

Automated Intelligent Cueing Device to Improve Ambient Gait Behaviors for Patients with Parkinson's Disease

A Dissertation Presented for the
Doctor of Philosophy
Degree

The University of Tennessee, Knoxville

Nader Naghavi

December 2020

Copyright © by Nader Naghavi, 2020
All Rights Reserved.

Acknowledgments

This material is based upon work supported by the National Science Foundation (Smart and Connected Health) under Grant No. 1502242.

I would like to express my gratitude to Dr. Espinoza–Wade, whom without his contagious passion and guidance this work would have not been possible. You have set an example of excellence as a researcher, mentor, instructor, and role model. I am also very grateful to Dr. Caleb Rucker for guiding and supporting me over the last year. I am thankful to Dr. Amir Sadovnik for his contribution and encouragement. I also have to thank the members of my committee, Dr. Jindong Tan and Dr. Subhadeep Chakraborty for their valuable advice and suggestions.

I am also thankful to my friends and colleagues Soheil Borhani, Reza Yazdanpanah and Aaron Miller, and lab–mates Fengpei Yuan, Sri Sadhu, Jake Childs and Vincent Aloï in DO307 and DO305 for all the supports and good times. I am also grateful for all my friends in ISA–UTK organization.

I thank with love to Maryam, my wife and best friend for more than a decade. You have supported, encouraged, and helped me get through this period in the most positive way. Special thanks to my parents, brothers and sisters who have always been steadfast and supportive.

Finally, I would like to thank Dr. Michelle Brewer and nurses in the University Neurology Cole Neuroscience Center for supporting this work. My sincere thanks go to all participants who trusted me and helped this study move forward.

Abstract

Freezing of gait (FoG) is a common motor dysfunction in individuals with Parkinson’s disease (PD). FoG impairs walking and is associated with increased fall risk. Although pharmacological treatments have shown promise during ON-medication periods, FoG remains difficult to treat during medication OFF state and in advanced stages of the disease. External cueing therapy in the forms of visual, auditory, and vibrotactile, has been effective in treating gait deviations. Intelligent (or on-demand) cueing devices are novel systems that analyze gait patterns in real-time and activate cues only at moments when specific gait alterations are detected. In this study we developed methods to analyze gait signals collected through wearable sensors and accurately identify FoG episodes. We also investigated the potential of predicting the symptoms before their actual occurrence.

We collected data from seven participants with PD using two Inertial Measurement Units (IMUs) on ankles. In our first study, we extracted engineered features from the signals and used machine learning (ML) methods to identify FoG episodes. We tested the performance of models using patient-dependent and patient-independent paradigms. The former models achieved 92.5% and 89.0% for average sensitivity and specificity, respectively. However, the conventional binary classification methods fail to accurately classify data if only data from normal gait periods are available. In order to identify FoG episodes in participants who did not freeze during data collection sessions, we developed a Deep Gait Anomaly Detector (DGAD) to identify anomalies (i.e., FoG) in the signals. DGAD was formed of convolutional layers and trained to automatically learn features from signals. The convolutional layers are followed by fully connected layers to reduce the dimensions of the features. A k-nearest neighbors (kNN) classifier is then used to classify the data as normal or FoG. The models identified 87.4% of FoG onsets, with 21.9% being predicted on average for each participant.

This study demonstrates our algorithm’s potential for delivery of preventive cues. The DGAD algorithm was then implemented in an Android application to monitor gait patterns of PD patients in ambient environments. The phone triggered vibrotactile and auditory cues on a connected smartwatch if an FoG episode was identified. A 6-week in-home study showed the potentials for effective treatment of FoG severity in ambient environments using intelligent cueing devices.

Table of Contents

1	Introduction	1
2	Background and Motivation	4
2.1	Parkinson’s disease	4
2.2	Freezing of Gait	4
2.3	External Cueing	7
2.4	FoG Detection	9
2.5	FoG Prediction	11
2.6	Research Design	12
3	Prediction of Freezing of Gait Using Statistical Inference and Lower-Limb Acceleration Data	14
3.1	Introduction	14
3.2	Background and Approach	14
3.3	Methods	16
3.3.1	Dataset	16
3.3.2	Feature Extraction	17
3.3.3	FoG Detection Algorithm	21
3.3.4	Annotation of Ground Truth	24
3.3.5	Online vs. Offline FoG Detection	26
3.3.6	Performance Evaluation	26
3.4	Results	29
3.4.1	Predictivity	29

3.4.2	Sensitivity	29
3.4.3	Specificity	29
3.5	Discussion	31
3.6	Conclusion	34
4	Addressing the Class Imbalance Problem in FoG Identification	35
4.1	Introduction	35
4.2	Background and Approach	36
4.3	Methods	38
4.3.1	Experiment Protocol	38
4.3.2	Dataset	38
4.3.3	Data Synthesis	41
4.3.4	Classifiers	45
4.3.5	Cost Sensitive Classification	47
4.3.6	Patient–Dependent and Patient–Independent Settings	47
4.4	Results	49
4.4.1	ADASYN Synthesized Datasets	49
4.4.2	Patient–Dependent Models	52
4.4.3	Patient–Independent Models	56
4.4.4	FoG Detection Latency	56
4.5	Discussion	62
4.6	Conclusion	64
4.7	Future Works	65
5	Addressing the One–Class Problem in FoG Identification	66
5.1	Introduction	66
5.2	Background and Approach	67
5.3	Methods	68
5.3.1	Dataset	68
5.3.2	Deep Gait Anomaly Detection (DGAD)	73
5.3.3	Transfer Learning	75

5.3.4	Performance Measures	77
5.4	Results	78
5.4.1	Effect of Transfer Learning	78
5.4.2	Effects of Data Augmentation and pre-FoG Windows	82
5.4.3	Specificity Threshold (θ_{spec})	82
5.4.4	Number of Model Outputs	88
5.4.5	Normal Gait Boundary Area ($\mathbf{R}$)	88
5.4.6	Step Size Optimization	94
5.4.7	Computational Complexity	94
5.5	Discussion	96
5.6	Conclusions	100
6	Mobile Application Development	102
6.1	Introduction	102
6.2	Real-time PDFoG System Development	102
6.2.1	Design Requirements	103
6.2.2	Android Application Development	104
6.2.3	User Interface	105
6.3	Computational Complexity	108
6.4	Prototype Evaluation	108
6.5	Conclusion	110
7	Cueing System Evaluation:	
	In-home Testing	111
7.1	Introduction	111
7.2	Study Design	111
7.2.1	Participants	112
7.2.2	Outcomes Measures	112
7.2.3	Participant Recruitment	114
7.3	Results	116
7.3.1	Gait and Fall Questionnaire (GFQ)	116

7.3.2	Usability Questionnaire (UQ)	116
7.3.3	Observations	116
7.4	Discussion	119
7.4.1	Gait and Fall Questionnaire	119
7.4.2	Usability Questionnaire	119
7.4.3	Observations and Comments	120
7.5	Conclusion	121
8	Conclusions and Future Works	122
8.1	Conclusions	122
8.2	Future Works	124
8.2.1	Recurrent Neural Networks	124
8.2.2	Activity Recognition	125
8.2.3	Larger Study	128
	Bibliography	129
	Appendix	149
A	Montreal Cognitive Assessment (MoCA)	150
B	MoCA - Instrusctions	151
C	Gait and Fall Questionnaire (GFQ)	155
D	Usability Questionnaire	157
E	User Manual	159
	Vita	173

List of Tables

3.1	Confusion matrix (NG: normal gait, FoG: freezing of gait).	28
3.2	Highest performing ($> 85\%$ specificity) combinations of sensor position, axis, window length, and feature sorted by descending specificity.	32
4.1	Feature set extracted from each window of data.	42
4.2	Classifiers performance in patient-dependent models using the original imbalanced dataset and equal misclassification cost (results show average performance of all seven participants).	53
4.3	Best performing classifiers in patient-dependent models (results show average performance of all seven participants).	54
4.4	Classifiers performance in patient-independent models using the original imbalanced dataset and equal misclassification cost (results show average performance of all seven participants).	57
4.5	Best performing classifiers in patient-independent models (results show average performance of all seven participants).	58
5.1	Datasets used for model training (values in parenthesis show the number of pre-FoG windows added to the anomalous class of training set).	70
5.2	Distribution of labels in participants' training and testing datasets.	72
5.3	Classification accuracy obtained using reference and target models trained with D4-A datasets.	80
5.4	Classification accuracy ($\theta_{spec} = 99\%$).	83
5.5	FoG onset prediction/detection ($\theta_{spec} = 99\%$).	84

