Detection and Prediction of Freezing of Gait in Parkinson's Disease using Wearable Sensors and Machine Learning

by

Scott Pardoel

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design 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	Erik Scheme Associate Professor, University of New Brunswick
Supervisors	Jonathan Kofman Associate Professor, University of Waterloo
	Edward Lemaire Adjunct Professor, University of Waterloo Professor, University of Ottawa
	Julie Nantel Adjunct Professor, University of Waterloo Associate Professor, University of Ottawa
Internal Member	Eihab Abdel-Rahman Professor, University of Waterloo
Internal-external Members	Michael Barnett-Cowan Associate Professor, University of Waterloo
	Andrew Laing Associate Professor, 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

Dr. Jonathan Kofman, Dr. Edward Lemaire and Dr. Julie Nantel, as Scott Pardoel's supervisors, assisted with project conceptualization, and creation of the data collection and analysis protocols. They also participated in writing and editing manuscripts submitted and published in peer-reviewed journals that served as the basis for Chapters 3 to 6 and editing all thesis chapters. The first draft was always written by Scott Pardoel, and he was involved in the editing process at all stages.

Gaurav Shalin assisted in the data collection. He also participated in editing manuscripts submitted and published in peer-reviewed journals that served as the basis for Chapters 3 to 5.

Dr. Julie Nantel and Scott Pardoel recruited all participants. Dr. Julie Nantel, Gaurav Shalin and Scott Pardoel performed all data collections. Scott Pardoel performed all data analyses presented in the thesis.

Abstract

Freezing of gait (FOG), is a brief episodic absence of forward body progression despite the intention to walk. Appearing mostly in mid-late stage Parkinson's disease (PD), freezing manifests as a sudden loss of lower-limb function, and is closely linked to falling, decreased functional mobility, and loss of independence.

Wearable-sensor based devices can detect freezes already in progress, and intervene by delivering auditory, visual, or tactile stimuli called cues. Cueing has been shown to reduce FOG duration and allow walking to continue. However, FOG detection and cueing systems require data from the freeze episode itself and are thus unable to prevent freezing. Anticipating the FOG episode before onset and supplying a timely cue could prevent the freeze from occurring altogether.

FOG has been predicted in offline analyses by training machine learning models to identify wearable-sensor signal patterns known to precede FOG. The most commonly used sensors for FOG detection and prediction are inertial measurement units (IMU) that include an accelerometer, gyroscope and sometimes magnetometer. Currently, the best FOG prediction systems use data collected from multiple sensors on various body locations to develop person-specific models. Multi-sensor systems are more complex and may be challenging to integrate into real-life assistive devices. The ultimate goal of FOG prediction systems is a user-friendly assistive device that can be used by anyone experiencing FOG. To achieve this goal, person-independent models with high FOG prediction performance and a minimal number of conveniently located sensors are needed.

The objectives of this thesis were: to develop and evaluate FOG detection and prediction models using IMU and plantar pressure data; determine if event-based or period of gait disruption FOG definitions have better classification performance for FOG detection and prediction; and evaluate FOG prediction models that use a single unilateral plantar pressure insole sensor or bilateral sensors.

In this thesis, IMU (accelerometer and gyroscope) and plantar pressure insole sensors were used to collect data from 11 people with FOG while they walked a freeze provoking path.

A custom-made synchronization and labeling program was used synchronize the IMU and plantar pressure data and annotate FOG episodes. Data were divided into overlapping 1 s windows with 0.2 s shift between consecutive windows. Time domain, Fourier transform based, and wavelet transform based features were extracted from the data. A total of 861 features were extracted from each of the 71,000 data windows.

To evaluate the effectiveness of FOG detection and prediction models using plantar pressure and IMU data features, three feature sets were compared: plantar pressure, IMU, and both plantar pressure and IMU features. Minimum-redundancy maximum-relevance (mRMR) and Relief-F feature selection were performed prior to training boosted ensembles of decision trees.

The binary classification models identified Total-FOG or Non-FOG states, wherein the Total-FOG class included windows with data from 2 s before the FOG onset until the end of the FOG episode. The plantar-pressure-only model had the greatest sensitivity, and the IMU-only model had the greatest specificity. The best overall model used the combination of plantar pressure and IMU features, achieving 76.4% sensitivity and 86.2% specificity.

Next, the Total-FOG class components were evaluated individually (i.e., Pre-FOG windows, freeze windows, and transition windows between Pre-FOG and FOG). The best model, which used plantar pressure and IMU features, detected windows that contained both Pre-FOG and FOG data with 85.2% sensitivity, which is equivalent to detecting FOG less than 1 s after the freeze began. Models using both plantar pressure and IMU features performed better than models that used either sensor type alone.

Datasets used to train machine learning models often generate ground truth FOG labels based on visual observation of specific lower limb movements (event-based definition) or an overall inability to walk effectively (period of gait disruption based definition). FOG definition ambiguity may affect FOG detection and prediction model performance, especially with respect to multiple FOG in rapid succession. This research examined the effects of defining FOG either as a period of gait disruption (merging successive FOG), or based on an event (no merging), on FOG detection and prediction. Plantar pressure and lower limb acceleration data were used to extract a set of features and train decision tree ensembles. FOG was labeled using

an event-based definition. Additional datasets were then produced by merging FOG that occurred in rapid succession. A merging threshold was introduced where FOG that were separated by less than the merging threshold were merged into one episode. FOG detection and prediction models were trained for merging thresholds of 0, 1, 2, and 3 s. Merging had little effect on FOG detection model performance; however, for the prediction model, merging resulted in slightly later FOG identification and lower precision. FOG prediction models may benefit from using event-based FOG definitions and avoiding merging multiple FOG in rapid succession.

Despite the known asymmetry of PD motor symptom manifestation, the difference between the more severely affected side (MSS) and less severely affected side (LSS) is rarely considered in FOG detection and prediction studies. The additional information provided by the MSS or LSS, if any, may be beneficial to FOG prediction models, especially if using a single sensor. To examine the effect of using data from the MSS, LSS, or both limbs, multiple FOG prediction models were trained and compared. Three datasets were created using plantar pressure data from the MSS, LSS, and both sides together. Feature selection was performed, and FOG prediction models were trained using the top 5, 10, 15, 20, 25 or 30 features for each dataset. The best models were the MSS model with 15 features, and the LSS and bilateral features with 5 features. The LSS model reached the highest sensitivity (79.5%) and identified the highest percentage of FOG episodes (94.9%). The MSS model achieved the highest specificity (84.9%) and the lowest false positive (FP) rate (2 FP/walking trial). Overall, the bilateral model was best. The bilateral model had 77.3% sensitivity, 82.9% specificity, and identified 94.3% of FOG episodes an average of 1.1 s before FOG onset. Compared to the bilateral model, the LSS model had a higher false positive rate; however, the bilateral and LSS models were similar in all other evaluation metrics. Therefore, using the LSS model instead of the bilateral model would produce similar FOG prediction performance at the cost of slightly more false positives. Given the advantages of single sensor systems, the increased FP rate may be acceptable. Therefore, a single plantar pressure sensor placed on the LSS could be used to develop a FOG prediction system and produce performance similar to a bilateral system.

Acknowledgements

Special thanks to Dr. Jonathan Kofman, Dr. Edward Lemaire and Dr Julie Nantel for the continued support and guidance they have provided as my thesis supervisors, especially during the COVID-19 pandemic. I also greatly appreciate the feedback of the examination committee members: Dr. Eihab Abdel-Rahman, Dr. Michael Barnett-Cowan, Dr. Andrew Laing and Dr. Erik Scheme. Thank you to the University of Waterloo and the University of Ottawa for the opportunity to conduct my research in an environment composed of excellent researchers with world-class engineering facilities.

I would like to thank the participants who took part in the study.

Thank you to my labmate Gaurav who was a great help through the data collection process and provided valuable insights regarding machine learning techniques.

Special thanks to Danielle Taillon and Wesley Ellis for graciously providing accommodation during the data collection as well as their invaluable support.

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada, the Ontario Ministry of Colleges and Universities, Microsoft Canada; Waterloo Artificial Intelligence Institute, and Network for Aging Research, at University of Waterloo; and the University of Waterloo. The financial support is greatly appreciated.

Dedication

To my supervisors, Dr. Jonathan Kofman, Dr. Edward Lemaire and Dr. Julie Nantel, who have been constant sources of guidance and encouragement.

To Laura, my family, and my friends Danielle and Justin who have provided continual support. I could not have done it without you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AP: Anterior-posterior

CNN: Convolutional neural network

COP: Centre of pressure

DBS: Deep brain stimulation

DFT: Discrete Fourier transform

DWT: Discrete wavelet transform

EEG: Electroencephalography

FFT: Fast Fourier transform

FI: Freeze index

FOG: Freezing of gait

GRF: Ground reaction force

GSR: Galvanic skin response

ID: Identification delay

IMU: Inertial measurement unit

LOFO: Leave one freezer out (cross validation)

LOPO: Leave one person out (cross validation)

LSS: Less severely affected side (side of the body least affected by PD)

ML: Medial-lateral

MSS: More severely affected side (side of the body most affected by PD)

MT: Merging threshold

MTD: Model trigger decision

NN: Neural network

PD: Parkinson's disease

SVM: Support vector machine

Chapter 1
Introduction

Parkinson's disease (PD) is a progressive neurodegenerative condition that presents various symptoms, including rigidity, bradykinesia (slowed movements), postural instability and tremor [1]. As the disease progresses, motor symptoms can worsen and additional symptoms such as freezing of gait (FOG) [2] may develop. A FOG episode is a sudden loss of forward body progression despite the intention to move and is often described as a sensation of having one's feet glued to the floor. FOG severely reduces mobility and can cause falls that result in serious injury [3,4]. Long term effects of FOG include fear of falling, decreased functional mobility, and loss of independence [5–7]. Interventions are needed to reduce the severity and occurrence of freezing to enable safe mobility and thereby enhance quality of life.

The precise cause of FOG is uncertain; however, triggers that increase the likelihood of freezing have been identified. These triggers include turning, walking through narrow spaces, stressful situations (such as walking in a crowd), and divided attention [8]. Cueing in the form of external auditory, visual, or tactile stimuli has improved gait parameters [9] and reduced freeze episode incidence [10]. Rhythmic cues delivered once a freeze occurs have been helpful in breaking the freeze and allowing continuation of walking [11,12]. However, constant cueing throughout the day may be distracting when a person is not walking, and cueing with a pre-set rhythm that is not matched to the intended stepping rhythm may induce FOG [13]. A preferred cueing method is one that is adaptive to the person's walking dynamics and the transient need for assistance.

FOG detection systems based on signals acquired during the freeze episode have been used to trigger cues in order to end the freeze and facilitate resumption of walking [14–16]. However, these approaches do not prevent FOG. A system that can predict FOG prior to onset is needed so that a preventative cue can be delivered.

Although FOG is typically unpredictable by visual observation, gait data from wearable sensors such as inertial measurement units (IMU) (i.e., accelerometer, gyroscope, and sometimes magnetometer) have been used to identify differences between normal PD walking characteristics and characteristics preceding FOG (Pre-FOG gait). Using features (variables

used in machine learning) calculated with data from Pre-FOG and normal PD gait, machine learning models can be trained to classify new data as either normal PD gait or Pre-FOG. Thus, FOG can be predicted by training a model to recognize data from immediately before a FOG episode. Data from body-worn IMU [17–25] and electroencephalography (EEG) [26,27] sensors have been used to predict FOG [17–28] based on the identification of movement patterns and brain activity known to be associated with Pre-FOG.

During straight line walking, changes in cyclic gait parameters such as stride length and cadence have been linked to imminent FOG [29,30]. Stride length and cadence can be easily measured using IMU sensors, and gait anomalies that indicate FOG can be identified as disruptions of the steady-state. For example, FOG has been detected by comparing stride length and cadence based features to specific thresholds [31,32]. However, distinguishing between normal PD gait and FOG is more challenging during activities such as turning and voluntary slowing (e.g., prior to stopping) [19]. During turning, cadence changes are not necessarily related to FOG, for instance, cadence can vary significantly in PD populations due to turning direction [33]. In contrast, plantar pressure features such as foot centre-of-pressure path may be useful for identifying Pre-FOG regardless of walking activity. Plantar pressure is a common and informative measure in PD gait analysis [34–42] and has recently been used for FOG detection [43–45]. Features based on deviations from normal centre of pressure movement have also been used in the fall-risk assessment of healthy elderly adults [46,47]. Since a complex interaction exists between postural stability and freezing [48], plantar pressure data may include subtle parameters linked to FOG that would be difficult to detect using IMU or EEG data, such as weight transfer changes between feet or foot centre of pressure movement [49]. Therefore, plantar pressure sensors may open new avenues in predicting FOG. In addition, in-shoe sensor integration may lead to a self-contained shoe-based system that is less obtrusive and thereby enhances end-user compliance.

FOG characteristics can vary considerably between individuals and between FOG episodes for the same individual. Therefore, developing a single model capable of predicting FOG for many individuals (person-independent) is challenging and previous models have had inadequate prediction performance [17,19]. FOG prediction models optimized for a particular

individual (person-specific models) have shown good prediction performance [25,50]; however, person-specific models may not generalize well to other individuals and may not be suitable for a prediction system intended to be used by a broad population of freezers (people with PD who experience FOG). Therefore, there is a need to improve the FOG prediction performance of person-independent FOG prediction systems.

Currently, the best FOG prediction systems use data collected from multiple sensors on various body locations. These multi-sensor systems must overcome challenges such as sensor synchronization and wireless data streaming between the sensors and a processing unit. Multi-sensor systems may also be cumbersome, and time consuming to don and doff, which could contribute to low user compliance (i.e., users abandoning the system). To reduce system complexity and improve wearability (comfort and user-friendliness), the number of body-worn sensors in a FOG prediction system should be minimized. FOG prediction models using single sensor input, such as a single shank-mounted accelerometer [21] or waist-mounted IMU [19] have been developed; however, prediction performance was worse than models using multiple sensor inputs. FOG prediction is a relatively new area of study and additional research is required to determine the feasibility of single sensor FOG prediction systems.

Sensor location may also be important since movement symptoms associated with PD can manifest asymmetrically and commonly affect one side of the body more severely. The more severely affected side (MSS) and less severely affected side (LSS) are person specific and do not necessarily correspond to the dominant leg or hand. Despite the presence of a more severe side, the distinction between more and less severely affected sides is rarely considered in FOG detection and prediction studies. Thus, there is a need to determine whether the MSS or LSS is preferable for instrumentation in a single-sensor FOG prediction system.

To further improve wearability, sensor integration should also be considered. A single plantar pressure sensor insole would be simpler than a multi-sensor system and could be integrated into regular footwear to facilitate donning and doffing.

Various model development methods have been reported in the literature for FOG detection and prediction. Frequently, detection and prediction models are set up as supervised machine-learning classifiers [51] that utilize labeled datasets containing both FOG and Non-

FOG (i.e., normal PD gait) data. Since the models learn to distinguish between classes based on the labels assigned to the training data (e.g., FOG class or Non-FOG class), accurate dataset labelling is essential. Since FOG characteristics can vary considerably [52], several FOG definitions have been used for dataset labeling. Differing definitions can result in multiple distinct datasets being produced from the same input data, with each dataset containing a different number of FOG episodes. FOG definition differences are especially apparent for data containing multiple FOG episodes in rapid succession.

Two main FOG definitions have been used: event-based definitions [14,16,53–57] and periods of gait disruption [32,58–61]. Event-based definitions have a very specific onset (e.g., foot fails to leave the ground) and termination (e.g., foot leaves the ground), and multiple consecutive FOG episodes separated by a few steps would be labeled as many separate freezes. In contrast, the "periods of gait disruption" definitions are more general and relate to functional locomotion. For example, cessation of "effective stepping" [58] does not specify exact onset and termination timing. Accordingly, multiple FOG episodes in quick succession could be considered as a single period of disrupted gait. Currently, evidence is lacking to support the decision to use an "event-based" or "period of gait disruption" approach for labeling FOG. Given the importance of ground truth labeling in classification studies, and the difference in how FOG episodes in rapid succession are handled, the possible impact of using an "event-based" or "period of gait disruption" definition should be investigated.

1.1 Rationale

Given the negative impact FOG can have on mobility, the development of assistive devices to predict and prevent FOG is very important. Although FOG is visually unpredictable, slight changes in gait characteristics have been observed in the data immediately preceding FOG. Pre-FOG walking patterns [17,62] measured using wearable IMUs have been used to predict FOG in offline analyses from data collected during walking tasks [17–25]. Currently, the best FOG prediction systems use data collected from multiple sensors on various body locations to develop person-specific models. Multi-sensor systems are more complex and may be challenging to integrate into real-life assistive devices. The ultimate goal of FOG prediction

systems is a user-friendly assistive device that can be used by anyone with PD experiencing FOG. To achieve this goal, person-independent models with high FOG prediction performance and using a minimal number of conveniently located sensors must first be developed.

Plantar pressure data has been widely used in gait and balance studies including in PD populations [34–42], and may be useful in measuring subtle changes in gait associated with an upcoming freeze that would be difficult to detect with other sensors. Plantar pressure data has been used for FOG detection [43,44]; however, it has never been used for FOG prediction. Algorithms further developed for FOG-PD gait, incorporating plantar pressure, may be able to differentiate between normal movements and gait disturbances indicating imminent FOG better than the current IMU based systems.

FOG prediction is a relatively new area of study and there is a lack of consensus regarding several important aspects of prediction model development. FOG is highly variable and different FOG definitions exist to emphasize certain characteristics of freezing. Different definitions can lead to confusion regarding what data are considered FOG, especially for FOG episodes that occur in rapid succession. The impact of the two most common types of FOG definitions (event-based definitions, and periods of gait disruption) used during dataset labeling has never been specifically examined in the context of FOG detection and prediction model performance.

Despite the known asymmetry of PD motor symptom manifestation, the difference between the MSS and LSS is rarely considered in FOG detection and prediction studies. The additional information provided by the MSS or LSS, if any, may be beneficial to FOG prediction models, especially if using a single sensor. Given the asymmetry present in PD gait, and the advantages of single-sensor FOG prediction systems, research is needed to determine if there is a preferred leg for instrumentation.

FOG can greatly limit the mobility of people with PD. Wearable FOG prediction systems could help reduce the negative effects of FOG. However, current FOG prediction systems have inadequate prediction performance and use inconsistent FOG definitions that make system evaluation and comparison challenging. This thesis addresses the need for a FOG prediction system that could be integrated into a wearable device to assist people with PD who

experience FOG. This thesis also guides future development of FOG prediction systems by examining previously under-utilized plantar pressure sensors and by critically examining assumptions and definitions used in existing FOG prediction methods.

1.2 Objectives

The objectives of this research were to:

1) Develop FOG detection and prediction models using IMU and plantar pressure data.
    a) Determine which features are most useful for FOG detection and prediction using plantar-pressure data alone, IMU data alone, and combined IMU and plantar pressure data.
    b) Compare the performance of models based on plantar-pressure data alone, IMU data alone, and combined IMU and plantar pressure data, for FOG detection.
    c) Compare the performance of models based on plantar-pressure data alone and plantar pressure data combined with IMU data, for FOG prediction.
2) Determine if event-based or "period of gait disruption" FOG definitions lead to better classification performance for FOG detection and prediction.
3) Evaluate FOG prediction models that use a single unilateral plantar pressure insole sensor and models that use bilateral sensors.
    a) Determine if models using plantar pressure data from a single foot can predict FOG with performance comparable to models that use plantar pressure data collected from both feet.
    b) Determine if models using plantar pressure data from either the more severely affected side (MSS) or the less severely affected side (LSS) produce better classification results for FOG prediction.

1.3 Contributions

This research improves the field of FOG detection and prediction using wearable sensors in several important ways. In this thesis, the following contributions were made:

1. **Compared, plantar pressure, IMU, and combined plantar pressure and IMU feature-based models for FOG detection.** Plantar-pressure sensors contributed useful

information in FOG detection models and can be effectively used for FOG identification. The research results indicated that plantar pressure sensors may be used alone in FOG detection and do not require the simultaneous use of IMU sensors.

2. **Developed a plantar-pressure based FOG detection system.** Only one study prior to this research [64] and one in parallel [43] detected FOG from plantar pressure data; however, FOG detection performance was not evaluated with respect to time after FOG onset. A FOG detection system that detects the freeze immediately after onset would be more useful than a model that only detects the end of the episode. In this research, a newly developed model was able to detect the transition between Pre-FOG and FOG, thus the system detected the very beginning of a freeze episode, which had never been done using plantar-pressure features. Detecting freeze onset facilitates early cueing and perhaps early resumption of walking.
3. **Developed the first plantar-pressure based FOG prediction system for people with Parkinson's disease[1].** Prior to this study, plantar pressure had never been used in a wearable-sensor based FOG prediction system. Plantar-pressure based model performance was found to be comparable to IMU-based models for FOG prediction. FOG is a debilitating walking disturbance and prediction of FOG allows time for FOG cueing to prevent FOG. A plantar pressure based system can be integrated into regular footwear and therefore be easier to use than systems that require multiple sensors at various body locations. This ease of use may lead to high user compliance of a wearable FOG prediction and cueing system.
4. **Compared the performance of models using 'event-based' and 'period of gait disruption' FOG definitions, and determined that merging multiple FOG episodes that occur in rapid succession had little effect on FOG detection and was detrimental to prediction.** In existing FOG detection and prediction literature, no consensus existed for defining FOG as specific gait events (no merging) or as a period

[1] A study on FOG prediction using plantar pressure data [63] was conducted by our research team from the Movement Performance Laboratory, University of Ottawa and Intelligent Human Machine Systems Laboratory, University of Waterloo in parallel with the research described in this thesis.

of gait disruption (merging successive FOG). In this research it was determined that for detection, there was little difference between merged and non merged cases. However, prediction model performance was adversely affected by merging FOG episodes. Therefore, FOG prediction models should be trained using event-based FOG definitions (e.g., foot leaves or fails to leave the ground) that consider successive FOG episodes separately.

5. **Demonstrated that data from one plantar pressure sensor can be used effectively in a FOG prediction model.** By requiring only one plantar pressure sensor on one foot instead of sensors on two feet, a wearable system could be easier to use, less costly, and eliminate delays caused by wireless communication and the need for synchronization between sensors. This research demonstrated that single sensor plantar pressure based models performed as well as two sensor models in nearly all metrics. Therefore, single-limb plantar pressure systems are viable, and potentially more desirable for FOG prediction. Furthermore, models using data from the LSS predicted FOG further in advance than models using MSS data. However, data from the more severely affected limb resulted in fewer false positives. Therefore, the preferred limb for instrumentation may be person-specific and dependent upon their need for early predictions, false positive tolerance, and ability to recover from FOG independently.

Publications resulting from this research include:

- Pardoel S, Shalin G, Nantel J, Lemaire ED, Kofman J. Selection of plantar-pressure and ankle-acceleration features for freezing of gait detection in Parkinson's disease using minimum-redundancy maximum-relevance. Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society; 2020 July 20-24; Montreal, Canada. IEEE; 2020. p. 4034–7.
- Pardoel S, Kofman J, Nantel J, Lemaire ED. Wearable-sensor-based detection and prediction of freezing of gait in Parkinson's disease: a review. Sensors. 2019;19(23):5141.

- Pardoel S, Shalin G, Nantel J, Lemaire ED, Kofman J. Early detection of freezing of gait during walking using inertial measurement unit and plantar pressure distribution data. Sensors. 2021;21(6):2246.
- Pardoel S, Shalin G, Lemaire ED, Kofman J, Nantel J. Grouping successive freezing of gait episodes has neutral to detrimental effect on freeze detection and prediction in Parkinson's disease. PLOS ONE. 2021. (Accepted)
- Pardoel S, Kofman J, Nantel J, Lemaire ED. Prediction of freezing of gait in Parkinson's disease using unilateral and bilateral plantar pressure data. (in preparation).

1.4 Thesis Outline

Chapter 1 introduced the field of wearable FOG identification systems, provided background information, and stated the rationale, objectives, and contributions of this research. Chapter 2 provides an overview of Parkinson's disease, freezing of gait, wearable sensors, features, and machine learning algorithms used for FOG detection and prediction and underlines the limitations of existing FOG identification methods and directions for improvement. Chapter 3 provides the data collection methodology including participant details, data collection hardware and protocols, and the dataset processing steps. Chapter 4 compares FOG detection models developed using IMU and plantar-pressure features. Chapter 5 analyses the effect of grouping FOG episodes in rapid succession, including the impact on detection and prediction model performance. Chapter 6 develops and compares FOG prediction models using plantar pressure data collected from one and two feet and examines the effect of using MSS and LSS instrumentation. Chapter 7 presents the thesis conclusions and outlines possible directions for future work.

Chapter 2

Literature Review

2.1 Parkinson's Disease

Parkinson's disease affects approximately 67,500 to 100,000 Canadians in the later stages of life, [65,66], with 85% of people with PD over the age of 65 [67]. Symptoms appear at an average age of 64.4 years, with medical diagnosis 1.9 years later [65]. The number of people with PD is expected to rise given the aging population in Canada [68] and elsewhere.

2.2 Freezing of Gait

In moderate to advanced PD, locomotion deteriorates into a flexed upper body posture with small shuffling steps and an anteriorly shifted centre of mass. Rigidity and slowed movement (bradykinesia) lead to decreased walking speed and step length, poor balance, increased gait variability, and in some cases FOG [36,69–72].

Freezing of gait (FOG) is a complex and highly variable phenomenon defined as "a brief, episodic absence or marked reduction of forward progression of the feet despite the intention to walk" [73]. FOG is often described as the sensation of having one's feet glued to the floor and an inability to initiate the next step. FOG becomes increasing likely as PD progresses [72,74], and is experienced by approximately 68% of individuals with advanced PD [75]. Although typically lasting only a few seconds [52], freezes can lead to falls [3,4,76] that can have immediate and lasting effects [5]. Injury, reduced mobility, fear of future falls, and decreased independence are all linked to FOG and can all contribute to a reduced quality of life [6,7,77,78]. With the cause of PD uncertain [74,79], no imminent cure, and the negative effect of reduced mobility on quality of life, research into assistive devices to improve mobility for people with FOG is important.

FOG symptom management commonly involves medication, exercise regimes, cueing devices, and in some cases, deep brain stimulation (DBS). Pharmaceutical options vary, but the most widely used medication is Levodopa (L-dopa) [1,80]. L-dopa can immediately and dramatically improve coordination and motor function for people with PD. However, the

effects are temporary and due to the degenerative nature of the disease, medication becomes less effective over time necessitating multiple doses per day [81]. Furthermore, an estimated 40% of people develop medication-induced motor fluctuations after 4-6 years of L-dopa treatment [82].

In cases where medication is ineffective or side effects from high dosages are intolerable, DBS may be used. DBS employs surgically implanted electrodes to apply voltage to specific brain structures, most commonly the subthalamic nucleus (STN), and despite its functional mechanisms being uncertain [83], STN-DBS is a well-established treatment option in individuals with tremor and rigidity as main motor symptoms [84–86]. However, the effectiveness of STN-DBS to improve postural instability and gait disorders, and more specifically to manage FOG, especially ON state FOG (i.e., when the person is on medication), remains unclear [87].

Given that FOG is resistant to medication, that FOG can occur when a person with PD is both ON and OFF medication [88,89], and that DBS requires further investigation [83], physical interventions such as cueing and gait training [90] are of particular interest.

2.3 FOG Identification

For this research, FOG identification includes both detection and prediction. FOG detection is the recognition of the freeze episode after it has begun based on the classification of data collected during the FOG episode. In contrast, FOG prediction is the recognition of the freeze episode before onset based on the classification of data collected prior to the FOG onset (Pre-FOG data).

The current gold standard in FOG detection is video analysis by a clinician or movement specialist. However, since FOG is elusive in clinical settings [25] and people cannot feasibly be under prolonged video monitoring, wearable systems have been developed to detect FOG so that treatment can be assessed and adjusted [91,92]. Real-time systems have also been used to detect freezing and activate a cue in an attempt to decrease freeze duration [14,16,31,93].

2.3.1 FOG Identification Methods Using Wearable Sensors

Current FOG identification methods vary in complexity, with the simplest methods directly comparing variables from wearable sensors to thresholds [32,53,94–99]. Threshold methods tend to have poorer detection performance but faster processing time, making them potentially useful in real-time systems [16,31,100–103]. To improve threshold-based classification, features that can better differentiate between normal PD gait and FOG (or Pre-FOG, the period just before the freeze) are desired. Features are numeric values calculated from sensor signals that capture a specific characteristic of the data. Features for FOG identification include Fourier transform based features [53,58,94–96,104–107], wavelet transform based features [27,98,101–103,108–112], various time domain features, and the widely used freeze index [94]. Features are discussed further in Section 2.4.

To further improve FOG detection performance, machine learning techniques can be used, such as neural networks [60,110,113–124], decision trees [18,54,55,93,121,125–127], random forests [117,126,128], naïve Bayes [126,128], nearest neighbour [126], and support vector machines [108,111,119,129–132]. Most machine learning models used for FOG classification are supervised models. In supervised classification, data labeled as specific classes (e.g., FOG or Non-FOG), are used to train a classification model. Unlabeled test data are then given to the trained model, which classifies the data into the appropriate classes. A FOG detection model would thus classify data as FOG or Non-FOG based on previous training examples (dataset with class labels). In addition, anomaly detection [133] and unsupervised machine learning have been used for FOG identification [134] and are discussed in Section 2.3.1.4.

The best machine learning method for FOG detection using wearable sensors has yet to be determined, but some of the best classifiers in the literature have been convolutional neural networks (CNN), CNN combined with long short-term memory recurrent neural networks, support vector machines, random forest, and AdaBoosted decision trees, as summarized in Table 2.1. A more complete list of recent research involving FOG detection using wearable sensors can be found in Table A.1 (Appendix A).

Table 2.1: Summary of top methods from studies that compared different machine learning classifiers for FOG detection using wearable sensors.

Machine learning methods tested	Best method	Second best	Third Best	Source
Random forests, decision trees, naive Bayes, *k*-nearest neighbour (KNN-l) (KNN-2), multilayer perceptron, boosting (AdaBoost) and bagging with pruned decision trees	AdaBoosted decision tree (1 s window)	Random forest (1 s window)	Bagging with decision tree (1 s window)	[126]
Naïve Bayes, random forest, decision trees, random tree	Random forest (1 s window)	Decision tree (1 s window)	Random tree (1 s window)	[128]
k-nearest neighbour, random forest, logistic regression, naïve Bayes, multilayer perceptron, SVM	Support vector machine (1.6 s window)	Random forest (1.6 s window)	Multilayer perceptron (1.6 s window)	[130,135]
CNN, decision trees with bagging, AdaBoosting, LogitBoost, RUSBoost, RobustBoost, SVM	CNN (2.56 s window)	Support vector machine (2.56 s window)	RUSBoost (2.56 s window)	[121]
Denoising autoencoder, CNN, CNN-LSTM, one-class SVM, SVM, random forest, AdaBoosted decision tree ensembles	CNN-LSTM (3.2 s window)	Random Forest (3.2 s window)	CNN (3.2 s window)	[122]

KNN: *k*-nearest neighbour, CNN: convolutional neural network, CNN-LSTM: convolutional neural network combined with long short-term memory neural network, SVM: support vector machine.

2.3.1.1 Decision Trees

Decision tree classifiers are a series of binary selections that form branches resembling a tree structure. At each decision node, a feature value is compared to a threshold, which determines the next decision node. When no decision nodes remain, the sequence stops, and a class label is assigned to the sample. Figure 2.1 shows an example of a binary decision tree with five decision nodes (N1-N5) that could be used to classify an input sample as Class 0 (Non-FOG), or Class 1 (FOG). Each input sample has four features, X1: Dominant frequency (Hz), X2: Cadence (steps/s), X3: Stride length (m) and X4: Freeze index.

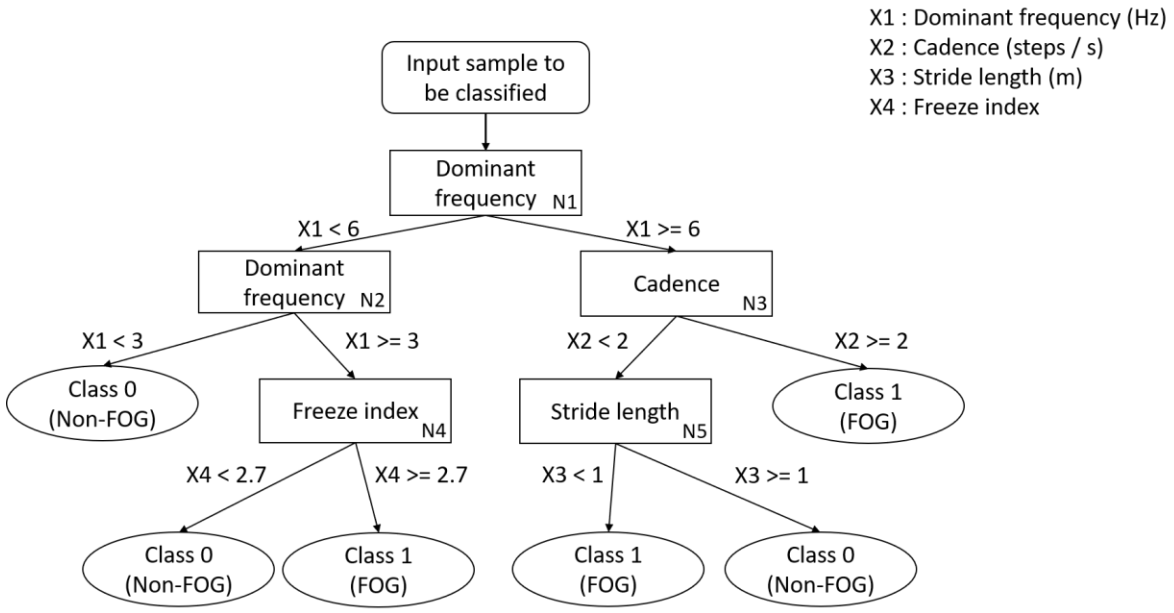

Figure 2.1: Example of a decision tree classifier, where X1-X4 are the input features.

The node thresholds and branch topology are set during model training by optimizing a split criterion, Gini impurity minimization is typically used (Equation 2.1).

$$Gini\ impurity = 1 - \sum_{i} p^2(i) \tag{2.1}$$

where i is the class index and p is the fraction of training samples in each subset belonging to class i. A Gini impurity of 0 indicates a perfect split.

To improve performance and reduce overfitting due to excessively deep trees, ensemble methods can be used. Ensemble methods train multiple weak-learners and make the final classification decision based on the majority vote of the weak learners. Boosting can also improve performance. AdaBoosting (adaptive boosting) repeatedly retrains the classifier, placing increasing importance on incorrectly classified training examples [19,136,137]. LogitBoosting (logistic boosting) [138], RUSBoosting (random undersampling boosting) [139], and RobustBoosting [140] are extensions of AdaBoosting that can further improve performance [121]. Decision tree based models for FOG detection include ensembles of trees and boosting techniques [121,126,128], with performance results ranging from 66.25% to

98.35% sensitivity and 66.00% to 99.72% specificity [18,54,55,93,121,125–128]. Decision-tree based models have also been used in FOG prediction [18,19], achieving 83.8% sensitivity and 82.1% specificity [19].

Boosted decision tree ensembles have performed well for FOG detection and have a number of advantages. First, unlike other machine learning algorithms such as support vector machines and neural networks, the inner workings of decision tree models can be easily understood. Second, the input features do not require normalization since decision node thresholds are feature specific. Finally, ensemble methods help prevent overfitting, which may occur with a single, deeper classifier. For these reasons, ensembles of decision trees were used in this research.

2.3.1.2 Support Vector Machines

Support vector machines (SVM) are binary (two class) classifiers that trace a hyperplane to separate the data points from each class (Figure 2.2).

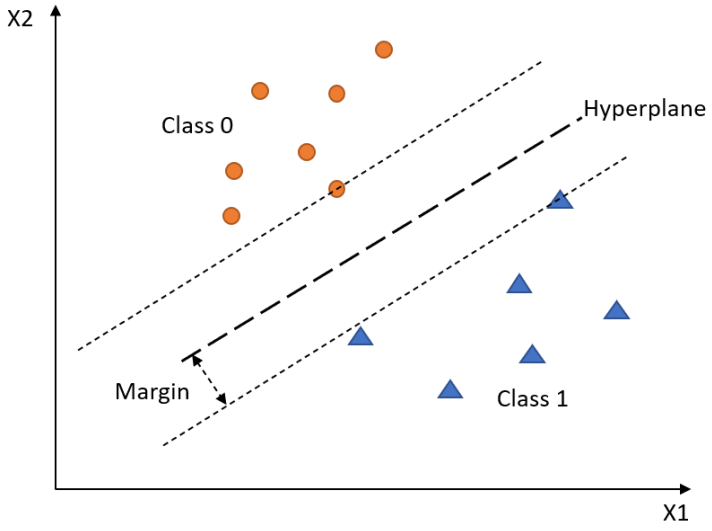

Figure 2.2: Example of SVM classifier hyperplane, where X1 and X2 are two features.

