer, KC Srijana, Walter Maetzler, Clemens Becker, et al. Fall risk in relation to individual physical activity exposure in patients with different neurodegenerative diseases: a pilot study. *The Cerebellum*, 18(3):340–348, 2019.

[178] Jennifer M Srygley, Talia Herman, Nir Giladi, and Jeffrey M Hausdorff. Self-report of missteps in older adults: a valid proxy of fall risk? *Archives of physical medicine and rehabilitation*, 90(5):786–792, 2009.

[179] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*, pages 127–140, 2015.

[180] Erik E Stone and Marjorie Skubic. Unobtrusive, continuous, in-home gait measurement using the microsoft kinect. *IEEE Transactions on Biomedical Engineering*, 60(10):2925–2932, 2013.

[181] Marcin Straczkiewicz, Nancy W Glynn, and Jaroslaw Harezlak. On placement, location and orientation of wrist-worn tri-axial accelerometers during free-living measurements. *Sensors*, 19(9):2095, 2019.

[182] Samuel Stuart, Lucy Parrington, Rosie Morris, Douglas N Martini, Peter C Fino, and Laurie A King. Gait measurement in chronic mild traumatic brain injury: A model approach. *Human movement science*, 69:102557, 2020.

[183] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going

deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[184] Naoto Takayanagi, Motoki Sudo, Yukari Yamashiro, Sangyoon Lee, Yoshiyuki Kobayashi, Yoshifumi Niki, and Hiroyuki Shimada. Relationship between daily and in-laboratory gait speed among healthy community-dwelling older adults. *Scientific reports*, 9(1):1–6, 2019.

[185] Kenneth Taylor, Brenda Reginatto, Matthew R Patterson, Dermot Power, Yusuke Komaba, Kazuho Maeda, Akihiro Inomata, and Brian Caulfield. Context focused older adult mobility and gait assessment. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 6943–6946. IEEE, 2015.

[186] Kenneth Taylor, Brenda Reginatto, Matthew R Patterson, Dermot Power, Yusuke Komaba, Kazuho Maeda, Akihiro Inomata, and Brian Caulfield. Context focused older adult mobility and gait assessment. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 6943–6946. IEEE, 2015.

[187] Bugra Tekin, Artem Rozantsev, Vincent Lepetit, and Pascal Fua. Direct prediction of 3d body poses from motion compensated sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 991–1000, 2016.

[188] Miguel Terroso, Natacha Rosa, Antonio Torres Marques, and Ricardo Simoes. Physical consequences of falls in the elderly: a literature review from 1995 to 2010. *European Review of Aging and Physical Activity*, 11(1):51–59, 2014.

[189] Mary E Tinetti. Performance-oriented assessment of mobility problems in elderly patients. *Journal of the American Geriatrics Society*, 34(2):119–126, 1986.

[190] Mary E Tinetti and Chandrika Kumar. The patient who falls:"it's always a trade-off". *JAMA*, 303(3):258–266, 2010.

[191] Andreas Tobola, Franz J. Streit, Chris Espig, Oliver Korpok, Christian Sauter, Nadine Lang, Björn Schmitz, Christian Hofmann, Matthias Struck, Christian Weigand, Heike Leutheuser, Björn M. Eskofier, and Georg Fischer. Sampling rate impact on energy consumption of biomedical signal processing systems. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pages 1–6, 2015.

[192] Katerina Trajanoska, Lotta J Seppala, Carolina Medina-Gomez, Yi-Hsiang Hsu, Sirui Zhou, Natasja M van Schoor, Lisette CPGM de Groot, David Karasik, J Brent Richards, Douglas P Kiel, et al. Genetic basis of falling risk susceptibility in the uk biobank study. *Communications biology*, 3(1):1–10, 2020.

[193] Erica Twardzik, Kate Duchowny, Amby Gallagher, Neil Alexander, Debra Strasburg, Natalie Colabianchi, and Philippa Clarke. What features of the built environment matter most for mobility? using wearable sensors to capture real-time outdoor environment demand on gait performance. *Gait & posture*, 68:437–442, 2019.

[194] Tzutalin. Labelimg. git code. https://github.com/tzutalin/labelImg, 2015.

[195] Kimberley S Van Schooten, Mirjam Pijnappels, Sietse M Rispens, Petra JM Elders, Paul Lips, Andreas Daffertshofer, Peter J Beek, and Jaap H Van Dieen. Daily-life gait quality as predictor of falls in older people: a 1-year prospective cohort study. *PLoS one*, 11(7):e0158623, 2016.