5.6	Average FoG identification performance using D2-A datasets for training and varied specificity thresholds.	90
5.7	Average FoG identification performance using D4-A datasets and $\theta_{spec} = 99\%$ obtained with different number of model outputs.	91
5.8	Model performace using different step sizes for training and testing. False Positive (FP) rate is calculated by the number of FPs per 5 minutes.	95
5.9	Average FoG identification performance using conventional patient-independent models and DGAD target models.	98
6.1	Equipment model used to realize PDFoG system.	107
7.1	Demographic information of the participants (bolded rows show participants who finished the in-home study).	115
7.2	Scores for each section of the Usability Questionnaire. The scores show normalized scores (total score divided by the number of questions in each section).	117

List of Figures

2.1	(a) Distribution of FoG duration; (b) Number and category of FoG events (initiation, turning, or obstacle) relative to time after LD administration [9].	6
2.2	Three external cueing modalities. Information may be perceived and processed through multiple brain regions (BG–basal ganglia, SMA–supplementary motor area, and cerebellum [57].	8
3.1	Windowing approach to segment acceleration data into windows of equal length.	18
3.2	Vertical acceleration collected from a shank–worn sensor and corresponding power spectra in a participant with advanced PD. (a) Minimal activity was observed during volitional standing; (b) and (c) High–frequency components in the 2–6 Hz band were observed during FoG; (d) Highly tuned components at the stride (~ 1 Hz) and step (~ 2 Hz) frequencies (with higher frequency harmonics) were observed during normal walking [9].	19
3.3	Acceleration data and corresponding freeze index for each axis of the shank sensor. Areas labeled as pre–FoG are three seconds before FoG onset.	22
3.4	Grouping windows of data to predict/detect FoG (T is the starting point of any arbitrary window, W is the window length, and S is the step size). (a) Windows of data used to identify FoG/preFoG at point T_1 ; (b) Windows of data used to identify FoG/preFoG at the next window, point T_2	23
3.5	Post–processing of the labeled FoG signal to annotate the ground truth (F_s and F_e are the start and end of the FoG episode; T_p and T_d are prediction and detection duration; and P_i s are the identified FoG episodes).	25

3.6	Flow chart comparing the online and offline FoG prediction algorithms including all parameters.	27
3.7	Predictivity, sensitivity, and specificity versus window length for the three selected features (MV refers to majority voting as using all the three sensors; (F) and (V) refer to forward and vertical axes of the sensors).	30
4.1	Experiment setting. Number of obstacles in the object area varied between 0, 1, and 2. The width of the walking path in the object area (w) varied between 150% and 100% of participant shoulder width.	39
4.2	Distribution of FoG events duration.	40
4.3	The process of creating samples from acceleration signals. (a) Extracting features from six successive windows at time $t=T$ (left) and the next time step, $t=T+1$ (right). The red highlighted area shows the FoG labeled period using recorded videos and the green boxes show segments of data used to extract features from acceleration signal. (b) Aggregating arrays of features from different combinations of sensor-axis.	43
4.4	t-SNE visualization of the original dataset and new synthesized data using ADASYN ($\beta = 1$). <i>FoG-ADASYN</i> : new synthesized samples for FoG class, <i>FoG-original</i> : FoG samples from all participants, <i>NG</i> : normal gait.	46
4.5	Cost sensitive classification approaches.	48
4.6	Sensitivity, specificity, and F-score of SVM classifier using synthesized dataset.	50
4.7	Violin plot of SVM classifier using original and synthesized datasets for patient-dependent and -independent models. Solid black lines and white circles show average and median of performance, respectively.	51
4.8	Performance measures of classifiers for patient-dependent models using synthetic data and misclassification cost.	55
4.9	Performance measures of classifiers for patient-independent models using synthetic data and misclassification cost.	59
4.10	Acceleration signal collected from the right ankle sensor and the corresponding labeled and detected events using <i>ClsfBagging</i> with $\beta = 0.2$ and $C_{FoG} = 2$	60

4.11	Average FoG detection latency of <i>ClsfBagging</i> in patient-dependent models with $\beta = 0.2$ and $C_{FoG} = 2$. Negative values of time represent duration before FoG onset.	61
5.1	Adding pre-FoG samples to the FoG class in the training datasets.	71
5.2	(a) Structure of the reference and target models. The blocks show the output shape of each layer, and the values below show the number and size of kernels in convolutional layers or units in fully connected layers. (b) Input layer of the model (LA and RA correspond to left and right ankle, respectively) . . .	74
5.3	Ideal representation of model output. All NG and FoG samples are clustered in the NG and FoG boundary areas.	76
5.4	FoG onset prediction, detectio, and identification periods.	79
5.5	Mapped features from testing sets using reference and target models trained with D4-A datasets.	81
5.6	Average performance of the target models using different datasets for training ($\theta_{spec} = 99\%$).	85
5.7	FoG detection latency ($\theta_{spec} = 99\%$). Negative and positive latency correspond to prediction and detection of FoG, respectively.	86
5.8	Testing templates and the classification boundary using $\theta_{spec} = 99\%$	87
5.9	Effect of θ_{spec} on the testing sample classification using D4-A trained models.	89
5.10	Average performance of the target models using different number of output nodes (models are trained using D4-A datasets and $\theta_{spec} = 99\%$).	92
5.11	ROC curve using D4-A training datasets and $\theta_{spec} = 99\%$ (AUC = 0.82). . .	93
5.12	Circular heatmap of sensitivity and specificity obtained in FoG detection studies using patient-dependent models (CLSF: individual classifiers, ENSM: ensemble classifiers, CNN: convolutional neural networks, AD: anomaly detectors).	99
6.1	PDFoG system (version 1).	106
6.2	PDFoG system (version 2).	109

7.1	In-home FoG cueing intervention study design.	113
8.1	Structure of the suggested LSTM–DGAD network. First row shows training process for the current DGAD network. Second and third rows show training process for the LSTM layers. The fourth row shows the suggested network structure to predict FoG. Black solid outlines indicate that the layers are not trainable.	126
8.2	An activity recognition (AR) algorithm analyzes the received data to identify if the user is walking. The FoG identification algorithm analyzes the data only if the user is recognized as walking.	127

Chapter 1

Introduction

Parkinson’s disease (PD) affects nerve cells in a part of the brain that controls muscle movement. Among the neurological disorders, PD is the most common disorder leading to gait disturbance and falls [1]. Injuries related to such impairments impose heavy costs on both individual and healthcare systems, increase dependency and social isolation, and result in reduced quality of life [2–4]. Despite advances in pharmaceutical treatments and surgical techniques, impaired gait still persists, therefore the development of novel non-invasive rehabilitation approaches that work in conjunction with current treatments is a necessary step to address these difficulties [5–7].

Freezing of gait (FoG) is a paroxysmal symptom commonly seen among patients in advanced stages of the disease. There are three different sub-types of FoG: 1) suddenly becoming incapable to start walking or continue moving forward; 2) a complete absence of movement; and, 3) shuffling forward with very short steps [8]. FoG is a common cause of falls in patients with PD and the associated physical and psychosocial consequences have a great impact on the patients’ quality of life. The FoG episodes are transient and generally last for a few seconds [9], but the frequency and duration tend to increase with the disease progression [10]. The episodic and paroxysmal nature of the symptom makes it an extremely difficult task to measure and manage the FoG events. Standard clinical subjective measures rely on the patients’ or caregivers’ responses to questionnaires, such as Unified Parkinson’s Disease Rating Scale (UPDRS), Activities of Daily Living (ADL) part 14, and, freezing of gait questionnaire (FOG-Q). During the past decade, objective detection and quantitative

evaluation of the FoG events have become feasible due to the development of wearable sensors and signal processing methods [11, 12].

Exercise has shown positive effects on PD motor impairments in the short term [13, 14]. Increased cognitive engagement during the exercise through provision of cueing can also enhance the effects [15, 16]. External cueing refers to the provision of stimuli repeatedly regardless of the patients state or immediately after the detection of FoG. Cues, in the form of temporal or spatial stimuli, regulate and facilitate repetitive movements (e.g. walking) by providing a discrete motor target. Cueing have shown immediate positive effects on gait parameters (e.g. velocity, stride, and cadence) and FoG severity [17, 18], and the effects were consolidated after a follow-up period [19]. In Chapter 2 we discuss the FoG symptom, pharmacological treatments, and novel external cueing interventions in more detail.