In SVM classifiers, the hyperplane separating the classes is found by maximizing the margin, which is the distance between the hyperplane and the nearest samples of each class (the support vectors) (Figure 2.2). New data points are classified based on the side of the hyperplane where they occur. If the training data classes are not easily separable, a kernel can

be used to transform the data into a dimension where the data can be linearly separable [137]. Machine learning classifiers including nearest neighbours, random forests, logistic regression, naïve Bayes, multilayer perceptron, and support vector machines have been compared, using waist-mounted accelerometer data acquired from 15 people with PD performing walking tasks and activities of daily living in their homes [130]. The machine learning methods were also compared using feature sets from three other studies [16,126,128]. The mean sensitivity and specificity was highest when using a support vector machine, regardless of the feature set [130]. This suggests that support vector machines may be well suited for FOG detection. SVM classifiers for FOG detection achieved 74.7%–99.73% sensitivity and 79.0%–100% specificity [108,111,119,129–132]. SVM classifiers have also been used in FOG prediction [23,25], and reached 89.2% sensitivity [23].

The largest drawback of SVM classifiers is that they perform poorly when the classes are not separable (i.e., when there is overlap between classes). The data can be made separable using kernel functions; however, it is difficult to determine which kernel function to use. Furthermore, since the inner workings of the models are difficult to understand, especially as dimensionality increases, model tuning can be challenging. Despite being challenging to tune, SVM classifiers have been shown to work well for FOG detection and prediction studies and were used in this research.

2.3.1.3 Neural Networks

Neural networks (NN) are made up of interconnected layers of nodes inspired by the structure of neurons in the brain [141]. NN have been frequently used in FOG detection and prediction studies. For FOG detection, NN model performance achieved 72.2%–99.83% sensitivity and 48.4%–99.96% specificity [60,110,113–124]. Neural networks for FOG prediction tended to perform slightly worse, reaching up to 86% sensitivity, 80.25% specificity, and 89% precision [26,27,135]. Different NN subtypes have been used in FOG detection and prediction, such as convolutional [121–124,142] and recurrent [20,21,122,123] networks. Convolutional neural networks (CNN) have become popular in numerous applications, including medical image analysis, in part due their ability to recognize local patterns within images and because feature

extraction prior to classification is not required [143,144]. A convolutional neural network performed well for FOG detection [123], achieving 95.1% sensitivity and 98.8% specificity. Recurrent neural networks have recently been used for FOG prediction due to their applicability to time-series data [20,21]. Recurrent neural networks utilize previous data in addition to current inputs during classification [145], thus giving the network "memory" to help recognize sequences [146]. Long short-term memory networks, a type of recurrent neural networks, have been used in FOG detection [63,122,123] and FOG prediction [20]. For prediction, a long short-term memory network achieved over 90% accuracy when predicting FOG up to 5 s in advance [20].

Neural networks have been shown to work well in FOG detection and prediction. However, NN require large amounts of data and time for training. Furthermore, similar to SVM, the inner workings of NN models are obscure which makes model tuning difficult. This research did not use NN models; however, CNN were used in a parallel research project [63] as part of a larger project at the Movement Performance laboratory, University of Ottawa and Intelligent Human Machine Systems laboratory, University of Waterloo, which this research is also part of.

2.3.1.4 Unsupervised and Semi-Supervised models

Since freezing manifests differently for each person, person-specific models usually outperform person-independent models [19,55,119,126,129]. However, person-specific models may not generalize well to other individuals, which is important for a system intended to be used by a broad population of freezers (people with PD who experience FOG). Furthermore, in practice, it is difficult to obtain enough data to develop a model for an individual. To address the lack of data, unsupervised learning has been attempted. Unsupervised methods do not rely on experts labelling FOG episodes. Instead, clustering techniques are used to define the classes [134]. Alternatively, an anomaly detection approach can be used to define the normal class and then identify abnormalities, such as FOG, that do not conform to that class [18,142]. Although unsupervised FOG detection would not require

data labelling, few studies have used this approach, and unsupervised model performance has been worse than supervised models [142].

Recently, transfer learning, and semi-supervised learning have been used to create partly personalized FOG detection methods without requiring large amounts of data. Transfer learning was used by training a neural network first using grouped data from multiple people then adding an additional network layer that was trained using an individual's data [20]. Semi-supervised learning methods use labeled data to train a base classifier before updating the model in an unsupervised manner [107,116,120]. Semi-supervised methods reduce the need for labeled data and preserve the generalization ability from a multi-person data set, while allowing person-specific tuning. Semi-supervised learning theoretically combines the advantages of both supervised and unsupervised learning. When applied to FOG detection, model classification performance achieved 89.2%–95.9% sensitivity [107,120] and 93.1%–95.6% specificity [107,120]. However, it is unclear if the reported evaluation metrics were calculated using a FOG-episode or window based approach [107,120]. The distinction between window-based and FOG episode based evaluation is discussed in Section 2.5.1. In addition, a 2 s tolerance was used that allowed classifications up to 2 s before or after the FOG episode to count as correct [120]. Therefore, although transfer and semi-supervised machine learning methods seem promising, the value of these methods for FOG detection and prediction remains unclear.

2.3.2 FOG Prediction During Walking

Five main FOG types have been identified in the literature: start hesitations (FOG during gait initiation), turn hesitation (FOG while turning), hesitation in tight quarters (FOG while passing through narrow spaces), destination-hesitation (FOG when approaching a target), and open space hesitation (FOG without apparent cause while walking in open space) [52,147]. Generally, FOG during gait initiation can be detected but not predicted due to a lack of data preceding the freeze. In contrast, FOG types that occur during walking usually have several steps preceding FOG onset. The walking immediately preceding FOG has been studied, and differences have been found between normal PD walking gait and Pre-FOG gait. For example,

prior to FOG, stride length decreases while cadence increases, often leading to festination (small, shuffling, ineffective steps) [29,30,74]. To quantify the differences between normal PD gait and Pre-FOG, various features have been calculated (Section 2.4). Using these features, machine learning models have been trained to predict FOG by detecting Pre-FOG gait [17–28]. There is no consensus regarding the precise beginning of Pre-FOG gait, and Pre-FOG labels are usually assigned to the data preceding FOG onset based on an assumed duration, as discussed next.

2.3.2.1 Prediction Approach and Pre-FOG Duration

Prediction models are typically developed by labeling the Pre-FOG gait data and training the model to detect these Pre-FOG data. FOG episodes are generally unpredictable by human observers and labelling the start of Pre-FOG visually is not possible. Instead, a FOG episode is visually identified and data prior to the freeze are labeled as Pre-FOG using a fixed period. Pre-FOG durations of 5 s [21,26,27] or 2 s [17] have been set prior to model training. Alternatively, models have been tested using multiple Pre-FOG durations ranging from 1 to 6 s, and the best performing Pre-FOG duration was selected [18,20,22,23].

Optimal Pre-FOG segment duration is difficult to determine. If Pre-FOG walking is assumed to be a degradation of gait leading to FOG (threshold theory [79]), data closest to the freeze would resemble FOG and the data farther from the freeze would resemble typical PD walking without freezing. Using long Pre-FOG durations has produced more false positives, presumably due to similarities between Pre-FOG and normal PD gait [19]. Thus, short Pre-FOG durations may perform better for FOG prediction models since data are taken closer to FOG onset and are likely more distinct from typical walking [20]. Relatively short Pre-FOG durations of 2-3 s have had better classification accuracy than longer Pre-FOG durations [23]. However, short Pre-FOG durations can result in later predictions, which limits the time in which a preventative cue can be activated [19]. In FOG prediction models, the choice of Pre-FOG duration appears to be a trade-off between early prediction and classification accuracy.

FOG is known to manifest differently in different individuals and the Pre-FOG duration that results in the best model performance has differed across participants [17–19,28]. Models

with person-specific Pre-FOG durations have performed better than fixed duration models [19]. However, the data in the study were collected while participants were OFF medication and the dataset included only 37 voluntary stops and 185 windows of pre-stop data. In addition, cadence was used to determine the optimal Pre-FOG duration [19], but it is unclear if cadence is an appropriate input to identify optimal Pre-FOG duration since cadence naturally varies during daily walking (e.g., during turning, stopping, obstacle avoidance, etc.). Further study with a larger population of freezers is required. In this thesis, fixed Pre-FOG durations were used; however, future studies may benefit from person-specific Pre-FOG duration tuning.

Various sensors have been used for FOG detection and prediction, with IMUs being the most common [51,148] (Table A.1). An IMU consists of an accelerometer, gyroscope, and sometimes a magnetometer. IMUs track movement of the limbs or body, and have been used for FOG prediction [17–22,24,25]. However, more sensitive sensor systems may improve prediction performance by detecting subtle movements or physiological parameters that cannot be easily measured with IMUs. For example, FOG prediction has been done by tracking an individual's emotional state. FOG is known to be exacerbated by stressful situations and correlations have been found between FOG and physiological variables such as heart rate [149] and galvanic skin response (GSR) [28]. GSR is a measure of the skin's electrical resistance, and decreases in response to increased perspiration, which is a physiological indicator of stress. Unlike gait-based methods, emotional state measurement may be applicable to all types of freezes (e.g., walking, turning, gait initiation), potentially making FOG prediction models more versatile. A multivariate Gaussian distribution approach that used GSR features predicted FOG with 71% accuracy, an average of 4.2 s before the freeze [28]. The predictions were made offline with the future goal of real-time implementation. Although FOG prediction using GSR was promising [28], the relevance of GSR was different in each participant. For some people, GSR was closely correlated with imminent FOG, whereas for others GSR indicators were delayed or absent [28]. In addition, it is unclear whether the usefulness of GSR would change following a cue or prolonged system use. Physiological changes in the brain, measured using electroencephalography (EEG), have also been used for FOG identification [26,27,60,115]. EEG signals combined with a Bayesian neural network have been used to predict FOG with

86% sensitivity and 80% specificity [26]. Despite promising results, EEG based systems are likely not applicable to an everyday wearable device due to the intricate system of electrodes used for data collection and related wearability issues. Simpler, more user-friendly systems are needed.

2.4 Features used for FOG Detection and Prediction

A variety of features in the time domain and frequency domain have been used in FOG detection and prediction (Table A.2). While most of these features were previously established in non-PD applications [46,150–152], custom features have been created specifically to detect FOG; such as FOG criterion (FOGC) [32,153], GaitScore [154], FOG detection on glasses (FOGDOG) [31], *k* value [59,99,155–158], *R* value [159], freeze index [94], *K* freeze index [160], and multichannel freeze index [160].

Feature calculation from wearable sensor data is typically done using data windows. Windowing involves segmenting a set of discrete data points into smaller subsets for processing. For FOG identification, window lengths range from 0.2 to 32 s [117,161], [102,103], with the most common window length being 1 s. Long windows with many sample points are desirable for calculating frequency-based features involving the discrete Fourier transform (DFT), since the number of sample points in the input signal determine the output frequency bin resolution. However, long windows decrease the temporal resolution and do not accurately represent short events within the window. In addition, long windows with many data points may be slower to process and may introduce unwanted lag between data acquisition and classification. Studies comparing multiple window lengths have found that, in general, 1-4 s windows are preferable [96,98,126,132,135,161].

2.4.1 Freeze Index

The freeze index (FI) is the most widely used feature in FOG classification studies. FOG often presents as uncontrolled shaking of the lower-limbs with little or no forward body progression [74]. Using plantar force sensors, ground reaction force (GRF) signal frequency content has been determined to be within 0-3 Hz during normal walking. In contrast, during freezing, leg shaking increased the GRF signal's frequency range [34]. The freeze index (FI) was defined

as the ratio of the area under the power spectral density (PSD) curve for the freezing band (3-8 Hz) to the area under the locomotion band (0-3 Hz), based on shank acceleration along the shin's longitudinal axis [94] (Equation 2.2).

$$FI = \frac{Area\ under\ the\ PSD\ curve\ in\ freeze\ band}{Area\ under\ the\ PSD\ curve\ in\ locomotion\ band} \quad (2.2)$$

FOG detection using the FI and person-specific thresholds has had good performance, with sensitivity of 73.1% and specificity of 81.6% [16,162]. The FI has been used with various sensor locations and has been combined with a movement threshold to differentiate walking from standing still [16]. The FI is the most well-known and widely used FOG classification feature (Table A.2) and has become a comparator for new detection methods [94]; hence, the FI was used in this research.

2.4.2 Time Domain Features

A variety of time domain features (detailed in Table A.2) have been used for FOG detection and prediction. Simple features such as mean [18,28,54,55,57,100,108,118,125,126,129–131,135] and standard deviation [17,18,23,25,28,54,55,57,100,108,116,118,120,125,126,129–131,135] have been widely used to quantify signal characteristics from multiple sensor types and locations. Features have also been used to measure movement of specific body parts (e.g., trunk angular velocity [17], foot velocity [163], medial-lateral angular jerk of the shank [23]). Stride parameter features have been used to examine walking; such as, cadence [19,58,100,104,164], step length [58,100,108] and symmetry (left-right cross correlation of medial-lateral angular velocity [17,57,61]). Other features were focused less on typical gait measurements. For example, entropy has been used to quantify the randomness in acceleration, gyroscope and EEG signals associated with freezing [18,19,23,27,57,114,125–128,132,134]. Other features combined gait variables, such as FOGC (freezing of gait criterion) [32] and FOGDOG (freezing of gait detection on glasses) [31], which used cadence and stride length to detect the progressive shortening and quickening of steps that has been observed prior to FOG [29,30,74]. An advantage of time domain features is that they are typically easily understood,

which facilitates model interpretation. However, some FOG characteristics, such as trembling of the limbs, can be better measured using frequency-based features.

Time domain features are usually calculated using either fixed-duration windows or windows normalized to steps such that each window contains one step or stride [19,23,56,64]. The distinction between fixed-duration windows and step normalized windows, (where each step is processed when the step is completed), is potentially very important, especially for real-time systems where both classification performance and classification speed are critical. For example, calculating stride duration from a window that is synchronized to the beginning and end of the stride (approximately 1 s) could result in the delayed detection of an event that occurred during the stride. Other features such as step length, cadence, cadence variation, stride peaks, and FOGC may share the limitations of step-synchronized windows, depending on the feature calculation method. In contrast, features extracted from fixed duration windows can be calculated as soon as the data window is available, which is independent of the gait cycle and determined by the size of the shift between consecutive sliding windows. Using a sliding window shift that is shorter than a stride allows a finer resolution in the time domain since multiple classifications can be made within each stride. For this reason, the features used in this research are fixed duration windows rather than windows synchronized to gait.

2.4.3 Frequency Based Features

For FOG detection and prediction, frequency based features are used to quantify signal characteristics present in the frequency domain and include standard deviation in frequency domain [56,129,132,135], spectral density center of mass [27,56,114,118,119,129,130,135], peak amplitude and corresponding frequency [95,132,165], power of the signal in specific frequency bands [16,25,54,55,93,105,118,119,125,126,131,165–167,169], and the freeze index [15–17,53,93,94,96,100,104–108,116,118–120,125,126,132,162,164,166,167]. While Fourier transforms are typically used to convert signals from the time domain to frequency domain, Fourier transform limitations have led to the increased use of wavelet approaches [27,98,101–103,108–112].

2.4.3.1 Fourier Transform

The discrete Fourier transform (DFT) approximates a given signal through summation of scaled sine waves (Equation 2.3). This is used to examine the distribution of the signal power across different frequency bands. For FOG detection and prediction, DFT has been calculated in a given window from acceleration, derivative of knee angle or angular velocity in the sagittal plane and FI [56,58,94,132]. The DFT has also been used to preprocess data windows prior to being classified using a CNN [121,122].

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{i2\pi kn}{N}} \qquad k = 0, \dots, N-1 \tag{2.3}$$

where *x* is an input sequence of *N-1* equally spaced samples and *X* is the output of the same length. Since DFT represents a given signal as a summation of constant sine waves, signals that are short relative to the length of the window are not well represented. Large time windows are desirable since more data points creates higher resolution in the frequency domain. However, if the window size is too large, entire FOG episodes can be missed, since short periods of higher frequencies that may correspond to FOG, are difficult to represent using constant sine waves. The resulting frequency spectrum will contain mostly the lower frequencies, which are relatively unchanging within the window, and are therefore well represented in the frequency domain [16]. Larger windows also require more processing, which would be a disadvantage in a real-time system. Generally, window lengths of 1-5 s are used [16,126,130], which is considered a good balance between resolution in the frequency domain, resolution in the time domain (being able to detect short freezes), and computation time.

Despite window length limitations, the DFT remains widely used in FOG detection [14–16,94,104,121,122,125,126,164,168], especially for the freeze index [15–17,19,24,25,53,57,61,93,94,96,100,104–108,116,118–120,125,126,132,164,166,167,169], and is included in this research.

2.4.3.2 Wavelet Transform

Wavelet transforms (WT) use short wave segments called wavelets instead of constant sine waves. Through the choice of starting wavelet, shifting in time, and scaling in frequency, wavelets could better capture sudden movements sometimes present in FOG, and have been used in FOG detection systems [98,101,102,108,110,111,164].

The discrete wavelet transform (DWT) uses low-pass and high-pass filters to generate vectors of coefficients. The low pass filter generates the approximation coefficients, and the high-pass filter gives the detail coefficients. Mean, minimum, maximum and variance of the coefficient vectors have performed well as FOG identification features [27,101–103,108–112]. Therefore, WT features are used in this research.

2.4.4 Feature Sets and Feature Selection

Individual FOG detection features tend to be tuned to specific types of freezes (e.g., FOG during straight line walking) [32]. To be applicable to more walking conditions, and more representative of the wide range of FOG manifestations, a larger set of features can be considered. Time domain features have been used to quantify gait parameters such as step length [58,100,108] cadence [164], asymmetry [18], and peak limb angular velocity [116,120], whereas frequency domain features can capture small movements characteristic of FOG, such as trembling in specific frequency bands [94]. Combining multiple features from the time and frequency domains can provide the benefits of both feature groups. However, an excessive number of features or complex features requiring many calculation stages may induce unacceptable delays when computing power is limited, as in many wearable systems. Using a minimal number of easily calculated features is desirable; however, too few or overly simple features may adversely impact classification performance and classifier generalizability. To address the delicate balance of classification performance and classification speed, feature selection algorithms can be used to determine the best features from a larger set. Algorithms such as Relief-F [170] or correlation-based approaches can be used to rank features according to their relevance so that the least relevant can be eliminated [171]. Feature selection is commonly used in FOG identification studies [17,18,26–28,55,60,109,111,112,114,115,118,119]. In this thesis,

a set of features in the time and frequency domains, were calculated, followed by feature selection to eliminate the least useful features and reduce the number of input features.

2.5 Limitations and Challenges of Current FOG Identification Methods

2.5.1 FOG Detection Methods

Some FOG detection methods presented in Section 2.3.1 and Table A.1 achieved excellent performance, with sensitivity and specificity above 95% [22,43,56,123,126,132,155]. Despite high sensitivity and specificity, FOG detection studies have used different performance metrics, which complicates performance comparisons. For example, a FOG detection system used to trigger a real-time cue during walking might emphasize freeze onset detection. This detection system might attempt to classify every data point or window as FOG or Non-FOG [16,162,166,167]. Incorrect classification of individual windows would influence performance metrics such as sensitivity and specificity. In contrast, a long-term monitoring system may treat each freeze occurrence as a binary event and evaluate whether or not the FOG episode was successfully detected [43,129,131]. In episode-based evaluations, the model is not required to identify each window correctly; instead, some incorrectly classified windows may be ignored so long as the episode is detected. Thus, compared to window-based evaluations, FOG episode evaluations can lead to better performance results for the same model. Model performance evaluations can also be influenced by various experimental procedures and underlying definitions. Two examples are ignoring FOG shorter than 3 s [128], and calculating specificity using only data from participants who did not freeze during testing, which may not reflect the model's true performance when used on freezers, since gait from people who froze would likely generate more false positives [132]. Differences in evaluation metrics and procedures make comparisons of FOG detection method performance more difficult.

When evaluating a classification system, ideally, different data should be used for training and testing to prevent model performance overestimation that can occur when the model is evaluated with data previously used in model training [21,26,27,60,93,109,112–114,121,132,160,172]. Cross-validation is often used when dataset size is limited [16–18,54,55,116,119,120,125–131,134,142,162,166,167]. For FOG research, leave-one-person-

out (LOPO) cross-validation is the most common. In LOPO cross validation, model training uses data from all but one participant and model testing uses data from the remaining participant. The process is repeated for each participant and the model performance results for each person are averaged. Some FOG identification studies, often more preliminary in nature, used ad hoc optimization to tune parameters and set thresholds [28,43,96,98,105,156–158,161,173]. Ad hoc optimization, although useful for initial system assessment, is not a good indicator of classifier performance, and should be followed by a more robust evaluation scheme, such as cross-validation.

For people with PD, the antiparkinsonian medication state (ON or OFF) can have a substantial effect on motor control, gait patterns, physical abilities, and FOG. Freezing occurs more frequently in the OFF state than the ON state. In the OFF state, smaller shuffling steps are common, whereas in the ON state, many people can walk fairly normally. Since freezing occurs more frequently when OFF medication, FOG identification models are often developed using OFF medication walking data due to the increased number of FOG episodes available for training. However, during daily walking, individuals are typically ON medication to enable their activities of daily living. Therefore, using OFF medication data to train a FOG identification model that will be used during the ON medication state may be detrimental to classification performance [23]. Given that medication is needed in PD management, and any wearable cueing device would primarily be used while participants are ON medication, in this research, data collection was performed while participants were in the ON medication state.

Following data collection, FOG episodes are typically visually identified and labeled. These labels serve as the ground truth for model training and system validation. Even though FOG is a well-defined clinical phenomenon [73], the criteria for defining the beginning and end of FOG episodes are not always described in FOG identification studies. Differing FOG definitions make comparisons between studies difficult.

Based on the FOG detection literature, this research used sensitivity and specificity as evaluation metrics, leave-one-out cross validation, and data collected while participants were in the ON medication state.

2.5.2 FOG Prediction Methods

FOG prediction studies have utilized various machine learning methods including decision trees [18,19], support vector machines [23,25], *k*-nearest neighbours [22,23,25], linear discriminant analysis [17,22,23], neural networks [20,21,26,27] and statistical tests [24]. While good classification performance has been reported [22–24,26], existing FOG prediction models are somewhat preliminary in nature and real-time prediction has not been reported. Additionally, FOG prediction studies are generally not evaluated in terms of a real-life application. For instance, sensitivity and specificity are often reported for Pre-FOG window classification; however, the number of episodes predicted, how early the predictions occurred, and the number of false positives are frequently absent or under reported. Therefore, there is uncertainty regarding how far in advance FOG can be predicted. Predictions as early as 6 steps prior to FOG onset for OFF medication participants and 4 steps prior to FOG onset for ON medication participants have been reported [23]. However, the data were collected during the walking portion of a 7 m timed up and go test that involved mostly straight-line walking and a single turn. The prediction performance for more complex walking involving obstacles and multiple turns is therefore unknown. Other studies, that specifically calculated the delay between FOG identification and freeze onset, showed that the majority of predictions were made 0.5 s before onset [44], and that earlier classifications accounted for fewer predictions [25]. However, many studies did not differentiate between Pre-FOG duration and prediction time. As previously stated, FOG prediction is typically done by detecting Pre-FOG data. Crucially, Pre-FOG duration does not necessarily reflect the prediction time. Thus, if a model is trained using data from 5 s before FOG, this does not necessarily mean that the model can predict FOG 5 s before FOG onset [20,26,27]. In this thesis, the delay between FOG predictions and FOG onset was calculated to quantify how far in advance predictions were made.

Current FOG prediction models are mostly person-specific [19,20,24,28], or use a cross validation procedure in which a single person's data is in both training and testing datasets [22]. Tuning classifiers to individuals has been shown to improve FOG identification performance [19,55,119,126,129]; however, this may be challenging for real-life wearable

device implementation since a calibration session would be required to collect person-specific data before the model could be tuned to the individual. More importantly, current models are unable to achieve real-time FOG prediction. Therefore, models should first be person-independent, so that robust and versatile algorithms can be created. Once real-time prediction has been achieved, personalization could be implemented to tune the models to boost performance. In this research, the models are person-independent, and personalization could be investigated in future work.

2.6 Potential Improvements for FOG Detection and Prediction

2.6.1 Plantar Pressure Measurement for FOG Prediction

Plantar pressure is a common and informative measure in PD gait analysis [34–42] and has recently been used for FOG detection [43–45]; however, plantar pressure has not been explored in FOG prediction. Ground reaction force (GRF) from pressure sensors under the foot, usually under heel and forefoot, have been used to calculate stride parameters that are related to FOG [36,39,41,42,174]. Rapid GRF changes caused by shaking of the legs during freezing have been useful in characterizing FOG [34,36], and contributed to creating the FI [94]. With multiple pressure sensors under each foot, more descriptive features, such as the centre of pressure, can be calculated to detect abnormal walking [38] or to identify specific activities such as gait initiation and termination [175]. Recently, high resolution pressure sensing insoles have been used in the fall-risk assessment of healthy elderly adults [46,47]. Features extracted from the pressure sensing insoles, such as deviations of the centre of pressure (COP) path [46,47], may be applicable to FOG identification and may be especially useful in differentiating between normal PD gait and imminent FOG. Features extracted from wearable in-shoe plantar-pressure sensors that have been used to characterize or detect FOG in PD, and asses elderly fall risk [46,47] are presented in Table A.3.

While gait parameters such as cadence and stride length have been used to predict FOG offline in steady state walking, these measures may be less relevant to FOG prediction during more transient activities such as turning or voluntary slowing. In contrast, foot centre of pressure path features could potentially be used to identify Pre-FOG regardless of the walking

activity. Plantar pressure data has been used in models that detected differences between fallers and non-fallers in healthy elderly adults [46], and has been used in FOG detection [43–45]. Therefore, plantar pressure data may be helpful for FOG prediction [44,176], and are used in this research.

2.6.2 FOG Definition and Merging

FOG detection and prediction models are frequently set up as supervised machine-learning classifiers [51] that utilize training datasets containing both FOG and Non-FOG data (i.e., steps without freezing). Therefore, accurate manual labelling of the dataset as FOG or Non-FOG is essential. Unfortunately, FOG characteristics can vary considerably between individuals and between FOG episodes for the same individual. FOG can occur with small shuffling steps, trembling in place, or with a complete lack of movement (akinetic) [52]. The FOG definition "an episodic inability (lasting seconds) to generate effective stepping…" has been used by other researchers [53,58] and encompasses the shuffling, trembling in place, and akinetic FOG subtypes [52]. However, the definition relies on subjective judgement of "effective" walking and, even when performed by experts, visual FOG assessment is prone to inter-rater discrepancies, especially between different clinical teams [177]. Despite this, expert assessments likely capture the majority of gait deviations and are sufficient for FOG detection, as evidenced by good detection performance of the resulting models [27,32,58–60,95,96,99–102,104,108,115,117,118,120,125,128,131,153,158,159,161,163,165]. However, FOG prediction cannot be approached the same way since the period before a freeze cannot be easily identified visually. Instead, FOG prediction ground truth is typically identified by selecting a period of walking data immediately before FOG onset (as described in Section 2.3.2.1). Models are trained to differentiate between this Pre-FOG gait, FOG episodes, and normal PD walking [51]. Appropriate ground truth labelling can improve the model training dataset and allow reproducibility and comparison between different studies.

Table 2.2 presents various definitions used for FOG ground truth labeling in FOG detection and prediction studies. Key phrases such as "episodic inability to generate effective stepping" [58], or "stop in alternating left-right stepping" [14,54,55], can be subjective and

leave room for ambiguity regarding what is considered an "effective" step, especially when activities other than straight line walking are performed, where normal "alternating left-right stepping" is intentionally disrupted (e.g., changing speed or direction, obstacle avoidance). Ambiguity also occurs for festination and small shuffling steps, which are a common FOG subtype [52] and may not be considered as freezes according to some definitions [53,57]. Table 2.2 also presents definitions used in FOG detection and prediction studies that are more specific or encompass multiple FOG subtypes. The definition used by [57] lists different ways a freeze might present (e.g., no foot movement, heel lifting while toes stay on the ground, irregular turning rhythm while the pivot foot stays on the ground [57]), whereas [32,56,153,159] use multiple FOG labels according to different types or severities of FOG instances.

Table 2.2: FOG definitions in FOG detection and prediction studies.

FOG Definitions	**Source**
"The beginning of a FOG event was detected when the gait pattern (i.e., alternating left–right stepping) was arrested, and the end of FOG was defined as the point in time at which the pattern was resumed" (authors reference [52])	[16]
"…the moment of arrested gait pattern, i.e., stop in alternating left-right stepping, as start of a FOG episode, and the instant when the patient resumed a regular gait pattern as end of FOG"	[14,54,55]
"…an episodic inability to generate effective stepping" (authors reference [178])	[58]
"… an unintentional and temporary phenomenon where the feet failed to progress" (authors reference [52,178,179])	[53]
"… an absolute cessation or marked reduction of forward progression of the feet despite the intention to walk" (authors reference [73])	[60]
"… paroxysmal interruption of stride or marked reduction in forward feet progression"	[59]
"… an epoch of time in which patients suddenly became unable to make a turn inside a taped 1 m^2 box on the floor, despite the intention to do so" (authors reference [60])	[115]
"…when the gait pattern (alternating right and left steps) was arrested or if it appeared as if they were trying unsuccessfully to initiate or continue locomotion/turn. The end of an episode was defined as the time when an effective step had been performed and followed by continuous locomotion."	[61]
Definitions including subtypes	
"(1) slight modification of the gait with no falling risk (green); (2) main gait modification with falling risk (orange); (3) FOG gait is blocked with or without festination (red)."	[32,153]
"… an intention to walk without movement of the feet, or as heel lifting while toes stay on the ground, or an irregular turning rhythm while the pivot foot stays on the ground" (authors reference [52,58,180])	[57]
"… each stride is classified at the output as one of the six types: normal, $short^+$ (similar to, but shorter than 'normal' strides), $short^-$ (very short forward movements, up to 20	[56]

cm, with frequencies of the movement in the low (locomotor) band), FOG^+ (FOG with knee trembling/tremor), FOG^- (FOG with complete motor block), and progressive shortening of stride while turning (PST).” *	
No definition provided, however, a distinction is made between trembling in place and shuffling forward FOG subtypes.	[159]

* Locomotor band refers to the 0 – 3 Hz frequency range.

The definitions in Table 2.2 can be broadly grouped as event-based [14,16,53–57] or periods of gait disruption [32,58–61,115,153]. The event-based definitions focus on specific behaviors of the limbs, such as cessation of foot advancement [53] or failure of the stepping foot to leave the ground [57]. Event-based definitions have a very specific onset (e.g., foot fails to leave the ground) and termination (e.g., foot leaves the ground); however, shuffling FOG or multiple consecutive FOG episodes separated by a few steps would be labeled as many separate freezes that may be more appropriately classified as a single FOG episode. In contrast, the “periods of gait disruption” definitions are more general and relate to functional locomotion. For example, cessation of “effective stepping” [58] does not specify exact onset and termination timing. Accordingly, shuffling FOG and multiple FOG episodes in quick succession could be considered as a single period of disrupted gait.

In FOG detection and prediction studies, FOG episodes are labeled and datasets are subjected to various assumptions (e.g., ignoring short FOG [57,128]) and pre-processing steps (e.g., merging FOG episodes [57] or window homogeneity requirements [121]) to refine which frames or data windows are considered as FOG. Since very short duration FOG can be difficult to detect using automatic systems [58] or could be considered a minor gait disturbance, some researchers exclude FOG episodes shorter than 1 s [57], or shorter than 3 s [128]. In addition to explicitly eliminating FOG episodes based on a duration, short FOG can also be excluded by using a low temporal resolution (e.g., labels applied at one second intervals or longer [113,163]). Similarly, some FOG episodes can be excluded through windowing. If the windows are required to be homogeneous (i.e., composed entirely of data with the same label) then all FOG episodes shorter than the chosen window duration are excluded. In many cases, the chosen window length is a compromise between being short enough to capture brief FOG episodes and long enough for specific feature calculations, such as FI [51,94].

Excluding short FOG may overlook periods of multiple FOG in rapid succession. For example, a person may freeze, take a few ineffective steps while attempting to resume normal walking, then freeze again. According to an event-based FOG definition, multiple FOG episodes in quick succession would be labeled as individual FOG episodes with a few steps in between. If a low temporal resolution for labelling is used (i.e., labels applied at long time intervals), a minimum FOG duration is imposed, or windows are required to be homogeneous, entire sequences of short FOG episodes may be excluded or labeled as normal gait. However, multiple short FOG episodes may be relevant gait disturbances that should be detected and considered in a cueing system. A FOG definition based on a period of gait disruption would consider a sequence of multiple short FOG episodes as a single FOG occurrence. Combining many short FOG episodes into one FOG occurrence would be less likely to result in discarded data due to windowing or the labeling interval.

Various approaches can be used to merge multiple FOG episodes that occur in quick succession. For instance, FOG episodes separated by less than a specific time threshold can be merged [57]. Alternatively, windows could be considered to be FOG if they contain at least 50% FOG data; therefore, as the window moves through the data, two FOG episodes separated by a short Non-FOG period, such as one or two small steps, could result in the windows all being labeled as FOG [121].

Currently, evidence is lacking to support the decision to use an "event-based" or "period of gait disruption" approach for classifying FOG. This research investigated the effect of merging successive FOG on freeze detection and prediction in PD. The outcomes can help guide the development of appropriate classification models for wearable FOG identification systems.

2.6.3 Single Sensor Instrumentation for FOG Prediction

Accelerometers and gyroscopes are the most commonly used sensor type for FOG detection and prediction and have demonstrated accurate detection using wearable sensors [51,148]. To improve wearability, researchers have developed systems that can use everyday devices and clothing such as smartphones [104,125,164,168], smartwatches [55], pants [95,165] and shoes

[43,161]. Smart-device and clothing-integrated systems are less cumbersome than systems requiring multiple sensors on various body parts and may lead to higher user satisfaction and compliance. A drawback of smartphone and clothing-based systems is signal noise that may be generated by sensor movement relative to the body. In-shoe sensors, such as plantar pressure sensors, are less susceptible to this type of noise since there is little relative movement between the shoe and foot.

FOG prediction systems would benefit from increased wearability and simplicity. Attempts have been made to reduce prediction system complexity by using a single sensor input, such as a single shank-mounted accelerometer [21] or a waist-mounted IMU [19]. However, additional research is required to determine the feasibility of single sensor FOG prediction systems. A single sensor system would have several advantages compared to a multi sensor version such as eliminating the need for sensor synchronization, reducing the number of sensors worn, reducing the amount of data to acquire and process, and ultimately allowing simpler classifiers that could run on local microprocessors. Research is needed to determine if single sensor FOG prediction systems could produce models comparable in performance to multi-sensor systems.

Parkinson's disease movement symptoms manifest asymmetrically and commonly affect one side of the body more severely. The more severe side (MSS) and less severe side (LSS) are person specific and do not correspond to the dominant leg or hand. Despite the presence of a more severe side, the difference between MSS and LSS limbs is rarely considered in FOG detection and prediction studies. The distinction between the MSS and LSS may not be necessary for FOG detection systems since both the MSS and LSS provide sufficient information for FOG detection, as evidenced by single sided FOG detection studies that identified FOG without considering MSS and LSS [94,110,125,161] or utilized data from the Daphnet dataset [94], which consists of data from the waist, and left leg only [51]. Unlike FOG detection studies, FOG prediction studies frequently assume that walking patterns gradually degrade prior to FOG. This degradation is difficult to identify and FOG prediction studies produce lower sensitivity and specificity than FOG detection models using similar methods [44] (Table A.1). The additional information provided by the MSS or LSS, if any, may be

beneficial to FOG prediction models, especially if using a single sensor. Given the asymmetry present in PD gait and the advantages of single-sensor FOG prediction systems, there is a need to determine if there is a preferred leg for instrumentation. This research aimed to determine whether single MSS, LSS, or bilateral plantar-pressure data are most useful for FOG prediction.

2.7 Summary

Automatic detection and prediction systems that use data from wearable sensors could reduce FOG incidence and severity by identifying FOG in real-time and providing an assistive cue. FOG detection and cueing after freeze onset has been reported [14,16,31,93]; however, to prevent FOG, the episodes must be predicted. FOG prediction is more challenging than detection and has only been performed in offline analyses [17–28]. To achieve the eventual goal of a real-time FOG prediction system, FOG prediction systems, which are currently, mostly IMU-based, must first be improved. Plantar pressure data may improve FOG prediction systems by providing features that represent subtle gait parameters that would be difficult to detect with IMUs or other wearable sensors.

Datasets used to train machine learning models often use ground truth FOG labels generated based on visual observation of specific lower limb movements or an overall inability to walk effectively. FOG definition ambiguity may affect model performance, especially with respect to multiple FOG in rapid succession. There is a need to determine whether merging FOG in rapid succession is beneficial to FOG identification systems.

Simplicity and ease of use are important factors for wearable assistive devices. FOG prediction systems need to be simplified before an everyday wearable FOG prediction system can be created. Eliminating unnecessary sensors and integrating the sensors into footwear may be an effective way to produce a simpler and more user-friendly wearable FOG prediction system. Research is needed to determine whether plantar pressure sensor insoles can be used to predict FOG and determine if it is preferable to instrument the more severely affected limb, the less severely affected limb, or both.

Chapter 3
Methodology

3.1 Overview

Chapter 3 describes the data collection, pre-processing, and cross validation steps that were common to all analyses in this thesis. The additional, analysis-specific, steps are explained in Chapters 4-6, including model training setup and model evaluation.

To fulfill the thesis objectives, a dataset containing labeled FOG episodes was required. Since no publicly available FOG dataset included data from the desired sensor systems, a new dataset was collected from a group of participants with Parkinson's disease (Section 3.2). Plantar pressure and IMU data were collected while participants walked a freeze-provoking path (Sections 3.3 and 3.4). Once collected, IMU and plantar pressure data were synchronized. FOG instances were visually annotated and Pre-FOG labels were assigned to the data immediately preceding each FOG episode (Section 3.5). Data were divided into overlapping 1 s windows and grouped into target and non-target classes in preparation for binary classifier development (Section 3.6). A set of diverse features including time-domain, Fourier transform, and wavelet transform based features were calculated from each window (Section 3.7.1). The importance of each feature was assessed using mRMR and Relief-F feature ranking (Section 3.7.2), and the top ranked features were used to train models using a modified LOPO cross-validation (Section 3.8). The final steps of model development and evaluation varied between analyses and are described in Chapters 4-6.