[196] Kimberley S Van Schooten, Mirjam Pijnappels, Sietse M Rispens, Petra JM Elders, Paul Lips, and Jaap H van Dieën. Ambulatory fall-risk assessment: amount and quality of daily-life gait predict falls in older adults. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 70(5):608–615, 2015.

[197] Kimberley S Van Schooten, Sietse M Rispens, Petra JM Elders, Paul Lips, Jaap H van Dieën, and Mirjam Pijnappels. Assessing physical activity in older adults: required days of trunk accelerometer measurements for reliable estimation. *Journal of aging and physical activity*, 23(1):9–17, 2015.

[198] Igor Vasiljevic, Ayan Chakrabarti, and Gregory Shakhnarovich. Examining the impact of blur on recognition by convolutional networks. *arXiv preprint arXiv:1611.05760*, 2016.

[199] Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, and Luc Van Gool. Temporal segment networks: Towards good practices for deep action recognition. In *European Conference on Computer Vision*, pages 20–36. Springer, 2016.

[200] Yangang Wang, Yebin Liu, Xin Tong, Qionghai Dai, and Ping Tan. Outdoor markerless motion capture with sparse handheld video cameras. *IEEE transactions on visualization and computer graphics*, 24(5):1856–1866, 2018.

[201] Aner Weiss, Marina Brozgol, Moran Dorfman, Talia Herman, Shirley Shema, Nir Giladi, and Jeffrey M Hausdorff. Does the evaluation of gait quality during daily life provide insight into fall risk? a novel approach using 3-day accelerometer recordings. *Neurorehabilitation and neural repair*, 27(8):742–752, 2013.

[202] Aner Weiss, Marina Brozgol, Nir Giladi, and Jeffrey M Hausdorff. Can a single lower trunk body-fixed sensor differentiate between level-walking and stair descent and ascent in older adults? preliminary findings. *Medical engineering & physics*, 38(10):1146–1151, 2016.

[203] Aner Weiss, Talia Herman, Nir Giladi, and Jeffrey M Hausdorff. Objective assessment of fall risk in parkinson's disease using a body-fixed sensor worn for 3 days. *PloS one*, 9(5):e96675, 2014.

[204] Aner Weiss, Ilan Shimkin, Nir Giladi, and Jeffrey M Hausdorff. Automated detection of near falls: algorithm development and preliminary results. *BMC research notes*, 3(1):1–8, 2010.

[205] Mike P Wilson. Development of satra slip test and tread pattern design guidelines. In *Slips, stumbles, and falls: pedestrian footwear and surfaces*. ASTM International, 1990.

[206] Weipeng Xu, Avishek Chatterjee, Michael Zollhoefer, Helge Rhodin, Pascal Fua, Hans-Peter Seidel, and Christian Theobalt. Mo2cap2: Real-time mobile 3d motion capture with a cap-mounted fisheye camera. *arXiv preprint arXiv:1803.05959*, 2018.

[207] Jia Xue, Hang Zhang, and Kristin Dana. Deep texture manifold for ground terrain recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 558–567, 2018.

[208] Jia Xue, Hang Zhang, Kristin Dana, and Ko Nishino. Differential angular imaging for material recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 764–773, 2017.

[209] Aron Yu and Kristen Grauman. Fine-grained visual comparisons with local learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 192–199, 2014.

[210] Christopher Zach, Thomas Pock, and Horst Bischof. A duality based approach for realtime tv-l 1 optical flow. In *Joint Pattern Recognition Symposium*, pages 214–223. Springer, 2007.

[211] Allen Zhang, Kelvin CP Wang, Baoxian Li, Enhui Yang, Xianxing Dai, Yi Peng, Yue Fei, Yang Liu, Joshua Q Li, and Cheng Chen. Automated pixel-level pavement crack detection on 3d asphalt surfaces using a deep-learning network. *Computer-Aided Civil and Infrastructure Engineering*, 32(10):805–819, 2017.

[212] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

APPENDICES

Appendix A

sEMG responses to different perturbation types

Table A.1: Biomedical principles and sEMG responses to different perturbation types. BFEM, RFEM, TA, GAS, and GMED refer to biceps femoris, rectus femoris, tibialis anterior, gastrocnemius, and gluteus medius muscles, respectively.