Administration of externally cued exercise by specialized healthcare professionals usually requires cumbersome clinical setups [20]. Considering the increasing prevalence of PD [21] and the predicted shortage of physiotherapists [22], development of new ways to deliver cueing seems essential. Recent studies have been focused on using wearable sensors to capture gait signals and identify FoG episodes, which enables provision of cueing immediately after recognition of FoG. In this study we aimed to design an intelligent cueing device that monitors the gait pattern using wearable sensors and triggers cueing when an episode of freezing is detected. The cueing parameters will be set for each participant by an expert clinician to match their normal gait. The device will then be evaluated in the participants' homes to investigate the effectiveness of the intervention on FoG symptom. This study is one of the first to evaluate the effectiveness of a personalized intelligent on-demand cueing system in ambient environments.

Recent studies have shown encouraging results for detection of FoG events using machine learning techniques [8, 23–25]. However, it is hard to capture intrinsic properties embedded in time series data due to the random onset and duration of FoG events [26]. In addition, classification of biomedical signals raises some additional challenges; e.g., each individual has a unique normal gait pattern, variations from which may be due to freezing or just deviations from their optimal normal pattern. Recently, deep learning (DL) techniques and in particular convolutional neural networks (CNN) have obtained impressive success in

many time-series analysis due to their strong capabilities for automatically learning features [27, 28]. Using convolutional kernels with different sizes at different layers, the CNN can learn most fundamental features that needed to be designed manually otherwise.

In this study we aimed to extract features from signals collected via accelerometers placed on ankles, thighs, and waist, and to identify the best combination of sensor location and features to detect FoG occurrence (Chapter 3). Next, we aimed to improve the classification accuracy using ensemble classifiers and oversampling methods to address the imbalance in the dataset. We also investigated the performance of patient-dependent and patient-independent models in classifying gait signals (Chapter 4). However, these methods required data from both normal gait and freezing of gait classes, while half of the participants in our study did not exhibit freezing and only normal gait data was available for model training. To address the one-class issue, we introduced a new deep gait anomaly detection algorithm to identify FoG episodes. We also investigated the ability of the models to predict FoG onset before actual occurrence of the symptom (Chapter 5). This study demonstrates the potential of our algorithm to accurately identify FoG episodes and deliver of preemptive cueing to help patients prevent or overcome FoG.

Chapter 2

Background and Motivation

2.1 Parkinson’s disease

Parkinson’s disease (PD) results from degeneration of the dopamine-synthesizing neurons and is the second most common neurodegenerative disease among elders, preceded by Alzheimer’s [29]. PD is clinically characterized by motor symptoms, including resting tremor, slowness of movement (bradykinesia), hastening of the gait (festination), paucity of spontaneous movements (akinesia), reduced arm swing, and poor postural stability, as well as non-motor features, including cognitive impairment, hallucination and depression. Gait impairment is the most incapacitating symptom among patients with PD [30], as it negatively affects mobility and independence, and may result in emotional stresses and reduced quality of life [2–4]. The prevalence of PD in individuals over age of 50 is estimated to double by 2030 [21], which underlines the need for development of treatments for gait impairment in this relatively large population.

2.2 Freezing of Gait

Freezing of gait (FoG) is described as brief episodes of inability to step forward or as taking extremely short steps when initiating gait or turning [31]. FoG is commonly regarded as a feature of akinesia, an extreme form of bradykinesia [32]. The majority of FoG episodes last less than 30 seconds [9] (Figure 2.1a). This incapacitating motor dysfunction is highly

affected by environmental stimuli, cognitive input, medication and anxiety [33, 34]. As a result, FoG occurs more frequently at home than in the clinic, in complete darkness, and in other settings that require greater cognitive load like dual-tasking situations [35–38]. Schaafsma *et al.* described five subtypes of freezing when: initiating gait, turning, walking through tight hallways, terminating gait, and open space hesitation [39]. Figure 2.1b compares the number FoG episodes in three categories of initiation, turning, and walking through obstacles [9]. Clinical assessment of FoG severity is usually done using scales such as Unified Parkinson’s Disease Rating Scale (UPDRS) and FoG questionnaire (FoG-Q) [40]. However, these subjective scales lack validation of the frequency (i.e. number of events per day), time onset and the duration of the FoG events.

Levodopa (LD) and dopamine agonist (DA) are the most common pharmacological treatments for PD patients in early stages of the disease and suffering from impaired activities of daily living. LD provides the greatest benefit for patients with PD+FoG by improving stride length and gait velocity, and reducing FoG duration [41, 42], whereas DA may provoke more FoG episodes in early stages of disease [43]. Patients are considered to be “ON” when medication is working and “OFF” when it is no longer effective. ON and OFF phases can be quite distinct from each other, with an individual feeling energetic and able to move with reduced tremor and stiffness in the ON state, and unable to arise from a chair unassisted in the OFF state. Although LD is quite effective in early stages of the disease, its effect on parkinsonian symptoms wears off over time, as the disease progresses and treatment continues, and the effective duration may shorten to 2 hours [5]. Furthermore, long-term administrations of LD may induce involuntary, erratic movements, called dyskinesia [44]. On the other hand, drugs for non-motor symptoms could interfere with the effectiveness of LD and aggravate motor symptoms [45]. FoG happens more frequently during the OFF state [12] (Figure 2.1b).

For many patients with concurrent FoG symptoms and cognitive disorder, the efficacy of medication therapy is poor and deep brain stimulation (DBS) is often prescribed [7]. Non-randomized studies with low sample sizes have shown that DBS can improve FoG and the effect can last for at least 1 year, however, the risk of aggravating other symptoms still

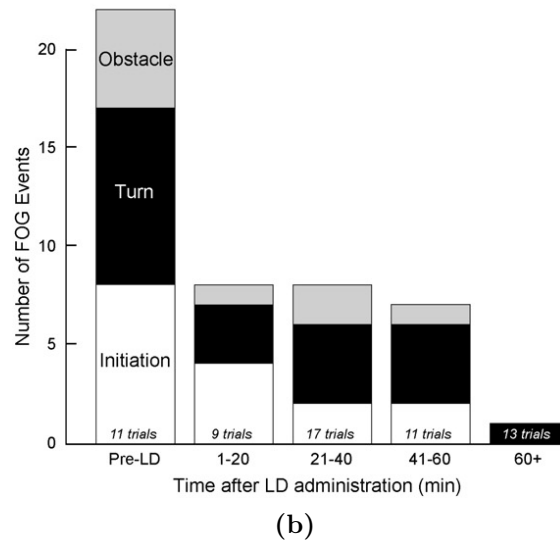
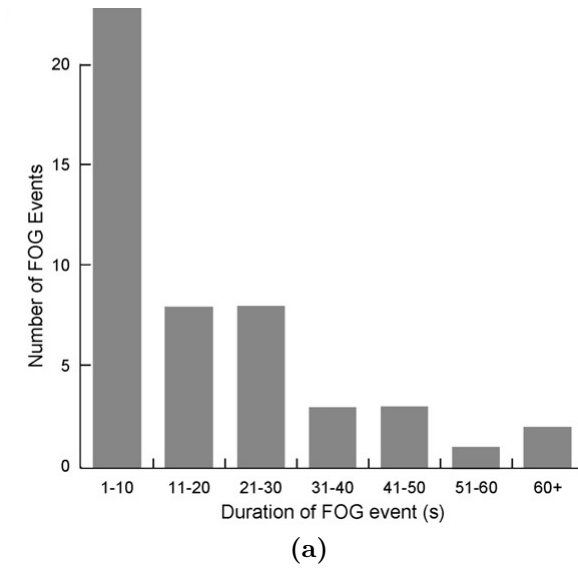


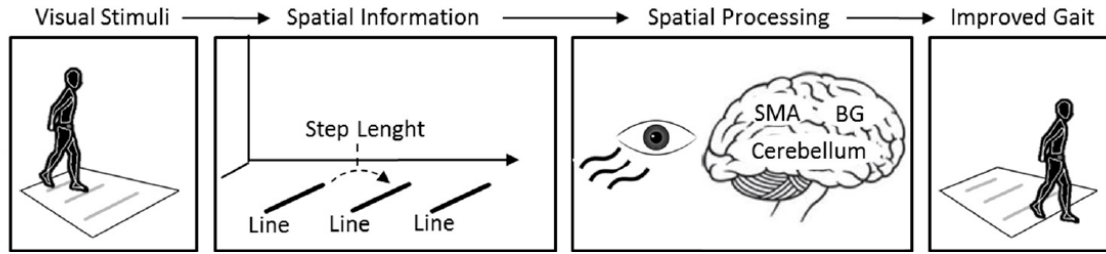
Figure 2.1: (a) Distribution of FoG duration; (b) Number and category of FoG events (initiation, turning, or obstacle) relative to time after LD administration [9].

exists [6, 7]. Therefore, effective non-pharmacologic treatments are also required to relieve symptoms and improve mobility and quality of life.