3.2 Participants

A convenience sample of people with PD were recruited through the Ottawa-Outaouais region Parkinson's community. To be eligible for this research, participants were required to have a confirmed diagnosis of PD, experience freezing at least once a week, and be able to walk independently (without a walking aid). Participants must not have undergone deep brain stimulation or have conditions other than PD that impaired their ability to walk. Eleven males

volunteered to participate in the study. Participants had a mean age of 72.7 ± 5.5 years and mean time since PD diagnosis of 10.5 ± 4.8 years (Table 3.1).

Data were collected during a single visit to the Movement Performance Laboratory at the University of Ottawa. Participants were on their normal antiparkinsonian medication schedule and dosage. Data collection was typically scheduled in the hours prior to the participant's next dose so that the medication would be wearing off during testing and FOG would be more likely to occur. Participants were assessed using the New Freezing of Gait Questionnaire (NFOG-Q) [181], Self-Reported Fall Questionnaire, and the Movement Disorder Society Unified Parkinson's Disease Rating Scale motor examination (UPDRS III) [182], (Table 3.1). Participants were also asked whether their PD symptoms predominantly affected the right or left side of their body and which direction tended cause freezes more frequently when turning. Ethics approval was obtained from the University of Ottawa (H-05-19-3547) and University of Waterloo (40954), and all participants provided informed written consent.

The participant details, questionnaire responses, and the results of the data collection and labeling (Section 3.5) are presented in Table 3.1.

Table 3.1: Participant information and questionnaire responses.

Participant	Age (Years)	More severely affected side	Time since PD diagnosis (years)	NFOG-Q score	UPDRS III score	Number FOG episodes	Average FOG duration (s)
P01	67	Right	16	14	10	49	0.69
P02	80	Left	11	21	20	35	2.64
P03	71	Left	11	17	13	14	1.06
P04	64	Left	10	4	18	0	0
P05	70	Right	14	20	13	0	0
P06	68	Left	19	22	29	10	4.23
P07	78	Right	5	15	16	221	1.52
P08	70	Right	12	17	20	24	1.51
P09	80	Left	10	18	18	9	0.75
P10	80	Left	2	4	15	0	0
P11	72	Right	5	19	20	0	0
Mean **SD**	**72.7** **5.5**		**10.5** **4.8**	**15.5** **5.9**	**17.5** **4.8**		

3.3 Biosignal Measurement, Sensors and Equipment

3.3.1 IMU Sensors

IMU data were collected using the Shimmer3 (Shimmer, Dublin, Ireland) system (Figure 3.1a). Four wireless sensor units (65 mm x 32 mm x 12 mm, 31 g) recorded 3-axis acceleration (± 4 g) and gyroscope data (± 500 dps). Sensor units were placed on the medial side of each shank, just above the malleolus, and lateral side of each thigh, just above the knee (Figure 3.1 b). IMU data were collected at 512 Hz (downsampled to 100 Hz in post processing) and streamed to a computer running the ConsensysPRO v1.6.0 software for processing. Synchronization between the four sensor units was done automatically in the ConsensysPRO software.

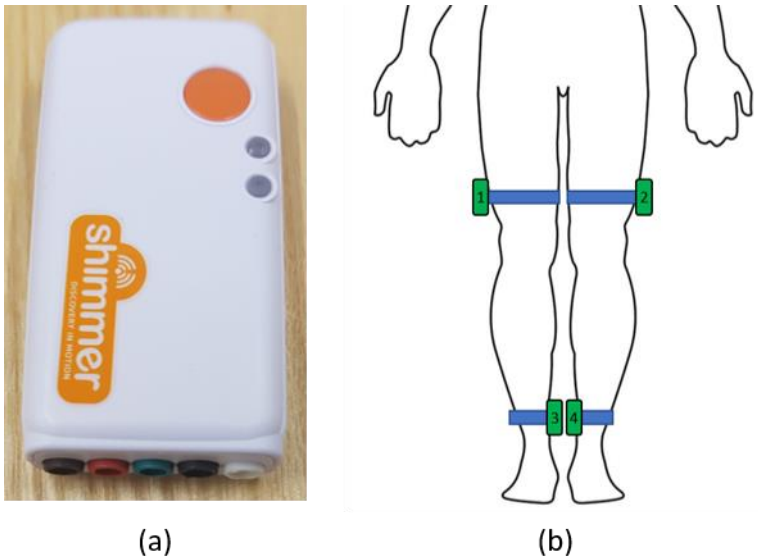

Figure 3.1: Shimmer IMU system: (a) sensor unit, (b) locations on the body.

3.3.2 Plantar Pressure Sensing Insoles

Plantar pressure was measured bilaterally with F-Scan insole sensors (Tekscan, Boston, MA, USA) [183] (Figure 3.2). These insoles are thin (< 1 mm) and flexible with a resolution of 3.9 pressure sensing cells per cm^2 [184]. A new pair of insoles was used for each participant and trimmed to fit inside their regular shoes. Insole sensors were connected to a receiver unit strapped to the lateral shank (Figure 3.2 c), which was tethered to a laptop computer. Plantar pressure data were collected at 100 Hz and recorded using the FScan Research v7.50 software.

(a) (b) (c)

Figure 3.2: F-Scan insole plantar-pressure sensors: (a) insole sensor, (b) sensor inside running shoe, (c) sensor system when worn.

3.3.3 Calibration and Equilibration

The IMU sensors were calibrated using the Shimmer 9DoF Calibration software. Accelerometer calibration consisted of sequentially placing the sensor such that each axis was aligned with the direction of gravity. The gyroscope calibration involved slowly rotating the sensor around each axis. IMU calibration was performed once prior to data collection and was not repeated between participants.

The FScan insoles were equilibrated prior to participant arrival and calibrated once installed into the participant's shoes. Equilibration used a pressurized air bladder to apply a uniform force across the entire insole; the software then adjusted the constants for each sensor cell to equalize the output pressure. The equilibration was performed at 138, 276, and 414 kPa [185]. The FScan step calibration was performed immediately prior to the walking trials. Participants were weighed using a digital scale. Next, the participants were asked to stand with all their weight on a single foot and then shift to stand on the other foot. This was done for both feet.

3.4 Data Collection Protocol

For all walking trials, the participants wore the Shimmer IMU sensors (4 sensor units), and the FScan plantar pressure sensors (2 insoles), as described in Section 3.3. Participants completed up to 30 self-paced walking trials following a pre-determined path approximately 25 m in length. The walking task path started and ended in a seated position and involved multiple 90° and 180° turns, stops, starts and a narrow passageway leading to a dead end (Figure 3.3).

The walking task path was designed to include freeze inducing situations (gait initiation, turning, narrow halls and turning in a confined space) interspersed with straight walking to allow gait to stabilize and reach steady state before the next obstacle was encountered. Participants were told that in the event of a freeze, they should recover independently (without assistance) and continue walking until the end of the full path. The turn in the hallway could be performed in either direction, according to whichever was more likely to elicit FOG for each participant (Section 3.2).

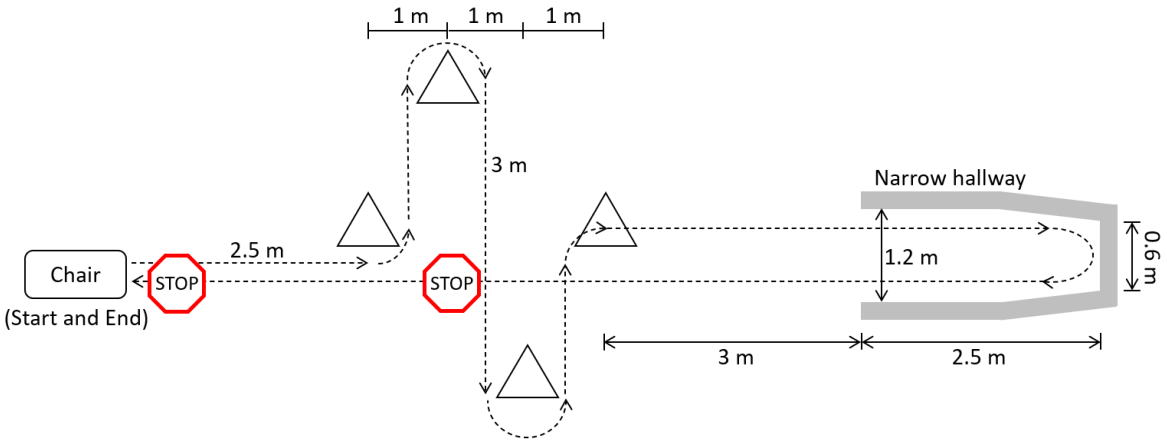

Figure 3.3: Walking task path. The box on the left-hand side is the chair at the beginning and end of the trial. The triangles are cones, and the grey zone delimits a narrow hallway. Red octagons indicate momentary voluntary (2 s) stops.

Five baseline trials were performed followed by 25 additional trials while performing a cognitive and motor task simultaneously to increase the likelihood of freezing. The cognitive task consisted of continuously saying words out loud beginning with a specific letter provided by the researcher prior to each trial. The words could not be proper nouns, could not be repeated, and could not use the same root (e.g., tea, teapot, teacup). A standardized list of letters

was used for the cognitive test. The motor task consisted of carrying a plastic tray with objects on it, without letting the objects fall. Initially a small pyramid of three wooden blocks was used. To increase task difficulty, an empty paper cup (weighted slightly with a coin) was placed on the tray instead of the blocks (Figure 3.4). Alternatively, some participants carried the tray in one hand and a sealed water bottle in the other hand.

The walking trials were video recorded using a smartphone camera (30 Hz). During data collection, FOG episodes were identified, and an offline labeling process was used to refine the FOG onset and termination times (Section 3.5). For all trials, the researcher walked with the participant to assist in the case of loss of balance and thus prevent a fall. The participants were allowed to rest between trials for as long as they required.

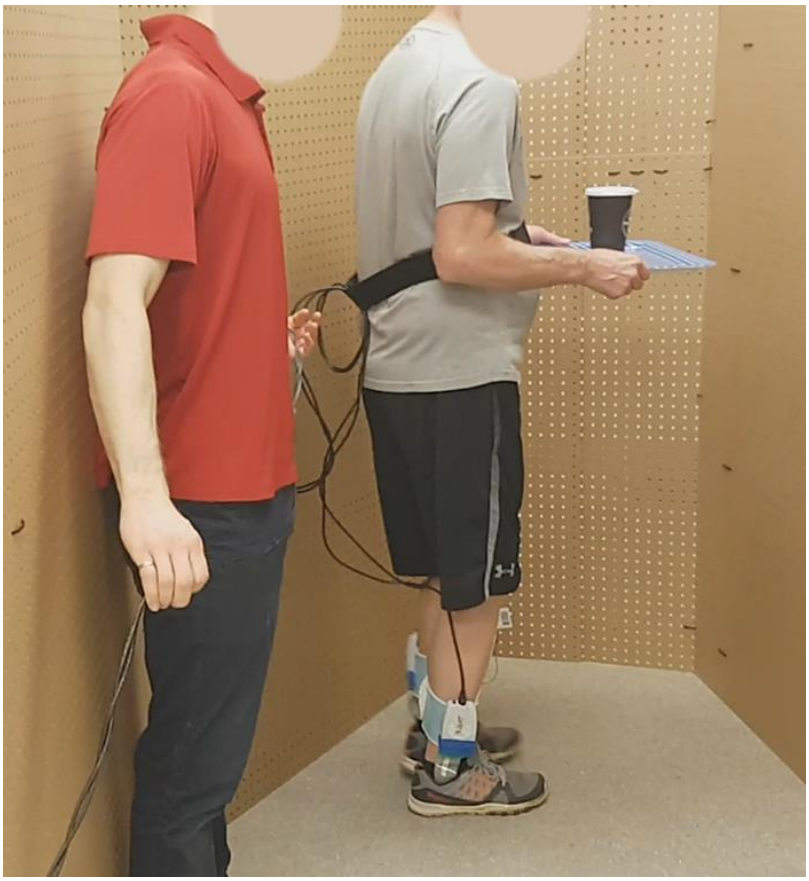

Figure 3.4: Example of walking trial with participant turning to the right in narrow hallway.

3.5 Synchronization and FOG Labeling

All data processing was done using MATLAB R2019b (or newer) with the Statistical and Machine Learning and the Deep Learning toolboxes (formerly neural network toolbox) and App Designer (MathWorks, MA, USA). Following data collection, a custom program was used for data labeling and synchronization. The program imported and displayed the video and plantar pressure and IMU data simultaneously, which allowed the data to be visually inspected and facilitated synchronization (Figure 3.5).

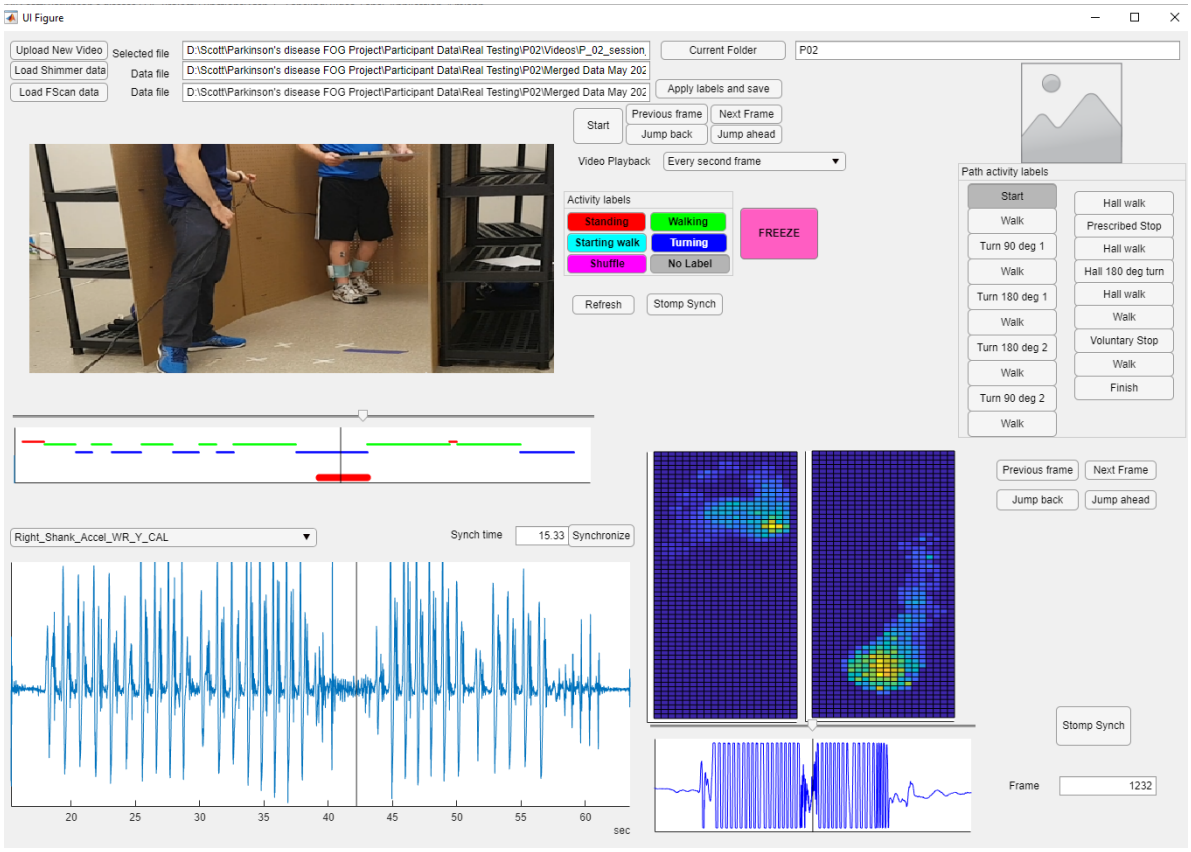

Figure 3.5: Sample display from the synchronization and labeling program.

Plantar pressure, IMU data, and video were synchronized using a single leg stomp performed at the beginning of each trial. Lifting the foot off the ground produced a period of zero plantar pressure. The first frame of non-zero plantar pressure data was used as the stomp event. For the IMU, the stomp was identifiable as the maximum positive amplitude at the start of the trial for vertical acceleration from the shank sensor on the stomping foot. For the video,

the stomp was the frame in which the shoe contacted the ground. Synchronization was confirmed using multiple heel strike events throughout the walking trial.

Offline labeling was performed to annotate the onset and termination of each FOG episode. Labeling was performed by researcher S Pardoel. In cases of uncertainly, a second labeler (Dr. J Nantel) was consulted. Each video frame was labeled as FOG or Non-FOG. The video labels were transferred to the synchronized plantar pressure and IMU data using linear interpolation to the closest timestamp. The beginning of a freeze was defined as "the instant the stepping foot fails to leave the ground despite the clear intention to step". The end of the freeze was defined as "the instant the stepping foot begins or resumes an effective step". For example, a step was considered effective the instant the heel lifted from the ground, provided that it was followed by a smooth toe off with the entire foot lifting from the ground and advancing into the next step without loss of balance. As a special case, if a person froze, stopped trying to advance, and remained standing, the instant that the participant stopped trying to advance was considered the end of the freeze. This was determined by the complete absence of foot movement and known FOG characteristics such as trembling of the knee, medial-lateral weight shifting, or attempt at shuffling.

Following FOG and Non-FOG labeling, Pre-FOG labels were applied to all data within the 2 s period immediately prior to the onset of each freeze episode. For two FOG episodes less than two seconds apart, data between the two FOG episodes were labeled as Pre-FOG. The 2 s Pre-FOG duration was chosen since this time represents approximately two strides and was sufficient for FOG prediction in previous studies [17,186]. Furthermore, 2 s to 3 s Pre-FOG durations have led to better Pre-FOG classification accuracy than longer Pre-FOG durations [23].

Activity labels were also applied to the data to identify standing, walking, and turning, using the same methods as the FOG and Non-FOG labeling. Once synchronization and labeling were complete, the data were trimmed and exported from the synchronization and labeling program. All data without an activity label (i.e., data that was not standing, walking, or turning) were discarded.

3.6 Windowing and Class Creation

3.6.1 Windowing

The labeled data were split into 1 s windows with a 0.2 s shift between consecutive windows (Figure 3.6 a). A window could contain a combination of Pre-FOG, FOG, and Non-FOG data.

3.6.2 Target and Non-Target Class Creation

Prior to classifier model development, windows were grouped into target and non-target classes and models were trained to differentiate between the classes. Whether a given data window was assigned to the target or non-target class varied by analysis. All class composition explanations are in reference to Figure 3.6.

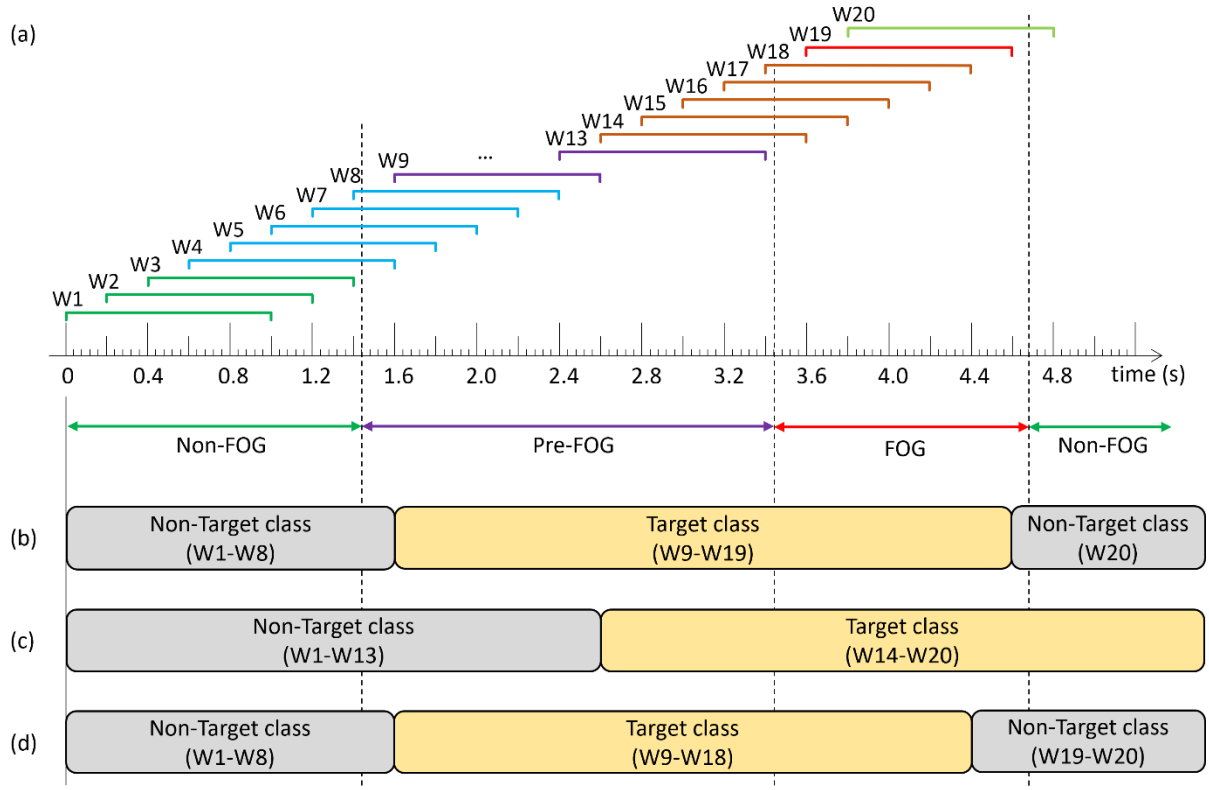

Figure 3.6: Diagram of data windowing and target class compositions: (a) windows W1-W3 contain: Non-FOG data only, W4-W8: Non-FOG and Pre-FOG data, W9-W13: Pre-FOG data only, W14-W18: Pre-FOG and FOG data, W19: FOG data only, and W20: FOG and Non-FOG data, (b) class composition for models in Chapters 4 and 6, (c) class composition for detection models in Chapter 5, (d) class composition for prediction models in Chapter 5.

In Chapter 4, the objective was to develop a single model that could predict and detect FOG. Therefore, the target class included data windows containing purely Pre-FOG data (W9-W13), purely FOG data (W19), and windows containing both Pre-FOG and FOG data (W14-W18). The non-target class included all other windows (W1-W8, W20) (Figure 3.6 a,b).

In Chapter 5, different target and non-target classes were used to develop separate detection and prediction models. For the detection models, the target class included all windows that contained any FOG data. Thus, windows containing Pre-FOG and FOG (W14-W18), purely FOG (W19), or FOG and Non-FOG data (W20) were included in the target class. The non-target class contained all other windows which included purely Non-FOG data (W1-W3), Non-FOG and Pre-FOG data (W4-W8), and purely Pre-FOG data (W9-W13) (Figure 3.6 a,c).

For the prediction models in Chapter 5, the target class contained the windows beginning anytime during the 2 seconds prior to FOG onset. Thus, windows containing purely Pre-FOG data (W9-W13) and windows containing Pre-FOG and FOG data (W14-W18) were included in the target class. Windows that contained purely Non-FOG data (W1-W3), Non-FOG and Pre-FOG (W4-W8), purely FOG data (W19), or FOG and Non-FOG data (W20) were in the non-target class for the FOG prediction models (Figure 3.6 a,d).

For all FOG prediction models in Chapter 6, target and non-target class composition was the same as Chapter 4. Thus, the target class included data windows containing purely Pre-FOG data (W9-W13), both Pre-FOG and FOG data (W14-W18), and purely FOG data (W19). The non-target class included all other windows (W1-W8, W20) (Figure 3.6 a,b).

3.7 Feature Extraction and Feature Selection

In this thesis, features were calculated from the data windows and used to train FOG detection and prediction models. These features differed according to the analysis-specific methods; however, the same set of starting features, and feature selection methods were utilized.

3.7.1 Feature Extraction

The features used in this research were based on [45] (Table 3.2). Features were grouped by time domain (n=13), Fast Fourier transform (n=8), and discrete wavelet transform (Haar

mother wavelet) (n=14). The fast Fourier transform (FFT) is an efficient algorithm for computing the discrete Fourier transform. For the FFT and DWT categories, 38 signal inputs were used: total ground reaction force (GRF); position, velocity, and acceleration of foot centre of pressure (COP) in anterior-posterior and medial-lateral directions; ankle and thigh acceleration in anterior-posterior, vertical, and medial-lateral directions; and ankle and thigh angular velocity in anterior-posterior, vertical, and medial-lateral directions. COP velocity and acceleration are the first and second derivatives of COP position, respectively. All features were calculated separately for the left and right sides, with the exception of "number of weight shifts" that required data from both feet.

Table 3.2: Features extracted from windowed data.

Feature	Feature description	Source	Number of input parameters	Total features
Time domain features (n=13)				
Number, duration, length of COP reversals	Number, length, duration of centre of pressure (COP) path anterior-posterior direction reversals per window (n=3)	[47]	2	6
Number, duration, length of COP deviations	Number, length, duration of medial-lateral COP deviations per window. Deviation is the first derivative of medial-lateral COP exceeding a threshold of ± 0.5 mm/window (n=3)	[47]	2	6
CV of COP position, velocity, acceleration	Anterior-posterior and medial-lateral coefficients of variation (CV) of COP position, velocity, and acceleration (n=6)	[47]	2	12
Number of weight shifts	Number of times the majority of total GRF (>50%) changed foot (n=1)	-	1	1
			Total computed features	**25**
Fast Fourier Transform (FFT) features (n=8)				
Total power in FFT signal	Power in FFT signal per window as sum of squared amplitude (n=1)	[105]	38	38
Dominant frequency	Frequency bin with highest amplitude per window (n=1)	[132]	38	38
Max, min, mean	Maximum, minimum, and mean amplitude of FFT signal (n=3)	[132]	38	114
Power in locomotion, freeze bands	Power under FFT curve in locomotion band (0.5-3 Hz) and freeze band (3-8 Hz) (n=2)	[94]	38	76

Freeze index	Ratio of power in freeze band (3-8 Hz) and locomotion band (0.5-3 Hz), (n=1)	[94]	38	38
			Total computed features	**304**
Discrete Wavelet Transform features (n=14), Haar mother wavelet				
Variance of coefficients	Variance of the detail and approximation coefficient vectors (n=2)	[187]	38	76
Max, min, mean	Maximum, minimum, mean of detail and approximation coefficient vectors (n=6)	[187]	38	228
Max, min, mean energy	Maximum, minimum, mean energy of detail and approximation coefficient vectors (n=6)	[187]	38	228
			Total computed features	**532**

In total, 861 individual features were extracted from the over 71,000 data windows [44]; 528 features were calculated from accelerometer and gyroscope data, and 333 features calculated from plantar pressure data (GRF; COP position, velocity, acceleration). Before calculating the COP, GRF values less than 5% of the two-foot total were set to 0, since the limb was in swing and the small pressures were not relevant to FOG.

3.7.2 Feature Selection

An excessively large number of features can increase computational cost and increase the risk of model overfitting. To reduce feature set size and improve model performance, feature selection methods can be used to identify the most relevant features for FOG classification.

Filter feature selection methods compare features according to a specific metric (e.g., correlation with the output class) and rank features according to their usefulness [171]. The most useful features can then be used as inputs to any classifier. Wrapper methods incorporate the classifier into the feature selection process and, in their simplest form, repeatedly train and test the classifier using different combinations of features from the initial set. The best feature set for that classifier and set of parameters is chosen based on the classifier performance [171]. In this thesis, filter type feature selection was used because the method is independent of the classifier and the highest ranking features can be used with any machine learning classifier, which makes the feature selection results applicable to future FOG identification studies.

Minimum-redundancy maximum-relevance (mRMR) is an established feature selection method in machine learning studies, including PD studies, with input signals such as speech patterns [188], kinematics of handwriting (in-air movement in handwriting) [189], GRF [190], and body acceleration [50]. The mRMR algorithm is a multivariate approach that selects features such that the mutual information between a feature and class is maximized, while the pairwise information between features is minimized [191]. In this thesis, mRMR using mutual information was used for feature ranking. The result was a ranked list of features that are correlated with the target class, with each feature contributing different information.

Relief-F feature ranking incorporates interactions between features [170] and has been used in activity monitoring situations with plantar pressure data collected during walking [47]. Relief-F assigns a score to each feature according to that feature's estimated quality [170]. The Relief-F algorithm first sets all feature scores to 0. Then, for m randomly selected training instances, the k nearest samples from each class are used to update the feature scores, according to a distance metric (e.g., Euclidean, Manhattan). In this research, Relief-F was performed with $m = 2000$ random selections and $k = 200$ nearest neighbours. The k and m values were chosen based on preliminary testing that showed no changes in feature ranks with higher values.

Both the mRMR and Relief-F feature ranking algorithms sort the features according to estimated importance. The top ranked features are selected according to a desired number of features. In this thesis, the desired number of features varied according to the analysis being performed, as described in Chapters 4 and 6.

3.8 Leave-One-Freezer-Out Cross Validation

Leave-one-freezer-out (LOFO) cross validation was used to evaluate all models in this thesis. The typical leave-one-person-out cross validation trains a model using the data from all but one person, then tests the model using the held-out person's data. In FOG classification studies, it is common for some individuals to experience FOG in normal living but not during the in-laboratory data collection. Thus, if a person who did not freeze during testing was held out as the test subject, the corresponding test data would be entirely from the non-target (Non-FOG) class. This is problematic since a model cannot truly be evaluated using data from only the

negative class. In some studies, the model is assumed to have 100% sensitivity for these individuals [57,128]; however, this assumption can skew overall model performance results. The LOFO method avoids this issue since only participants who froze during data collection are involved with model testing, while participants who did not experience FOG are always included in the training set.

Chapter 4

Early Freezing of Gait Detection using Plantar Pressure and IMU Data

4.1 Introduction

In this chapter IMU and plantar pressure data were combined for early FOG detection. The binary classification models identified Total-FOG or Non-FOG states, wherein the Total-FOG class included data windows from 2 s before the FOG onset until the end of the FOG episode. Three feature sets were compared: plantar pressure features, inertial measurement unit (IMU) features, and both plantar pressure and IMU features (PP-IMU).

4.2 Methodology

4.2.1 Windowing

The windowing used in this chapter is explained in Section 3.6. In addition, in this chapter each window was assigned a label corresponding to the data it contained. The window labels were:

- Pre-FOG (purely Pre-FOG data)
- FOG (purely FOG data)
- Pre-FOG-Transition (Pre-FOG and FOG data)
- Non-FOG (Non-FOG data, Non-FOG and Pre-FOG data, and FOG and Non-FOG data)

Another combined label was generated as Total-FOG, which contained Pre-FOG, Pre-FOG-Transition, and FOG windows.

4.2.2 Feature Extraction and Selection

The starting feature set described in Section 3.7.1 was used for this analysis. Three feature groups were compared: plantar pressure, inertial measurement unit (IMU), and both plantar pressure and IMU features (PP-IMU). Feature selection was performed separately for each of the feature sets using both the mRMR and Relief-F algorithms (Section 3.7.2). For feature

selection, the target class was composed of all windows with the Total-FOG label and the non-target class contained the Non-FOG windows. The top 5, 10, 15, 20, 25, 50, 75, and 100 features according to both the Relief-F and mRMR feature selection methods were used for model training and comparison.

4.2.3 Classification Model Development

For classification model development, decision tree ensembles and support vector machines were used. In pilot testing, different tree ensemble boosting techniques were examined. RUSBoosting performed better than bagging, LogitBoosting and AdaBoosting approaches. This is likely due to the dataset being highly imbalanced, which can negatively affect classifier performance. FOG occurs infrequently during walking and FOG datasets generally contain several times more Non-FOG samples than FOG samples [25]. RUSBoosting randomly undersamples the majority class (non-target class) so that the number of samples matches the minority class (target class). Note that undersampling is only done during model training and not during testing; therefore, the class imbalance in the testing data is unaffected. Support vector machine models with linear, polynomial (3^{rd} and 5^{th} order), and radial basis function kernels were also evaluated. SVM models with various hyperparameter configurations were tested, including the box constraint parameter that determined the amount of allowable misclassifications. The MATLAB classifier optimization function was used to set the box constraint value. Class imbalance was addressed by setting the cost function to be proportional to the class size. The SVM models had good specificity but low sensitivity. The poor performance of the SVM models may indicate that the dataset used in this research is not separable. To improve SVM model performance, additional model tuning could be performed, or a custom kernel function could be designed to create separation between the classes for the current dataset; however, this would be challenging and time consuming. Instead, since RUSBoosted decision tree ensembles consistently outperformed the SVM models, the SVM models were abandoned in favour of decision tree ensembles.

The classifier selected for all subsequent model development and comparison was a RUSBoosted decision-tree ensemble. The base decision trees (n=100) were tested with

maximum depths of 5 or 10 decision splits, and with the top 5, 10, 15, 20, 25, 50, 75, and 100 features according to both the Relief-F and mRMR feature selection methods, from each of the three feature groups.

Five test cases were used during LOFO analysis (Table 4.1). The target and non-target classes for the five test cases were defined as different groupings of the labeled windows. For each cross-validation fold, the model was trained only once using Case 1 (target class: Total-FOG, non-target class: Non-FOG), then evaluated on each of the five test cases. Case 1, where the target class was Total-FOG windows and included Pre-FOG, Pre-FOG-transition, and FOG windows, was based on the goal of a clinically relevant cueing system, where real-time cueing would be activated before or during a freeze. For Cases 2, 4, and 5, the target class contained a single label. This was done to evaluate the model's ability to recognize each of the labels individually. For Case 3, the Pre-FOG and Pre-FOG-Transition windows were grouped to form the target class, to examine the feasibility of using these two labels in future model development. This target class (Case 3) contained windows from the beginning of Pre-FOG data until, at most, one second into the FOG event; therefore, detection of windows in this target class would be either prediction of a freeze or detection of freeze episode initiation.

Table 4.1: Target and non-target class composition for each test case.

	Target class	**Non target class**
Case 1	Total-FOG: Pre-FOG, Pre-FOG-Transition, FOG	Non-FOG
Case 2	Pre-FOG	Non-FOG
Case 3	Pre-FOG, Pre-FOG-Transition	Non-FOG
Case 4	Pre-FOG-Transition	Non-FOG
Case 5	FOG	Non-FOG

4.3 Results

Table 4.2 presents the total number of windows for each label from each participant. Table 4.3 presents the LOFO cross-validation results for plantar pressure, IMU, and both plantar pressure and IMU (PP-IMU) features.

Table 4.2: Number of data windows of each label extracted from each participant.

Participant	Window labels			
	Pre-FOG	Pre-FOG-Transition	FOG	Non-FOG
P01	217	166	7	3721
P02	178	171	294	5188
P03	66	62	17	6884
P04	0	0	0	2635
P05	0	0	0	5331
P06	52	49	162	9368
P07	725	1303	766	6572
P08	75	126	84	4848
P09	44	30	5	6848
P10	0	0	0	6034
P11	0	0	0	9039
Label total	**1357**	**1907**	**1335**	**66468**

Table 4.3: Top performing RUSBoosted ensembles of decision trees. Target class is Total-FOG (Case 1).

Held out test participant	Plantar pressure features		IMU features		PP-IMU features	
	Relief-F, 5 features, 5 splits		mRMR, 25 features, 5 splits		Relief-F, 10 features, 5 splits	
	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)
P01	69.7	83.7	68.2	84.0	70.0	86.0
P02	71.7	86.7	67.0	90.7	70.6	87.9
P03	68.3	89.7	54.5	96.1	61.4	92.9
P06	93.9	89.5	73.4	93.5	93.2	90.2
P07	72.8	80.3	34.8	92.1	68.7	78.9
P08	89.5	79.6	70.9	92.3	82.1	87.6
P09	79.7	72.5	64.6	92.2	88.6	79.7
Mean	**78.0**	**83.2**	**61.9**	**91.6**	**76.4**	**86.2**
SD	**9.4**	**5.7**	**12.4**	**3.4**	**10.8**	**4.8**

Sens: sensitivity, Spec: specificity.

Performance was very similar for the plantar-pressure features model (sensitivity 78.0%, specificity 83.2%) and the PP-IMU features model (sensitivity 76.4%, specificity 86.2%) (Table 4.3). The IMU-features model had the lowest sensitivity (61.9%) but the highest specificity (91.6%). The best number of features and best feature ranking method differed for each group of features (Table 4.3). The best plantar-pressure features model used the top 5 Relief-F features. The best IMU-features model used the top 25 mRMR features. The best PP-IMU features model used the top 10 features according to Relief-F rankings. For all models,

decision trees with 5 splits outperformed decision trees with 10 splits. The best model from Table 4.3 used combined features (i.e., PP-IMU features model), with 76.4% sensitivity and 86.2% specificity for the Total-FOG target class (Pre-FOG, Pre-FOG-Transition, FOG). The features used in the PP-IMU model are presented in Table 4.4.

Table 4.4: Top 10 features (according to Relief-F) used in the PP-IMU features model.