Study	CBR type	Methodology	Onset of perturbation	Reaction and Latency
[22]	Slip	Oily vinyl tile		increased knee flexion and hip extension moment (leading foot) $25\% - 45\%$ of stance (190-350ms after heel contact)
[43, 166]	Trip (elevating strategy)		5-25% of stride duration	BFEM (flex the swing leg) 64 msec after impact RFEM (extend the knee to touch down beyond the obstacle) 154 msec TA (ankle dorsiflexion) 75msec after impact
[43, 166]	Trip (lowering strategy)		55-75% of stride duration	RFE (push-off) 62msec BFEM (87msec) TA and Soleus (ankle dorsiflexion) by 60 msec Soleus (ankle plantar flexion for push-off) 110 msec

[43, 166]	Trip (both lowering and elevating)		30-50% of stride duration	
[10]		Treadmill deceleration		deceleration evoked a bilateral TA activation; acceleration evoked an ipsilateral gastrocnemius (GAS) and contralateral TA activation (latency in either condition and on both sides was 65-75 ms, duration about 150 ms)
[37]	Stumbles	treadmill, random short impulses (acceleration and deceleration)		acceleration was compensated for by a strong ipsilateral GAS and contralateral TA activation, and deceleration by a bilateral TA activation. In both muscles the responses appeared with a latency of about 70 msec and lasted for about 150 msec.
[165]		releasing obstacles (using electromagnet) to block the forward swinging foot	its early swing phase	mean latencies of 76 ms in both the ipsilateral BFEM and RFEM. During the perturbed swing, increased flexion in the knee occurred to lift the foot over the obstacle.
[67]	Lateral perturbation			full body stability is recovered in a maximum duration of one stride (T_{Stride} ≈1.13s) in response to a lateral perturbation
[66]	Lateral perturbation			medium response from GMED (100-150ms) long responses from GMED (170-250ms) late action (270-1,000 ms) after a perturbation
[25]	12 different directions			EMG responses of erector spinae, GMED, tensor fascia lata, RFEM, vastus medialis, BFEM, medial GAS, soleus, peroneus, and TA during self-selected speed walking occur 100-ms after the onset of the perturbation exerted in 12 different directions
[129, 126]	Lateral crossover and sidestep	pushes to shoulder		the index corresponding to the maximum value of the total acceleration (ACC) at the sternum (argmax($SVA_{ACC,Sternum}$), $SVA_{ACC,Sternum} = \sqrt{ACC_{ML}^2 + ACC_V^2 + ACC_{AP}^2}$) is a reliable indicator of the onset of a CBR event.

Appendix B

The preliminary test results obtained for Models 1 and 2 before detecting and filtering the possibly-noisy ROIs.

Table B.1: The preliminary test results obtained for Models 1 and 2 before detecting and filtering the possibly-noisy ROIs.

		Model 1			Model 2		
	#ROIs	Sensitivity	Specificity	#FP's	Sensitivity	Specificity	#FP's
FIVR D1	190	0.00	100	0	0.00	99.47	1
FIVR D2	203	0.00	99.00	2	0.00	86.63	27
FIVR D3	197	100	100	0	100	99.48	1
FIVR D4	198	0.00	98.98	2	100	98.98	2
FIVR D5	200	0.00	100	0	0.00	100	0
FIVR D6	201	100	93.50	13	100	83.00	34
FIVR D7	194	100	100	0	100	100	0
MAGFRA-W D1	173	0.00	98.26	3	100	100	0
		3 out of 8 detected		20	5 out of 8 detected		65

Appendix C

Does a Single Trunk-Mounted IMU Provide Sufficient Information on the Properties of Walking Surfaces In the Wild? (A preliminary investigation)

A preliminary study was conducted to examine k-fold and leave-one-subject-out (LOSO) cross-validation performance of: 1) a binary IMU-based classifier to differentiate between stairs and level walking, and 2) a three-class IMU-based classifier to distinguish between three level surfaces (i.e., grass, stone, and flat/even), using an open access dataset [104] as described below. The generalizability of the terrain identification model(s) with the satisfactory LOSO performance was further assessed by conducting a case series using two older adults' (OAs) data from the Multimodal Ambulatory Gait and Fall Risk Assessment in the Wild (MAGFRA-W) dataset [124].