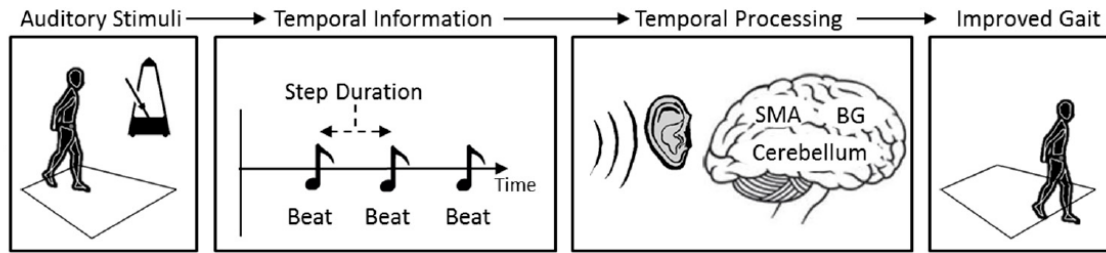
2.3 External Cueing

Automaticity in gait of individuals with PD is hampered when dual tasking, and even more during FoG [46]. It is thought that motor dysfunctions in PD result from limited resources and less automaticity of motor plans due to the damage to the basal ganglia [47]. Although rates of motor learning in PD patients can be slower than their age-matched individuals, their motor performance can be improved with rehabilitation using motor learning principles [48]. To improve automaticity, behavioral self-management strategies are conducted to shift patients' habitual motor control to a target-specific tasks [49, 50]. Patients are instructed to allocate their attention to gait by specific self-prompting instructions (internal cues) or make use of devices (external cues) or a combination of these. External cues circumvent the use of the defective internal trigger originating in the basal ganglia and enable planned movements [51]. External spatial cues (e.g., strips placed or laser beams projected on the floor) can be customized for each patient based on their stride length and show patients *where* to put their next step (Figure 2.2a). On the other hand, temporal cues (e.g., auditory metronome or vibrotactile feedback) are customized based on cadence and inform users *when* a step should be taken (Figure 2.2b and 2.2c). Such non-invasive, non-pharmacological interventions provide patients with discrete targets for execution of movement to preserve normal gait and prevent FoG, or implement rescue instructions once FoG has occurred [52, 53]. Studies have shown that rehabilitation with external cueing resulted in overall positive effects on FoG severity as well as improved gait velocity and stride length [49, 54–56].

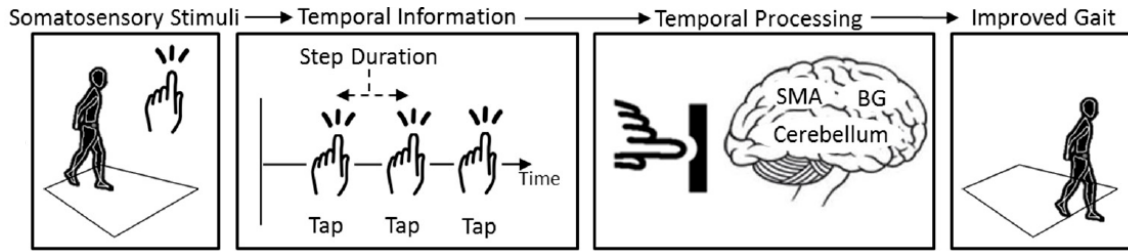
Frazzitta *et al.* investigated the effects of visual and auditory external cueing on PD participants with FoG symptoms [58]. Participants received cueing therapy daily for 20 minutes and demonstrated statistically significant improvements in FOG-Q score after four weeks. Nieuwboer *et al.* delivered cueing training in the home of 153 individuals with PD [1]. Cueing was provided through three modalities: (1) auditory (a beep triggered through an earpiece); (2) visual (light flashes triggered through a light-emitting diode attached to a pair



(a) Visual cueing: parallel lines marked on the floor at a fixed distance apart



(b) Auditory cueing: beeps set at the normal walking cadence



(c) Somatosensory cueing: taps or vibrations set at the normal walking cadence

Figure 2.2: Three external cueing modalities. Information may be perceived and processed through multiple brain regions (BG–basal ganglia, SMA–supplementary motor area, and cerebellum [57]).

of glasses); and (3) somatosensory (pulsed vibrations triggered by a miniature cylinder worn under a wristband). The results showed that severity of freezing was reduced by 5.5% in participants with FoG symptoms. Kadivar *et al.* compared a battery of clinical assessments after a 6-week training session and 4 weeks follow-up in two groups of eight participants practicing with rhythmic auditory stimulation stepping (RAS group) and no-cue stepping (no RAS group) [48]. Results suggested that the RAS group significantly improved FoG symptoms (as measured by FOG-Q) and maintained improvements above baseline values for at least 4 weeks after practice termination.

2.4 FoG Detection

“Always-on” cueing is defined as a paradigm in which stimulus is delivered repeatedly to the user regardless of any prior or imminent FoG episodes. However, individuals with PD are known to adapt to interventions provided continuously, thus reducing the effect of cueing [49, 59]. Therefore, it is ideal to deliver an external stimulus only when it is contingent on symptom onset. This requires the development of an integrative system capable of automatically detecting FoG episodes. A variety of methods for such an approach include using data captured from electrocardiography (ECG) systems [60], electromyography (EMG) systems [61, 62], 3D motion systems [63, 64], foot pressure sensors [65, 66], and Inertial Measurement Units (IMUs) [8, 9]. FoG detection methods using the last category mainly employ spectral analysis of accelerometer signals collected by wearable sensors placed on lower parts of the body (ankle and knee). Moore *et al.* used an IMU mounted around shank, and defined a freeze index (FI) using power spectrum analysis of accelerometer signals [9]. Applying an individually calibrated threshold to the FI indices, they could recognize 89% of FoG events, while classifying 10% of events incorrectly as FoG. Bächlin *et al.* obtained higher sensitivity (88.6%) and specificity (92.4%) in detecting FoG using FI and a second power threshold to discriminate volitional standing from recognized FoG [12].

Tripolitia *et al.* calculated the entropy of signals captured from six accelerometers and two gyroscopes placed on wrists, ankles, waist and chest of sixteen participants to train four classification algorithms (Naïve Bayes, Random Forests, Decision Trees and Random Tree)

to detect FoG events [8]. They compared different configurations of sensors in order to find a set of sensors optimal for FoG episode detection. Their proposed methodology was able to detect FoG events with 81.94% sensitivity and 98.74% specificity with the Random Forests classification algorithm using signals from all sensors.

Martín *et al.* distinguished FoG from normal gait using data collected from 21 PD participants wearing a single waist-worn triaxial accelerometer [24]. A total of 55 features were extracted from windows of 3.2 seconds with 50% overlap, and used to train and test generic and personalized SVM classifiers. The results showed sensitivity and specificity of 74.7% and 79.0% for generic, and 88.1% and 80.1% for personalized models. Tahafchi *et al.* used temporal, spatial, and physiological features to train a SVM classifier to identify freezing episodes [25]. Data were collected using inertial sensors attached to the thigh, shank, and foot, and non-invasive surface EMG sensors attached to quadriceps/tibialis muscles of PD participants. They detected 90% of the FoG events correctly, while identifying 8% of the non-FoG data incorrectly as FoG.

Mazilu *et al.*, tested different supervised machine learning algorithms on detecting FoG events using 3-D acceleration signals collected from the ankle, knee, and hip of 10 PD participants [23]. A correlation-based feature subset selection was used to choose only the most discriminative features. They compared results from two different approaches: “patient-dependent”, in which both training and testing data were from the same participant, and “patient-independent” utilizing leave-one-out cross validation (LOOCV). Their results for patient-dependent models showed average sensitivity, specificity, and F1 (see section 3.3.6) of 99.54%, 99.96% and 99.75%, respectively, using Random Forest classifiers. However, the average performance for patient-independent models resulted in much lower sensitivity and specificity (66.25% and 95.38%, respectively).

Using more recent techniques, Camps *et al.* applied a deep learning method to detect FoG episodes in home environments [27]. Their algorithm employed an eight-layered one-dimensional convolutional neural network and spectral window stacking as data representation to combine information from both the prior and current signal windows. They used data from a single IMU placed on the waist of thirteen participants to train the model and tested the model on data from four other participants (not included in the training

set). The model detected FoG episodes with sensitivity and specificity of 91.9% and of 89.5%, respectively. Finally, Xia *et al.* implemented a deep convolutional neural network to detect FoG events [28]. The system segmented one-dimensional acceleration signals into windows of 4 seconds and realized automatic feature learning in order to discriminate FoG from normal gait, thus, removing the need for extracting hand-crafted features and time-consuming feature selection. They reached average sensitivity and specificity of 99.64% and 99.99%, respectively, using patient-dependent, and 74.43% and 90.60% using patient-independent models.