Feature rank	Feature description
1	Dominant frequency of COP velocity in AP direction for right leg
2	Dominant frequency of COP velocity in AP direction for left leg
3	Dominant frequency of COP velocity in ML direction for right leg
4	Dominant frequency of thigh acceleration in AP direction for left leg
5	Number of AP COP path reversals for left leg
6	Number of AP COP path reversals for right leg
7	Minimum WT dC of COP position in AP direction for right leg
8	Dominant frequency of thigh acceleration in AP direction for right leg
9	Mean energy of WT aC of COP position in AP direction for right leg
10	Mean WT aC of COP position in AP direction for right leg

AP: anterior-posterior, ML: medial-lateral, WT: wavelet transform, aC: approximation coefficient, dC: detail coefficient.

The results for Cases 2-5 are presented in Table 4.5, Table 4.6, and Table 4.7. The specificity results for Cases 2-5 were constant across cases, since specificity is based on the non-target class (true negatives and false positives), which was unchanged across test cases.

Table 4.5: Target class test cases for PP-IMU features model, using top 10 features according to Relief-F. Column headers are the target class label(s), as defined in Table 4.1.

Held out test participant	Pre-FOG (Case 2)		Pre-FOG and Pre-FOG-Transition (Case 3)		Pre-FOG-Transition (Case 4)		FOG (Case 5)	
	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)
P01	52.5	86.0	69.5	86.0	91.6	86.0	100.0	86.0
P02	23.0	87.9	49.0	87.9	76.0	87.9	96.3	87.9
P03	37.9	92.9	57.8	92.9	79.0	92.9	88.2	92.9
P06	73.1	90.2	84.2	90.2	95.9	90.2	98.8	90.2
P07	48.8	78.9	64.5	78.9	73.2	78.9	79.9	78.9
P08	69.3	87.6	78.6	87.6	84.1	87.6	90.5	87.6
P09	81.8	79.7	87.8	79.7	96.7	79.7	100.0	79.7
Mean	**55.2**	**86.2**	**70.2**	**86.2**	**85.2**	**86.2**	**93.4**	**86.2**
SD	**19.3**	**4.8**	**13.2**	**4.8**	**8.9**	**4.8**	**7.0**	**4.8**

Sens: sensitivity, Spec: specificity.

Table 4.6: Target class test cases for plantar-pressure features model, using top 5 features according to Relief-F. Column headers are the target class label(s), as defined in Table 4.1.

Held out test participant	Pre-FOG (Case 2)		Pre-FOG and Pre-FOG-Transition (Case 3)		Pre-FOG-Transition (Case 4)		FOG (Case 5)	
	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)
P01	52.5	83.7	69.2	83.7	91.0	83.7	100.0	83.7
P02	23.6	86.7	49.6	86.7	76.6	86.7	98.0	86.7
P03	43.9	89.7	64.1	89.7	85.5	89.7	100.0	89.7
P06	76.9	89.5	85.1	89.5	93.9	89.5	99.4	89.5
P07	36.7	80.3	62.9	80.3	77.4	80.3	99.2	80.3
P08	82.7	79.6	88.1	79.6	91.3	79.6	92.9	79.6
P09	70.5	72.5	78.4	72.5	90.0	72.5	100.0	72.5
Mean	**55.3**	**83.2**	**71.0**	**83.2**	**86.5**	**83.2**	**98.5**	**83.2**
SD	**20.5**	**5.7**	**12.7**	**5.7**	**6.4**	**5.7**	**2.4**	**5.7**

Sens: sensitivity, Spec: specificity.

Table 4.7: Target class test cases for IMU features model, using top 25 features according to mRMR. Column headers are the target class label(s), as defined in Table 4.1.

Held out test participant	Pre-FOG (Case 2)		Pre-FOG and Pre-FOG-Transition (Case 3)		Pre-FOG-Transition (Case 4)		FOG (Case 5)	
	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)
P01	53.5	84.0	67.9	84.0	86.7	84.0	85.7	84.0
P02	16.3	90.7	43.8	90.7	72.5	90.7	94.6	90.7
P03	31.8	96.1	49.2	96.1	67.7	96.1	94.1	96.1
P06	44.2	93.5	62.4	93.5	81.6	93.5	80.2	93.5
P07	17.8	92.1	35.5	92.1	45.4	92.1	32.9	92.1
P08	65.3	92.3	66.2	92.3	66.7	92.3	82.1	92.3
P09	50.0	92.2	62.2	92.2	80.0	92.2	100.0	92.2
Mean	**39.8**	**91.6**	**55.3**	**91.6**	**71.5**	**91.6**	**81.4**	**91.6**
SD	**17.2**	**3.4**	**11.6**	**3.4**	**12.7**	**3.4**	**20.9**	**3.4**

Sens: sensitivity, Spec: specificity.

4.4 Discussion

Comparing the different models in Table 4.3, and the same test cases across Table 4.6 and Table 4.7, the plantar-pressure features model reached higher sensitivity than the IMU-features model. However, the IMU-features model achieved higher specificity for all cases. This indicates that plantar pressure may identify FOG related patterns that the IMU sensors cannot;

however, plantar pressure sensors may produce more false positives. Thus, including features from both sensor systems is recommended.

The PP-IMU features model was selected as the best overall model. Further analysis from the additional four test cases (Table 4.5) showed that just over half the Pre-FOG windows were correctly identified. If this model were used to trigger an assistive cue, identifying 55.2% of the Pre-FOG windows before the FOG occurs would be helpful, but may result in many missed opportunities to avoid a freeze (i.e., assuming that an appropriate cue can mitigate or avoid an upcoming freeze episode). For Pre-FOG-Transition, sensitivity was 85.2% using plantar pressure and IMU data, indicating that most transition windows between Pre-FOG and the freeze would be identified; therefore, a cue could be administered within the first second of the FOG episode. When Pre-FOG and Pre-FOG-Transition windows were combined, model sensitivity decreased to 70.2%. Hence, including Pre-FOG adversely affected freeze-event recognition. FOG window classification using plantar pressure and IMU data was highly effective (93.4% sensitivity), indicating that few FOG windows were missed. In practice, the freeze identification model would perform very well as a FOG detection system, with a cue administered during the freeze if the Pre-FOG or transition states were missed. A similar analysis in [25] predicted 66.7% of the freeze episodes within 2 s of onset and detected 97.4% of the episodes between 2 s before and 4 s after FOG onset. These results were based on the number of FOG episodes, which may contribute to the better performance compared to results presented in this analysis, where results were based on decisions for each window.

The PP-IMU features model sensitivity was 76.4%, indicating that approximately 24% of the target-class windows were missed by the model. Other FOG prediction research [50] reported higher sensitivity (93%), although as in [20,25], the performance metrics were calculated based on FOG episodes, rather than windows. Thus, the sensitivity results are not directly comparable to this window-based analysis.

The PP-IMU features model specificity was 86.2%, indicating that approximately 14% of the non-target classifications were false positives. In an intelligent cueing device, this could result in false cues during walking, which may lead to reduced user compliance, depending on the type of cue. To ensure that the cueing system is effective and is used as intended, the

number of false cues should be minimized in future research. For example, a decision threshold could be implemented such that consecutive classifications are required to trigger a cue (Chapters 5 and 6). In addition, minimalistic or variable cues could be used such that false positives are better tolerated by the user. For instance, cue intensity or magnitude could begin at an almost imperceptible level and increase with successive positive FOG predictions. While 90% or greater specificity would be ideal, specificity below this threshold is common in the FOG prediction literature. Specificities of 67.0% [17], 80.25% [26], and recently 86% [23] have been reported.

The analysis outcomes could be applicable to a wearable freeze-detection system that is localized to the shoe. PP-IMU features model performance was only slightly better than the plantar-pressure features model. While improvements could be made to plantar pressure features model sensitivity, the plantar pressure model performed very well as a detection system, detecting 98.5% of the FOG windows. Including IMU features in the PP-IMU features model was primarily to improve specificity. If the plantar pressure features model specificity could be improved by other means, then the IMU sensors could be excluded. Models using only plantar pressure sensors are explored further in Chapter 6.

4.5 Conclusion

The combination of accelerometer, and plantar pressure data gave the best FOG identification results. The best decision tree ensemble model was built using 10 features and achieved 76.4% sensitivity and 86.2% specificity when classifying 1 s windows of Total-FOG data (data from 2 s before FOG onset until the end of the FOG episode). This model detected the transition between Pre-FOG gait and FOG with 85.2% sensitivity, which corresponds to detecting FOG less than one second after the freeze began. Furthermore, the FOG windows were detected with 93.4% sensitivity, indicating that few FOG windows were missed.

If the best model were applied in a wearable cueing device that helps avoid or break out of a freeze, this system would have a 70.2% chance of identifying FOG windows from before or within 1 second of the FOG onset. If this transition phase is missed, the cue would likely be applied during the freeze since the model was able to detect 93.4% of FOG windows.

While the model using both plantar-pressure and IMU features to detect Total-FOG had 86% specificity (i.e., 14% false positive rate, which is common in FOG prediction studies), higher specificity is preferred in practice. To address this, a cueing threshold could be implemented such that a cue is only triggered if multiple consecutive positive classifications are obtained, as explored in Chapters 5 and 6.

Chapter 5
Merging Multiple FOG in Rapid Succession

5.1 Introduction

This chapter examined whether merging multiple freezes that occur in rapid succession could improve FOG detection and prediction model performance. Plantar pressure and lower limb acceleration data were used to extract a feature set and train decision tree ensembles. The event-based definition of FOG was used in generating the non-merged dataset. Additional datasets were produced by merging FOG episodes separated by less than a merging threshold into a single episode. FOG detection and prediction models were trained for merging thresholds of 0, 1, 2, and 3 s.

5.2 Methodology

5.2.1 Merging

Consecutive freezes were merged into a single freeze if the time between the beginning of a FOG episode and the end of the previous FOG episode was less than a merging threshold. All data between the two freezes were relabeled as FOG, thus forming a single longer FOG episode. Merging thresholds of 0, 1, 2, and 3 s were used to create separate datasets for model development.

5.2.2 Windowing and Target Class Creation

For each merging threshold dataset (0, 1, 2, 3 s), windowing (Section 3.6.1) and feature extraction (Section 3.7.1) were performed. The target and non-target class composition was different for the detection and prediction models developed in this chapter (Section 3.6.2).

5.2.3 Detection and Prediction Model Development

The detection and prediction models in this chapter used identical starting parameters and differed only by the target and non-target class composition (Section 3.6.2). This chapter used the input features (Table 4.4) and model architecture found to be best in Chapter 4 (Table 4.5).

To examine the effect of FOG merging on FOG identification performance, the detection and prediction models were trained repeatedly, using identical model parameters but with different dataset merging thresholds (0, 1, 2, 3 s). LOFO cross validation was used for all models.

5.2.4 Model Evaluation

The trained models were evaluated using windows and FOG episodes. The window-based evaluation compared each window classification to the ground truth label and calculated sensitivity and specificity. While sensitivity and specificity are useful measures, this evaluation does not necessarily reflect a model's ability to act as a timely trigger for a cueing system since a model may only detect freeze windows and trigger a cue at the end of a FOG episode. Therefore, the FOG-episode-based evaluation determined if and when each episode was detected by the model. To avoid cues caused by misclassified windows, three consecutive positive target class classifications were required to generate a model trigger decision (MTD) (i.e., three previous windows had to be classified as belonging to the target class, Figure 5.1). Each MTD would correspond to a cue if used in a real-time intelligent cueing system. For each FOG episode, a MTD target zone was defined as the period between the start of Pre-FOG (2 s prior to the FOG episode) and FOG termination since a cue within this target zone would be helpful to either prevent or mitigate a FOG episode. If a MTD occurred within the MTD target zone, then the corresponding FOG episode was successfully identified.

Identification delay (ID) was defined as the time between FOG onset and a successful MTD identification. A positive ID indicates that the FOG episode was identified after it began (FOG detection) and a negative ID indicates FOG episode identification before episode onset (FOG prediction). If a MTD target zone contained multiple MTD, then the FOG episode was identified multiple times and only the first MTD was used to calculate the ID.

Each MTD was considered to be either a true positive (TP) (within the MTD target zone), or a false positive (FP) (outside the MTD target zone). MTD precision was calculated as the number of true positive MTD divided by the total number of positive MTD (TP+FP).

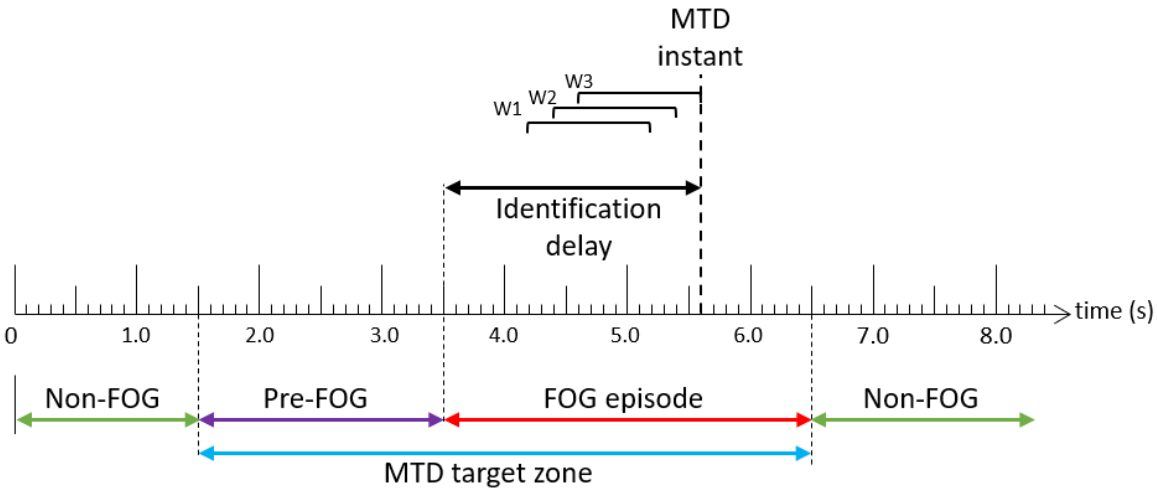

Figure 5.1: Model trigger decision diagram. Three consecutive windows classified as being part of the target class (W1-W3) results in a model trigger decision (MTD), where the MTD instant corresponds to the end of the third window. FOG is successfully identified if there is a MTD instance within the MTD target zone. The time difference between the FOG onset and the MTD instant is the identification delay.

As a final step in the model, a no-cue interval of 2.5 s was implemented. During the no-cue interval, any MTD generated by the system would be ignored. In a real cueing system, a person receiving a cue should react to the stimulus and modify their gait. The no-cue interval ensures that the person has time to respond to the cue, and that the system has time to reassess their gait before another cue is given.

5.3 Results

The number of FOG episodes experienced by each participant for different merging thresholds is presented in Table 5.1. Merging FOG episodes reduced the number of FOG episodes, primarily for participants P07 and P08.

Table 5.1: Number of FOG episodes for each participant for different merging thresholds.

Participant	Number of FOG episodes				Reduction in number of episodes by merging with MT = 3 s
	MT = 0 s	MT = 1 s	MT = 2 s	MT = 3 s	
P01	49	48	48	48	1
P02	35	35	35	35	0
P03	14	14	13	13	1
P04	0	0	0	0	-
P05	0	0	0	0	-
P06	10	10	10	10	0
P07	221	171	118	87	134
P08	24	16	14	14	10
P09	9	9	9	7	2
P10	0	0	0	0	-
P11	0	0	0	0	-

MT: merging threshold.

For window-based FOG detection (Table 5.2), sensitivity and specificity averages across all participants changed little ($\leq \pm$ 1%) due to merging (mean sensitivity: 83.4% for MT = 2 s, compared to 82.4% for MT= 0 s; mean specificity: 87.9% for MT = 2,3 s, compared to 88.3% for MT = 0 s). This included participants P07 and P08, who had the largest reduction in number of FOG episodes due to merging (Table 5.1). For the prediction models (Table 5.3), mean sensitivity decreased slightly as the merging threshold increased (68.4% for MT = 2 s, from 73.4% for MT = 0 s). Mean specificity was highest (82.8%) for MT = 2 s and lowest (80.9%) for MT = 3 s.

Table 5.2: Window-based FOG detection model performance for various merging thresholds.

Participant	MT = 0 s		MT = 1 s		MT = 2 s		MT = 3 s	
	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)
P01	88.1	88.2	88.3	88.3	89.5	87.3	88.8	87.4
P02	81.0	90.2	81.4	90.4	80.6	90.1	81.0	90.2
P03	70.6	93.1	72.0	93.0	74.8	93.1	73.4	93.0
P06	90.6	90.7	93.8	90.3	90.6	90.6	93.8	90.3
P07	64.9	86.9	65.1	86.4	63.2	86.6	61.3	86.8
P08	87.2	87.2	87.2	87.2	86.6	87.3	87.0	87.0
P09	94.4	81.6	93.1	80.8	98.6	80.4	94.1	80.4
Mean	**82.4**	**88.3**	**83.0**	**88.1**	**83.4**	**87.9**	**82.8**	**87.9**
SD	**10.1**	**3.4**	**10.0**	**3.6**	**10.8**	**3.7**	**11.0**	**3.7**

MT: merging threshold, Sens: sensitivity, Spec: specificity.

Table 5.3: Window-based FOG prediction model performance for various merging thresholds.

Participant	MT = 0 s		MT = 1 s		MT = 2 s		MT = 3 s	
	Sens (%)	**Spec (%)**	**Sens (%)**	**Spec (%)**	**Sens (%)**	**Spec (%)**	**Sens (%)**	**Spec (%)**
P01	70.5	81.7	68.8	80.6	63.1	82.0	67.3	79.0
P02	55.3	83.5	57.3	83.1	58.2	83.4	59.0	82.3
P03	61.7	93.4	60.2	93.9	57.4	94.3	63.1	93.1
P06	87.1	88.7	85.1	89.1	82.2	90.7	82.2	90.3
P07	72.2	67.5	68.6	67.0	66.2	66.7	65.6	64.7
P08	77.6	84.3	71.2	84.5	69.4	85.6	67.2	82.8
P09	89.2	75.6	87.8	76.1	82.4	76.7	78.6	73.7
Mean	**73.4**	**82.1**	**71.3**	**82.1**	**68.4**	**82.8**	**69.0**	**80.9**
SD	**11.5**	**7.9**	**10.7**	**8.1**	**9.6**	**8.4**	**7.7**	**9.0**

MT: merging threshold, Sens: sensitivity, Spec: specificity.

FOG episode-based evaluation results are presented in Table 5.4 and Table 5.5. For the FOG detection model (Table 5.4), the mean percentage of correctly identified FOG episodes increased from 91.3% for 0 s merging threshold to 93.3% for 2 s merging threshold. For the prediction model (Table 5.5), the mean percentage of correctly identified FOG episodes increased from 94.0% (0 s threshold) to 95.9% (3 s threshold). For the detection model, the highest percentage of correctly identified FOG episodes occurred using a 2 s merging threshold. For prediction, the highest percentage was achieved with a 3 s merging threshold.

For the detection model, changing merging thresholds from 0 s to 3 s, led to FOG identification (earliest MTD) occurring 0.21 s later (changing from -0.4 s to -0.19 s). When changing merging threshold from 0 s to 2 s, which had the best percentage of correctly identified FOG episodes, the mean ID occurred 0.14 s later (-0.4 to -0.26 s). For the prediction model, changing merging thresholds from 0 s to 3 s led to the FOG identification (earliest MTD) occurring 0.08 s later (changing from -0.56 s to -0.48 s). For both detection and prediction models, a negative ID indicated FOG prediction since FOG identification was before FOG onset.

Table 5.4: Episode-based FOG detection model performance for various merging thresholds.

Participant	MT = 0 s		MT = 1 s		MT = 2 s		MT = 3 s	
	EI (%)	Average ID (s)	EI (%)	Average ID (s)	EI (%)	Average ID (s)	EI (%)	Average ID (s)
P01	91.8	0.02	91.7	0.04	93.8	0.01	93.8	0.03
P02	85.7	0.48	85.7	0.49	85.7	0.47	85.7	0.47
P03	71.4	-0.34	71.4	-0.32	84.6	-0.13	76.9	-0.18
P06	100.0	-0.35	100.0	-0.41	100.0	-0.35	100.0	-0.41
P07	90.0	-0.72	90.1	-0.62	89.0	-0.21	88.5	-0.08
P08	100.0	-1.09	100.0	-0.73	100.0	-0.53	100.0	-0.56
P09	100.0	-0.83	100.0	-0.78	100.0	-1.10	100.0	-0.58
Mean	**91.3**	**-0.40**	**91.3**	**-0.33**	**93.3**	**-0.26**	**92.1**	**-0.19**
SD	**9.7**	**0.50**	**9.7**	**0.43**	**6.4**	**0.45**	**8.2**	**0.34**

MT: merging threshold, ID: identification delay, EI: episodes identified as a percentage of the total number of FOG episodes for each participant, ID: identification delay (positive delay indicates FOG identified after onset, negative delay indicates FOG identified before onset).

Table 5.5: Episode-based FOG prediction model performance for various merging thresholds.

Participant	MT = 0 s		MT = 1 s		MT = 2 s		MT = 3 s	
	EI (%)	Average ID (s)	EI (%)	Average ID (s)	EI (%)	Average ID (s)	EI (%)	Average ID (s)
P01	95.9	-0.02	95.8	0.00	89.6	0.04	91.7	-0.01
P02	94.3	0.30	94.3	0.27	97.1	0.27	100.0	0.30
P03	78.6	-0.33	64.3	-0.49	76.9	-0.26	92.3	-0.28
P06	100.0	-0.49	100.0	-0.49	100.0	-0.59	100.0	-0.61
P07	97.3	-1.17	97.1	-1.01	95.8	-0.83	94.3	-0.76
P08	91.7	-1.15	100.0	-0.81	100.0	-0.72	92.9	-1.08
P09	100.0	-1.10	100.0	-1.12	100.0	-0.98	100.0	-0.92
Mean	**94.0**	**-0.56**	**93.1**	**-0.52**	**94.2**	**-0.44**	**95.9**	**-0.48**
SD	**6.9**	**0.55**	**11.9**	**0.47**	**7.9**	**0.43**	**3.6**	**0.46**

MT: merging threshold, ID: identification delay, EI: episodes identified as a percentage of the total number of FOG episodes for each participant, ID: identification delay (positive delay indicates FOG identified after onset, negative delay indicates FOG identified before onset).

The number of true positive (TP) and false positive (FP) MTD for each participant are presented in Table 5.6 and Table 5.7. Detection model precision did not change with merging, with 40.3% precision for a 3 s merging threshold. Prediction model precision decreased from 19.4% to 14.3% as the merging threshold increased from 0 s to 3 s.

Table 5.6: MTD precision for the FOG detection model.

Participant	MT = 0 s			MT = 1 s			MT = 2 s			MT = 3 s		
	TP	**FP**	**PR (%)**	**TP**	**FP**	**PR (%)**	**TP**	**FP**	**PR (%)**	**TP**	**FP**	**PR (%)**
P01	324	231	58.4	323	227	58.7	330	245	57.4	321	236	57.6
P02	436	346	55.8	443	337	56.8	434	346	55.6	436	343	56.0
P03	79	268	22.8	82	276	22.9	87	270	24.4	81	271	23.0
P06	221	575	27.8	233	608	27.7	221	570	27.9	232	604	27.8
P07	1147	409	73.7	1128	437	72.1	1131	407	73.5	1196	414	74.3
P08	211	391	35.0	213	381	35.9	206	374	35.5	206	373	35.6
P09	62	797	7.2	61	844	6.7	67	851	7.3	71	836	7.8
Total	2480	3017		2483	3110		2476	3063		2543	3077	
Mean SD			**40.1 21.6**			**40.1 21.5**			**40.2 21.2**			**40.3 21.5**

MT: merging threshold, TP: true positive (MTD within MTD target zone), FP: false positive (MTD outside MTD target zone), PR: precision (PR= TP/ (TP+ FP) ×100).

Table 5.7: MTD precision for the FOG prediction model.

Participant	MT = 0 s			MT = 1 s			MT = 2 s			MT = 3 s		
	TP	**FP**	**PR (%)**	**TP**	**FP**	**TP**	**FP**	**PR (%)**	**TP**	**TP**	**TP**	**FP**
P01	171	383	30.9	162	377	30.1	137	338	28.8	154	429	26.4
P02	117	705	14.2	123	721	14.6	126	699	15.3	132	781	14.5
P03	41	205	16.7	40	154	20.6	32	157	16.9	42	213	16.5
P06	73	728	9.1	70	649	9.7	63	474	11.7	67	535	11.1
P07	998	1565	38.9	793	1641	32.6	508	1691	23.1	365	1863	16.4
P08	120	437	21.5	78	413	15.9	64	342	15.8	66	492	11.8
P09	48	1008	4.5	46	883	5.0	44	832	5.0	32	1010	3.1
Total	1568	5031		1312	4838		974	4533		858	5323	
Mean SD			**19.4 11.2**			**18.3 9.4**			**16.7 7.1**			**14.3 6.5**

MT: merging threshold, TP: true positive (MTD within MTD target zone), FP: false positive (MTD outside MTD target zone), PR: precision (PR= TP/ (TP+ FP) ×100).

Table 5.8 presents the result of using a 2.5 s no-cue interval after each triggered cue. The no-cue interval was applied to the models with the highest precision (i.e., detection model with 3 s merging threshold, prediction model with 0 s merging threshold). For detection model episode identification, the no-cue interval did not change the percent of identified FOG. For the prediction model, the no-cue interval reduced the mean percentage of identified FOG minimally (94% to 93.8%). This decrease was due solely to participant P07 for whom the percentage of identified FOG episodes decreased from 97.3% to 96.4%. The no-cue interval

reduced the number of false positive and true positive MTD for both the detection and prediction models. The no-cue interval decreased mean detection model precision from 40.3% (MT = 3 s) to 31.8% and increased mean prediction model precision from 19.4% (MT = 0 s) to 30.6%.

Table 5.8: MTD precision for FOG prediction and detection models using a 2.5 s no-cue interval between consecutive cues.

Participant	Detection Model MT = 3 s (2.5 s no-cue interval)				Prediction model MT = 0 s (2.5 s no-cue interval)			
	EI (%)	TP	FP	PR (%)	EI (%)	TP	FP	PR (%)
P01	93.8	45	40	52.9	95.9	47	51	48.0
P02	85.7	45	73	38.1	94.3	50	76	39.7
P03	76.9	11	67	14.1	78.6	12	54	18.2
P06	100.0	23	108	17.6	100.0	24	109	18.0
P07	88.5	187	98	65.6	96.4	222	164	57.5
P08	100.0	25	67	27.2	91.7	23	64	26.4
P09	100.0	10	132	7.0	100.0	11	165	6.3
Total		346	585			389	683	
Mean	**92.1**			**31.8**	**93.8**			**30.6**
SD	**8.2**			**19.9**	**6.8**			**17.0**

MT: merging threshold, ID: identification delay, EI: episodes identified as a percentage of the total number of FOG episodes for each participant, TP: true positive (MTD within MTD target zone), FP: false positive (MTD outside MTD target zone), PR: precision (PR= TP/ (TP+ FP) ×100).

Figure 5.2 shows an example walking session with MTD TP and FP. Without a no-cue interval (Figure 5.2 a), the first FOG episode was detected at the beginning of the episode (leftmost green circle at approximately 26 s). The second FOG was predicted approximately 1 s before FOG onset (multiple MTD starting at approximately 44 s), and MTD occurred in groups of consecutive windows for both the TP MTD (green) and FP MTD (red). When the no-cue interval was used (Figure 5.2 b), there was also successful FOG identification at the beginning of the first episode (TP MTD at 26 s), successful FOG prediction (TP MTD at 44 s), and the number of false positive MTD was reduced from 15 (Figure 5.2 a) to 2 (Figure 5.2 b).

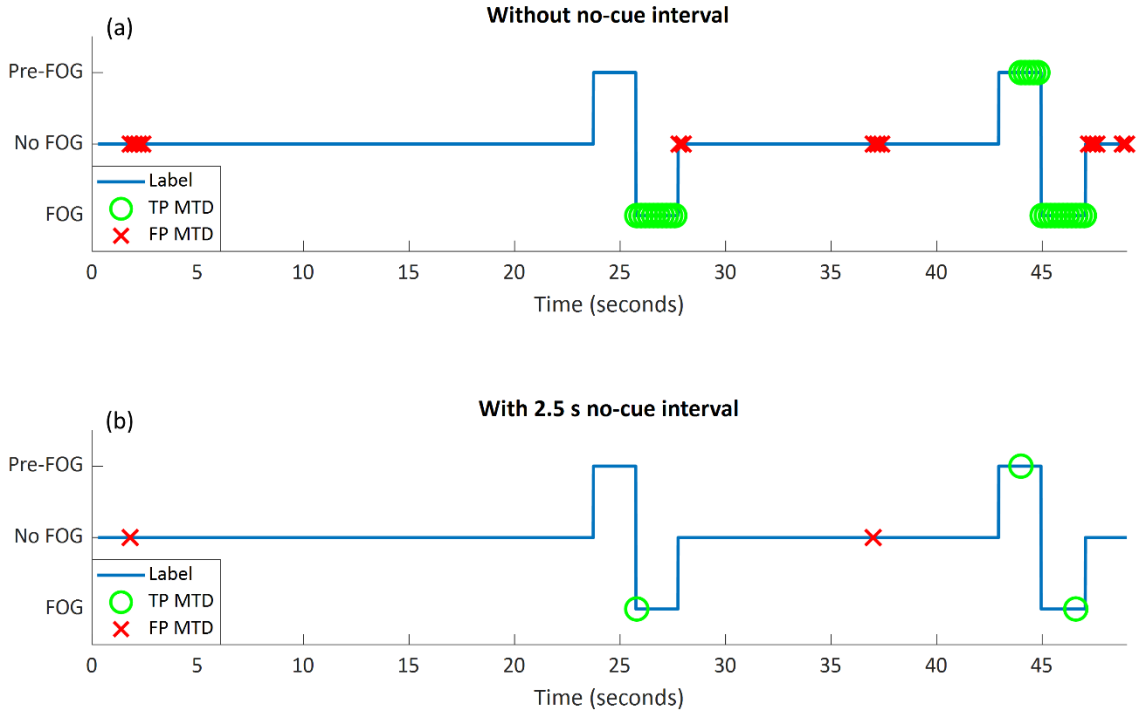

Figure 5.2: Example session of walking data classification and freeze identification: (a) without no-cue interval, (b) with 2.5 s no-cue interval. TP MTD: true positive model trigger decision (MTD within MTD target zone), FP MTD: false positive model trigger decision (MTD outside MTD target zone).

5.4 Discussion

The best performing FOG detection model used a 2 s merging threshold, whereas the best prediction model had a 0 s merging threshold (i.e., no merging). For the window-based evaluation, model performance was similar for all detection models, across merging thresholds, and there was a slight difference in performance for prediction models. Model performance was similar to other person-independent FOG detection [57,121,126,128,130] and prediction [17,23,26,28] models in the literature.

For the FOG episode-based analysis, the percentage of successfully identified FOG episodes increased slightly due to FOG-episode merging for both the detection (+ 2.0%) and prediction (+ 1.9%) models. The prediction model with a 3 s merging threshold outperformed the detection model by identifying 95.9% of FOG episodes. For all merging thresholds of the

detection and prediction models, FOG episodes were identified prior to the FOG onset; therefore, both detection and prediction models were able to predict FOG.

The earliest predictions occurred without merging (0 s merging threshold). Individual participant FOG identification was as early as 1.09 s before FOG onset for the detection model (P08, Table 5.4), and 1.10 s to 1.17 s before FOG onset for the prediction model (P09-P07, Table 5.5). When averaged across participants, the earliest identifications were 0.40 s before FOG onset for the detection model and 0.56 s before FOG onset for the prediction model, which both occurred with no merging. The FOG identification was closer to freeze initiation when the merging threshold was 3 s for detection (0.19 s before FOG onset) and 2 s for prediction (0.44 s before FOG onset). Therefore, a merging threshold of 3 s for detection and 2 s for prediction would provide the shortest time for preventative cueing. Merging FOG episodes may not be beneficial in a preventative cueing system since merging led to later FOG identifications but similar FOG identification percentages.

For the detection model, less than ± 0.2% differences in MTD precision were found between merging thresholds. For the prediction model, increasing the merging threshold from 0 s to 3 s decreased the number of true positive MTD from 1568 to 858 and increased the number of false positives from 5031 to 5323, resulting in a 5.1% decrease in precision. This may be the result of having fewer data windows in the target class during training due to merging. Also, there were more FP compared to TP, for both detection and prediction models.

Models tended to produce grouped zones of MTD (Figure 5.2 a), likely because of the 80% overlap between consecutive windows, where data in successive windows were similar and lead to the same classification. To reduce FP, a larger shift between windows may be helpful [57]; however, this would decrease the temporal resolution of a cueing protocol. The 2.5 s no-cue interval greatly reduced the number of false positive MTD (5323 to 585 for detection, 5031 to 683 for prediction) by excluding consecutive FP MTD after the first MTD in the group. As shown in Figure 5.2, a TP MTD near the end of a FOG episode can eliminate FP MTD immediately after the end of the FOG episode, since the FP MTD would fall within the no-cue interval. For detection, the no-cue interval had no effect on the percentage of identified FOG episodes. For prediction, the no-cue interval only affected the percentage of

identified FOG episodes for participant P07, and this was only a 0.9% difference. P07 had many short FOG in rapid succession. FOG episodes that began within a no-cue interval were considered to be successfully detected, whereas Pre-FOG data for subsequent short freezes within the no-cue interval were ignored. Therefore, models using the no-cue interval may miss FOG episodes that otherwise would have been predicted from the Pre-FOG data. However, these missed episodes do not necessarily indicate decreased model performance, since in a real application, if a cue were given, the subsequent (missed) episode may never occur. In this analysis, 2.5 s was considered to be enough time for the person to respond to the cue and for the model to collect additional data to inform the next classification. Further study is required to determine the time required for gait to adjust following a cue, if the time is person or FOG-subtype specific, and whether subsequent FOG episodes can be avoided. The results could then be used as relevant parameters for personalized FOG cueing systems. For example, the user's reaction to the cue could be the input of a secondary classifier that is trained using post-cue data. The secondary classifier could determine if the gait parameters are stabilizing and cueing can be stopped, or that gait remains abnormal, and cueing should be continued or modulated.

For the FOG detection model, merging successive FOG episodes did not substantially improve performance and no-merging resulted in the earliest MTD. For the prediction model, FOG episode merging increased the percentage of identified FOG episodes, but slightly decreased window-based sensitivity and specificity, decreased model precision, and resulted in less time between identification and freeze onset. The improvement in percentage of identified FOG episodes was at the cost of identifications being made later. For a FOG prediction model intended to be used in a cueing system, where early detection of FOG may be important, the merging of FOG episodes could be detrimental.

5.5 Conclusion

This chapter examined the effects of defining FOG either as a period of gait disruption (merging successive FOG), or based on an event (no merging), on FOG detection and prediction. For detection, defining FOG as a period of gait disruption produced minimal changes in performance; therefore, expert labeling based on periods of ineffective gait is likely

sufficient and labeling the onset and termination of each successive FOG episode within a larger period of gait disruption may not be required. Prediction model performance was adversely affected by increasing the merging threshold, specifically in terms of precision. Therefore, FOG prediction models should be trained using event-based FOG definitions (e.g., foot leaves or fails to leave the ground) that consider successive FOG episodes separately.

Chapter 6

FOG Prediction using Unilateral and Bilateral Plantar Pressure Data

6.1 Introduction

In this chapter, FOG prediction models developed using data from one plantar pressure insole worn unilaterally and two insoles worn bilaterally are compared. Three datasets were created using features from the more severely affected side (MSS) of the body, the less severely affected side (LSS), and both sides. Feature selection was performed, and FOG prediction models were trained using each dataset. The best MSS, LSS and bilateral-limb models were identified and compared.

6.2 Methodology

6.2.1 Feature Extraction, Feature Selection and Target Class Creation

Windowing and feature extraction were performed as described in Sections 3.6.1 and 3.7.1. For all prediction models developed in this chapter, the target class included windows containing purely Pre-FOG data, purely FOG data, and windows containing both Pre-FOG and FOG data (Section 3.6.2, Figure 3.6 a,b).

For each data window, 166 unilateral features and 1 bilateral feature (number of weight shifts) were extracted from the plantar pressure data, resulting in 333 features (Section 3.7.1). Relief-F ranking was used to determine the best features. For the bilateral-limb models, all 333 features were ranked. For the unilateral models, separate datasets were created with 166 MSS or 166 LSS features. Relief-F feature ranking was then performed for both the MSS and LSS datasets.

6.2.2 Prediction Model Development

All prediction models developed in this chapter used the same parameters and training methods. The only difference between the models was the input dataset and the number of input features. Separate prediction models were trained using 5, 10, 15, 20, 25 and 30 features from each of the MSS, LSS, and bilateral datasets.

Each data window was classified using a binary classification model. In this chapter, RUSBoosted decision tree ensembles (100 trees with 5 splits each, Section 4.2.3) were trained and validated using LOFO cross validation.

6.2.3 Model Evaluation

The trained models were evaluated using windows and FOG episodes. The window-based evaluation compared each window classification to the ground truth label and calculated sensitivity and specificity. Episode based evaluation was similar to the evaluation in Chapter 5, where three consecutive windows classified as belonging to the target class resulted in a model trigger decision (MTD) (Figure 6.1). If a MTD occurred within the MTD target zone, then the corresponding FOG episode was successfully identified. The identification delay (ID) was the time between FOG onset and a successful MTD identification. However, in this chapter, the MTD target zone was defined differently and was specific to each FOG episode.