C.1 Material and Methods

C.1.1 Dataset

Young healthy adults

An open access dataset of 30 young healthy adults (15 males, 15 females, age = 23.5±4.2 yrs), who wore six IMUs (MTw Awinda, Xsens, Enschede, Netherlands) with a sampling frequency of 100 Hz (±160 m/s^2, ±2000 deg/s) was considered here [104]. The participants walked over nine outdoor surfaces: grade (up-, down-, and cross-slopes), level (paved, stone, grass), and stairs (material: cement, up and down), at self-selected speed. All participants walked over/through the same environment 6 times resulting in 180 trials per environment (16.4±4.2 s per trial). In the present study, the inertial data collected by a lower back-mounted IMU during 540 level walking (flat/even, stone, grass) and 360 stair negotiation (up and down) trials were considered.

While data captured from multiple IMUs or a single IMU placed on other anatomical locations (e.g., right thigh) may lead to superior terrain classification accuracies [72], considering trunk has been the most common anatomical location for IMU placement for free-living gait and fall risk assessment in the literature [122], and to achieve a less obtrusive setup for longitudinal studies, data collected from a single trunk-mounted IMU were considered for investigation.

OAs (for case series)

To assess the generalizability of the terrain identification models developed using young adults' data to OAs' data, a subset of MAGFRA-W dataset was taken into account. The out-of-lab data considered here were collected by a lower back-mounted IMU (Axivity, Newcastle upon-Tyne, UK; ±8 g, ±500 rad/s and 100 Hz) as well as a waist-mounted GoPro camera (GoPro Hero 5, 30fps, wide view) in public environments within Northumbria University, during which OAs navigated through different indoor and/or outdoor environments while walking alongside a researcher. The camera (providing gold/criterion standard data here) was centered at each OA's waist by means of a belt attachment to capture top-down views of feet and the regions around them. The subset considered here includes 2 OAs' data (mean age ≈76 yrs), who walked over multiple surfaces.

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.

C.1.2 Data preprocessing and feature extraction

All data were processed using MATLAB R2021. Unit conversion was performed to obtain comparable inertial data within both datasets. For each trial selected from the young adults' dataset, each of the 6 acceleration (ACC) and angular velocity ($Gyro$) signals was detrended (DC offset removal) separately. To compensate for possible orientation changes for the trunk-mounted IMU during longer data collection in the MAGFRA-W dataset, instead of detrending the full-length signals acquired for each participant, sliding windows with the length of $15s$ were applied to each signal and the overlapping data were detrended separately. Afterwards, each detrended signal in both datasets was filtered using a low-pass butterworth filter with the cut-off frequency and order of 6 Hz and 2, respectively (as suggested in [104]).

From each of the 540 level walking and 360 stair walking trials in young adults dataset, a 5-second epoch (a 6×600 matrix) was cropped from the six detrended and filtered signals and considered for feature extraction. A similar process was taken into account for the analysis of OAs' data as discussed in section C.1.3.

Previous research showed that compliance for free-living data collection, in terms of sensor placement (location and/or orientation) may be challenging for participants. In [181], 15.6% of participants who wore accelerometers for seven days did not follow the protocol for $\geq$ 1 day(s) resulting in miscalculations of physical activity. Considering the goal of surface identification in free-living longitudinal studies, and since the data investigated in the present study were acquired from two sources with different data collection protocols, signal vector amplitude of the acceleration (SVA_{ACC}) and angular velocity (SVA_{Gyro}) signals, rather than all 6 axes, were employed to generally compensate for potential sensor misalignment. Thus, for each epoch, only 2 signals (i.e., SVA_{Acc}, SVA_{Gryro}) were taken into account.

The following 20 features were extracted from the SVA_{ACC} and SVA_{Gyro} components of each segment (in accordance with the features considered in our previous works [125, 130]): 1) range, 2) root mean square (RMS), 3) mean, 4) variance, 5) skewness, 6) kurtosis, 7) number of peaks, 8) maximum autocorrelation, 9) integral, 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 of the signal derivative, and 20) the Shannon entropy of signal derivative.

C.1.3 Experiments

A random forest (RF) model structure was chosen due to parallel processing and demonstrated robustness against nonlinear relationships. Here, RF models were developed to address 1) stair vs level walking differentiation (binary classification, dataset: $X_{900\times40}$ matrix), and 2) flat/even vs grass vs stone differentiation (3-class classification, dataset: $X_{540\times40}$ matrix).