2.5 FoG Prediction

Although activation of cues after “*detection*” of FoG onset can help patients overcome freezing, it may result in cognitive overload by superimposing an external stepping rhythm and aggravate the FoG. Ginis *et al.* suggested that an optimal timing for delivering intelligent cues is before the actual onset of a potential FoG episode [49]. Such “*predictions*” would enable preemptive cueing and potentially reduce the likelihood of FoG occurrence [1].

Palmerini *et al.* removed FoG labeled samples and formed a binary dataset of 2-second pre-FoG and normal gait [67]. The linear discriminant analysis classifier identified 83% of the pre-FoG episodes on average in patient-dependent model, although only 67% of normal gait samples were correctly identified. Torvi *et al.* studied the performance of a deep domain adaptation algorithm to address variability in gait data and developed a prediction model for a particular patient by leveraging data from different subjects [68]. Their model reached 88% prediction accuracy within 1 second before FoG onset in patient-dependent mode. The prediction accuracy improved to 93% with the addition of transfer learning techniques. However, it is not clear if the prediction accuracy in this study was equivalent to sensitivity metric, or the accuracy of classification for both classes combined. Thus, no comparison with the other studies can be made.

2.6 Research Design

In this study, high-resolution, quantitative data obtained by wearable sensors will be used to identify the event onset and provide effective intervention. There are therefore three associated research objectives: (1) development of algorithms to predict FoG symptom onset; (2) implement the developed methods into an Android application for real-time prediction of FoG and provision of cueing; and, (3) evaluation of the cueing impact on frequency and severity of FoG. In the following we describe these objectives in detail.

The described studies above achieved high classification accuracy for FoG detection, especially with patient-dependent models which reduces the effect of heterogeneity in data from different participants. However, these studies seldom reported the detection latency (i.e., time associated with classification after FoG onset) and prediction capability (i.e., time associated with classification prior to FoG onset). This study aimed to evaluate the classification performance of different machine learning and deep learning methods to identify FoG. In Chapter 3, a new algorithm was employed to define the best combination of sensor position, axis, sampling window length and features to recognize FoG. We hypothesized that gait deterioration before FoG onsets can be discriminated from normal gait using statistical analysis of features extracted from successive windows of lower-limb acceleration signals. We defined a new performance measure, “predictivity”, to compare the number of correctly predicted FoG events among different combinations [69]. In Chapter 4, we investigated the ability of individual and ensemble classifiers to predict FoG. We also studied the effect of imbalanced class ratio (i.e., the relative infrequency of FoG occurrence when compared to normal gait behaviors) and classification cost on classifier performance [70, 71]. Using all these techniques, the detection and prediction remained challenging particularly for individuals who did not exhibit sufficient instances of freezing episodes for model training. In Chapter 5 we introduced a Deep Gait Anomaly Detector (DGAD) to predict FoG. The algorithm implements a transfer learning-based approach to train a reference model using FoG and normal gait samples of participants who experienced FoG during the clinical data collection sessions. The model is then re-trained using only normal gait data from a new participant to form a patient-specific model. We also evaluated the effect of augmenting

training data and adding pre-FoG segments (data occurring prior to the freezing onset) to the FoG training data on model performance.

With the encouraging results obtained in the experiments, we developed Android applications to collect and analyze data in real-time and provide vibrotactile and auditory cueing for patients (Chapter 6). We hope this technology can provide an effective intervention to accurately predict FoG events using wearable inertial sensor data, and ultimately help patients prevent FoG through external preemptive cueing. Finally, the effect of the intervention on FoG symptom will be evaluated by PD participants during a 6-week in-home study.

Chapter 3

Prediction of Freezing of Gait Using Statistical Inference and Lower-Limb Acceleration Data

3.1 Introduction

In this chapter we discuss feature extraction from the signals collected through sensors placed on ankles, thighs, and waist of participants with PD. First, we introduce features capable of discriminating FoG from normal gait. Previous studies have shown that using thresholds for these features can reveal FoG. Next, we discuss our new approach using feature values from successive segments of data to predict FoG before its occurrence. We expected that, using a statistical analysis, the feature values from pre-FoG periods can be distinguished from those of normal gait periods. Finally, we will discuss the results and ability of the proposed algorithm to accurately identify FoG.

3.2 Background and Approach

Moore *et al.* used an shank-worn Inertial Measurement Unit (IMU) and defined a freeze index (FI) using power spectrum analysis of accelerometer signals [9]. They Applied an individually calibrated threshold to the FI values and obtained 89% sensitivity and 90%

specificity (see section 3.3.6). Although the sensitivity was relatively high, the algorithm was highly patient-dependent and a global threshold value reduced the accuracy. Bächlin *et al.* proposed an FoG detection algorithm based on the freeze index and a second power threshold to discriminate standing events from recognized FoG, and achieved higher sensitivity (88.6%) and specificity (92.4%) [12]. Rezvanian and Lockhart used continuous wavelet transform (CWT) to define an index for identifying FoG events using a single accelerometer on the shank [72]. They defined an optimal threshold as the value that showed minimal false negative and false positive results among all ten PD participants, and achieved 82.1% and 77.1% sensitivity and specificity, respectively.

Freeze index has been widely used in different studies aimed to detect FoG episodes. Compared with features extracted from wavelet coefficients (WT), freeze index requires less memory and computation which makes it easier to implement in portable devices, with less computational processing power, used for real-time FoG detection. Sample entropy (SampEn), a complexity measure developed by Richman *et al.* [73], on the other hand, has not been used in studies with the PD population. Yentes *et al.* used SampEn to investigate gait characteristics of younger and older groups [74]. Riva *et al.* also used SampEn to assess fall risk using trunk acceleration data of healthy participants. In this study, we used FI, WT and SampEn features from lower limb accelerations to validate our hypothesis about identifying FoG episodes before they actually occur.

Some wearable systems including motion sensors have been developed for the automatic detection of FoG, and the subsequent provision of cues during FoG episodes to help the patients overcome motor block. However, these systems must wait for the patient to freeze and then, if detected, take proper actions. Predicting FoG, on the other hand, enables preventive cueing and potentially reduces the likelihood of FoG occurrence [1]. In this work we go a step further and propose a methodology for the automated prediction of FoG events using wearable inertial sensors. The hypothesis of this study is based on the statistical analysis of feature vectors during normal gait and before FoG onset to identify the premonitory state, i.e. when the normal gait is about to transform into freezing. In cases where there is a statistically significant difference between the feature values of successive windows, FoG is about to occur. We selected various feature vectors on which we performed

statistical analyses in an effort to predict FoG. In addition to common sensitivity and specificity measures, we defined *predictivity* as an evaluation of the predictor performance using different combinations of sensor position, sensor axis, sampling window length and features.

3.3 Methods

3.3.1 Dataset

The data set used in our analyses was derived from the DAPHNet project dataset, which consists of sensor data extracted from seven male and three female advanced PD participants ($66. \pm 4.8$ years, Hoehn and Yahr score in ON 2.6 ± 0.65) [12]. Three tri-axial accelerometers were placed at the shank (just above the ankle), thigh (just above the knee), and lower back to collect acceleration in three axes of horizontal forward (perpendicular to the frontal plane), vertical (perpendicular to the transverse plane), and horizontal lateral (perpendicular to the sagittal plane). Data from these custom sensors were sampled at 64 Hz and transmitted to a wearable computer via Bluetooth.

Participants were tested in OFF medication state (at least 12 hours after taking anti-Parkinsonian medication), except two participants who had frequent FoG episode during the ON state. The participants were asked to perform three walking tasks (10–15 minutes each): 1) walking in a straight line in the laboratory hallway including 180° turns; 2) random walking including stops and 360° turns in a larger hall; and, 3) tasks simulating activities of daily living including walking through doorways to enter and leave rooms and the kitchen, fetching coffee, and walking with the cup of water in hand as a dual tasking activity. Participants walked at their own natural pace while a therapist walked closely for safety reasons. The experiments were video-taped and the files were used by physiotherapists to determine onset and duration of FoG episodes.

A total of five hundred minutes of data were collected, among which a total of 237 freezing events from eight participants (0–66 per subject, 23.7 ± 20.7 , ranging from 0.5 to 40.5 seconds long, 7.3 ± 6.7 s) were recognized by the physiotherapists. Fifty percent of the FoG episodes

lasted less than 5.4 seconds, and 93.2% were less than 20 seconds long. The labeled data were used as the ground truth in our FoG performance evaluations.