In the literature, Pre-FOG gait has been identified 3 steps prior to onset [29], and predictions have been reported 4-5 s in advance [20,28]. Furthermore, model classification target zones have been defined as 8 s prior to FOG onset [50]. Therefore, the MTD target zone in this analysis was initially defined as the period beginning 6 s prior to FOG onset until the end of the FOG episode. The period between the beginning of the MTD and FOG onset is the prediction target zone (Figure 6.1). To define the MTD target zone for a given FOG, the prediction target zone was initially set to 6 s. If there was another FOG, a stand to walk transition, or a turn to walk transition within the 6 s period prior to FOG onset, then the prediction target zone was shortened to exclude these turning, standing, or FOG data. This was done to ensure that false positives caused by the end of the previous FOG episode, turn to walk transitions, or stand to walk transitions were not mistakenly interpreted as predictions of the upcoming FOG. To ensure that the turning data were not included in the MTD target zone, a 1 s delay was used so that the prediction target zone started 1 s after the end of the turn. Similarly, for transitions from standing to walking, a 1 s delay was used to remove periods of gait initiation from the MTD target zone (Figure 6.1).

Each MTD was considered to be either a true positive (within the MTD target zone) or a false positive (outside the MTD target zone). The MTD false positive rate was calculated as the total number of false positive MTD per trial for each participant, and as the average FP rate across all participants. While freezing during gait initiation is a known FOG manifestation [4,73], for this analysis, FP MTD that occurred during standing or gait initiation were ignored. Gait initiation was defined as the first second of walking after standing. As a final step in model development, a 2.5 s no-cue interval was used.

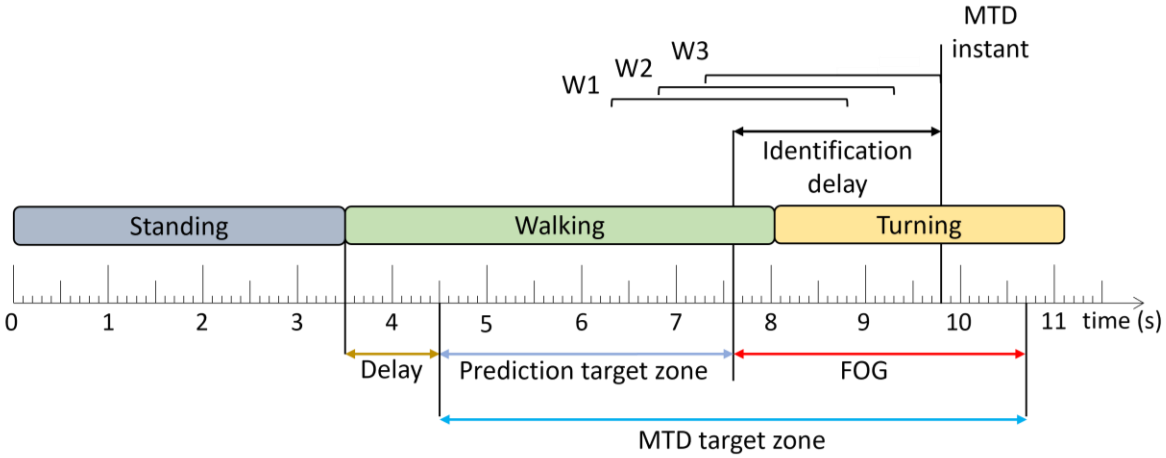

Figure 6.1: Model trigger decision diagram. Three consecutive windows classified as the target class (W1-W3) results in a model trigger decision (MTD), where the MTD instant corresponds to the end of the third window. FOG is successfully identified if there is a MTD instant within the MTD target zone. The time difference between FOG onset and MTD instant is the identification delay (ID). The period between the beginning of the MTD target zone and the FOG onset is the prediction target zone.

6.3 Results

FOG prediction model performance for each number of features is presented in Figure 6.2. Overall, the highest sensitivity (79.5%) was for the LSS model with 5 features. The LSS model had the highest sensitivity for 5, 10, 15 and 25 features. The bilateral model had the highest sensitivity for 20 (74.6%) and 30 (66.7%) features.

Specificity for all MSS, LSS, and bilateral models ranged between 81.3% and 88.0%. The highest overall specificity (88.0%) was for the bilateral model with 30 features. The LSS (87.5%) and MSS (83.9%) models also had high specificity using 30 features.

The highest percentage of identified FOG episodes ranged from 90.3% to 94.9% for all models that used 5, 10 or 15 features. For increasing numbers of features, the percentage of identified FOG decreased for all models. Overall, the highest percent of identified FOG (94.9%) was for the LSS models with 5 or 10 features.

The LSS and bilateral models produced similar identification delays using 5, 10, 15 and 20 features. Overall, the earliest identifications were for the bilateral and LSS models with 5 features, which both had a -1.1 s ID. For all models that used 5 or 10 features, the ID values were between -0.9 s and -1.1 s.

The MSS models had the lowest average false positive rate per walking trial for all number of features, and the LSS models had the highest FP rates. Overall, the lowest false positive rate was for the MSS model using 30 features (1.0 FP/trial). The highest false positive rate was for the LSS using 5 or 10 features (3.4 FP/trial).

Overall, using more features tended to increase specificity, decrease sensitivity, decrease percentage of identified FOG episodes, and decrease number of false positives per trial. Increasing the number of features resulted in later predictions for the bilateral and MSS models.

To select the ideal number of features, the different models were ranked for each evaluation metric, then the summation of ranks was calculated and the number of features with the smallest sum was selected. For instance, the MSS model with 5 features was the third best MSS model for sensitivity, fifth best for specificity, third best for percentage of identified FOG episodes, best (first ranked) for ID, and fifth best for FP rate. These ranks (3, 5, 3, 1, 5) were summed to produce a summed score of 17 for the MSS model with 5 features. This ranking was done for the MSS, LSS, and bilateral models (Table 6.1).

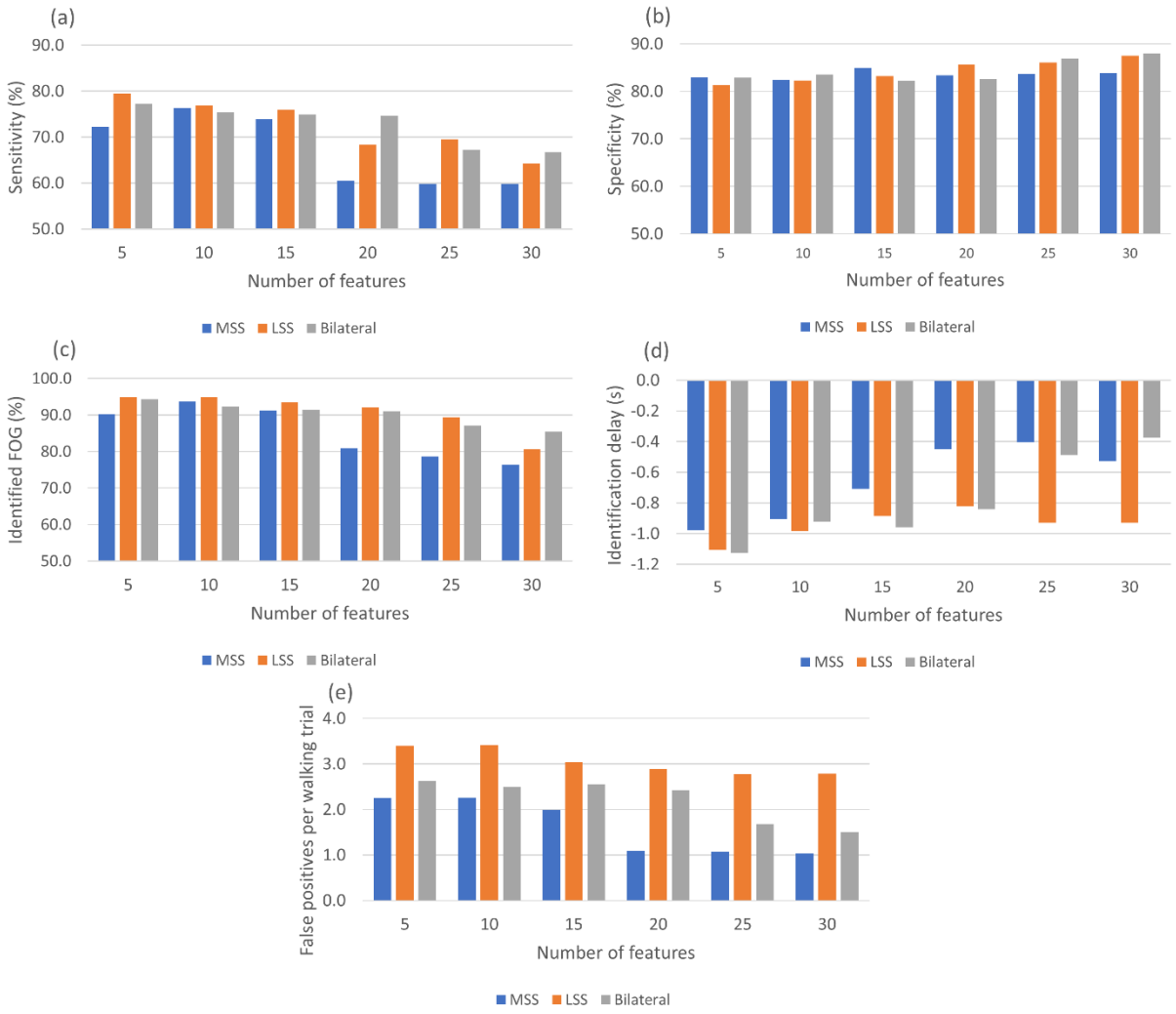

Figure 6.2: FOG prediction model performance: a) sensitivity, b) specificity, c) episodes identified as a percentage of the total number of FOG episodes for each participant, d) average identification delay, e) average number of false positives per walking trial.

Table 6.1: Summed ranks for each combination of dataset and number of features.

Dataset	**Number of input features**					
	5	**10**	**15**	**20**	**25**	**30**
MSS	17	16	12*	20	22	18
LSS	15*	16	19	21	16	18
Bilateral	13*	14	19	20	19	20

* indicates model selected as best.

According to the ranking, the best MSS model used 15 features, and the best LSS and bilateral models both used 5 features. The features used in the best models are presented in Table 6.2. To examine model performance for each participant, the cross-validation results for the best MSS, LSS, and bilateral models are presented in Table 6.3 and Table 6.4.

Table 6.2: Features used for the best MSS, LSS and bilateral models.

MSS **15 features**	**LSS** **5 features**	**Bilateral** **5 features**
Number of AP COP path reversals	Number of AP COP path reversals	Dominant frequency of COP velocity AP for right leg
Dominant frequency of COP velocity AP	Power in freeze band (3-8 Hz) of COP velocity AP	Number of AP COP path reversals for left leg
Dominant frequency of COP velocity ML	Dominant frequency of COP velocity AP	Number of AP COP path reversals for right leg
Mean energy of WT aC of COP position AP	Power in freeze band (3-8 Hz) of COP position AP	Dominant frequency of COP velocity ML for right leg
Number of ML COP path deviations	Dominant frequency of COP acceleration AP	Mean energy of WT aC of COP position AP for right leg
Mean WT aC of COP position AP		
Power in freeze band (3-8 Hz) of COP velocity AP		
Mean WT aC of COP velocity AP		
Dominant frequency of COP acceleration ML		
Power in freeze band (3-8 Hz) of COP position AP		
Dominant frequency of COP acceleration AP		
Mean WT dC of GRF		
Max energy of WT aC of COP position AP		
Power in freeze band (3-8 Hz) of COP position ML		
Mean duration of AP COP path reversals		

AP: anterior-posterior, ML: medial-lateral, WT: wavelet transform, aC: approximation coefficient, dC: detail coefficient.

Table 6.3: Sensitivity and specificity results for the best MSS, LSS and bilateral models.

Participant	MSS 15 features		LSS 5 features		Bilateral 5 features	
	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)
P01	59.2	87.6	77.9	73.3	69.7	81.8
P02	73.7	87.9	70.0	86.6	71.7	86.7
P03	60.7	89.3	73.8	87.5	68.3	89.5
P06	84.0	88.3	95.1	88.6	93.5	89.6
P07	57.2	83.6	64.5	79.7	68.8	81.0
P08	88.8	80.7	92.6	85.2	89.1	79.1
P09	93.7	76.9	82.3	68.2	79.7	72.3
Mean	**73.9**	**84.9**	**79.5**	**81.3**	**77.3**	**82.9**
SD	**14.1**	**4.3**	**10.5**	**7.3**	**9.7**	**5.8**

Sens: sensitivity, Spec: specificity.

Table 6.4: Episode based model performance for the best MSS, LSS and bilateral models.

Participant	MSS 15 features			LSS 5 features			Bilateral 5 features		
	EI (%)	ID (s)	FPR (FP/trial)	EI (%)	ID (s)	FPR (FP/trial)	EI (%)	ID (s)	FPR (FP/trial)
P01	85.7	-0.6	1.5	87.8	-0.7	1.4	89.8	-1.2	1.8
P02	97.1	-0.3	0.9	91.4	-1.0	1.1	94.3	-1.1	1.1
P03	78.6	-0.7	2.2	85.7	-0.7	2.5	85.7	-0.7	2.2
P06	100.0	-0.1	2.2	100.0	-0.8	2.0	100.0	-0.6	1.7
P07	77.4	-0.3	1.7	99.1	-1.0	6.8	90.0	-0.8	3.5
P08	100.0	-1.1	1.7	100.0	-1.5	3.9	100.0	-1.4	3.2
P09	100.0	-1.9	3.8	100.0	-2.1	6.1	100.0	-2.0	5.0
Mean	**91.3**	**-0.7**	**2.0**	**94.9**	**-1.1**	**3.4**	**94.3**	**-1.1**	**2.6**
SD	**9.6**	**0.6**	**0.8**	**5.9**	**0.5**	**2.1**	**5.5**	**0.4**	**1.2**

ID: identification delay, EI: episodes identified as a percentage of the total number of FOG episodes for each participant, FPR: false positive rate.

6.4 Discussion

The overall best model for FOG prediction was the bilateral model, with 77.3% sensitivity, 82.9% specificity, -1.1 s ID, 94.3% of FOG episodes identified, and 2.6 false positives per walking trial. Compared to the bilateral model, the MSS model had 3.4% lower sensitivity and identified 3% fewer FOG episodes, 0.4 s later. Thus, the MSS model identified fewer FOG episodes and identifications were made later. The LSS model had similar sensitivity, specificity, percentage of identified FOG episodes, identification delay and 0.8 more false positives per walking trial than the bilateral model; thus, the bilateral model had similar

prediction performance but fewer false positives. Therefore, the bilateral model is recommended.

Between the two single-limb models, the LSS model had higher prediction performance than the MSS model. LSS model sensitivity and specificity were comparable to other single-sensor FOG prediction studies in the literature [19,21,26,27]. The best LSS model performed better for FOG prediction than a similar tree based algorithm (AdaBoosted C4.5 decision tree) that used data from a single waist mounted IMU [19]. Compared to a FOG prediction model that used EEG signals, the LSS had lower sensitivity (79.5% compared to 85.86%), and similar specificity (81.3% compared to 80.25%) [26]. However, a single plantar pressure sensor could be integrated into regular footwear and could therefore be used in a simpler and much more user-friendly wearable system than EEG sensors.

For the bilateral model, sensitivity and specificity were lower than for models in existing literature where sensors were worn on both the limbs [23]. A model using gyroscope data from the shins predicted FOG with 84.1% sensitivity and 85.9% specificity [23]. However, the model was developed using data from only 35 FOG episodes.

Other models in the literature achieved even higher sensitivity and specificity [22,24,50]. For example, a person-specific model using an ensemble of 9 SVM classifiers and data from 3 IMU sensors reported 93% sensitivity and 87% specificity [50]. Using the same dataset, a 3 class (Pre-FOG, FOG, Non-FOG) k-nearest neighbours classifier achieved 94.1% sensitivity and 97.1% specificity. However, these systems were not person-independent or used multiple sensors on various parts of the body, and thus are not directly comparable to the system analyzed in this thesis.

Compared to other models in the literature, the LSS model FOG episode identification performance was very good. The LSS model identified 94.9% of episodes, which is similar to [50] where, 94% of episodes were identified, and was only slightly worse than a person-specific model used in [25] that identified 97.4% of episodes. The best MSS, LSS, and bilateral models in this thesis all identified more than 91% of the FOG episodes. Furthermore, for the LSS and bilateral models, the average identification delay was -1.1 s. Thus, if used as part of a real-life cueing system, the LSS or bilateral models would cue most of the FOG episodes,

with identifications made just over 1 s prior to FOG onset. Although earlier predictions are generally considered to be preferable, the time needed for a user to respond to a cue and alter their gait to avoid freezing is largely unknown. It is thus unclear whether FOG identifications being made 1.1 s prior to onset is sufficient for use in a real-life system. Future testing involving FOG prediction and cueing is required to determine how far in advance the FOG predictions must be.

The comparison of models that used plantar pressure data from the MSS and LSS was a primary goal of this analysis. The LSS model had higher sensitivity, earlier FOG identifications, and identified a higher percentage of FOG episodes than the MSS model. The better prediction performance of the LSS may be explained by an increased role of this limb in balance and postural stability during walking. Differences between the MSS and LSS have been identified in various motor tasks [192], and participants with PD (with and without FOG) preferentially adjusted the positioning of their least affected limb in order to retain balance after slipping [193]. Therefore, the LSS limb may also be preferentially used for stability during walking, similar to how amputees rely on the intact limb for stability and balance [194]. Postural stability and FOG are intricately related [48] and dual-task walking (a common trigger for FOG) can negatively affect stability in freezers [195]. Furthermore, stability and postural control in PD can be assessed using COP [196,197]. COP based features that indicate postural instability may also indicate upcoming FOG. Therefore, if participants are preferentially using the less severely affected limb for stability control when walking, the link between instability and FOG may lead to the LSS being the more informative limb for FOG prediction. The connection between postural stability, FOG, and the preferential use of the LSS or MSS for stability control during walking should be further investigated.

The best MSS model had the highest specificity, lowest false positive rate, and latest predictions compared to the LSS and bilateral models. Therefore, the MSS predicted FOG less in advance but resulted in fewer false positive MTD. The best MSS model had a false positive rate of 2.0 FP per walking trial. In addition to the number of FP per walking trial, the FP frequency was estimated using the duration of each walking trial. Once averaged for all walking trials and all participants, the best MSS model produced one FP approximately every

35 s of walking. Similarly, one FP was produced approximately every 26 s for the bilateral model and every 23 s for the LSS model. However, since this research used a specially designed freeze inducing walking path, fewer false positives may be experienced during daily walking.

In a real-life setting, which limb to use may depend on the person, their FOG history, and the intervention (cueing) approach. For someone who tends to recover independently following a freeze, minimizing false positives may be more important than early cueing. Thus, instrumenting the MSS may be preferable since it had higher specificity and fewer false positives. In contrast, for someone who frequently experiences loss of balance and potential falls when freezing, collecting data from the LSS may be preferable, since the FOG episodes would be identified earlier and with higher sensitivity. For this person, a late or missing cue may be more disruptive to overall walking than the increased number of false positives. In addition, the type of cue may also influence the decision to instrument the MSS or LSS limb. When using a minimalistic cue, FP may be better tolerated and thus permit the use of the LSS model since false positive cues are less of a concern. However, an intense or potentially bothersome cue may be best used with MSS instrumentation to reduce unnecessary cueing.

While the LSS model performance was similar to the bilateral model, the bilateral model is recommended for FOG prediction since it produced fewer false positives. LSS false positive frequency was one FP every 23 s, compared to one FP every 26 s for the bilateral model. If implemented in a real-life cueing system, a 3 s difference in FP frequency may be imperceptible to the user. Single sensor systems can potentially be simpler, less expensive, and more user-friendly than systems with multiple sensors and the use of a single sensor instead of two may be more important than a slight increase in FP rate. Therefore, models that use plantar pressure data from the LSS may be preferable to models that use plantar pressure data from both feet, in some situations.

6.5 Conclusion

This chapter compared FOG prediction models that used plantar pressure data collected from the more severely affected side, the less severely affected side, and both sides together.

RUSBoosted ensembles of decision trees were trained. The best models were MSS with 15 features, and LSS and bilateral with 5 features. The LSS model had higher sensitivity and identified a higher percentage of FOG episodes more in advance of the FOG onset, compared to the MSS. The MSS model had higher specificity and fewer false positives. In a system that uses a single plantar pressure sensor, the decision to instrument the LSS or MSS may be person specific. For someone who tends to recover independently from FOG, instrumenting the MSS may be preferable since there would be fewer false positives. However, for someone who experiences loss of balance during freezing, cueing earlier may be more important than minimizing false positives, thus instrumenting the LSS may be preferable.

LSS and bilateral model performance was similar for all evaluation metrics except the false positive rate. The LSS model had a higher FP rate than the bilateral model. Therefore, in terms of prediction performance, using plantar pressure data from both feet is recommended. However, since the difference in FP rate between the LSS and bilateral models was small, the advantages of a single sensor system may outweigh the increase in FP rate. In practice, using a single-sided plantar-pressure based FOG prediction system could enhance wearability and compliance since fewer sensors would need to be worn.

Chapter 7
Conclusion

In this thesis, FOG detection and prediction models that used data from IMU and plantar pressure insole sensors were developed. The plantar pressure based models detected and predicted FOG with classification performance comparable to models that used both IMU and plantar pressure data. Merging multiple FOG episodes that occur in rapid succession had little effect on detection model performance but was detrimental to prediction. Models that used plantar pressure data collected from the LSS and from both feet were found to have similar FOG prediction performance.

All thesis objectives were met:

Objective 1: Develop FOG detection and prediction models using IMU and plantar pressure data.

a) **Determine which features are most useful for FOG detection and prediction using plantar-pressure data alone, IMU data alone, and combined IMU and plantar pressure data.**

FOG detection models were developed using features extracted from plantar pressure, IMU, and IMU and plantar pressure data. The model that used only plantar pressure features produced the highest sensitivity (78.0%), whereas the model that used only IMU features produced the highest specificity (91.6%). The best overall performance was achieved by the model that used 10 features from the combination of IMU and plantar pressure data. Of the top 10 features, only two were calculated from IMU data; the remaining eight features were calculated using the plantar pressure data. Separately, FOG prediction models were trained using features extracted from plantar pressure data. The prediction models performed well, achieving over 77% sensitivity and 82% specificity. Therefore, plantar pressure features were useful for FOG detection and prediction.

From the top 10 features used in the FOG detection model, eight were from the frequency domain and two from the time domain. Similarly, the plantar pressure features used for prediction were from both time domain and frequency domain. The dominant frequency of

COP velocity feature and the number of COP path reversals in the AP direction feature ranked highly for all detection and prediction models. The mean energy of the wavelet transform approximation coefficient vector of COP position in the AP direction also ranked well for multiple models. Therefore, features from the time domain and frequency domain (both FFT and WT) performed well for FOG detection and prediction.

b) Compare the performance of models based on plantar-pressure data alone, IMU data alone, and combined IMU and plantar pressure data, for FOG detection.

All FOG detection models that used IMU data alone, plantar pressure data alone, or both IMU and plantar pressure data, were able to detect FOG. However, the combination of IMU and plantar pressure data gave the best results. The best model was a RUSBoosted ensemble of decision trees built using 10 features from the IMU and plantar pressure data. The model achieved 76.4% sensitivity and 86.2% specificity when classifying 1 s windows of Total-FOG data (i.e., data from 2 s before FOG onset until the end of the FOG episode). Furthermore, the FOG windows were detected with 93.4% sensitivity.

A separate RUSBoosted decision tree ensemble FOG detection model was trained using the same set of 10 features from the IMU and plantar pressure data. The model was trained to detect windows containing FOG data and achieved 83.4% sensitivity, 87.9% specificity, and detected 93.3% of FOG episodes.

c) Compare the performance of models based on plantar-pressure data alone and plantar pressure data combined with IMU data for FOG prediction.

For FOG prediction, two different models were developed, both using RUSBoosted decision tree ensembles. The first model used 10 features extracted from IMU and plantar pressure data, and the second model used 5 features extracted from bilateral plantar pressure data. The FOG prediction model with features from IMU and plantar pressure data achieved 73.4% sensitivity, 82.1% specificity and identified 94% of FOG episodes 0.56 s in advance. The FOG prediction model that used only plantar pressure data achieved 77.3% sensitivity, 82.9% specificity, and identified 94.3% of the FOG episodes 1.1 s in advance. The model using only 5 features extracted from plantar pressure data had higher sensitivity and earlier detections, while the

specificity and percentage of identified FOG episodes were similar between models. Therefore, the model that used 5 plantar pressure features was better than the model that used IMU and plantar pressure based features. The comparison of these prediction models demonstrated that the data from IMU sensors is not necessary to achieve good FOG prediction performance.

Objective 2: Determine if event-based or "period of gait disruption" FOG definitions lead to better classification performance for FOG detection and prediction.

Detection models developed using merged and unmerged data had similar performance. Therefore, when training a FOG detection system, labeling groups of multiple FOG in succession as a single period of gait disruption is sufficient (i.e., each small freeze does not need to be independently labeled). Since ground truth labeling is typically a time-consuming manual task, using the period of gait disruption (merged) approach is more efficient and recommended for FOG detection model development.

Prediction models developed using merged data had worse performance compared to models developed using unmerged data. In particular, using merged data decreased model precision. Therefore, FOG prediction models should be trained using event-based FOG definitions (e.g., foot leaves or fails to leave the ground) that consider successive FOG episodes separately.

Objective 3: Evaluate FOG prediction models that use a single unilateral plantar pressure insole sensor and models that use bilateral sensors.

a) **Determine if models using plantar pressure data from a single foot can predict FOG with performance comparable to models that use plantar pressure data collected from both feet.**

The FOG prediction model that used plantar pressure data from both feet had 77.3% sensitivity, 82.9% specificity, -1.1 s ID, and identified 94.3% of the FOG episodes with a false positive rate of 2.6 FP per walking trial. Compared to the bilateral model, the LSS model had similar sensitivity, specificity, percentage identified FOG episodes, identification delay and 0.8 more false positives per walking trial.

Bilateral and LSS model performance was similar. The LSS produced a slightly higher false positive rate than the bilateral model. In practice, the advantages of single sensor systems may outweigh the increase in false positives. Therefore, a single plantar pressure sensor placed on the LSS could be used to develop a FOG prediction system and produce performance similar to a bilateral system.

b) Determine if models using plantar pressure data from either the more severely affected side (MSS) or the less severely affected side (LSS) produce better classification results for FOG prediction.

When comparing the FOG prediction models developed using data from the MSS and LSS, the LSS model had higher sensitivity and identified a higher percentage of FOG episodes more in advance. The MSS model had higher specificity and fewer false positives. Therefore, in a system that uses a single plantar pressure sensor, the decision to instrument the LSS or MSS may be person specific. For someone who tends to recover independently from FOG, the MSS may be preferable since there would be fewer false positives. For someone who experiences loss of balance during freezing, cueing earlier may be more important than minimizing false positives, thus the LSS may be preferable.

7.1 Future Work

The models and analyses performed in this thesis can be continued and improved in multiple ways. In this thesis, data were collected from only 11 participants. In the FOG detection and prediction literature fewer than 12 participants is common (Table A.1) and only five FOG detection studies had more than 25 participants [57,59,61,99,124]. For prediction, the largest participant pools were 18 participants [17,25]. While the sample size used in this thesis could be considered average, future research should use data from larger groups of freezers. In addition, as women tend to be underrepresented in FOG research, a targeted recruitment strategy should be implemented to address this disparity. More participants will help with model generalization and FOG subtype analysis. For instance, a larger participant pool would allow a more complete understanding of FOG manifestations and analysis of different FOG subtypes leading to FOG-subtype-specific models. Compared to fully personalized models,

FOG-subtype-specific models could be easier to implement since they would not require large amounts of data from a single individual for training. Therefore, models personalization according to FOG-subtype may improve performance compared to fully person-independent models, while avoiding challenges associated with fully person-specific systems.

An important step in developing real-life FOG prediction systems is implementing real-time prediction models. To date, FOG prediction models have not been tested in real-time. The models developed in this thesis, especially the model using data from only one plantar pressure sensor on the LSS, could be integrated into a real-time wearable system. Once a FOG prediction system has been implemented in real-time, cueing systems could be investigated to help prevent or mitigate FOG episodes.

Evaluation of a real-time prediction and cueing system would also allow additional analyses and improvements such as the no-cue interval duration analysis described in Section 5.4. Once the prediction and cueing systems have been tested in real-time, new data could be used to study how gait parameters change following a cue. These data could be used to create models that can identify whether gait parameters are stabilizing and cueing can be stopped, or that gait remains abnormal and cueing should be continued or modulated.

The connection between postural stability, FOG, and the preferential use of the LSS or MSS for stability control during walking should be further investigated. Additional experiments specifically designed to evaluate FOG prediction systems that use plantar pressure data from the LSS and MSS are needed. Moreover, additional factors could be explored such as FOG subtype and medication state, to determine the effect on FOG prediction using plantar pressure data from the LSS and MSS.

For FOG prediction models developed in Chapter 6, most false positives occurred during turning, indicating that differentiating between a turn and a freeze is challenging for the model. Furthermore, features used by the MSS, LSS, and bilateral models predominantly included COP movements in the AP direction. During turning, the COP path includes movement in the ML direction. Thus, the lack of COP features from the ML direction may have contributed to the high number of FP observed during turning and could be investigated in more depth.

To further reduce FP during turning, an activity recognition algorithm could be developed to identify turning. Subsequently, activity recognition and FOG prediction models could be trained to run jointly. Alternatively, the output from the activity recognition model could be used as an input to the FOG prediction model.

Finally, high resolution plantar pressure sensors were used in this research. The most useful features were calculated using COP movements, which can be calculated with a much lower resolution insole sensor. Thus, as part of future research, the models developed in this work could be validated on a simpler insole system with fewer pressure-sensing cells. Using lower resolution plantar pressure sensors would reduce cost and simplify the system, which are both desirable for real-life wearable systems.

References

1. Jankovic J. Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry. 2008 Apr;79(4):368–76.

2. Chaudhuri KR, Healy DG, Schapira AH. Non-motor symptoms of Parkinson's disease: diagnosis and management. Lancet Neurol. 2006 Mar;5(3):235–45.

3. Okuma Y. Freezing of gait and falls in Parkinson's disease. J Park Dis. 2014;4(2):255–60.

4. Bloem BR, Hausdorff JM, Visser JE, Giladi N. Falls and freezing of gait in Parkinson's disease: a review of two interconnected, episodic phenomena. Mov Disord. 2004 Apr 21;19(8):871–84.

5. Masud T, Morris RO. Epidemiology of falls. Age Ageing. 2001;30 Suppl 4:3–7.

6. Young WR, Mark Williams A. How fear of falling can increase fall-risk in older adults: applying psychological theory to practical observations. Gait Posture. 2015 Jan;41(1):7–12.

7. Ward-Griffin C, Hobson S, Melles P, Kloseck M, Vandervoort A, Crilly R. Falls and fear of falling among community-dwelling seniors: the dynamic tension between exercising precaution and striving for independence. Can J Aging. 2004;23(4):307–18.

8. Rahman S, Griffin HJ, Quinn NP, Jahanshahi M. The factors that induce or overcome freezing of gait in Parkinson's disease. Behav Neurol. 2008;19(1):127–36.

9. Canning CG. The effect of directing attention during walking under dual-task conditions in Parkinson's disease. Parkinsonism Relat Disord. 2005 Mar;11(2):95–9.

10. Barthel C, Nonnekes J, van Helvert M, Haan R, Janssen A, Delval A, et al. The laser shoes: a new ambulatory device to alleviate freezing of gait in Parkinson disease. Neurology. 2018 Jan;90(2):e164–71.

11. Nieuwboer A. Cueing for freezing of gait in patients with Parkinson's disease: a rehabilitation perspective. Mov Disord. 2008;23 Suppl 2:S475–81.

12. Ginis P, Nackaerts E, Nieuwboer A, Heremans E. Cueing for people with Parkinson's disease with freezing of gait: a narrative review of the state-of-the-art and novel perspectives. Ann Phys Rehabil Med. 2018 Nov;61(6):407–13.

13. Moreau C, Defebvre L, Bleuse S, Blatt JL, Duhamel A, Bloem BR, et al. Externally provoked freezing of gait in open runways in advanced Parkinson's disease results from motor and mental collapse. J Neural Transm. 2008 Oct;115(10):1431–6.

14. Mazilu S, Blanke U, Hardegger M, Tröster G, Gazit E, Hausdorff JM. GaitAssist: a daily-life support and training system for Parkinson's disease patients with freezing of gait. Proceedings of the Annual ACM conference on Human factors in computing systems; 2014 Apr 26 - May 1; Toronto, Canada. ACM Press; 2014. p. 2531–40.

15. Jovanov E, Wang E, Verhagen L, Fredrickson M, Fratangelo R. deFOG - a real time system for detection and unfreezing of gait of Parkinson's patients. Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine; 2009 Sept 3-6; Minneapolis, USA. IEEE; 2009. p. 5151–4.

16. Bachlin M, Plotnik M, Roggen D, Maidan I, Hausdorff JM, Giladi N, et al. Wearable assistant for Parkinson's disease patients with the freezing of gait symptom. Trans Inf Technol Biomed. 2010 Mar;14(2):436–46.

17. Palmerini L, Rocchi L, Mazilu S, Gazit E, Hausdorff JM, Chiari L. Identification of characteristic motor patterns preceding freezing of gait in Parkinson's disease using wearable sensors. Front Neurol. 2017 Aug 14;8(1):394.

18. Mazilu S, Calatroni A, Gazit E, Roggen D, Hausdorff JM, Tröster G. Feature learning for detection and prediction of freezing of gait in Parkinson's disease. In: Perner P, editor. Machine Learning and Data Mining in Pattern Recognition. Berlin, Heidelberg: Springer; 2013. p. 144–58.

19. Zhang Y, Yan W, Yao Y, Ahmed JB, Tan Y, Gu D. Prediction of freezing of gait in patients with Parkinson's disease by identifying impaired gait patterns. IEEE Trans Neural Syst Rehabil Eng. 2020 Mar;28(3):591–600.

20. Torvi VG, Bhattacharya A, Chakraborty S. Deep domain adaptation to predict freezing of gait in patients with Parkinson's disease. Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA); 2018 Dec 17-20; Orlando, USA. IEEE; 2018. p. 1001–6.

21. Zia J, Tadayon A, McDaniel T, Panchanathan S. Utilizing neural networks to predict freezing of gait in Parkinson's patients. Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility; 2016 Oct 23-26; Reno, USA. Association for Computing Machinery; 2016. p. 333–4.

22. Demrozi F, Bacchin R, Tamburin S, Cristani M, Pravadelli G. Toward a wearable system for predicting freezing of gait in people affected by Parkinson's disease. IEEE J Biomed Health Inform. 2020 Sept;24(9):2444–51.

23. Borzì L, Mazzetta I, Zampogna A, Suppa A, Olmo G, Irrera F. Prediction of freezing of gait in Parkinson's disease using wearables and machine learning. Sensors. 2021 Jan 17;21(2):614.

24. Naghavi N, Wade E. Prediction of freezing of gait in Parkinson's disease using statistical inference and lower-limb acceleration data. IEEE Trans Neural Syst Rehabil Eng. 2019 May;27(5):947–55.

25. Naghavi N, Miller A, Wade E. Towards real-time prediction of freezing of gait in patients with Parkinson's disease: addressing the class imbalance problem. Sensors. 2019 Sept;19(18):3898.

26. Handojoseno AMA, Naik GR, Gilat M, Shine JM, Nguyen TN, Ly QT, et al. Prediction of freezing of gait in patients with Parkinson's disease using EEG signals. Stud Health Technol Inform. 2018;246:124–31.

27. Handojoseno AMA, Shine JM, Nguyen TN, Tran Y, Lewis SJG, Nguyen HT. Analysis and prediction of the freezing of gait using EEG brain dynamics. IEEE Trans Neural Syst Rehabil Eng. 2015 Sept;23(5):887–96.

28. Mazilu S, Calatroni A, Gazit E, Mirelman A, Hausdorff JM, Troster G. Prediction of freezing of gait in Parkinson's from physiological wearables: an exploratory study. IEEE J Biomed Health Inform. 2015 Nov;19(6):1843–54.

29. Nieuwboer A, Dom R, De Weerdt W, Desloovere K, Fieuws S, Broens-Kaucsik E. Abnormalities of the spatiotemporal characteristics of gait at the onset of freezing in Parkinson's disease. Mov Disord. 2001 Nov;16(6):1066–75.

30. Chee R, Murphy A, Danoudis M, Georgiou-Karistianis N, Iansek R. Gait freezing in Parkinson's disease and the stride length sequence effect interaction. Brain. 2009 Aug;132(Pt 8):2151–60.

31. Ahn D, Chung H, Lee H-W, Kang K, Ko P-W, Kim NS, et al. Smart gait-aid glasses for Parkinson's disease patients. IEEE Trans Biomed Eng. 2017 Oct;64(10):2394–402.

32. Coste CA, Sijobert B, Pissard-Gibollet R, Pasquier M, Espiau B, Geny C. Detection of freezing of gait in Parkinson disease: preliminary results. Sensors. 2014 Apr 15;14(4):6819–27.

33. Spildooren J, Vercruysse S, Meyns P, Vandenbossche J, Heremans E, Desloovere K, et al. Turning and unilateral cueing in Parkinson's disease patients with and without freezing of gait. Neuroscience. 2012 Apr 5;207:298–306.

34. Hausdorff JM, Balash Y, Giladi N. Time series analysis of leg movements during freezing of gait in Parkinson's disease: akinesia, rhyme or reason? Phys Stat Mech Its Appl. 2003 Apr 15;321(3–4):565–70.