Based on initial tests, an RF model with 19 trees (RF_{19}) showed satisfactory results on all validation datasets. MATLAB defaults were used for other parameters of the classifier.

For models' performance assessment, 10-fold, leave-one-trial-out (LOTO), and LOSO cross-validation measures were reported. The LOSO measures represent the average of results across all participants. For the binary classifier, samples corresponding to stairs were considered as 'positives', and the subsequent cross-validation metrics were obtained based on this consideration. For the three-class classifier, per-class accuracies were reported.

Validation based on OAs' data

A sliding window with the the length of 5 s and stride of 3 s was applied to the SVA_{ACC} and the corresponding SVA_{Gyro} signals attributed to two older participants. This resulted in 123 and 88 segments from participants A and B, respectively. Each segment was considered for feature extraction (as discussed in C.1.2).

RF_{19} models were trained using all samples from young adult's dataset (e.g., 540 level walking and 360 stair walking samples to train the binary classifier). To indicate that the results are not impacted by the inherent model's randomness, a 'confidence score' was defined. For each segment, the average of outputs (e.g., 1: Stair, 0: Level) from 20 RF_{19}'s was defined as the segment's confidence score. Subsequently, a segment is considered as 'stair walking' if its confidence score is $\geq$0.9 (i.e., at least 18 out of 20 RF_{19} models classified the segment as stair). The centers of these segments are highlighted by circles in Fig. C.1.

C.2 Results

Using young adults' data and LOTO cross-validation, the accuracy of 100% was achieved for the binary classifier (stair walking vs level walking). Moreover, walking patterns over flat/even, grass-, and stone-covered surfaces were detected with the LOTO accuracies of