3.3.2 Feature Extraction

In this study, a battery of literature-based features were extracted from the lower body acceleration data to identify FoG incidents. We compared results obtained from three features: freeze index, wavelet mean, and sample entropy. In order to extract features, we used sliding windows of duration 2, 2.5, 3, 3.5, and 4 seconds, and a delay between windows of 0.5 second (Figure 3.1). Bächlin *et al.* also used sampling size of 4 seconds with the same step size in the original DAPHNet study to calculate FI of acceleration [12].

Freeze Index

Moore *et al.* discovered components in the 3–8 Hz band of power spectra from shank’s movement during FoG, which were not present during normal walking [9]. In order to discriminate FoG from normal walking, they defined a freeze index (FI) as the power in 3–8 Hz (“freeze” band) divided by the power in 0.5–3 Hz (“locomotor” band) of the Fast Fourier Transform (FFT) of the acceleration signals. They identified windows of data with FI values above a freeze threshold as FoG events. FI has been widely used to detect freezing events because it reflects the high-frequency components in the freeze band related to trembling of the lower limb that are not apparent when standing or walking (Figure 3.2). In this study, we used the same frequency bands as [9] to find FI for each segment of signals.

Wavelet Mean

wavelet transform is a standard tool that shows how the power amplitude of a specific frequency in a signal changes over time. This detailed time–frequency analysis enables localization of the transient states of a signal in time, e.g. from normal gait to freezing. Rezvanian and Lockhart used the sum of wavelet coefficients in freeze and locomotion bands over a sampled window of acceleration data to discriminate FoG from non-FoG events [72].

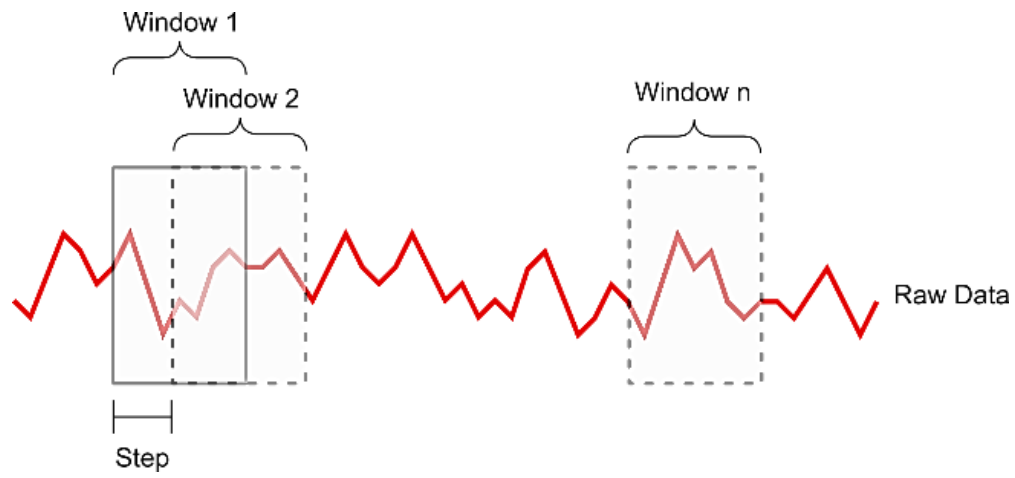


Figure 3.1: Windowing approach to segment acceleration data into windows of equal length.

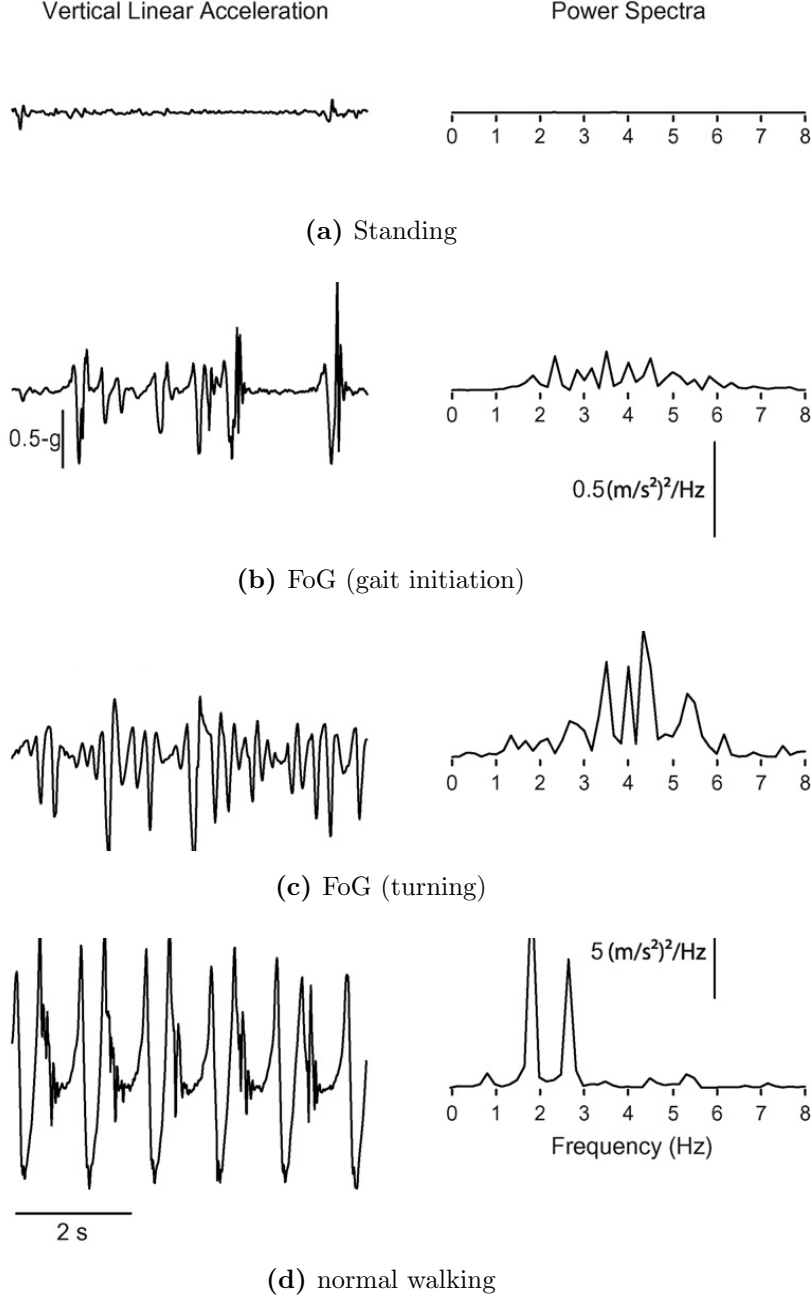


Figure 3.2: Vertical acceleration collected from a shank-worn sensor and corresponding power spectra in a participant with advanced PD. (a) Minimal activity was observed during volitional standing; (b) and (c) High-frequency components in the 2–6 Hz band were observed during FoG; (d) Highly tuned components at the stride (~1 Hz) and step (~2 Hz) frequencies (with higher frequency harmonics) were observed during normal walking [9].

Sneha Baby *et al.* also used mean of absolute values of the detail coefficients in each sub-band as a feature to train different classifiers to distinguish PD patients from controls [75]. In this study, we used mean of detail coefficients of Discrete Wavelet Transform (DWT) decomposition from level three (with mid-pseudo-frequency of $f_s=8$ Hz) which includes the freeze band discussed earlier. We chose the Daubechies wavelet of order four (db4) as the mother wavelet due to its simplicity and continuous first-order derivative [72, 76, 77].

Sample Entropy

Tripoliti *et al.* used entropy values of acceleration signals to train different classifiers to detect FoG events [8]. They suggested that because FoG is a symptom happening with a non-linear pattern, a non-linear feature such as entropy may more effectively distinguish FoG from normal gait. In this study, we used Sample Entropy (SampEn), a modification of entropy, which has been used to analyze physiological variability in human gait [78]. $\text{SampEn}(m, r, N)$ is a measure of repeatability or predictability within a time series and is derived from approaches developed by Richman and Moorman [73]. $\text{SampEn}(m, r, N)$ is the negative natural logarithm of the conditional probability that if two sequences of m points have distance $< r$, then two sequences of $m + 1$ points also have distance $< r$. Let's define $B_m(r)$ as the probability that two sequences will match for m points, and $A_m(r)$ as the probability that two sequences will match for $m + 1$ points:

$$B_m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \quad (3.1)$$

$$A_m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r) \quad (3.2)$$

where $B_i^m(r)$ is $\frac{1}{N-m-1}$ times the number of vectors $x_m(j)$ within r of $x_m(i)$, and $A_i^m(r)$ is $\frac{1}{N-m-1}$ times the number of vectors $x_{m+1}(j)$ within r of $x_{m+1}(i)$, where j ranges from 1 to $N - m$, and $j \neq i$ to exclude self-matches. Then:

$$\text{SampEn}(m, r, N) = -\ln \frac{A^m(r)}{B^m(r)} \quad (3.3)$$

In this study, we used suggestions of Tochigi *et al.* for entropy analysis of tri-axial leg acceleration signal waveforms, and set the parameters as: $m = 2$ and $r = 0.2 \times \sigma$, where σ is the standard deviation [78].