35. Plotnik M, Giladi N, Hausdorff JM. Bilateral coordination of walking and freezing of gait in Parkinson's disease. Eur J Neurosci. 2008 Apr;27(8):1999–2006.

36. Hausdorff JM, Cudkowicz ME, Firtion R, Wei JY, Goldberger AL. Gait variability and basal ganglia disorders: stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease. Mov Disord. 1998 May;13(3):428–37.

37. Maculewicz J, Kofoed LB, Serafin S. A technological review of the instrumented footwear for rehabilitation with a focus on Parkinson's disease patients. Front Neurol. 2016 Jan 20;7(1).

38. Bae J, Kong K, Byl N, Tomizuka M. A mobile gait monitoring system for abnormal gait diagnosis and rehabilitation: a pilot study for Parkinson disease patients. J Biomech Eng. 2011 Apr;133(4):041005.

39. Plotnik M, Giladi N, Balash Y, Peretz C, Hausdorff JM. Is freezing of gait in Parkinson's disease related to asymmetric motor function? Ann Neurol. 2005 May;57(5):656–63.

40. Jenkins ME, Almeida QJ, Spaulding SJ, van Oostveen RB, Holmes JD, Johnson AM, et al. Plantar cutaneous sensory stimulation improves single-limb support time, and EMG activation patterns among individuals with Parkinson's disease. Parkinsonism Relat Disord. 2009 Nov;15(9):697–702.

41. Kimmeskamp S, Hennig EM. Heel to toe motion characteristics in Parkinson patients during free walking. Clin Biomech. 2001 Nov;16(9):806–12.

42. Shah J, Pillai L, Williams DK, Doerhoff SM, Larson-Prior L, Garcia-Rill E, et al. Increased foot strike variability in Parkinson's disease patients with freezing of gait. Parkinsonism Relat Disord. 2018 Aug;53:58–63.

43. Marcante A, Di Marco R, Gentile G, Pellicano C, Assogna F, Pontieri FE, et al. Foot pressure wearable sensors for freezing of gait detection in Parkinson's disease. Sensors. 2021;21(1):128.

44. Pardoel S, Shalin G, Nantel J, Lemaire ED, Kofman J. Early detection of freezing of gait during walking using inertial measurement unit and plantar pressure distribution data. Sensors. 2021;21(6):2246.

45. Pardoel S, Shalin G, Nantel J, Lemaire ED, Kofman J. Selection of plantar-pressure and ankle-acceleration features for freezing of gait detection in Parkinson's disease using minimum-redundancy maximum-relevance. Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society; 2020 July 20-24; Montreal, Canada. IEEE; 2020. p. 4034–7.

46. Howcroft J, Lemaire ED, Kofman J. Wearable-sensor-based classification models of faller status in older adults. PLOS ONE. 2016 Apr 7;11(4):e0153240.

47. Howcroft J, Kofman J, Lemaire ED. Feature selection for elderly faller classification based on wearable sensors. J Neuroeng Rehabil. 2017 May 30;14(1):47.

48. Nantel J, Bronte-Stewart H. The effect of medication and the role of postural instability in different components of freezing of gait (FOG). Parkinsonism Relat Disord. 2014 Apr;20(4):447–51.

49. Rahimi F, Xian SY, Delrobaei M, Jog M. Characteristics of gait freezing: possibilities for rehabilitation. Proceedings of the 6th International IEEE/EMBS Conference on Neural Engineering; 2013 Nov 6-8; San Diego, USA. IEEE; 2013. p. 1594–7.

50. Arami A, Poulakakis-Daktylidis A, Tai YF, Burdet E. Prediction of gait freezing in Parkinsonian patients: a binary classification augmented with time series prediction. IEEE Trans Neural Syst Rehabil Eng. 2019 Sept;27(9):1909–19.

51. Pardoel S, Kofman J, Nantel J, Lemaire ED. Wearable-sensor-based detection and prediction of freezing of gait in Parkinson's disease: a review. Sensors. 2019 Nov 24;19(23):5141.

52. Schaafsma JD, Balash Y, Gurevich T, Bartels AL, Hausdorff JM, Giladi N. Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson's disease. Eur J Neurol. 2003 Jul;10(4):391–8.

53. Zach H, Janssen AM, Snijders AH, Delval A, Ferraye MU, Auff E, et al. Identifying freezing of gait in Parkinson's disease during freezing provoking tasks using waist-mounted accelerometry. Parkinsonism Relat Disord. 2015 Nov;21(11):1362–6.

54. Mazilu S, Blanke U, Troster G. Gait, wrist, and sensors: detecting freezing of gait in Parkinson's disease from wrist movement. Proceedings of the International Conference on Pervasive Computing and Communication Workshops; 2015 Mar 23-27; St. Louis, USA. IEEE; 2015. p. 579–84.

55. Mazilu S, Blanke U, Calatroni A, Gazit E, Hausdorff JM, Tröster G. The role of wrist-mounted inertial sensors in detecting gait freeze episodes in Parkinson's disease. Pervasive Mob Comput. 2016 Dec;33:1–16.

56. Djuric-Jovicic MD, Jovicic NS, Radovanovic SM, Stankovic ID, Popovic MB, Kostic VS. Automatic identification and classification of freezing of gait episodes in Parkinson's disease patients. IEEE Trans Neural Syst Rehabil Eng. 2014 May;22(3):685–94.

57. Reches T, Dagan M, Herman T, Gazit E, Gouskova NA, Giladi N, et al. Using wearable sensors and machine learning to automatically detect freezing of gait during a FOG-provoking test. Sensors. 2020 Aug 10;20(16):4474.

58. Delval A, Snijders AH, Weerdesteyn V, Duysens JE, Defebvre L, Giladi N, et al. Objective detection of subtle freezing of gait episodes in Parkinson's disease. Mov Disord. 2010 Aug 15;25(11):1684–93.

59. Suppa A, Kita A, Leodori G, Zampogna A, Nicolini E, Lorenzi P, et al. L-DOPA and freezing of gait in Parkinson's disease: objective assessment through a wearable wireless system. Front Neurol. 2017 Aug 14;8:406.

60. Handojoseno AMA, Gilat M, Ly QT, Chamtie H, Shine JM, Nguyen TN, et al. An EEG study of turning freeze in Parkinson's disease patients: the alteration of brain dynamic on the motor and visual cortex. Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2015 Aug 25-29; Milan, Italy. IEEE; 2015. p. 6618–21.

61. Mancini M, Shah VV, Stuart S, Curtze C, Horak FB, Safarpour D, et al. Measuring freezing of gait during daily-life: an open-source, wearable sensors approach. J Neuroeng Rehabil. 2021 Jan 4;18(1):1.

62. Ferster ML, Mazilu S, Tröster G. Gait parameters change prior to freezing in Parkinson's disease: a data-driven study with wearable inertial units. Proceedings of the 10th EAI International Conference on Body Area Networks; 2015 Sept 28-30; Sydney, Australia. EAI; 2015. p. 159–66.

63. Shalin G. Prediction and detection of freezing of gait in Parkinson's disease using plantar pressure data [MASc thesis]. University of Waterloo; 2021.

64. Popovic MB, Djuric-Jovicic M, Radovanovic S, Petrovic I, Kostic V. A simple method to assess freezing of gait in Parkinson's disease patients. Braz J Med Biol Res. 2010 Sept;43(9):883–9.

65. Wong SL, Gilmour H, Ramage-Morin PL. Parkinson's disease: prevalence, diagnosis and impact. Statistics Canada Health Reports. 2014.

66. Canadian Institute for Health Information. The burden of neurological diseases, disorders and injuries in Canada. Ottawa: CIHI. 2007.

67. Parkinson's Society Canada. Parkinson's disease: social and economic impact. Health Canada. 2003.

68. Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. The Lancet. 2009 Oct 3;374(9696):1196–208.

69. Sofuwa O, Nieuwboer A, Desloovere K, Willems A-M, Chavret F, Jonkers I. Quantitative gait analysis in Parkinson's disease: comparison with a healthy control group. Arch Phys Med Rehabil. 2005 May;86(5):1007–13.

70. Baltadjieva R, Giladi N, Gruendlinger L, Peretz C, Hausdorff JM. Marked alterations in the gait timing and rhythmicity of patients with *de novo* Parkinson's disease. Eur J Neurosci. 2006 Sep;24(6):1815–20.

71. Hausdorff JM. Gait dynamics in Parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. Chaos. 2009 Jun;19(2):026113.

72. Macht M, Kaussner Y, Möller JC, Stiasny-Kolster K, Eggert KM, Krüger H-P, et al. Predictors of freezing in Parkinson's disease: a survey of 6,620 patients. Mov Disord. 2007 May 15;22(7):953–6.

73. Nutt JG, Bloem BR, Giladi N, Hallett M, Horak FB, Nieuwboer A. Freezing of gait: moving forward on a mysterious clinical phenomenon. Lancet Neurol. 2011 Aug;10(8):734–44.

74. Snijders AH, Nijkrake MJ, Bakker M, Munneke M, Wind C, Bloem BR. Clinimetrics of freezing of gait. Mov Disord. 2008;23 Suppl 2:S468-74.

75. Ge H-L, Chen X-Y, Lin Y-X, Ge T-J, Yu L-H, Lin Z-Y, et al. The prevalence of freezing of gait in Parkinson's disease and in patients with different disease durations and severities. Chin Neurosurg J. 2020 May 14;6:17.

76. Okuma Y, Silva de Lima AL, Fukae J, Bloem BR, Snijders AH. A prospective study of falls in relation to freezing of gait and response fluctuations in Parkinson's disease. Parkinsonism Relat Disord. 2018 Jan;46:30–5.

77. Adkin AL, Frank JS, Jog MS. Fear of falling and postural control in Parkinson's disease. Mov Disord. 2003 May;18(5):496–502.

78. Bloem BR, Grimbergen YAM, Cramer M, Willemsen M, Zwinderman AH. Prospective assessment of falls in Parkinson's disease. J Neurol. 2001 Nov;248(11):950–8.

79. Heremans E, Nieuwboer A, Vercruysse S. Freezing of gait in Parkinson's disease: where are we now? Curr Neurol Neurosci Rep. 2013 Jun;13(6):350.

80. Yahr MD, Duvoisin RC, Schear MJ, Barrett RE, Hoehn MM. Treatment of parkinsonism with levodopa. Arch Neurol. 1969 Oct;21(4):343–54.

81. Hely MA, Morris JG, Reid WGJ, Trafficante R. Sydney multicenter study of Parkinson's disease: non-L-dopa–responsive problems dominate at 15 years. Mov Disord. 2005 Feb;20(2):190–9.

82. Ahlskog JE, Muenter MD. Frequency of levodopa-related dyskinesias and motor fluctuations as estimated from the cumulative literature. Mov Disord. 2001 May;16(3):448–58.

83. Chiken S, Nambu A. Mechanism of deep brain stimulation: inhibition, excitation, or disruption? Neuroscientist. 2016 Jun;22(3):313–22.

84. The Deep-Brain Stimulation for Parkinson's Disease Study Group; Obeso JA, Olanow CW, Rodriguez-Oroz MC, Krack P, Kumar R, Lang AE. Deep-brain stimulation of the subthalamic nucleus or the pars interna of the globus pallidus in Parkinson's disease. N Engl J Med. 2001 Sep 27;345(13):956–63.

85. Kringelbach ML, Jenkinson N, Owen SLF, Aziz TZ. Translational principles of deep brain stimulation. Nat Rev Neurosci. 2007 Aug;8(8):623–35.

86. Roper JA, Kang N, Ben J, Cauraugh JH, Okun MS, Hass CJ. Deep brain stimulation improves gait velocity in Parkinson's disease: a systematic review and meta-analysis. J Neurol. 2016 Jun;263(6):1195–203.

87. Barbe MT, Tonder L, Krack P, Debû B, Schüpbach M, Paschen S, et al. Deep brain stimulation for freezing of gait in Parkinson's disease with early motor complications. Mov Disord. 2020 Jan;35(1):82–90.

88. Nonnekes J, Timmer MHM, de Vries NM, Rascol O, Helmich RC, Bloem BR. Unmasking levodopa resistance in Parkinson's disease. Mov Disord. 2016 Nov;31(11):1602–9.

89. Giladi N, Treves TA, Simon ES, Shabtai H, Orlov Y, Kandinov B, et al. Freezing of gait in patients with advanced Parkinson's disease. J Neural Transm. 2001;108(1):53–61.

90. Van de Weijer SCF, Hommel ALAJ, Bloem BR, Nonnekes J, De Vries NM. Promising non-pharmacological therapies in PD: targeting late stage disease and the role of computer based cognitive training. Parkinsonism Relat Disord. 2018 Jan;46 Suppl 1:S42–6.

91. Tzallas AT, Tsipouras MG, Rigas G, Tsalikakis DG, Karvounis EC, Chondrogiorgi M, et al. PERFORM: a system for monitoring, assessment and management of patients with Parkinson's disease. Sensors. 2014 Nov 11;14(11):21329–57.

92. Ossig C, Antonini A, Buhmann C, Classen J, Csoti I, Falkenburger B, et al. Wearable sensor-based objective assessment of motor symptoms in Parkinson's disease. J Neural Transm. 2016;123:57–64.

93. Mazilu S, Blanke U, Hardegger M, Troster G, Gazit E, Dorfman M, et al. GaitAssist: a wearable assistant for gait training and rehabilitation in Parkinson's disease. Proceedings of the IEEE International Conference on Pervasive Computing and Communication Workshops; 2014 Mar 24-28; Budapest, Hungary. IEEE; 2014. p. 135–7.

94. Moore ST, MacDougall HG, Ondo WG. Ambulatory monitoring of freezing of gait in Parkinson's disease. J Neurosci Methods. 2008 Jan 30;167(2):340–8.

95. Zhao Y, Tonn K, Niazmand K, Fietzek UM, D'Angelo LT, Ceballos-Baumann A, et al. Online FOG identification in Parkinson's disease with a time-frequency combined algorithm. Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics; 2012 Jan 5-7; Hong Kong, China. IEEE; 2012. p. 192–5.

96. Moore ST, Yungher DA, Morris TR, Dilda V, MacDougall HG, Shine JM, et al. Autonomous identification of freezing of gait in Parkinson's disease from lower-body segmental accelerometry. J Neuroeng Rehabil. 2013 Feb 13;10:19.

97. Tay A, Yen SC, Lee PY, Wang CY, Neo A, Phan SW, et al. Freezing of gait (FoG) detection for Parkinson disease. Proceedings of the 10th Asian Control Conference; 2015 May 31-Jun 3; Kota Kinabalu, Malaysia. IEEE; 2015. p. 1–6.

98. Rezvanian S, Lockhart TE. Towards real-time detection of freezing of gait using wavelet transform on wireless accelerometer data. Sensors. 2016 Apr 2;16(4):475.

99. Kita A, Lorenzi P, Rao R, Irrera F. Reliable and robust detection of freezing of gait episodes with wearable electronic devices. Sensors. 2017;17(6):1899–908.

100. Pepa L, Capecci M, Ciabattoni L, Spalazzi L, Ceravolo MG. An unobtrusive expert system to detect freezing of gait during daily living in people with Parkinson's disease. Proceedings of the 2nd International Multidisciplinary Conference on Computer and Energy Science, SpliTech; 2017 July 12-14; Split, Croatia. IEEE; 2017. p. 1–5.

101. Punin C, Barzallo B, Huerta M, Bermeo A, Bravo M, Llumiguano C. Wireless devices to restart walking during an episode of FOG on patients with Parkinson's disease. Proceedings of the Second Ecuador Technical Chapters Meeting; 2017 Oct 16-20; Salinas, Ecuador. IEEE; 2017. p. 1–6.

102. Punin C, Barzallo B, Clotet R, Bermeo A, Bravo M, Bermeo JP, et al. A non-invasive medical device for Parkinson's patients with episodes of freezing of gait. Sensors. 2019 Feb;19(3):737.

103. Barzallo B, Punin C, Llumiguano C, Huerta M. Wireless assistance system during episodes of freezing of gait by means superficial electrical stimulation. Proceedings of the World Congress on Medical Physics and Biomedical Engineering; 2018 June 3-8; Prague, Czech Republic. Springer; 2019. p. 865–70.

104. Capecci M, Pepa L, Verdini F, Ceravolo MG. A smartphone-based architecture to detect and quantify freezing of gait in Parkinson's disease. Gait Posture. 2016 Oct;50:28–33.

105. Bächlin M, Hausdorff JM, Roggen D, Giladi N, Plotnik M, Tröster G. Online detection of freezing of gait in Parkinson's disease patients: a performance characterization. Proceedings of the Fourth International Conference on Body Area Networks; 2009 Apr 1-3; Los Angeles, USA. ICST; 2009.

106. Wang K, Li B, Gu D, Dai K, Zhou L. A smartphone based system for freezing of gait monitoring for Parkinson's disease patients. Proceedings of the 9th IEEE International Conference on Communication Software and Networks; 2017 May 6-8; Guangzhou, China. IEEE; 2017. p. 1529–33.

107. Pham TT, Nguyen DN, Dutkiewicz E, McEwan AL, Leong PHW. Wearable healthcare systems: a single channel accelerometer based anomaly detector for studies of gait freezing in Parkinson's disease. Proceedings of the IEEE International Conference on Communications; 2017 May 21-25; Paris, France. IEEE; 2017. p. 1–5.

108. Tahafchi P, Molina R, Roper JA, Sowalsky K, Hass CJ, Gunduz A, et al. Freezing-of-gait detection using temporal, spatial, and physiological features with a support-vector-machine classifier. Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2017 July 11-15; Jeju, South Korea. IEEE; 2017. p. 2867–70.

109. Assam R, Seidl T. Prediction of freezing of gait from Parkinson's disease movement time series using conditional random fields. Proceedings of the 3rd ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health; 2014 Nov 4-7; New York, USA. 2014. p. 11–20.

110. El-Attar A, Ashour AS, Dey N, El-Kader HA, Abd El-Naby MM, Shi F. Hybrid DWT-FFT features for detecting freezing of gait in Parkinson's disease. In: Jain L.C. Zhao X., Balas V.E. Shi F., editors. Frontiers in Artificial Intelligence and Applications, IOS Press, 2019. p. 117–26.

111. Ashour AS, El-Attar A, Dey N, El-Naby MMA, El-Kader HA. Patient-dependent freezing of gait detection using signals from multi-accelerometer sensors in Parkinson's disease. Proceedings of the Cairo International Biomedical Engineering Conference; 2018 Dec 20-22; Cairo, Egypt. IEEE; 2019. p. 171–4.

112. Nivya Venu A, Lisa C. Prediction of freezing of gait (FOG) episodes in Parkinson's disease patients by gait analysis. Int J Innov Res Sci Eng Technol. 2016;5(7):13679–84.

113. Cole BT, Roy SH, Nawab SH. Detecting freezing-of-gait during unscripted and unconstrained activity. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2011 Aug 30-Sept 3; Boston, USA. IEEE; 2011. p. 5649–52.

114. Ly QT, Handojoseno AMA, Gilat M, Nguyen N, Chai R, Tran Y, et al. Identifying montages that best detect the electroencephalogram power spectrum alteration during freezing of gait in Parkinson's disease patients. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2016 Aug 16-20; Orlando USA. IEEE; 2016. p. 6094–7.

115. Ly QT, Handojoseno AMA, Gilat M, Chai R, Martens KAE, Georgiades M, et al. Detection of turning freeze in Parkinson's disease based on S-transform decomposition of EEG signals. Proceedings of the Annual International Conference of the Engineering in Medicine and Biology Society; 2017 July 11-15; Jeju, South Korea. IEEE; 2017. p. 3044–7.

116. Mikos V, Heng C-H, Tay A, Yen S-C, Chia NSY, Koh KML, et al. A neural network accelerator with integrated feature extraction processor for a freezing of gait detection system. Proceedings of the Asian Solid-State Circuits Conference; 2018 Nov 5-7; Tainan, Taiwan. IEEE; 2018. p. 59–62.

117. Djuric-Jovicic M, Jovicic NS, Milovanovic I, Radovanovic S, Kresojevic N, Popovic MB. Classification of walking patterns in Parkinson's disease patients based on inertial sensor data. Proceedings of the 10th Symposium on Neural Network Applications in Electrical Engineering; 2010 Sept 23-25; Belgrade, Serbia. IEEE; 2010. p. 3–6.

118. Saad A, Zaarour I, Guerin F, Bejjani P, Ayache M, Lefebvre D. Detection of freezing of gait for Parkinson's disease patients with multi-sensor device and Gaussian neural networks. Int J Mach Learn Cybern. 2017 Dec;8:941–54.

119. Oung QW, Basah SN, Muthusamy H, Vijean V, Lee H, Khairunizam W, et al. Objective evaluation of freezing of gait in patients with Parkinson's disease through machine learning approaches. Proceedings of the International Conference on Computational Approach in Smart Systems Design and Applications; 2018 Aug 15-17; Kuching, Malaysia. IEEE; 2018. p. 1-7.

120. Mikos V, Heng C-H, Tay A, Chia NSY, Koh KML, Tan DML, et al. Real-time patient adaptivity for freezing of gait classification through semi-supervised neural networks. Proceedings of the 16th IEEE International Conference on Machine Learning and Applications; 2017 Dec 18-21; Cancun, Mexico. IEEE; 2017. p. 871–6.

121. Camps J, Samà A, Martín M, Rodríguez-Martín D, Pérez-López C, Moreno Arostegui JMM, et al. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. Knowl-Based Syst. 2018 Jan 1;139:119–31.

122. Sigcha L, Costa N, Pavón I, Costa S, Arezes P, López JM, et al. Deep learning approaches for detecting freezing of gait in Parkinson's disease patients through on-body acceleration sensors. Sensors. 2020 Mar 29;20(7):1895.

123. Li B, Yao Z, Wang J, Wang S, Yang X, Sun Y. Improved deep learning technique to detect freezing of gait in Parkinson's disease based on wearable sensors. Electronics. 2020 Nov 14;9(11):1919.

124. Shi B, Yen SC, Tay A, Tan DML, Chia NSY, Au WL. Convolutional neural network for freezing of gait detection leveraging the continuous wavelet transform on lower extremities wearable sensors data. Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society; 2020 July 20-24; Montreal, Canada. IEEE; 2020. p. 5410–5.

125. Kim H, Lee HJ, Lee W, Kwon S, Kim SK, Jeon HS, et al. Unconstrained detection of freezing of Gait in Parkinson's disease patients using smartphone. Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2015 Aug 25-29; Milan, Italy. IEEE; 2015. p. 3751–4.

126. Mazilu S, Hardegger M, Zhu Z, Roggen D, Troester G, Plotnik M, et al. Online detection of freezing of gait with smartphones and machine learning techniques. Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare; 2012 May 21-24; San Diego, USA. IEEE; 2012. p. 123–30.

127. Tsipouras MG, Tzallas AT, Tripoliti E, Rigas G, Bougia P, Fotiadis DI, et al. On assessing motor disorders in Parkinson's disease. Proceedings of the Second International ICST Conference, MobiHealth; 2010 Oct 18-20; Ayia Napa, Cyprus. Springer; 2011. p. 35–8.

128. Tripoliti EE, Tzallas AT, Tsipouras MG, Rigas G, Bougia P, Leontiou M, et al. Automatic detection of freezing of gait events in patients with Parkinson's disease. Comput Methods Programs Biomed. 2013 Apr;110(1):12–26.

129. Rodríguez-Martín D, Samà A, Pérez-López C, Català A, Moreno Arostegui JM, Cabestany J, et al. Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer. PLOS ONE. 2017 Feb 15;12(2):e0171764.

130. Samà A, Rodríguez-Martín D, Pérez-López C, Català A, Alcaine S, Mestre B, et al. Determining the optimal features in freezing of gait detection through a single waist accelerometer in home environments. Pattern Recognit Lett. 2018 Apr;105:135–43.

131. Rodríguez-Martín D, Pérez-López C, Samà A, Català A, Moreno Arostegui J, Cabestany J, et al. A waist-worn inertial measurement unit for long-term monitoring of Parkinson's disease patients. Sensors. 2017 Apr 11;17(4):827–827.

132. Ahlrichs C, Samà A, Lawo M, Cabestany J, Rodríguez-Martín D, Pérez-López C, et al. Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients. Med Biol Eng Comput. 2016 Jan;54(1):223–33.

133. Pham TT, Nguyen DN, Dutkiewicz E, McEwan AL, Leong PHW. An anomaly detection technique in wearable wireless monitoring systems for studies of gait freezing in Parkinson's disease. Proceedings of the International Conference on Information Networking; 2017 Jan 11-13; Da Nang, Vietnam. IEEE; 2017. p. 41–5.

134. Li B, Zhang Y, Tang L, Gao C, Gu D. Automatic detection system for freezing of gait in Parkinson's disease based on the clustering algorithm. Proceedings of the 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference; 2018 May 25-27; Xi'an, China. IEEE; 2018. p. 1640–4.

135. Rodriguez-Martín D, Sama A, Pérez-López C, Catala A, Mestre B, Alcaine S, et al. Comparison of features, window sizes and classifiers in detecting freezing of gait in patients with Parkinson's disease through a waist-worn accelerometer. Proceedings of the 19th International Conference of the Catalan Association for Artificial Intelligence; 2016 Oct 19-21; Barcelona, Spain. IOS Press; 2016. p. 127–36.

136. Wu X, Kumar V, Ross QJ, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. Knowl Inf Syst. 2008;14(1):1–37.

137. Bishop CM. Pattern Recognition and Machine Learning. 1st ed. Springer; 2006.

138. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. Ann Stat. 2000 Apr;28(2):337–407.

139. Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: improving classification performance when training data is skewed. Proceedings of the 19th International Conference on Pattern Recognition; 2008 Dec 8-11; Tampa USA. IEEE; 2008. p. 1–4.

140. Freund Y. A more robust boosting algorithm. arXiv preprint arXiv:0905.2138; 2009.

141. Egmont-Petersen M, Ridder D De, Handels H. Image processing with neural networks - a review. Pattern Recognit. 2002;35(10):2279–301.

142. Mohammadian Rad N, Van Laarhoven T, Furlanello C, Marchiori E. Novelty detection using deep normative modeling for IMU-based abnormal movement monitoring in Parkinson's disease and autism spectrum disorders. Sensors. 2018;18(10):3533.

143. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018 Aug;9(4):611–29.

144. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural networks: a review. J Med Syst. 2018 Nov 8;42(11):226.

145. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.

146. Sak H, Senior A, Beaufays F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv:1402.1128; 2014.

147. Fahn S. The freezing phenomenon in parkinsonism. Adv Neurol. 1995;67:53–63.

148. Silva de Lima AL, Evers LJW, Hahn T, Bataille L, Hamilton JL, Little MA, et al. Freezing of gait and fall detection in Parkinson's disease using wearable sensors: a systematic review. J Neurol. 2017 Aug;264(8):1642–54.

149. Maidan I, Plotnik M, Mirelman A, Weiss A, Giladi N, Hausdorff JM. Heart rate changes during freezing of gait in patients with Parkinson's disease. Mov Disord. 2010 Oct 30;25(14):2346–54.

150. Howcroft J, Kofman J, Lemaire ED. Prospective fall-risk prediction models for older adults based on wearable sensors. IEEE Trans Neural Syst Rehabil Eng. 2017 Oct;25(10):1812–20.

151. Caby B, Kieffer S, de Saint Hubert M, Cremer G, Macq B. Feature extraction and selection for objective gait analysis and fall risk assessment by accelerometry. Biomed Eng Online. 2011 Jan 9;10:1.

152. Shany T, Redmond SJ, Narayanan MR, Lovell NH. Sensors-based wearable systems for monitoring of human movement and falls. IEEE Sens J. 2012;12(3):658–70.

153. Sijobert B, Denys J, Coste CA, Geny C. IMU based detection of freezing of gait and festination in Parkinson's disease. Proceedings of the 19th International Functional Electrical Stimulation Society Annual Conference; 2014 Sept 17-19; Kuala Lumpur, Malaysia. IEEE; 2014. p. 1-3.

154. Dvorani A, Jochner MCE, Seel T, Salchow-Hömmen C, Meyer-Ohle J, Wiesener C, et al. Inertial sensor based detection of freezing of gait for on-demand cueing in Parkinson's disease. IFAC-Pap. 2020;53(2):16004–9.

155. Lorenzi P, Rao R, Romano G, Kita A, Irrera F. Mobile devices for the real time detection of specific human motion disorders. IEEE Sens J. 2016 Dec;16(23):8220–7.

156. Lorenzi P, Rao R, Romano G, Kita A, Serpa M, Filesi F, et al. Smart sensing systems for the detection of human motion disorders. Procedia Eng. 2015 Jan;120:324–7.

157. Lorenzi P, Rao R, Romano G, Kita A, Serpa M, Filesi F, et al. Smart sensors for the recognition of specific human motion disorders in Parkinson's disease. Proceedings of the 6th IEEE International Workshop on Advances in Sensors and Interfaces; 2015 June 18-19; Gallipoli, Italy. IEEE; 2015. p. 131–6.

158. Kita A, Lorenzi P, Romano G, Rao R, Parisi R, Suppa A, et al. Smart sensing system for the detection of specific human motion symptoms of the Parkinson's disease. Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies; 2016; Rome, Italy. SciTePress; 2016. p. 152–9.

159. Mazzetta I, Zampogna A, Suppa A, Gumiero A, Pessione M, Irrera F. Wearable sensors system for an improved analysis of freezing of gait in Parkinson's disease using electromyography and inertial signals. Sensors. 2019 Feb 23;19(4):948.

160. Pham TT, Moore ST, Lewis SJG, Nguyen DN, Dutkiewicz E, Fuglevand AJ, et al. Freezing of gait detection in Parkinson's disease: a subject-independent detector using anomaly scores. IEEE Trans Biomed Eng. 2017 Nov;64(11):2719–28.

161. Kwon Y, Park SH, Kim JW, Ho Y, Jeon HM, Bang MJ, et al. A practical method for the detection of freezing of gait in patients with Parkinson's disease. Clin Interv Aging. 2014 Oct 8;9:1709–19.

162. Bachlin M, Plotnik M, Roggen D, Giladi N, Hausdorff JM, Troster G. A wearable system to assist walking of Parkinson's disease patients. Methods Inf Med. 2010;49(1):88–95.

163. Prateek GV, Skog I, McNeely ME, Duncan RP, Earhart GM, Nehorai A, et al. Modeling, detecting, and tracking freezing of gait in Parkinson disease using inertial sensors. IEEE Trans Biomed Eng. 2018 Oct;65(10):2152–61.

164. Pepa L, Ciabattoni L, Verdini F, Capecci M, Ceravolo MG. Smartphone based fuzzy logic freezing of gait detection in Parkinson's disease. Proceedings of the IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications; 2014 Sept 10-12; Senigallia, Italy. IEEE; 2014. p. 1-6.

165. Niazmand K, Tonn K, Zhao Y, Fietzek UM, Schroeteler F, Ziegler K, et al. Freezing of gait detection in Parkinson's disease using accelerometer based smart clothes. Proceedings of the IEEE Biomedical Circuits and Systems Conference; 2011 Nov 10-12; San Diego, USA. IEEE; 2011. p. 201–4.

166. Bächlin M, Plotnik M, Roggen D, Inbar N, Giladi N, Hausdorff JM, et al. Parkinson's disease patients' perspective on context aware wearable technology for auditive assistance. Proceedings of the 3rd International Conference on Pervasive Computing Technologies for Healthcare; 2009 Apr 1-3; London, UK. IEEE; 2009. p. 1-8.

167. Bachlin M, Roggen D, Troster G, Plotnik M, Inbar N, Meidan I, et al. Potentials of enhanced context awareness in wearable assistants for Parkinson's disease patients with the freezing of gait syndrome. Proceedings of the International Symposium on Wearable Computers; 2009 Sept 4-7; Linz, Austria. IEEE; 2009. p. 123–30.

168. Pepa L, Capecci M, Ciabattoni L, Spalazzi L, Ceravolo MG. A real-time fuzzy logic algorithm for freezing of gait management on a smartphone. Proceedings of the 7th International Conference on Consumer Electronics; 2017 Sept 3-6; Berlin, Germany. IEEE; 2017. p. 130–1.

169. Bächlin M, Plotnik M, Roggen D, Giladi N, Hausdorff JM, Tröster G. A wearable system to assist walking of Parkinson's disease patients benefits and challenges of context-triggered acoustic cueing. Methods Inf Med. 2010;49(1):88–95.

170. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Jason H. Relief-based feature selection: introduction and review. J Biomed Inform. 2018;85:189–203.

171. Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. Bioinformatics. 2007 Oct 1;23(19):2507–17.

172. Mazilu S, Blanke U, Roggen D, Tröster G, Gazit E, Hausdorff JM. Engineers meet clinicians: augmenting Parkinson's disease patients to gather information for gait rehabilitation. Proceedings of the 4th Augmented Human International Conference; New York, USA. ACM Press; 2013. p. 124–7.

173. Lorenzi P, Rao R, Romano G, Kita A, Irrera F. Mobile devices for the real-time detection of specific human motion disorders. IEEE Sens J. 2016;16(23):8220–7.

174. Hausdorff JM, Schaafsma JD, Balash Y, Bartels AL, Gurevich T, Giladi N. Impaired regulation of stride variability in Parkinson's disease subjects with freezing of gait. Exp Brain Res. 2003 Mar;149(2):187–94.

175. Novak D, Reberšek P, De Rossi SMM, Donati M, Podobnik J, Beravs T, et al. Automated detection of gait initiation and termination using wearable sensors. Med Eng Phys. 2013 Dec;35(12):1713–20.

176. Shalin G, Pardoel S, Nantel J, Lemaire ED, Kofman J. Prediction of freezing of gait in Parkinson's disease from foot plantar-pressure arrays using a convolutional neural network. Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society; 2020 July 20-24; Montreal, Canada. IEEE; 2020. p. 244–7.

177. Morris TR, Cho C, Dilda V, Shine JM, Naismith SL, Lewis SJG, et al. A comparison of clinical and objective measures of freezing of gait in Parkinson's disease. Parkinsonism Relat Disord. 2012 Jun;18(5):572–7.

178. Giladi N, Nieuwboer A. Understanding and treating freezing of gait in parkinsonism, proposed working definition, and setting the stage. Mov Disord. 2008;23 Suppl 2:S423-5.

179. Nonnekes J, Janssen AM, Mensink SHG, Oude Nijhuis LB, Bloem BR, Snijders AH. Short rapid steps to provoke freezing of gait in Parkinson's disease. J Neurol. 2014 Sept;261(9):1763–7.

180. Horak FB, Mancini M, Carlson-Kuhta P, Nutt JG, Salarian A. Balance and gait represent independent domains of mobility in Parkinson disease. Phys Ther. 2016 Sept;96(9):1364–71.

181. Nieuwboer A, Rochester L, Herman T, Vandenberghe W, Emil GE, Thomaes T, et al. Reliability of the new freezing of gait questionnaire: agreement between patients with Parkinson's disease and their carers. Gait Posture. 2009 Nov;30(4):459–63.

182. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord. 2008 Nov 15;23(15):2129–70.

183. Luo ZP, Berglund LJ, An KN. Validation of F-Scan pressure sensor system: a technical note. J Rehabil Res Dev. 1998 Jun;35(2):186–91.

184. F-Scan System | Tekscan [Internet]. [cited 2018 Jun 27]. Available from: https://www.tekscan.com/products-solutions/systems/f-scan-system

185. Howcroft J. Evaluation of wearable sensors as an older adult fall risk assessment tool [dissertation]. [Waterloo, ON]; 2016.

186. Pardoel S, Shalin G, Lemaire ED, Kofman J, Nantel J. Grouping successive freezing of gait episodes has neutral to detrimental effect on freeze detection and prediction in Parkinson's disease. PLOS ONE. 2021. (Accepted)

187. El-Attar A, Ashour AS, Dey N, Abdelkader H, Abd El-Naby MM, Sherratt SR. Discrete wavelet transform-based freezing of gait detection in Parkinson's disease. J Exp Theor Artif Intell. 2018 Sep.

188. Peker M, Sen B, Delen D. Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm. J Healthc Eng. 2015;6(3):281–302.

189. Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M. Analysis of in-air movement in handwriting: a novel marker for Parkinson's disease. Comput Methods Programs Biomed. 2014 Dec;117(3):405–11.

190. Alam MN, Garg A, Munia TTK, Fazel-Rezai R, Tavakolian K. Vertical ground reaction force marker for Parkinson's disease. PLOS ONE. 2017 May 11;12(5):e0175951.

191. Vergara JR, Estévez PA. A review of feature selection methods based on mutual information. Neural Comput Appl. 2014 Jan;24:175–86.

192. Louie S, Koop MM, Frenklach A, Bronte-Stewart H. Quantitative lateralized measures of bradykinesia at different stages of Parkinson's disease: the role of the less affected side. Mov Disord. 2009 Oct 15;24(13):1991–7.

193. Siragy T, Hill A, Nantel J. Recovery of dynamic stability during slips unaffected by arm swing in people with Parkinson's disease. PLOS ONE. 2021;16(4):e0249303.

194. Kendell C, Lemaire ED, Kofman J, Dudek N. Gait adaptations of transfemoral prosthesis users across multiple walking tasks. Prosthet Orthot Int. 2016 Feb;40(1):89–95.

195. Bekkers EMJ, Dockx K, Devan S, Van Rossom S, Verschueren SMP, Bloem BR, et al. The impact of dual-tasking on postural stability in people with Parkinson's disease with and without freezing of gait. Neurorehabil Neural Repair. 2018 Feb;32(2):166–74.