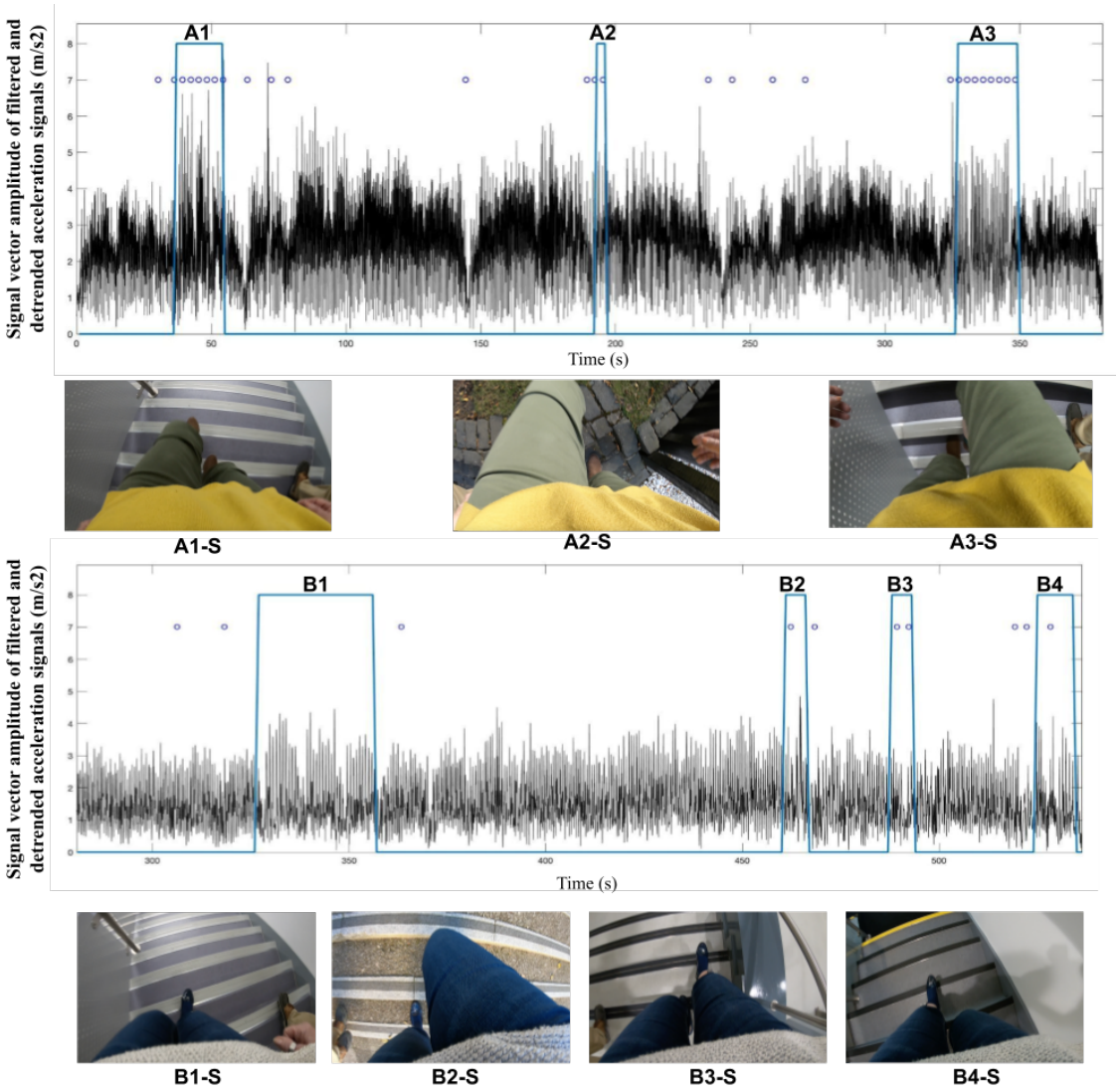


Figure C.1: Participant A and participant B walked over multiple surfaces. The starts and ends of stair walking episodes were annotated (A1 to A3, and B1 to B4). The center of each 5-second sliding window with the confidence score ≥ 0.9 was highlighted by a circle. Sample (S) frames corresponding to each stair walking activity is provided under each plot.

99.44%, 100% and 99.44%, respectively, with an overall 3-class classification accuracy: 99.63%).

The 10-fold cross-validations accuracies of 99.56% and 96.11% were achieved for the binary (stair vs level walking) and three-class (flat/even vs grass vs stone) classifiers, respectively.

For the binary classifer, the mean LOSO sensitivity, specificity, and accuracy of 81.11%, 88.15%, and 85.33% were obtained, respectively. However, for the three-class classifier, the mean LOSO accuracy dropped to 40.56% (flat/even: 46.11%, stone: 33.89%, grass: 41.67%).

The unsatisfactory LOSO performance of the 3-class classifer indicates the poor generalizability of the model even to the individuals within the same population (young adults). Thus, considering the classifier's anticipated poor performance when applied to OAs' data, only the stair detection model with promising performance was considered for further examination in the OA cases. To address this, the starts and ends of stair walking episodes were annotated (possible error$\approx \pm 2$ s) using the criterion standard data (sample frames as well as the annotations are provided in Fig. C.1). A1 (descending), A3 (ascending) and B1 (descending) show stair walking patterns captured over the same staircase (indoor, 2 subsequent staircases, each with 10 steps, a short episode of level walking on the flooring area between the staircases was recorded). In Fig. C.1, A2 (2 steps, outdoor), B2 (3 steps, outdoor), B3 (4 steps, indoor), and B4 (6 steps, indoor) show walking patterns over different indoor and outdoor steps/staircases with diverse materials (e.g., bricks).

For participant A, the stair negotiation episodes of A1, A2, and A3 were detected (the predictions with confidence scores ≥ 0.9 overlapped with the ground truth data as shown in Fig C.1). For Participant B, the stair negotiation episodes were partially detected (partial overlap between the predictions and ground truth, B1 was not detected). Multiple false positives were generated for both cases (circles out of the stair negotiation annotations in Fig. C.1).

C.3 Conclusions

Using a single trunk-mounted IMU data and random forest models, the feasibility of detecting different walking surfaces was discussed. Using young adults' data, high ($> 96\%$) 10-fold and LOTO cross-validation accuracies were obtained for both binary (stair vs gait walking) and three-class (flat/even, grass, gravel) classifiers. However, the LOSO cross-validation accuracies dropped to 85.33% and less than 50% for the binary and three-class classifiers, respectively. The considerable difference between the k-fold and LOSO cross-validation results highlights that k-fold measures may not reliably represent model's robustness against inter-participant differences that impact gait-related inertial data.

The binary classifier's satisfactory LOSO performance may be an indicator of the model's robustness against inter-participant differences (at least within the same population) suggesting that accurate differentiation between stair and level walking episodes using a trunk-mounted IMU data alone could be feasible. To further assess the generalizability of the stair walking detection model to other populations (OAs) and unseen environments (stairs with different materials and properties, e.g., riser, height and width, as well as indoor level walking surfaces) a case series was performed using 2 OAs' data, for which mixed results were achieved. On the other hand, the inferior LOSO cross-validation results for the three-class classifer (differentiation between level walking surfaces) indicated that the IMU-based models may not be robust enough against inter-participant differences associated with gait patterns. Thus, the integration of other sensor modalities may improve differentiation between different level walking surfaces.