3.3.3 FoG Detection Algorithm

Various studies have used FI thresholding methods to identify FoG episodes [9, 11, 12, 79]. Using this method, when the FI value during a window of time becomes greater than a set threshold, it is identified as an FoG event. In this study we investigate the accuracy of FoG detection using different features over multiple windows of data. Figure 3.3 shows the acceleration and corresponding FI for three axes of the sensor located on the shank of one participant. It suggests that FI remains almost constant during normal walking, but starts to grow before FoG occurrence (labeled as “*pre-FoG*”). Thus, we hypothesized that changes in FI values over successive windows of time not only can reflect FoG occurrence, but also can be used to predict FoG before its actual occurrence. This section describes how we used feature values of consecutive windows to predict FoG episode before, or detect FoG onset after, the actual FoG occurrence.

In this method, we calculate feature values from the most recent $2n$ windows. We then refer to the first n windows as *Group1* and the second n as *Group2*. Figure 3.4 shows the acceleration and FI values corresponding to each window as well as Group1 and Group2 windows used at time T_1 and its successive time step, T_2 . In this study we used a set of three consecutive windows per group ($n=3$). We chose this period having in mind that it is a parameter that could be variable with the context where each FoG occurs. However, six windows were assumed to represent an appropriate period of time to identify gait deterioration before an FoG episode.

In order to reduce false positives (incorrectly identified FoG events), gait windows that may represent stops were marked as “no-FoG”. Stop periods were identified using a method similar to one described in [80]. Inactive periods regarding the activities unrelated to the experimental protocol are labeled in the DAPHNet dataset. Here, we found a stop threshold for each participant equal to the maximum amplitude of acceleration during one of the labeled inactive periods. If acceleration data remained within the threshold for 1 second, the

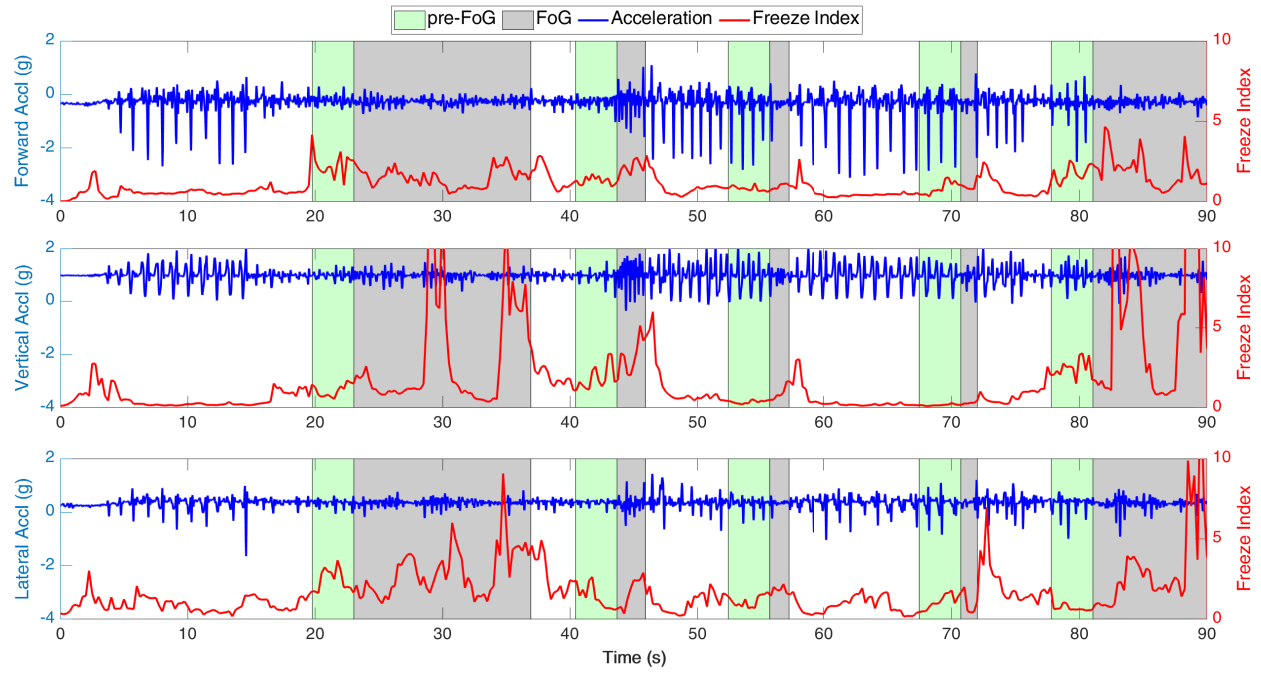


Figure 3.3: Acceleration data and corresponding freeze index for each axis of the shank sensor. Areas labeled as pre-FoG are three seconds before FoG onset.

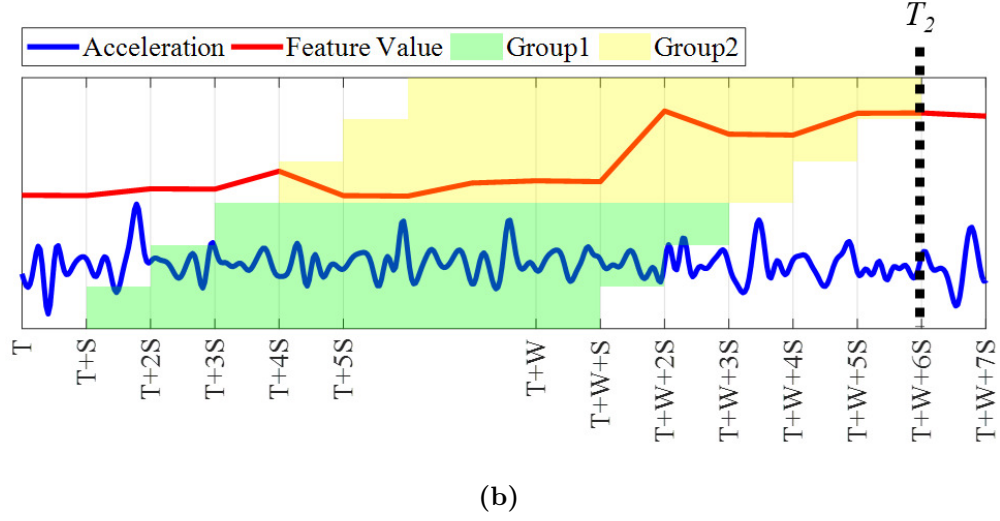
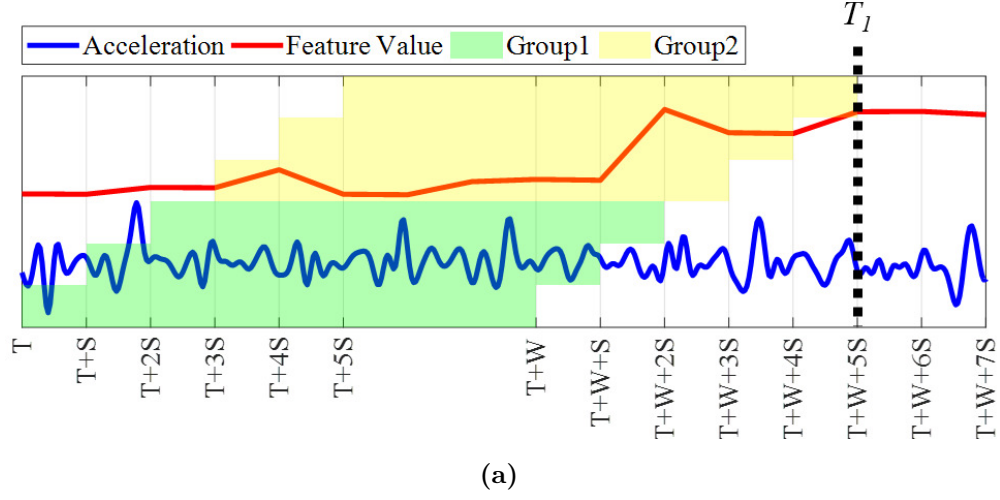


Figure 3.4: Grouping windows of data to predict/detect FoG (T is the starting point of any arbitrary window, W is the window length, and S is the step size). (a) Windows of data used to identify FoG/preFoG at point T_1 ; (b) Windows of data used to identify FoG/preFoG at the next window, point T_2 .

data were marked as “Stop.” Then, consecutive periods that were less than 1 second apart were merged together. This thresholding and merging method has been used by El-Gohary *et al.* to detect turns during gait in PD participants [81].