196. Kamieniarz A, Michalska J, Marszałek W, Stania M, Słomka KJ, Gorzkowska A, et al. Detection of postural control in early Parkinson's disease: clinical testing vs. modulation of center of pressure. PLOS ONE. 2021 Jan 12;16(1):e0245353.

197. Palakurthi B, Burugupally SP. Postural instability in Parkinson's disease: a review. Brain Sci. 2019 Sep 18;9(9):239.

198. Zabaleta H, Keller T, MartÍ Massó JF. Power spectral distribution analysis for detection of freezing of gait in patients with Parkinson's disease. Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering; 2008 Nov 23-27; Antwerp, Belgium. Springer; 2008. p. 2089–92.

199. Shine JM, Moore ST, Bolitho SJ, Morris TR, Dilda V, Naismith SL, et al. Assessing the utility of freezing of gait questionnaires in Parkinson's disease. Parkinsonism Relat Disord. 2012 Jan;18(1):25–9.

200. Roffo G, Melzi S, Cristani M. Infinite Feature Selection. Proceedings of the International Conference on Computer Vision; 2015 Dec 7-13; Santiago, Chile. IEEE; 2015. p. 4202–10.

201. Ziegler K, Schroeteler F, Ceballos-Baumann AO, Fietzek UM. A new rating instrument to assess festination and freezing gait in Parkinsonian patients. Mov Disord. 2010 Jun 15;25(8):1012–8.

202. Bikias T, Iakovakis D, Hadjidimitriou S, Charisis V, Hadjileontiadis LJ. DeepFoG: an IMU-based detection of freezing of gait episodes in Parkinson's disease patients via deep learning. Front Robot AI. 2021 May 7;8:537384.

203. REMPARK - Personal Health Device for the Remote and Autonomous Management of Parkinson's Disease [Internet]. [cited 2019 Oct 2]. Available from: https://rempark.cetpd.upc.edu/project

204. Sama A, Perez-Lopez C, Rodriguez-Martin D, Cabestany J, Arostegui JMM, Rodriguez-Molinero. A heterogeneous database for movement knowledge extraction in Parkinson's disease. Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning; 2016 Apr 24-26; Bruges, Belgium. ESANN; 2013. p. 413–8.

205. MASPARK [Internet]. [cited 2019 May 10]. Available from: https://www.aicos.fraunhofer.pt/en/our_work/projects/maspark.html

206. Niazmand K, Tonn K, Kalaras A, Fietzek UM, Mehrkens JH, Lueth TC. Quantitative evaluation of Parkinson's disease using sensor based smart glove. Proceedings of the 24th International Symposium on Computer-Based Medical Systems; 2011 June 27-30; Bristol, UK. IEEE; 2011. p. 1–8.

207. Nilsson J-O, Gupta AK, Handel P. Foot-mounted inertial navigation made easy. Proceedings of the International Conference on Indoor Positioning and Indoor Navigation; 2014 Oct 27-30; Busan, South Korea. IEEE; 2014. p. 24–9.

208. Zijlstra W, Hof AL. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait Posture. 2003 Oct;18(2):1–10.

209. Laudanski A, Yang S, Li Q. A concurrent comparison of inertia sensor-based walking speed estimation methods. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2011 Aug 30-Sept 3; Boston, USA. IEEE; 2011. p. 3484–7.

210. Madgwick SOH, Harrison AJL, Vaidyanathan R. Estimation of IMU and MARG orientation using a gradient descent algorithm. Proceedings of the International Conference on Rehabilitation Robotics; 2011 June 29-July 1; Zurich, Switzerland. IEEE; 2011. p. 1–7.

211. Cole BT, Roy SH, De Luca CJ, Nawab SH. Dynamic neural network detection of tremor and dyskinesia from wearable sensor data. Proceeding of the Annual International Conference of the IEEE Engineering in Medicine and Biology; 2010 Aug 31-Sept 4; Buenos Aires, Argentina. IEEE; 2010. p. 6062–5.

212. Moore ST, MacDougall HG, Gracies JM, Cohen HS, Ondo WG. Long-term monitoring of gait in Parkinson's disease. Gait Posture. 2007 Jul;26(2):200–7.

213. Khan AM, Lee YK, Kim T-S. Accelerometer signal-based human activity recognition using augmented autoregressive model coefficients and artificial neural nets. Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2008 Aug 20-25; Vancouver, Canada. IEEE; 2008. p. 5172–5.

214. Tochigi Y, Segal NA, Vaseenon T, Brown TD. Entropy analysis of tri-axial leg acceleration signal waveforms for measurement of decrease of physiological variability in human gait. J Orthop Res. 2012 Jun;30(6):897–904.

215. Kaminiski MJ, Blinowska KJ. A new method of the description of the information flow sources. Biol Cybern. 1991;65:203–10.

216. Koopman BO. Hamiltonian systems and transformation in Hilbert space. Proc Natl Acad Sci U S A. 1931 May;17(5):315–8.

217. Hua JC, Roy S, McCauley JL, Gunaratne GH. Using dynamic mode decomposition to extract cyclic behavior in the stock market. Phys Stat Mech Its Appl. 2016;448:172–80.

218. Phinyomark A, Limsakul C, Phukpattaranont P. A novel feature extraction for robust EMG pattern recognition. J Comput. 2009 Dec;1(1):71–80.

219. Phinyomark A, Nuidod A, Phukpattaranont P, Limsakul C. Feature extraction and reduction of wavelet transform coefficients for EMG pattern classification. Electron Electr Eng. 2012;122(6):27–32.

220. Challis RE, Kitney RI. Biomedical signal processing (in four parts). Part 3 the power spectrum and coherence function. Med Biol Eng Comput. 1991 May;29(3):225–41.

221. Shiavi R. Introduction to Applied Statistical Signal Analysis. 3rd ed. Burlington, USA: Elsevier; 2007.

222. Vinck M, Oostenveld R, Van Wingerden M, Battaglia F, Pennartz CMA. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. Neuroimage. 2011 Apr 15;55(4):1548–65.

223. Lachaux J-P, Rodriguez E, Martinerie J, Varela FJ. Measuring phase synchrony in brain signals. Hum Brain Mapp. 1999;8(4):194–208.

224. Zanotto D, Turchet L, Boggs EM, Agrawal SK. SoleSound: Towards a novel portable system for audio-tactile underfoot feedback. Proceedings of the 5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics; 2014 Aug 12-15; Sao Paulo, Brazil. IEEE; 2014. p. 193–8.

225. Sadeghi H, Allard P, Prince F, Labelle H. Symmetry and limb dominance in able-bodied gait: a review. Gait Posture. 2000 Sep;12(1):34–45.

Appendix A

Summary of FOG Detection and Prediction Studies and Features

Table A.1: Summary of recent FOG detection studies using wearable sensors.

Source	Studied Population	Walking Task Performed	Sensor Type and Location	FOG Detection Method	Features	Classifier Performance	Real Time
Moore 2008 [94]	11 FOG-PD (7 froze), ON and OFF, 46 episodes	Lab, straight walking, 180° turns, narrow doorways, obstacle avoidance.	IMU (1) left shank	Freeze index (FI) with person-specific thresholds. 6 s windows, detection based on FOG episode occurrences.	E	Detected 89.1% of episode occurrences, 10% false positives	No
Zabaleta 2008 [198]	4 FOG-PD, ON and OFF	Lab, sit to stand, 90° and 180° turns, figure-eight, doorway navigation, obstacle avoidance.	IMU (6) heels, shanks, thighs	Multivariate linear discriminant analysis, frequency-based features. Person-specific, detection based on classification of individual 3 s windows.	E	Area under ROC curve. Average of all participants: 0.937	No
Jovanov 2009 [15]	4 HC, 1 UFOG-PD	Lab, sit to stand and walking.	IMU (1) right knee	FI [94], 0.32 s windows (64 samples at 200 Hz).	E	-	Yes
Bachlin 2009-2010 [16,166,167, 169] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions to start, stop, and turn 360° in both directions. Simulated ADL (walk to room, return with glass of water)	Acc (3) left shank, left thigh, lower back	FI [94] with additional energy threshold to reduce false positives due to standing. 4 s windows with 0.5 s shift each step. Detection performance based on classification of windows with a 2 s tolerance.	E	Person-independent threshold: Sensitivity: 73.1% Specificity: 81.6%	Yes

Bachlin 2009 [105] *	10 FOG-PD (8 froze) 237 episodes	Lab, straight walking, 180° turns, randomly given instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Same methods as [16]. Improved offline through person-specific thresholds. Detection performance based on classification of windows with a 2 s tolerance.	E	Sensitivity: 88.6% Specificity: 92.8%	No
Delval 2010 [58]	10 HC, 10 NFOG-PD, 10 FOG-PD (5 froze), OFF, 20 episodes	Lab, 2 km/h treadmill, objects unexpectedly dropped on belt in front of participant.	CBMC, goniometers (2) knees	Compared stride features (e.g., step duration, step distance), and FI to person-independent thresholds, using 4.1 s windows.	E	Sensitivity: 75–83% Specificity: >95%	No
Djuric-Jovicic 2010 [117]	4 FOG-PD	Lab, sit to stand, straight walking through doorway, 180° turn, return to seat.	IMU (6) feet, shanks, thighs	Energy thresholds to detect movement, combined with NN for FOG detection. 0.2 s and 1.0 s windows. Classification performance based on number and duration of false detections.	E	Classification error up to 16%	No
Popovic 2010 [64]	9 FOG-PD (7 froze), ON, 24 episodes	Lab, sit to stand, straight walking through doorway, 180° turn, return to seat.	FSR in-shoe insole, Acc (6) feet, shanks, thighs	FSR signals to create single person-specific “normal step”. Pearson’s correlation coefficient (PCC) calculated for FSR signal of entire trial, then compared to a threshold.	E	-	No
Cole 2011 [113]	2 HC, 10 UFOG-PD, 107 episodes	Lab, unscripted ADL in mock apartment.	Acc (3) shin, thigh, forearm, EMG (1) shin	Stand vs sit detection, NN for FOG detection. Person-independent model, 2 s windows, detection performance calculated per 1 s segments.	E	Sensitivity: 82.9% Specificity: 97.3%	No
Tsipouras 2011 [127]	5 HC, 6 NFOG-PD, 5 FOG-PD	-	Acc (6) wrists, legs, chest, waist, Gyro (2) chest, waist	C4.5 decision tree, random forest, using 2 s windows.	E	Accuracy: Decision tree: 95.08% Random forest 96.11%	No

Niazmand 2011 [165]	6 FOG-PD (varying severity)	Lab, walk with 180° turns, with and without walking aid. Walking, 180° and 360° turns (both directions), doorways.	Instrumented pants, Acc (5) waist, thighs, shanks	Multi-stage, person-independent, threshold-based classification, identifies suspicious movement, then frequency feature used for classification, using 2 s windows.	E	Sensitivity: 88.3% Specificity: 85.3%	No
Zhao 2012 [95]	8 FOG-PD (6 froze), 82 episodes	Lab, 5-8 min random instructions (stand, walk, stop, turn).	Instrumented pants, Acc (5) waist, thighs, shanks (as in [165])	Time series, acceleration peaks detection (1.5 s windows) and frequency features via FFT (4 s windows), compared to person-independent thresholds.	E	Sensitivity: 81.7%	No
Mazilu 2012 [126] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, randomly given instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	AdaBoosted decision tree classifier best among several. Compared window sizes 1–4 s, 1 s was ideal. Detection performance based on classification of individual windows.	E	Person-specific: Sensitivity: 98.35% Specificity: 99.72% Person-independent: Sensitivity: 66.25% Specificity: 95.38%	No
Tripoliti 2013 [128]	5 HC, 6 NFOG-PD, 5 FOG-PD, ON and OFF, 93 episodes	Lab, rise from bed, walking tasks including doorways, 180° turns, and ADL.	Acc (4) ankles, wrists, IMU (2) waist, chest	Random forest classifier, 1 s windows. Person-independent detection performance based on classification of individual windows.	E	Sensitivity: 81.94% Specificity: 98.74%	No

Moore 2013 [96]	25 FOG-PD (20 froze), OFF, 298 episodes	Lab, TUG.	IMU (7) Lower back, thighs, shanks, feet	FI thresholds [94]. Compared different sensor locations, person-independent thresholds and window lengths. Detection performance based on classification of FOG episode occurrences and percentage of time frozen.	E	Lower back sensor, 10 s window: Sensitivity: 86.2% Specificity: 82.4%	No
Mazilu 2013 [18] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Person-specific decision tree, tested different feature sets and supervised vs unsupervised feature selection using principal component analysis (PCA). Detection performance based on classification of individual 1 s windows.	E, S	Unsupervised: Sensitivity: 77.7% Specificity: 87.56% Supervised: Sensitivity: 69.42% Specificity: 87.76%	No
Coste 2014 [32]	4 UFOG-PD, 44 episodes	Lab, corridor walk with dual task.	IMU (1) shank	Freezing of gait criterion (FOGC) feature, based on cadence and stride length, incorporating person-specific thresholds. Detection performance based on classification of FOG episode occurrences.	E	Sensitivity: 79.5%	No

Sijobert 2014 [153]	7 UFOG-PD, 50 episodes	Lab, corridor walk with dual task.	IMU (1) shank	FOGC [32], with person-specific thresholds. Detection performance based on classifying FOG episode occurrences. FOG episodes labeled as Green (n = 19, slight gait modification with no fall risk), Orange (n = 12, gait modification with fall risk) or red (n=19, FOG – blocked gait).	E	Correctly identified 26 of 31 FOG (orange and red)	No
Kwon 2014 [161]	20 FOG-PD (6 froze), ON, 36 episodes	Lab, repeated straight walk with 180° turns.	Acc (1) in shoe heel	Root mean square (RMS) of acceleration compared to person-specific threshold. 0.2–10 s windows. 3–4 s windows recommended.	E	Minimum of sensitivity or specificity: 85.8%	No
Pepa 2014 [164]	18 UFOG-PD, ON	Lab, 3 TUG variations: standard, with cognitive dual task, with manual dual task.	Acc (1) smartphone worn on belt at hip	Fuzzy logic model using frequency features, person-specific thresholds, 2.56 s windows. Detection performance based on classification of windows (sensitivity, specificity) and FOG episode occurrences (sensitivity) – distinction not indicated in results.	E	Sensitivity: 89% Specificity: 97%	No

Djuric-Jovicic 2014 [56]	12 FOG-PD, OFF	Lab, sit to stand, walk with 90° and 180° turns, multiple doorways.	IMU (2) shanks, FSR in-shoe insoles	Each stride is compared to a "normal" stride using spectral power, stride duration, and shank displacement. Custom rule-based method classified each stride based on person-specific thresholds.	E	FOG with tremor: Sensitivity: 99% Specificity: 100% FOG complete stop: Sensitivity: 100% Specificity: 100%	No
Assam 2014 [109] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Wavelet decomposition for feature extraction and conditional random fields for classification. Train/test for each person individually (person-specific model), compared 2.5, 4 and 8 s windows. Results for 3 participants, separately.	E, S	Best single participant results, with 4s window: Sensitivity: 65% Precision: 61.9%	No
Mazilu 2014 [93]	5 FOG-PD, 102 episodes	Lab, walking with turns and doorways.	IMU (2) ankles	Person-independent decision tree classifier (C4.5), multiple frequency-based input features, 2 s windows. Detection performance based on classifying FOG episode occurrences.	E	99 of 102 FOG detected	Yes

Mazilu 2015 [54] **	18 FOG-PD (11 froze), 182 episodes	Lab, walking tasks with cognitive and manual tasks. Straight walking, 180° and 360° turns, narrow spaces, hospital circuit with elevator, unexpected stops start, and turns.	IMU (2) wrists	Decision tree classifier (C4.5), features from wrist data, 3 s windows, person-specific detection performance based on classifying FOG episode occurrences.	E	Person-specific: Sensitivity: 90% Specificity: 83%	No
Zach 2015 [53]	23 FOG-PD (16 froze), OFF, 166 episodes	Lab, self-paced, fast walking, short steps, short fast steps, 360° turns both directions.	Acc (1) lower back	FI [94] compared to person-specific and person-independent thresholds, 2 s windows, detection performance based on classifying FOG episode occurrences.	E	Person-independent threshold: Sensitivity: 75% Specificity: 76%	No
Kim 2015 [125]	15 FOG-PD (9 froze), 46 episodes	Lab, hospital hallway, straight walk with 180° turns, also with dual tasks.	IMU (1) (smartphone) ankle, pants pocket, chest pocket, waist	AdaBoosted, person-independent, decision tree using 4 s windows. Compared different sensor locations, found waist best.	E	Smartphone on waist: Sensitivity: 86% Specificity: 91.7%	No
Handojoseno 2015 [60]	4 FOG-PD, OFF	Lab, TUG with 180° or 540° turns in both directions.	EEG, head	Person-independent NN to detect FOG during turning, 0.256 s windows, 1 s samples (117 normal turning, 224 FOG turning).	E, S	Sensitivity: 74.6% Specificity: 48.4%	No
Nivya Venu 2016 [112]*	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Wavelet decomposition used sub-band energies as features, continuous random field used for detection. 4 s windows. Person-independent detection performance based on classifying FOG episode occurrences.	E, S	Average of 3 participants test set: Sensitivity: 90.3% Precision: 95.8%	No

Martin 2016 [135] ****	6 FOG-PD, ON and OFF	Participant's home, 180° turns, doorways, walking outside, dual tasking and false positive test intended to create shaking resembling FOG (e.g. brushing teeth).	Acc (1) left hip	Different methods, feature sets, and window sizes compared. Best results from SVM. Detection performance based on classification of individual 1.6 s windows.	E	Sensitivity: 91.7% Specificity: 87.4%	No
Mazilu 2016 [55] **	18 FOG-PD (11 froze), 184 episodes	Lab, walking tasks with cognitive and manual tasks. Straight walking, 180° and 360° turns, narrow spaces and hospital circuit with elevator, unexpected stops start, and turns.	IMU (2) wrists	Decision tree classifier (C4.5) similar to [54], but fewer features and evaluation of single wrist input. 3 s windows, detection performance based on classifying FOG episode occurrences.	E, S	Person-specific: Sensitivity: 85% Specificity: 80% Person-independent: Sensitivity: 90% Specificity: 66%	No
Lorenzi 2016 [156–158,173]	16 UFOG-PD	Lab, walking through doorway, 180° turns.	IMU (2) shanks, IMU (1) side of head	Compared headset (combined with NN) and shin mounted IMUs. Shin method using custom *k*-index feature compared to person specific thresholds performed best.	E	From shin system: Sensitivity: 94.5% Specificity: 96.7%	No

Rezvanian 2016 [98] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Continuous wavelet transform computed ratio of frequency ranges, compared to person-independent threshold. Compared different window lengths, suggested 2 s windows for future real-time implementation.	E	Window 2 s: Sensitivity: 82.1% Specificity: 77.1% Window 4 s: Sensitivity: 84.9% Specificity: 81.01%	No
Ahlrichs 2016 [132] ***	20 FOG-PD (8 froze) ON and OFF, 209 episodes	Participant's home, 180° turns, doorways, walking outside, dual tasking and a false positive test intended to create shaking resembling FOG (e.g. brushing teeth).	Acc (1) waist	Person-independent SVM (linear kernel), best results with 3.2 s windows. Classified windows aggregated over 60 s and degree of confidence calculated and compared to threshold to determine whether a FOG episode was present during aggregation period.	E	Sensitivity: 92.3% Specificity: 100%	No
Capecci 2016 [104]	20 FOG-PD (16 froze), ON, 98 episodes	Lab, TUG test, cognitive or manual dual task.	IMU (1) smartphone at waist	Cadence and modified freeze index extracted and compared to person-specific thresholds. Detection performance based on classification of individual 3.56 s windows.	E	Sensitivity: 87.57% Specificity: 94.97%	No
Ly 2016 [114]	7 FOG-PD, OFF	Lab, TUG.	EEG, head	Person-independent NN, compared different features and number of EEG channel inputs. Data divided into 1 s segments (343 effective walking and 343 freezing).	E, S	Using all 32 channels: Sensitivity: 72.2% Accuracy: 71.46%	No

Pham 2017 [133] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Anomaly detection approach. Acceleration and spectral coherence features calculated for incoming window and "normal" reference. Person-independent thresholds used to classify FOG, "normal" reference updated with each Non-FOG window. Detection performance based on classification of individual 0.6 s windows.	E	Sensitivity: 87% Specificity: 94%	No
Pham 2017 [160] *	Development: 10 FOG-PD (8 froze), Test: 24 FOG-PD (OFF)	Lab, straight walking, 180° turns, random instructions and simulated ADL. Test: TUG, 180° and 540° turns in both directions.	Acc (3) left shank, left thigh, lower back IMU (7) foot, shank, thigh, lower back/hip	Development data from Daphnet*, test data from [199]. Several new features (including multichannel freeze index) presented and evaluated, detection used anomaly score compared to person-independent threshold to classify individual 3 s windows.	E, S	Freeze index using hip sensor X-axis: Sensitivity: 89% Specificity: 94%	No
Pham 2017 [107] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Freeze index and spectral coherence features used to generate average value used as threshold for FOG detection. Participant independent averages automatically updated during use. Detection performance based on classification of 0.6 s windows.	E	Sensitivity: 89.2% Specificity: 95.6%	No

Ahn 2017 [31]	10 HC, 10 FOG-PD, OFF, 42 episodes	Lab, TUG and 10 m walk tests.	IMU (1) in smart glasses	Custom FOG detection on glasses feature (FOGDOG), incorporated stride length and cadence, with person-specific thresholds, 1 s windows. Detection performance based on classifying FOG episode occurrences.	E	For PD participants: Sensitivity: 97% Specificity: 88%	Yes
Tahafchi 2017 [108]	2 FOG-PD	Lab, 6 min of walking turning and stepping in place.	EMG + IMU units (6) thighs, shanks, feet	SVM with Gaussian kernel, multiple time series and frequency features. 1 s windows.	E	Sensitivity: 90% Specificity: 92%	No
Suppa 2017 [59]	28 FOG-PD (25 froze), 152 episodes (102 OFF, 50 ON)	Lab, simulated home environment, TUG passing into narrow hall, turning both directions.	IMU (2) shins	*k* index from shin-mounted sensor compared to person-specific thresholds [155], with additional analysis of ON vs OFF states.	E	Sensitivity: 93.41% Specificity: 98.51%	No
Kita 2017 [99]	32 UFOG-PD (25 froze)	Lab, straight walking, through doorway, with 180° turn, and return.	IMU (2) shanks	Improvements on *k* index in [155], including new K_{swing}, *K'* features. Person-specific performance based on percentage of time frozen per trial.	E	Sensitivity: 93.41% Specificity: 97.57%	No
Rodriguez-Martin 2017 [129] ***, ****	21 FOG-PD, ON and OFF, 1321 episodes	Participant's home, 180° turns, doorways, walking outside, dual tasking and a false positive test intended to create shaking resembling FOG (e.g., brushing teeth).	IMU (1) left hip	SVM (radial basis function kernel), compared person-independent and person-specific models, using 3.2 s windows. Detection performance based on classifying FOG episode occurrences.	E	Person-independent: Sensitivity: 74.7% Specificity: 79.0% Person-specific: Sensitivity: 88.09% Specificity: 80.09%	No

Rodriguez-Martin 2017 [131] ***, ****	12 PD-FOG, 106 episodes	Participant's home, 180° turns, doorways, walking outside, dual tasking and a false positive test intended to create shaking resembling FOG (e.g., brushing teeth).	IMU (1) left hip	Same detection algorithm as [129], also using 3.2 s windows. Detection performance based on classifying FOG episode occurrences.	E	Sensitivity: 82.08% Specificity: 97.19%	Yes
Ly 2017 [115]	6 FOG-PD	Lab, TUG.	EEG, head	Person-independent Bayesian NN, to detect FOG during turns. Similar to [60], with addition of S-transform. Data divided into 1 s samples (204 normal turning, 204 FOG turning).	E, S	Sensitivity: 84.2% Specificity: 88.0%	No
Pepa 2017 [100]	20 UFOG-PD	Lab, TUG, with cognitive or manual dual task, sit, lay on bed, stand up and maintain upright posture, and run on a treadmill if able.	IMU (1) smartphone at waist	Fuzzy inference system compared to person-specific thresholds to detect periods of walking and FOG. 2.56 s windows (256 samples at 100 Hz). Detection performance based on classifying FOG episode occurrences, duration of FOG also examined.	E	FOG detection performance using ANOVA.	Yes
Wang 2017 [106]	9 UFOG-PD, OFF	Lab, gait initialization, narrow aisle, turning and dual tasks. One participant performed ADL in their home.	Acc (1) lower back	FI and RMS of acceleration. Both compared to person-specific thresholds and combined with an 'OR' statement. Detection performance calculated as percent time frozen per trial.	E	Sensitivity: 90.8% Specificity: 91.4%	No

Punin 2017 [101]	1 HC, 1 NFOG-PD, 6 FOG-PD, OFF, 27 episodes	Lab, stair climb and descent, straight walking and 180° turns.	IMU (1) right ankle	Discrete wavelet transform, compared to person-independent threshold. Detection performance based on classifying FOG episode occurrences.	E	Sensitivity: 86.66% Specificity: 60.61%	Yes
Saad 2017 [118]	5 FOG-PD ON, 64 episodes	Lab, straight walking, 180° turn, manual dual task or narrowed walking path. Clinic circuit including unscripted stops, starts, turns and doorways.	Acc (2) foot, shin, Goniometer (1) knee, Telemeters (IR proximity sensors) (2) upper and lower medial shank	Time and frequency domain features extracted from 2 s windows. Best features for each sensor identified. Person-independent, NN with Gaussian activation function used for detection. Defined average performance as mean of the fraction of FOG correctly identified and the fraction of Non-FOG correctly identified.	E, S	Average of all participants: Performance: 87%	No
Sama 2018 [130] ****	15 FOG-PD, ON and OFF	Participant's home, 180° turns, doorways, walking outside, dual tasking and a false positive test intended to create shaking resembling FOG (e.g. brushing teeth).	IMU (1) left hip	Compared multiple classifiers and feature sets, best results with SVM, using 1.6 s windows (64 samples at 40 Hz). Person-independent detection performance based on classifying FOG episode occurrences	E	Sensitivity: 91.81% Specificity: 87.45%	No

Prateek 2018 [163]	16 UFOG-PD (8 froze), 58 episodes	Lab, walking backwards, 180° turns, stepping over a board, walk a figure-eight loop, walk between sets of chairs placed close together.	IMU (2) heels	Detect instances of zero velocity or trembling, then, a point process filter computed probability of FOG based on foot position, orientation, and velocity. Detection performance based on classifying FOG episode occurrences, duration of FOG also examined.	E	Person-specific model, detected 47/58 FOG episode occurrences. Accuracy: 81.03%	No
Ashour 2018 [111] *	4 participants from Daphnet	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	SVM (linear kernel). Used infinite feature ranking [200] to reduce feature set. Person-specific detection performance based on classifying FOG episode occurrences.	E, S	1 patient top ranked (30 features) Accuracy: 94.4%	No
Camps 2018 [121] ****	21 FOG-PD, ON and OFF	Participant's home, 180° turns, doorways, walking outside, dual tasking and a false positive test intended to create shaking resembling FOG (e.g., brushing teeth).	IMU (1) left hip	1D CNN, 2.56 s windows stacked to combine current and previous windows. Person-independent detection performance based on classification of windows. Replicated other FOG detection methods and compared performance of models and feature sets.	-	CNN: Sensitivity: 91.9% Specificity: 89.5%	No

Oung 2018 [119] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Probabilistic NN, using time domain features (117) and frequency features (126), 4 s windows. Also examined SVM with RBF kernel. Person-specific and person-independent models compared.	E, S	Person-specific: Sensitivity: 99.83% Specificity: 99.96% Person-independent: Sensitivity: 87.71% Specificity: 87.38%	No
Li 2018 [134]	10 FOG PD, OFF, 281 episodes	Lab, straight walking (10 m and 100 m), 180° turns, narrow spaces.	Acc (1) lower back	Person-independent, unsupervised approach (training data not labeled). Mini batch *k* means clustering algorithm using acceleration entropy, 1 s windows. Once the centre of the FOG and Non-FOG classes were found, new data were classified based on which centre was closest.	E	Sensitivity: 92.4% Specificity: 94.9%	No
Mikos 2018 [116,120]	25 people, no other description provided (23 froze), 221 episodes	Lab, TUG and random walking.	IMU (2) ankles	Semi-supervised approach. NN, base training person-independent. Then unsupervised training during use improved performance.	E	Sensitivity: 95.9% Specificity: 93.1%	Yes

Rad 2018 [142] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Probabilistic anomaly detection approach using denoising autoencoder. Person-independent model trained to recognize normal gait (trained using Non-FOG data), 1 s windows. Compared CNN trained using Non-FOG (unsupervised) and FOG (supervised) data for comparison.	-	Proposed model: AUC: 77% Supervised model: AUC: 84%	No
El-Attar 2019 [110] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (1) left shank	Combined 1D discrete wavelet transform with FFT features, and used NN for classification. Person-specific detection performance based on classifying FOG episode occurrences.	E	Accuracy: 96.3%	No
Punin 2019 [102,103]	1 HC, 1 NFOG-PD, 6 FOG-PD, 27 episodes	Lab, straight walking, 180° turns, stair climbing.	IMU (2) back of ankles (distal posterior shank)	Discrete wavelet transform, signal energy compared to person-independent threshold using 32 s windows (256 samples at 8 Hz), updated every second. Detection performance based on classifying FOG episode occurrences.	E	Sensitivity: 60.61% Specificity: 86.66%	Yes

Mazzetta 2019 [159]	7 PD with varying disease severity, tested ON and OFF	Simulated apartment, TUG turning both ways, narrow hallways and doorways.	IMU/EMG devices shanks (tibialis anterior, gastrocnemius medialis)	Multi-stage thresholds using gyroscope and surface EMG. Gyro signal and threshold used to identify beginning and end of each step, then custom *R* feature compared to person-independent threshold distinguished FOG. Detection performance based on classifying individual steps.	E	False positive rate 5% False negative rate 2%	No
Reches 2020 [57]	71 PD tested ON and OFF	Lab, TUG with 360° turns both ways and a 180° through a doorway.	IMU (3) ankles and lower back	SVM with RBF, 3 s windows, Non-FOG windows 50% overlap, and 80% overlap for FOG windows. Window based evaluation.	E, S	Sensitivity: 84.1% Specificity: 83.4%	No
Dvorani 2020 [154]	4 UFOG-PD	Lab, 10 m straight line walking in hallway with 180°, sit to stand, 360° turns, passing through doorways [201].	IMU (1) foot	Acceleration and angular velocity compared to threshold to determine rest phase. Motion phase start detected using foot pitch and roll angles, motion phase end determined using jerk of the acceleration signal corresponding to heel strike. Once motion phase detected, GaitScore feature used for FOG detection. Evaluation based on identification of motion phases. Generalized model used for all participants.	-	Sensitivity: 97% Specificity: 87%	No

Shi 2020 [124]	63 FOG-PD	Lab, 7 m TUG test, walk through clinic.	IMU (3) ankles and neck (C7)	2D CNN, 4 s windows with 3 s overlap. Continuous wavelet transform on windowed data to produce scalograms. Train/test split was 50/13 participants.	-	Accuracy: 89.2% Sensitivity: 82.1% Specificity: 96%	No
Li 2020 [123]*	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Combination CNN and LSTM, with squeeze-and-excitation block, data augmentation was used to reduce dataset imbalance. 4 s windows, and 10 fold cross validation used.	-	Sensitivity: 95.1% Specificity: 98.8% AUC: 0.945	No
Sigcha 2020 [122] ***, ****	21 FOG-PD, ON and OFF, 1321 episodes	Participant's home, 180° turns, doorways, walking outside, dual tasking and a false positive test intended to create shaking resembling FOG (e.g., brushing teeth).	IMU (1) left hip	Compared many machine learning models, found a combination CNN LSTM was best. Used 3.2 s windows, 75% overlap. FFT was calculated for each window, groups of 4 windows were used as input. Used leave one out cross validation.	-	Sensitivity: 87.1% Specificity: 87.1% AUC: 0.939	No
Marcante 2021 [43]	20 PD tested ON and OFF, (53 of 140 walking trials contained FOG)	Lab, 2 minute walk test, 360° turns, TUG, and ADL.	Pressure sensing insoles with acc (2)	Threshold based FOG detection using 1 s windows and frequency-based features. FOG episode occurrence and duration used to produce binary decision "Did the participant experience FOG?".	E	Sensitivity: 96% Specificity: 94%	No

Mancini 2021 [61]	27 FOG-PD, 18 PD, 21 HC, tested OFF	Lab, 2 min walk test in 8m hallway, 1 min walk test with verbal dual-task.	IMU (8) shins, feet, wrists, sternum, lower back	Detect walking, using left/right correlation of angular velocity identify possible FOG, then use AP acceleration to calculate FI, compared to threshold. Results are for classification of Freezers vs Non-Freezers.	E	Rater 1 Sensitivity: 89% Specificity: 88% Rater 2 Sensitivity: 80% Specificity: 87%	No
Bikias 2021 [202] **	11 FOG-PD who froze, 180 episodes	Lab, walking with cognitive and manual tasks: straight, 180° and 360° turns, narrow spaces and hospital circuit involving elevator, unexpected stops start and turns.	IMU (1) wrist	3-class (FOG, walking with turns, stops) CNN classifier using linear and angular acceleration data as input. Leave one out cross validation, used 3 s windows, with 0.25 s overlap, window-based classification evaluation.	-	Sensitivity: 83% Specificity: 88%	No
FOG Prediction							
Mazilu 2013 [18] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Assumed duration of Pre-FOG class (1–6 s). 3 class decision tree classifier (Pre-FOG, FOG, not FOG) and 1 s window for feature extraction. Person-specific, prediction performance based on classification of individual windows.	E, S	1 participant with assumed 3 s Pre-FOG F1-score: 0.56	No

Mazilu 2015 [28] **	11 FOG-PD	Lab, walking with cognitive and manual tasks: straight, 180° and 360° turns, narrow spaces and hospital circuit involving elevator, unexpected stops start and turns.	Electrocardio-gram (1) (ECG) chest, galvanic skin response (1) (fingertip)	Assumed Pre-FOG duration (3 s) used for feature selection. Feature extraction used 3 s window. Multivariate Gaussian distribution used in anomaly detection model. Person-specific model for each individual. Instead of pre-defined Pre-FOG length, model decision threshold set manually. Prediction based on number of FOG episode occurrences.	E, S	SC data predicted 132/184 (71.3%) of FOG episode occurrences on average 4.2 s in advance, 71 false positives.	No
Handojoseno 2015 [27]	16 FOG-PD, 404 episodes	Lab, TUG.	EEG, head	Person-independent NN trained with 462, 1 s data segments for each class, tested on 172 segments. Extracted multiple frequency-based features using FFT and wavelets, multilayer perceptron NN for classification. Defined Pre-FOG as data between 5 s and 1 s prior to FOG.	E, S	Sensitivity: 86% Precision: 74.4%	No
Zia 2016 [21] *	3 chosen randomly from Daphnet	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (1) left shank	Person-specific layered recurrent NN. Detection applied to the 5 s prior to FOG. One participant had best results, trained on 9 episode occurrences, tested on 15.	-	Best participant: Sensitivity: 30% Precision: 89%	No

Palmerini 2017 [17] **	18 FOG-PD (11 froze), 180 episodes	Lab, walking with cognitive and manual tasks: straight, 180° and 360° turns, narrow spaces and hospital circuit involving elevator, unexpected stops start and turns.	IMU (3) ankles, lower back	Assumed Pre-FOG as 2 s before FOG. Features extracted from 2 s windows. Linear discriminant analysis to classify Pre-FOG vs normal gait windows. Person-independent model.	E, S	Sensitivity: 83% Specificity: 67%	No
Handojoseno 2018 [26]	16 FOG-PD	Lab, TUG.	EEG, head	Person-independent NN trained with 462, 1 s data segments for each class, tested on 172. Predict FOG by classifying data segment 5 s prior to freeze with Bayesian NN.	E, S	Sensitivity: 85.86% Specificity: 80.25%	No
Torvi 2019 [20] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	LSTM and RNN with 2 transfer learning approaches. Found best performance with LSTM, trained network then added person-specific final layer. Examined set Pre-FOG duration: 1, 3 and 5 s.	-	Predicted FOG up to 5 s in advance with >90% accuracy	No

Naghavi 2019 [24] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Window lengths (2-4 s), 2 s best. Personal 'stop threshold' to remove non-walking data. Used groups of 6 windows with 0.5 s shift. FOG onset identified when first three windows were statistically different (Kruskal–Wallis) from the next three. Identifications within 2s prior to FOG considered prediction. Defined 'Predictivity' as ratio of correctly predicted FOG events to all correctly identified events. Performance based on FOG episodes.	E	Sensitivity: 92.5% Specificity: 89.0% Predictivity: 88.8%	No
Naghavi 2019 [25]	18 PD (9 froze, 7 used in model training), 156 episodes	Lab, wide and tight 180° turns, straight walking, narrow hall, stops.	Acc (2), ankles (Vertical and AP axes)	1 s Pre-FOG data relabeled as FOG. Groups of 6, 2 s windows used for classification. Examined synthetic minority over-sampling technique (SMOTE), adaptive synthetic sampling (ADASYN), and misclassification cost. Used K-NN, SVM, decision tree, and MLP classifiers. Person-dependent models. Proposed ClsfBagging model (ensemble classifier with SVM, KNN and MLP trained using bagging).	E	Person-specific model identified 97.4% of episodes Predicted 66.7% of episodes (within 2 s before onset)	No

Arami 2019 [50]*	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Projected features forward in time then classified FOG with binary SVMs. Nine feature families were used, one SVM per family, then classification based on majority vote of SVMs. 10-fold cross validation 4 s windows with 0.5 s shift. Identifications correct if within ± 8 s of FOG onset.	E, S	Person-specific Sensitivity: 93% Specificity: 87% Predicted 94% of episodes 1.72 s in advance.	No
Zhang 2020 [19]	12 PD, tested OFF (304 episodes)	Lab, straight walking, turning 90° and 180°, doorway, narrow hall.	Acc (1) lower back	Step segmentation windows. Step based and conventional features compared, AdaBoosted models, personalized Pre-FOG duration. Window based evaluation.	E, S	Person-specific Sensitivity: 83.8% Specificity: 82.1% Person-independent Sensitivity: 72.7% Specificity: 78.9%	No
Demrozi 2020 [22] *	10 FOG-PD (8 froze), 237 episodes	Lab, straight walking, 180° turns, random instructions and simulated ADL.	Acc (3) left shank, left thigh, lower back	Three class k-NN, using 1-6 s windows, best was 2 s windows with 1 s overlap. Transformation matrix (Gaussian-kernel linear discriminant analysis) applied to windows prior to classification. Used Pre-FOG durations 2-4 s, 3-fold cross validation and window based evaluation.	-	Sensitivity: 94.1% Specificity: 97.1%	No

Borzi 2021 [23]	11 FOG-PD, tested ON (35 episodes) and OFF (34 episodes)	Lab, 7 m TUG test in simulated home environment.	Gyro (2) shins	Performed step segmentation on sagittal plane angular velocity. Features extracted from each step, wrapper-based feature selection. Used SVM, KNN, LDA, LR, best models were SVM and LDA. Tested Pre-FOG durations 2-5 s.	E, S	ON medication Sensitivity: 84.1% Specificity: 85.9% OFF medication Sensitivity: 85.5% Specificity: 86.3%	No

* Daphnet dataset originally collected by Bachlin et al. [16] (n = 10, 8 froze during testing). A total of 237 FOG episodes (8 participants OFF, 2 ON who claimed to freeze often while ON). Accelerometers on left shank, left thigh, and lower back.