Appendix D

Preliminary terrain type identification results using MINC-2500 and GTOS datasets

Table D.1: For *EgoTerrainNet*-Outdoor, MobileNetV2 was fine-tuned on images only from GTOS(-mobile) relevant classes with a relatively balanced distribution (pavement including asphalt, cement, stone-asphalt: 1309, Grass/Foliage including leaf, grass, dry leaf, turf: 1226, Soil: 1230, Gravel (pebble/Shale): 1266). For *EgoTerrainNet*-Indoor, all 2,500 images in 'Wood', 'Carpet' and 'Tiles' from MINC-2500 dataset were used to fine-tune the MobileNetV2. The validation accuracies of 99.20 and 87.56 for the outdoor and indoor versions were obtained, respectively. Confusion Matrices represent test results on MAGFRA-W datasets.

	Outdoor								Indoor				
		Pavement	Foliage	Gravel	Soil	Snow	Num	Acc		HFM	Tile	Wood	Num	Accuracy
P_1	Pavement	4161	1	19	920	0	5101	81.57	HFM	568	161	1349	2078	27.33
	Foliage	89	11	0	53	0	153	7.18	Tile	12	402	30	444	90.54
	Gravel	0	0	0	0	0	0	0	Wood	0	0	0	0	-
	Soil	20	4	0	15	0	39	38.46						
P_2	Pavement	-	-	-	-	-	-	-	HFM	98	1	7	106	92.45
	Foliage	-	-	-	-	-	-	-	Tiles	0	0	0	0	-
	Gravel	-	-	-	-	-	-	-	Wood	0	0	0	0	-

	Soil	-	-	-	-	-	-	-						
P_3	Pavement	47	3	1	0	0	51	92.16	HFM	48	2	10	60	80
	Foliage	1	7	0	0	0	8	87.50	Tiles	13	1	2	16	6.25
	Gravel	0	0	0	0	0	0	0	Wood	0	0	0	0	-
	Soil	3	0	0	0	0	3	0						
P_4	Pavement	1671	2	6	1672	0	3351	49.86	HFM	788	223	663	1674	47.07
	Foliage	14	23	0	8	0	45	51.11	Tiles	1	584	26	611	95.58
	Gravel	0	0	0	0	0	0	0	Wood	0	0	0	0	-
	Soil	0	0	0	0	0	0	0						
P_5	Pavement	1592	9	14	330	0	1945	81.85	HFM	1037	30	456	1523	68.08
	Foliage	28	51	0	28	0	107	47.66	Tiles	28	295	20	343	86.00
	Gravel	12	6	240	0	0	258	93.02	Wood	0	0	0	0	-
	Soil	0	0	0	0	0	0	0						
P_6	Pavement	74	2	2	0	0	78	94.87	HFM	234	3	24	261	89.65
	Foliage	21	8	0	0	0	29	27.58	Tiles	76	132	33	241	54.77
	Gravel	0	0	0	0	0	0	0	Wood	0	0	0	0	-
	Soil	0	0	0	0	0	0	0						
P_7	Pavement	1804	54	36	159	0	2053	87.87	HFM	455	91	533	1079	42.16
	Foliage	74	61	0	24	0	159	38.36	Tiles	6	212	8	226	93.80
	Gravel	33	9	309	0	0	351	88.03	Wood	0	2	45	47	95.74
	Soil	0	0	0	0	0	0	0						
P_8	Pavement	4956	200	0	1083	0	6239	79.43	HFM	243	93	91	427	56.90
	Foliage	12	139	0	0	0	151	92.05	Tiles	0	0	0	0	0
	Gravel	0	0	0	0	0	0	0	Wood	83	51	178	312	57.05
	Soil	0	0	0	0	0	0	0						
P_9	Pavement	-	-	-	-	-	-	-	HFM	604	130	253	987	61.19
	Foliage	-	-	-	-	-	-	-	Tiles	2	342	7	351	97.43
	Gravel	-	-	-	-	-	-	-	Wood	0	0	0	0	-
	Soil	-	-	-	-	-	-	-						