** CuPiD dataset originally collected by Mazilu et al. [172] (n = 18, 11 froze during testing). 180 FOG episodes (ON/OFF state not mentioned in original article, subsequently reported ON state [55]). [54] reported 182 FOG episodes and [55] reported 184 episodes. IMU (9) on wrists, thighs, ankles, feet, and lower back. Galvanic skin response sensor (1) on hand, ECG sensor (1) on chest, smartphone (1) in front pocket with integrated IMU, pressure sensing shoe insole (1), functional near-infra-red spectroscopy (fNIR) sensor on forehead.

*** REMPARK project (Personal Health Device for the Remote and Autonomous Management of Parkinson's Disease) [203,204]. Data collected by multiple researchers, in participant's homes in OFF and ON states. Waist worn IMU.

**** MASPARK project [205].

Abbreviations and acronyms: Feature extraction (E), selection (S), FOG: freezing of gait, HC: healthy control participants, FOG-PD: people with PD with FOG symptoms, NFOG-PD: people with PD with no FOG symptoms, UFOG-PD: FOG symptoms not reported, ON: on medication, OFF: off medication

Acc: accelerometer, EEG: electroencephalogram, EMG: electromyography, Gyro: gyroscope, IMU: inertial measurement unit, CBMC: camera-based motion capture

CNN: convolutional neural network, NN: neural network, RNN: recurrent neural network, LSTM: long short-term memory neural network, SVM: support vector machine, KNN: *k*-nearest neighbour, LDA: linear discriminant analysis, LR: logistic regression

ADL: activities of daily living, TUG: Timed Up and Go Test, AUC: area under ROC curve, FFT: fast Fourier transform, FI: freeze index [94], FOGC: freezing of gait criterion, FSR: force sensing resistor, GSR/SC: galvanic skin response/skin conductance, PCA: principal component analysis, PCC: Pearson correlation coefficient, PSD: power spectral density, RMS: root mean square, ROC: receiver operating characteristic.

Table A.2: Features extracted from wearable-sensor data and used for freezing of gait detection or prediction.

Feature Name	Sensor Type	Sensor Location	Feature Description	Source
Mean	Acc, Gyro GSR Goniometer Telemeters	Chest, wrist, lower back, waist, thigh, knee, shanks, ankle, foot, GSR: finger, Goniometers: knees, Telemeters: between shanks	Mean of signal within window and axis. Acceleration: 3D vector magnitude or 3 axes Gyro: Angular velocity 3D vector magnitude, or 3 axes GSR: Conductance, low-pass filtered at 0.9 Hz Goniometer: Knee angular rotation. Telemeter: Voltage output, spikes in signal indicate that legs are next to one another.	[18,28,54,55, 57,100,108, 118,125,126, 129–131,135]
Min, Max, Median, HarmMean, GeoMean, Trim mean, Mode, Range	Acc, GSR, Gyro	Shank, thigh, lower back, insole GSR: finger	Descriptive statistics within given window. Acceleration: 3D vector magnitude, or individual axes GSR: Conductance, low-pass filtered at 0.9 Hz Gyro: angular rotation of the shank or back	[18,23,28,43, 57,131,132]
Increment of mean values	Acc	Waist	Difference between mean of current window and mean of previous window for anterior/posterior acceleration.	[129,130,135]
Difference in means of different axes	Acc	Waist	Difference in acceleration mean values between axes for current window (X and Y, X and Z, Y and Z).	[130,135]
Number of peaks in a window	Acc	Instrumented pants, Acc (5) waist, thighs and shanks	Number of times relative acceleration signal [206] passes above a threshold during 1.5 s window. Normal reference set to 3. More than 3 peaks per 1.5 s considered possible FOG.	[95,165]
Duration of acceleration above threshold	Acc	Instrumented pants, Acc (5) waist, thighs and shanks	Time the relative acceleration signal [206] is above a threshold. Normal reference 0.85 s per 1.5 s window. Longer durations considered suspicious (possibly FOG).	[95,165]
Turning degrees	Gyro	Lower back	Angular rotation about vertical axis. Calculated as the integral of low pass filtered (1.5 Hz) angular velocity about the vertical axis.	[17]
Left-right cross-correlation	Gyro	Ankles, feet	Maximum cross-correlation between mediolateral angular velocity (de-trended), left and right ankles (0.25 to 1.25 s). Correlation between limbs in ML axis.	[17,57,61]

Left-Right average SD	Gyro	Ankles	Average between SD of mediolateral angular velocity (de-trended), of right and left ankles.	[17]
Stride similarity	Gyro	Shank	Stride similarity, is scalar value inversely proportional to the similarity between the current and previous strides, calculated using the dynamic time warping algorithm.	[23]
RMS	Acc, Gyro	Sole of shoe, shank, thigh, lower back, ankle, chest	Root mean square (RMS) of acceleration or angular velocity data in given window, for 3 axes. Also total across axes [57].	[18,19,23,57, 106,119,125, 161]
-	Acc, Gyro	Ankles, lower back	$\frac{\max\left[RMS(x_{RL}, y_{RL}, z_{RL}),\ RMS(x_{LL}, y_{LL}, z_{LL})\right]}{RMS(x_{back}, y_{back}, z_{back})}$ where *RL* and *LL* indicate right and left legs. Calculated for both lower back gyroscope and accelerometer data.	[57]
Inter quantile range	Acc, Gyro	Ankle, thigh, chest, and waist	Interquartile range of acceleration or angular velocity in given window, for 3 axes.	[125]
Standard deviation	Acc, Gyro GSR Goniometer (G) Telemeters (T)	Chest, lower back, waist, thigh, shanks, ankle, foot, wrist, GSR: finger G: knees T: between shanks	Standard deviation in given window. Acceleration: 3D vector magnitude or 3 axes Gyro: 3D vector magnitude of angular velocity, or 3 axes GSR: Conductance, low-pass filtered at 0.9 Hz Goniometer: Knee angular rotation. Telemeter: Voltage output, spikes in signal indicate that the legs are next to one another.	[17,18,23,25, 28,54,55,57, 100,108,116, 118,120,125, 126,129–131,135]
Variance	Acc, Gyro	Shanks, thigh, lower back, waist, ankle, chest	Variance in given window. Calculated for acceleration or angular velocity data in given window, for 3 axes. In [111] and [110], variance calculated for FFT signal and detail and approximation coefficients from discrete wavelet transform.	[18,19,110, 111,125,126]
Acceleration indicator (S_{AC})	Acc	Shank, thigh, lower back	Binary value, to detect acceleration in each axis $S_{AC} = sgn((X - (\bar{X} - \sigma))_+)$, where X is a set of acceleration data, $\bar{X}$ is mean of X, σ is standard deviation of X, and *sgn(a)* is a sign function of a while $(a)_+$ returns a only if $a \geq 0$, otherwise returns 0.	[133]
Zero velocity and Trembling event intervals (ZVEI, TREI)	Acc, Gyro	Heel	Direction of gravitational acceleration used to calculate ZVEI and TREI to determine if foot is stationary (zero velocity) or trembling, from all acceleration and angular velocity axes.	[163]

Foot speed	Acc, Gyro	Heel	Foot position, orientation, and velocity, from 3 axis acceleration and angular velocity [207].	[163]
Integral	Acc	Waist, shank, thigh, low back	Integral of acceleration in given window, for given axis.	[119,129,130, 135]
Angular jerk	Gyro	Shank	Angular jerk is the second derivative of the angular velocity. Calculated around the ML axis.	[23]
Normalized angular jerk	Gyro	Shank	Normalized angular jerk is the angular jerk normalized by the time duration of the input signal.	[23]
Kurtosis	Acc, Gyro	Waist, ankle, shank, thigh low back, ankle	Kurtosis within a given window, from all acceleration axes, angular velocity, acceleration 3D vector, or absolute value of harmonics in 0.04–0.68, 0.68–3 and 3–8 Hz frequency bands (calculated from FFT of 3D acceleration)	[18,57,125,129 –131,135]
Skewness	Acc, Gyro	Waist, shank, thigh, low back, ankle	Measure of signal asymmetry within a given window, from all axes of the acceleration, angular velocity, acceleration 3D vector magnitude, or absolute value of harmonics in 0.04–0.68, 0.68–3 and 3–8 Hz frequency bands (calculated from FFT of 3D acceleration, or angular velocity).	[18,57,129– 131,135]
Mean absolute Value	Acc	Shank, thigh, low back	$MAV = \frac{1}{N}\sum_{n=1}^{N}\lvert x_n\rvert$ For acceleration *x* within a window of *N* data points. Calculated for 3 axes.	[119]
Simple square interval	Acc	Shank, thigh, low back	$SSI = \sum_{n=1}^{N}\lvert x_n\rvert^2$ For acceleration *x* within a window of *N* data points. Calculated for 3 axes.	[119]
v-order 2 and 3	Acc	Shank, thigh, low back	$v2 = (\frac{1}{N}\sum_{i=1}^{N} x_i^2)^{\frac{1}{2}}, v3 = (\frac{1}{N}\sum_{i=1}^{N}\lvert x\rvert_i^3)^{\frac{1}{3}}$ For acceleration *x* within window of *N* data points. Calculated for 3 axes.	[119]

Waveform length	Acc	Shank, thigh, low back	$WL = \sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$ For acceleration *x* within window of *N* data points. Calculated for 3 axes.	[119]
Average amplitude change	Acc	Shank, thigh, low back	$AAC = \frac{1}{N} \sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$ For acceleration *x* within a window of *N* data points. Calculated for 3 axes.	[119]
Difference absolute standard deviation	Acc	Shank, thigh, low back	$DASDV = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N-1} (x_{n+1} - x_n)^2}$ For acceleration *x* within window of *N* data points. Calculated for 3 axes.	[119]
Maximum fractal length	Acc	Shank, thigh, low back	$MFL = log_{10}\left(\sqrt{\sum_{n=1}^{N-1} (x_n - x_{n+1})^2}\right)$ For acceleration *x* within window of *N* data points. Calculated for 3 axes.	[119]
Step length	Acc, CBMC	Waist, thigh, shank, foot	Distance (m) between consecutive footfalls of the same limb, measured as double integral of A/P acceleration or by camera-based motion capture.	[58,100,108]
Step duration	Gyro	Thigh, shank, ankle, foot	Duration (s) between consecutive footfalls of same limb (or contralateral limb [23]), calculated from angular velocity peaks (raw or filtered)	[23,56,58,108]
Cadence	Acc, Gyro	Feet, shank, thigh, waist, lower back	Number of steps in given time (e.g., steps/minute), from time between peaks in angular velocity, vertical acceleration, second harmonic of acceleration in frequency domain [104], or calculated as in [208].	[19,58,100, 104,164]
Cadence variation	Acc	Waist	Standard deviation of cadence, from last 3 windows.	[164]

Stride peaks	Gyro, Angular velocity	Shank (ankle)	Peak of low pass filtered (4th order Butterworth 10 Hz) angular velocity within gait cycle, in frontal plane.	[116,120]
Fraction of weight span	FSR	13 locations under foot	Maximum minus minimum total force relative to estimated body weight (defined as 90% of measured weight)	[43]
Peak height and width	Gyro, Angular velocity	Shanks	Height (with respect to zero) and half-power width of the positive portion of the signal peak. Peak height represents the maximum angular velocity reached in each step and Peak width is the proportional to the swing time.	[23]
Gait pattern variability	Acc	Lower back	Width of the dominant harmonic in the power spectrum	[19]
Zero Crossing rate, mean crossing rate	Acc	Shank, thigh, low back	Number of times acceleration signal changes between positive and negative. Number of times acceleration signal changes between below average and above average in a given window. Calculated for 3 axes.	[18]
Signal vector magnitude	Acc	Shank, thigh, low back	Summation of Euclidean norm over 3 axes over entire window, normalized by window length.	[18]
PCA	Acc Goniometer (G) Telemeters (T)	Waist, shank, thigh, low back G: knees T: between shanks	Principal component analysis, calculated from raw 3 axis acceleration data from all sensors, each acceleration axis within specific spectral bands, or used to decrease dimensionality of multi-sensor feature set.	[18,118,129, 130]
Normalized signal magnitude area (SMA)	Acc	Shank, thigh, low back	Acceleration magnitude summed over 3 axes normalized by window length.	[18,131]
Eigenvalues of dominant directions (EVA)	Acc	Shank, thigh, low back	Eigenvalues of covariance matrix of acceleration along all 3 axes.	[18]
Energy (time domain)	Acc, Gyro, EMG on tibialis anterior	Forearm, foot, shank and thigh, waist, EMG: on shin	Energy, where $x(n)$ is discrete signal in time domain, n sample index, T window length, and E signal energy: $E = \sum_{n=1}^{T} \lvert x(n) \rvert^2$ Calculated from each acceleration or angular velocity axis, or from surface EMG signal.	[113,117]

Average acceleration energy (AAE)	Acc	Shank, thigh, low back	Mean of acceleration signal energy over 3 axes.	[18]
Asymmetry coefficient	Acc	Shank, thigh, low back	The first moment of acceleration data in window divided by standard deviation over window. Calculated for 3 axes.	[18]
Freezing of gait criterion (FOGC)	Gyro, Acc	Shank	Cadence and stride length measure, for stride n $$FOGC_n = \frac{C_n\, L_{min}}{C_{max}(L_n + L_{min})}$$ where C_n is cadence, L_n stride length. Maximum cadence C_{max} set to 5 strides/s, and minimum stride length L_{min}= 5 cm. Cadence and stride parameters calculated from angular velocity and acceleration [209]	[32,153]
FOG detection on glasses (FOGDOG)	Acc	Head	$$FOGDOG = \frac{N_{step}}{N_{max}} \times \frac{(D_{ref} - D')}{D_{ref}}$$ where D' is cumulative forward distance travelled by person during window, D_{ref} pre-set normal forward distance travelled, N_{step} cadence (number of steps/s), N_{max} pre-set maximum normal cadence, forward distance from double integral of forward acceleration after correction for head tilt angle, step length from [210].	[31]
K index, and K' index	Gyro	Shank	Summation of absolute value of low pass filtered angular velocity of left and right shanks in sagittal plane: $$k = lowpass(\lvert\omega_{left}\rvert) + lowpass(\lvert\omega_{right}\rvert)$$ ω_{left} and ω_{right} are angular velocities in sagittal plane.	[59,99,155–158]
R value	EMG, Gyro, angular velocity	IMU/EMG devices on shanks (tibialis anterior and gastrocnemius medialis)	R value is calculated once for each stride. $$R = \frac{\max(ABS)}{sEMG\lvert_{t=t_{\max(ABS)}}}$$ ABS is absolute value of moving average angular velocity in sagittal plane, sEMG surface EMG signal, max (ABS) maximum ABS during a stride, $sEMG\lvert_{t=t_{\max(ABS)}}$ value of surface EMG at that instant.	[159]

GaitScore (Ω)	Acc, Gyro	Foot	$\Omega = \left\lvert\frac{\emptyset_{pitch,min}}{\gamma_{min}}\right\rvert \cdot \left\lvert\frac{\emptyset_{pitch,max}}{\gamma_{max}}\right\rvert \cdot \lambda$ where γ_{min} and γ_{max} are reference values. $\gamma_{min} = \begin{cases} \zeta_{min} & \zeta_{min} < \emptyset_{pitch,min} \\ \emptyset_{pitch,min} & \text{otherwise} \end{cases}$ $\gamma_{max} = \begin{cases} \zeta_{max} & \zeta_{max} < \emptyset_{pitch,max} \\ \emptyset_{pitch,max} & otherwise \end{cases}$ where ζ_{min} and ζ_{max} are thresholds. λ is a weight defied as: $\lambda = \begin{cases} 1 \text{ for } \text{sgn}(\emptyset_{pitch,max}) = -\text{sgn}(\emptyset_{pitch,min}) \\ \frac{\lvert\emptyset_{pitch,max} - \emptyset_{pitch,min}\rvert}{c(\lvert\emptyset_{pitch,max}\rvert + \lvert\emptyset_{pitch,min}\rvert)} \text{otherwise} \end{cases}$ where c is a tuning weight. Ω of one indicates healthy gait, Ω close to zero indicates pathological gait.	[154]
Ratio of height of first peak	EMG	EMG: shank (tibialis anterior)	Height of peak at origin in autocorrelation of filtered EMG signal, in a given window.	[113,211]
Lag of first peak (not at origin)	EMG	EMG: shank (tibialis anterior)	Autocorrelation of filtered EMG signal, in a given window.	[113,211]
Step regularity	Acc	Lower back	Amplitude of the first peak in the autocorrelation signal.	[19]
Pearson's correlation coefficient (PCC)	Acc, Gyro, FSR	Shanks, thighs, waist, ankles, lower back, FSR: under feet	Similarity between two signals, with *n* sample points, x_i, y_i, i^{th} value of *x* and *y* signals; means $\bar{x}$, $\bar{y}$ $PCC = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$ Calculated between acceleration or gyroscope axes or between FSR force of a step compared to template "normal" step.	[56,57,64,129–131,135]
Ground reaction force	FSR	Under heel, ball of foot	Sum of forces from all force sensing resistors (FSR) under a foot.	[64]
Shank displacement	Acc, Gyro	Shanks	Shank displacement (m) calculated from vertical acceleration and pitch angular velocity [212].	[56]
Change of the shank transversal orientation	Gyro	Shanks	Rotation angle in transversal plane, calculated as integral of angular velocity data about vertical axis, for each limb and each stride.	[56]

Auto regression coefficient	Acc	Waist	Four auto-regression coefficients obtained by Bourg method from acceleration in all 3 axes [213].	[129,130,135]
Entropy	Acc, Gyro, EEG	Acc: ankle, pants pocket, waist, wrists, chest, thigh Gyro: chest, waist, low back, shanks EEG: head	Shannon's entropy: $$H(x) = -\sum_{i=1}^{n} P(x_i) log_2 P(x_i)$$ where discrete variable *x* contains *n* values, *P* is probability (often defined from histogram), calculated from each axis of acceleration or angular velocity in time and frequency domains, or filtered EEG voltage from multiple scalp locations. Also calculated for specific frequency bands [57].	[18,19,23,27, 57,114,125–128,132,134]
Sample entropy	Acc	Shank, thigh, low back, ankle	$$SampEn(m,r,N) = -ln\left[\frac{A^m(r)}{B^m(r)}\right]$$ $$m = 2, r = 0.2 \times \sigma$$ where σ is standard deviation [214].	[24,25]
Direct transfer function	EEG	Head	Application of coherence directionality in multi-variate time series [215]. Signals from motor control regions: O1-T4 (visual), P4-T3 (sensorimotor affordance), Cz-FCz (motor execution) and Fz-FCz (motor planning). Data filtered band-pass (0.5–60 Hz), band-stop (50 Hz), then normalized with a z-transformation.	[26]
ICA Independent component analysis	EEG	Head	Independent component analysis, used to maximize separation between signal components. Signals from motor control regions: O1-T4 (visual), P4-T3 (sensorimotor affordance), Cz-FCz (motor execution) and Fz-FCz (motor planning). Data filtered bandpass (0.5–60 Hz), band-stop (50 Hz), then normalized with a z-transformation.	[26]
Raw FFT	Acc, gyro, Goniometer (G)	Waist, shank, G: knee joint	The output signal from FFT. Calculated using acceleration, derivative of knee angle or angular velocity in the sagittal plane, in given window.	[56,58,122, 132]

PSD bands	Acc, EEG, Goniometer (G), Telemeters (T)	Heels, shank and thighs, knee, shanks. G: knee T: between shanks	Specific frequency bands of power spectral distribution (PSD), generated by FFT, short-time FFT (SFFT), Z-transformation, or other method to convert time domain signal into frequency domain. Calculated from each acceleration and angular velocity axis, knee angular rotation, telemeter voltage, or filtered EEG voltages.	[60,114,118, 198]
Harmonic ratio	Acc	Lower back	Stability of walking calculated by acceleration signal frequency domain.	[19]
Ratio of peak frequencies	Goniometer	Knee angle	Computed from FFT of derivative of knee angle. Ratio of highest amplitude in 3–8 Hz divided by highest amplitude in 0.3–3 Hz.	[58]
Power in frequency domain	Acc	Ankle, shanks, thighs, waist, chest, wrists	Area under curve of power spectral density plot, between specific bands. From acceleration 3D vector magnitude or individual axes. Also, ratio of specific bands (similar to FI) [119].	[16,25,54,55, 93,105,118, 119,125,126, 131,165–167,169]
Freeze index (FI)	IMU (Acc), Goniometer (G), Telemeter (T)	Acc: Various locations and sensor orientations, G: knees, T: between shanks	Ratio of signal power in freeze band (3–8 Hz) and locomotion band (0–3 Hz) [94] $FI = \frac{Area\ under\ the\ PSD\ curve\ in\ freeze\ band}{Area\ under\ the\ PSD\ curve\ in\ locomotion\ band}$ Calculated from acceleration and angular velocity axes, 3D vector magnitude, knee angular rotation or telemeter voltage.	[15–17,19,24,25, 53,57,61,93,94, 96,100,104–108,116,118–120,125,126, 132,164,166, 167,169]

Multi-channel FI (FI_{MC})	Acc	Foot, shank, thigh, lower back/hip	Ratio of powers P_H to P_L (i.e., freeze and locomotor bands) that are summations of acceleration signal powers over N channels, where Matrix X of size $N \times M$ represents an N-channel recording session with M regularly spaced time samples $$FI_{MC} = \frac{P_H}{P_L}$$ $$P_H = \frac{1}{2f_s}\sum_{n=1}^{N}\left[\sum_{i=H_1+1}^{H_2}\left[P_{XX_n}(i)\right] + \sum_{i=H_1}^{H_2-1}\left[P_{XX_n}(i)\right]\right]$$ $$P_L = \frac{1}{2f_s}\sum_{n=1}^{N}\left[\sum_{i=L+1}^{H_1}\left[P_{XX_n}(i)\right] + \sum_{i=L}^{H_1-1}\left[P_{XX_n}(i)\right]\right]$$ where N is number of inputs, f_s sampling frequency, P_{XX}, power spectrum of signal x, $H_1 = \frac{3N_{FFT}}{f_s}$, $$H_2 = \frac{8N_{FFT}}{f_s}, L = \frac{0.5N_{FFT}}{f_s}$$	[160]
K freeze index (FI_K)	Acc	Foot, shank, thigh, lower back/hip	Freeze index from each acceleration signal axis, spectral analysis using the Koopman operator [216]. Koopman eigenvalues and eigenfunctions are considered frequencies *(λ)* and power *(K(λ))* [217]. $$FI_K = \frac{\sum_{\lambda=H_1+1}^{H_2} K(\lambda)}{\sum_{\lambda=L+1}^{H_1} K(\lambda)}$$ where $L = 0.5(2\pi), H_1 = 3(2\pi), H_2 = 8(2\pi)$	[160]
Total power	Acc, Gyro	Lower back, thigh, shank, ankle	$$TTP = \sum_{j=1}^{M} P_j$$ where P is the power spectrum of the acceleration signal for a window of length M [218,219]. Calculated for 3 axes. In [57] used gyroscope when turning, acceleration otherwise.	[57,119]

Mean power	Acc	Lower back, thigh, shank	$MNP = \frac{1}{M}\sum_{j=1}^{M} P_j$ where *P* is power spectrum of acceleration signal for window of length *M* [218,219]. Calculated for 3 axes.	[119]
Energy Derivative ratio (EDR)	Acc	Lateral waist	Derivative of vertical acceleration energy in 3–8 Hz band divided by derivative of energy in 0.5–3 Hz band.	[100,164]
Median frequency	Acc	Lower back, thigh, shank	$MDF = \frac{1}{2}\sum_{j=1}^{M} P_j$ where *P* is the power spectrum of acceleration signal for a window of length *M* [218,219]. Calculated for 3 axes.	[19,119]
Peak frequency	Acc	Lower back, thigh, shank	$PKF = max(P_j), j = 1, \ldots, M$ where *P* is power spectrum of acceleration signal for a window of length *M* [218,219]. Calculated for 3 axes.	[119]
Peak amplitude, Frequency of peak amplitude	Acc, Gyro, FSR	Waist, thighs, shanks, lower back, ankle Gyro: shanks, lower back, ankle	Maximum value in frequency domain and corresponding frequency bin. Calculated for [0.5–3 Hz] band and [3–8 Hz] band. In [95] relative acceleration signal is used, defined in [206]. In [23] angular velocity is used. Also calculated with total force curve from FSR and AP COP [43].	[19,23,43,57, 95,132]
Higher harmonics	Acc	Waist, shanks	3 frequency bins with highest peaks. Calculated for all acceleration axes.	[129,132,135]
Frequency standard deviation	Acc, Gyro	Waist, lower back, thighs, shanks, ankle FSR in-shoe insoles	Standard deviation of signal in specific frequency bands, e.g., 0.1–0.68 Hz, 0.68–3 Hz, 3–8 Hz, 8–20 Hz, 0.1–8 Hz. Calculated for 3 axes.	[57,129,132, 135]
Principal harmonic width	Gyro	Shanks	The half-power width of the principal harmonic (peak amplitude frequency).	[23]
Weighted power spectral peak	Gyro	Shanks	Product of the maximum value in the frequency domain and the corresponding bin.	[23]
Low power frequency	Gyro	Shanks	Ratio between the power in the 0-2 Hz band and the total signal power.	[23]

Spectral density centre of mass (COM)	Acc, EEG, Goniometer (G), Telemeters (T)	Acc: Waist, thigh, shank, foot, EEG: Head, G: knee, T: between shanks	*x(n)*, is amplitude of bin *n*, and *f(n)* is frequency of bin *n:* $$COM = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$$ Calculated from 3 axis acceleration signal, filtered EEG voltage calculated within specific frequency bands, knee angular rotation or telemeter voltage.	[27,114,119, 129,130,135]
1st 2nd 3rd spectral moments	Acc	Lower back, thigh, shank	$$SM1 = \sum_{j=1}^{M} f_j \times P_j \, , SM2 = \sum_{j=1}^{M} f_j^2 \times P_j \, , SM3 = \sum_{j=1}^{M} f_j^3 \times P_j$$ where *P* is power spectrum of acceleration signal for window of length *M* [218,219]. Calculated for 3 axes.	[119]
Spectral coherence	Acc, EEG	Lower back, thigh, shank, EEG: head	Calculated from 3D acceleration or filtered EEG data using Welch method [220] $$C_{xy}(\omega) = \frac{P_{xy}(\omega)}{\sqrt{P_{xx}(\omega) \times P_{yy}(\omega)}}$$ where ω is frequency, $P_{xx}(\omega)$ is power spectrum of signal *x*, $P_{yy}(\omega)$ is power spectrum of signal *y*, and $P_{xy}(\omega)$ is cross-power spectrum for signals *x* and *y*. Also used with wavelet power spectrum in[27]. EEG signal from 4 locations: O1-visual, P4-sensorimotor affordance, Cz-motor execution, and Fz-motor planning. Filtered bandpass (0.5–60 Hz).	[27,133,160, 160]
Max amplitude and number of peaks of spectral coherence	Acc	Foot, shank, thigh, lower back/hip	Maximum amplitude and number of peaks of spectral coherence feature [133].	[160]
Discrete wavelet transform (DWT)	Acc, EMG	Lower back, thigh, shank, EMG: quadriceps	Discrete wavelet transform, Decomposition coefficients (approximate and detail coefficients) used as features. Calculated from the acceleration 3D vector magnitude each axis individually, or the raw EMG signal.	[101,108,109, 112]
Select bands of the CWT	Acc	Lower back, thigh, shank	Continuous wavelet transform in specific ranges (0.5–3 Hz, 3–8 Hz), also ratio of signal in 0.5–3 Hz band divided by signal in 0.3–8 Hz. Calculated for 3 axes.	[98]

Ratio of peak amplitude in wavelet transform bands	Goniometer,	Knee, derivative of knee angle	Sinusoidal wavelet transform used to calculate ratio of peak amplitude in 3–8 Hz band divided by peak in 0.5–3 Hz band.	[58]
Wavelet mean	Acc	Lower back, thigh, shank	Mean of detail coefficients of DWT using Debauches wavelet.	[24]
S-transform, amplitude	EEG	Head	Maximum amplitude in theta (4–8 Hz), alpha (8–13 Hz), low beta (lβ, 13–21 Hz) and high beta (hβ, 21–38 Hz) bands. Total amplitude across all bands were extracted for a specific time. Electrodes placed: F3, F4, FC1, FC2, C3, C4, CP1, CP2, CZ, P3, P4, PZ and O1, O2, OZ (F = frontal, C = central, P = parietal, O = occipital and Z = midline). Data filtered band-pass filter (0.5–40 Hz), normalized with z-transformation.	[115]
Energy (frequency domain)	Acc, Gyro	Foot, shank, thigh, forearm, waist, chest, ankle	Summation of squared absolute value of signal, where *f(h)* is discrete signal in frequency domain, with frequency bins h=1 to H, and E is signal energy $E = \sum_{h=1}^{H} \lvert f(h) \rvert^2$ Calculated from 3 axis acceleration or angular velocity signal.	[56,91,100, 125,126,131, 164]
Min, max amplitude of FFT and DWT	Acc	Shank, thigh, low back	Minimum and maximum values of energy of frequency domain signal, for both FFT and DWT approximation and detail coefficients, as in [102,103]. Calculated from 3 axis acceleration signal.	[102,103,110, 111]
Cross-correlation	EEG	Head	$R_{xy}(k) = E[x(n)y(n+k)]$ where $x(n)$, and $y(n+k)$ are two signals and k is the number of time units that signal $y(n)$ lags $x(n)$, and $E[\cdot]$ is expectation operator. EEG signal from O1-visual, P4-sensorimotor affordance, Cz-motor execution, and Fz-motor planning. Filtered band-pass (0.5–60 Hz).	[27]

Cross power spectral density (CPSD)	EEG	Head	Cross power spectral density [221] $$P_{xy}(f) = \sum_{k=-\infty}^{\infty} R_{xy}(k)\, e^{-j2\pi fkT}$$ where R is cross correlation function. EEG signal from 4 locations: O1-visual, P4-sensorimotor affordance, Cz-motor execution, and Fz-motor planning. Filtered band-pass (0.5–60 Hz).	[27]
Weighted Phase Lag Index (WPLI)	EEG	Head	Weighted phase lag index [222]. EEG signal from 4 locations: O1-visual, P4-sensorimotor affordance, Cz-motor execution, and Fz-motor planning. Filtered band-pass (0.5–60 Hz).	[27]
Wavelet cross spectrum	EEG	Head	The wavelet cross spectrum $WCS_i(s)$, defined as $$WCS_{xyi}(s) = S(W_{xi}(s)W_{yi}^{*T}(s))$$ where x and y are two time series, i time shift index, s scale, S a smoothing operator, and W_{xi} and W_{yi} the wavelet transform coefficients. EEG signal from 4 locations: O1-visual, P4-sensorimotor affordance, Cz-motor execution, and Fz-motor planning. Filtered band-pass (0.5–60 Hz).	[27]
Phase locking value	EEG	Head	Phase locking value [223] $$PLV_t = \frac{1}{N}\left|\sum_{n=1}^{N} Ne^{j\theta(t,n)}\right|$$ where $\theta(t,n)$ is phase difference between signals which can be derived from the angles of their wavelet coefficients. EEG signal from 4 locations: O1-visual, P4-sensorimotor affordance, Cz-motor execution, and Fz-motor planning. Filtered band-pass (0.5–60 Hz).	[27]

Table A.3: Features extracted from in-shoe plantar pressure sensors to characterize gait or FOG in PD and features used to classify fall risk in healthy older adults.

Feature	Sensor and location	Description	Source
Stride time	FSR (2) under toe and heel	Duration between foot strike to ipsilateral foot strike	[34,36]
Swing time	FSR (2) under toe and heel	Time foot is in the air during a stride	[36]
Double support time	FSR (2) under toe and heel	Time both feet are in contact with ground during a stride	[36]
Stride time variability	FSR (2) under toe and heel	Coefficient of variation (CV) of stride time, CV=(SD/mean)*100%	[34]
Left, right swing time	FSR (2) under toe and heel	Total time each foot is in the air per stride, (left or right), averaged across all strides	[39]
Left, right swing variability	FSR (2) under toe and heel	Coefficient of variation of swing time, for left and right legs, calculated individually.	[39]
Short and long swing time	FSR (2) under toe and heel	Limb (right or left) labeled as short or long depending on which limb had larger mean swing time	[39]
Short and long swing time CV	FSR (2) under toe and heel	CV values of short swing time and long swing time	[39]
Gait asymmetry	FSR (2) under toe and heel	Absolute value of the natural logarithm of the short swing time divided by long swing time	[39]
Velocity	Instrument mat (GAITRite)	Body forward velocity	[40]
Step length	Instrument mat (GAITRite)	Distance between steps of same leg	[40]
Step length variability	Instrument mat (GAITRite)	Step-to-step variability of step length	[40]
Single limb support	Instrument mat (GAITRite)	Duration of single limb support phase	[40]
Ground reaction forces	PPS (4) or FSR (4) heel, toe and medial/lateral ball of foot, PSI	Summation of pressure readings of all sensors, normalized by bodyweight	[38,175,224]
COP	PPS (4) heel, toe and medial/lateral ball of foot, PSI	Centre of pressure of each foot	[38,175]
Fraction of weight span	13 locations under foot	Max – min total force relative to estimated body weight (defined as 90% of measured weight)	[43]

Dominant frequency	13 locations under foot	Dominant frequency in the Fourier domain, calculated using the total GRF or the COP position in AP direction.	[43]
Features extracted from pressure sensing insoles for fall risk assessment of elderly adults			
Cadence	PSI (900 cell)	Number of strides per second	[46]
Stride time	PSI (900 cell)	Time from one foot strike to the next consecutive strike of same foot	[46]
Swing time	PSI (900 cell)	Total time foot is in the air during each stride	[46]
Percent stance time	PSI (900 cell)	Time during which one foot is in contact with ground as % stride time	[46]
Percent double support time	PSI (900 cell)	Time during which both feet are in contact with ground, as % stride time	[46]
Stride time symmetry index	PSI (900 cell)	Symmetry index (SI) of right and left leg stride times. From [225] $SI = \frac{(X_R - X_L)}{0.5(X_R + X_L)} \times 100\%$ where X_R and X_L are stride time for right and left legs	[46]
CV stride time, stance time, swing time	PSI (900 cell)	Coefficient of variation (CV) of stride time, stance time, and swing time	[46]
COP path reversals (A/P), number, length, duration	PSI (900 cell)	COP path should advance monotonically. Number, length and duration of COP path direction reversals per stance.	[47]
COP path deviations (M/L), number, length, duration	PSI (900 cell)	Number, length and duration of mediolateral COP deviations per stance. Deviations defined as first derivative of COP ML signal exceeding a threshold of ± 0.5 mm/frame	[47]
Lateral COP position	PSI (900 cell)	Maximum distance from centre line of insole. Normalized by width of trimmed sensor	[47]
Coefficients of variation of COP trajectory	PSI (900 cell)	Anterior-posterior (AP) and (ML) coefficients of variation (CV) for stance phase COP path	[47]
Impulse of each phase of gait (I1, I2, I3, I4, I5, I6, I7)	PSI (900 cell)	Impulse calculated as area under the curve of GRF normalized by bodyweight. I1 (foot-strike to first peak), I2 (first peak to minimum), I3 (minimum to second peak), I4 (second peak to foot-off), I5 (foot-strike to minimum), I6 (minimum to foot-off), and I7 (foot-strike to foot-off)	[46]

FSR: force-sensing resistor, PPS: pneumatic pressure sensor, PSI: pressure sensing insole.