For each participant, sets of Group1 and Group2 windows were part of the statistical analysis. The starting point of a window (T_1 and T_2 in Figure 3.4) is marked as FoG onset if there is a statistically significant difference in the selected feature value between the Group1 and Group2 windows. The non-parametric Kruskal–Wallis test is chosen for statistical analysis, as the normally distributed assumption of parametric tests is violated when comparing groups of small samples. MATLAB (Release 2018a, The Math Works, Inc., Natick, MA, USA) was used for data preprocessing, feature extraction, and statistical analyses.

3.3.4 Annotation of Ground Truth

In this study our main objective was to predict FoG. This would potentially allow for applying preventing cues, reducing the likelihood of FoG occurrence [67, 82]. Figure 3.5a shows how the original labeled FoG signal (S_1) is post-processed to create the ground truth for assessment of the FoG detection algorithm. The FoG onset is shifted to left to the starting point of the last window with no FoG labeled sample (F_s). The end point (F_e) is the beginning of the first window with no FoG labeled sample (signal S_2). In the next step, only the FoG onset is kept and the other windows corresponding to FoG plus one window of data following the FoG end point are discarded (signal S_3). This part of data is discarded to make sure the comparison of post-FoG data with the following normal gait does not yield false positives. FoG events among the discarded data were also ignored as in Figure 3.5b.

T_p and T_d are the prediction and detection periods, respectively. The system was required to detect FoG onset in the range of 2 seconds before/after the actual FoG onset in order to report it as a correctly recognized predicted/detected freezing episode ($T_p = T_d = 2$). The same time range was used in [12] to calculate the accuracy of the FoG identification method. We considered 2 seconds to be the appropriate window length because it is long enough to capture data from two strides before the onset of freezing.

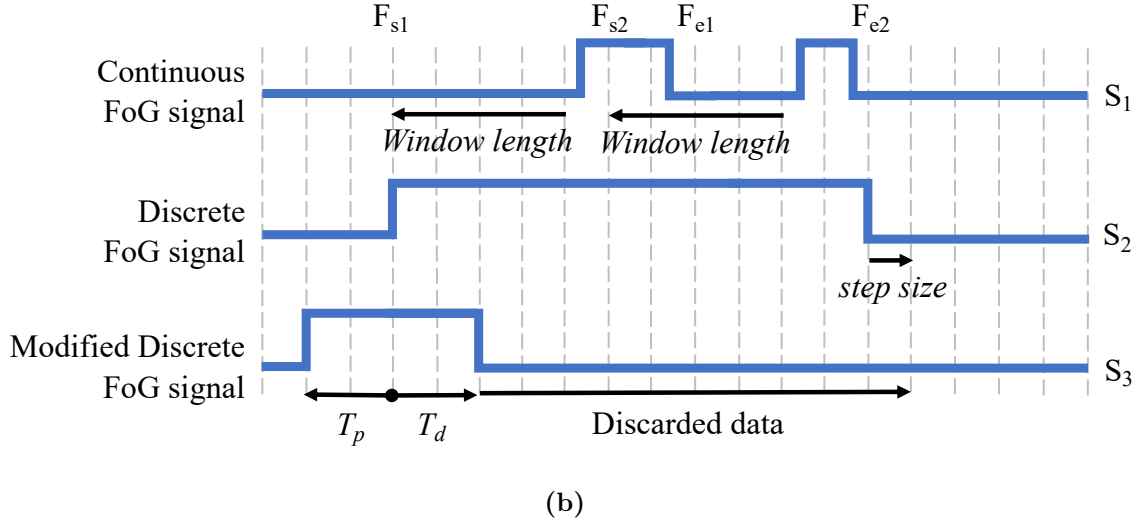
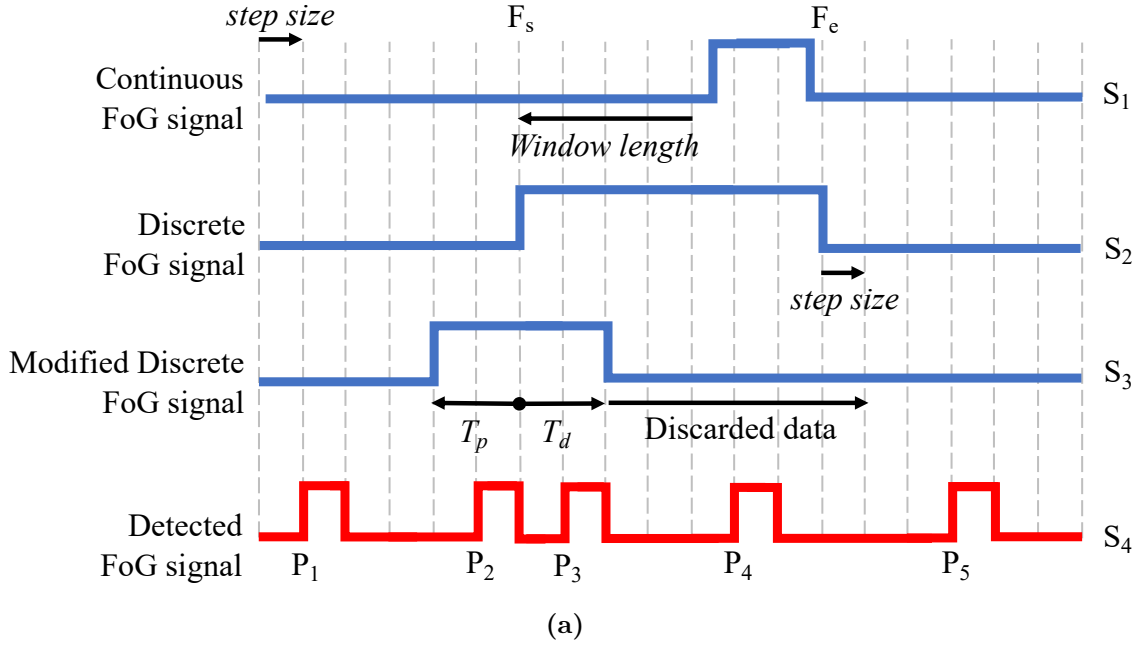


Figure 3.5: Post-processing of the labeled FoG signal to annotate the ground truth (F_s and F_e are the start and end of the FoG episode; T_p and T_d are prediction and detection duration; and P_i s are the identified FoG episodes).

In Figure 3.5a, signal S_4 is the detected FoG signal. P_1 is identified as a False Positive, while P_2 is a True Positive (a correctly predicted FoG). In this particular case, because the system has already predicted freezing, P_3 is ignored, otherwise it would be counted as a True Positive (correctly detected FoG). P_4 is discarded and P_5 is a False Positive.

3.3.5 Online vs. Offline FoG Detection

The proposed algorithm can be implemented in an online FoG prediction/detection system with minimal modifications. The flow chart of the online and offline FoG detection algorithms, including the parameters at each step, is depicted in Figure 3.6. The online FoG detection includes a cueing system to help patients overcome the freezing episodes [1, 83].

3.3.6 Performance Evaluation

In order to compare FoG detection performance between different combinations of sensor positions, sensor axes, window length, and feature, we calculated predictivity, sensitivity, specificity, and F-score for each case. We defined predictivity as the ratio of correctly predicted FoG events to all correctly identified events. In this case, the product of predictivity and sensitivity shows the ratio of correctly predicted events to all FoG events.

$$TP = TP_{pre} + TP_{det} \quad (3.4a)$$

$$Predictivity = 100 \times \frac{TP_{pre}}{TP} \quad (3.4b)$$

$$Sensitivity = 100 \times \frac{TP}{TP + FN} \quad (3.4c)$$

$$Specificity = 100 \times \frac{TN}{TN + FP} \quad (3.4d)$$

$$F - score = 100 \times \frac{2TP}{2TP + FP + FN} \quad (3.4e)$$

where TP_{pre} and TP_{det} are the number of correctly predicted and detected FoG events, respectively, TP is the total True Positives, TN is the total True Negatives, FP is the total False Positives and FN is the total False Negatives (Table 3.1).

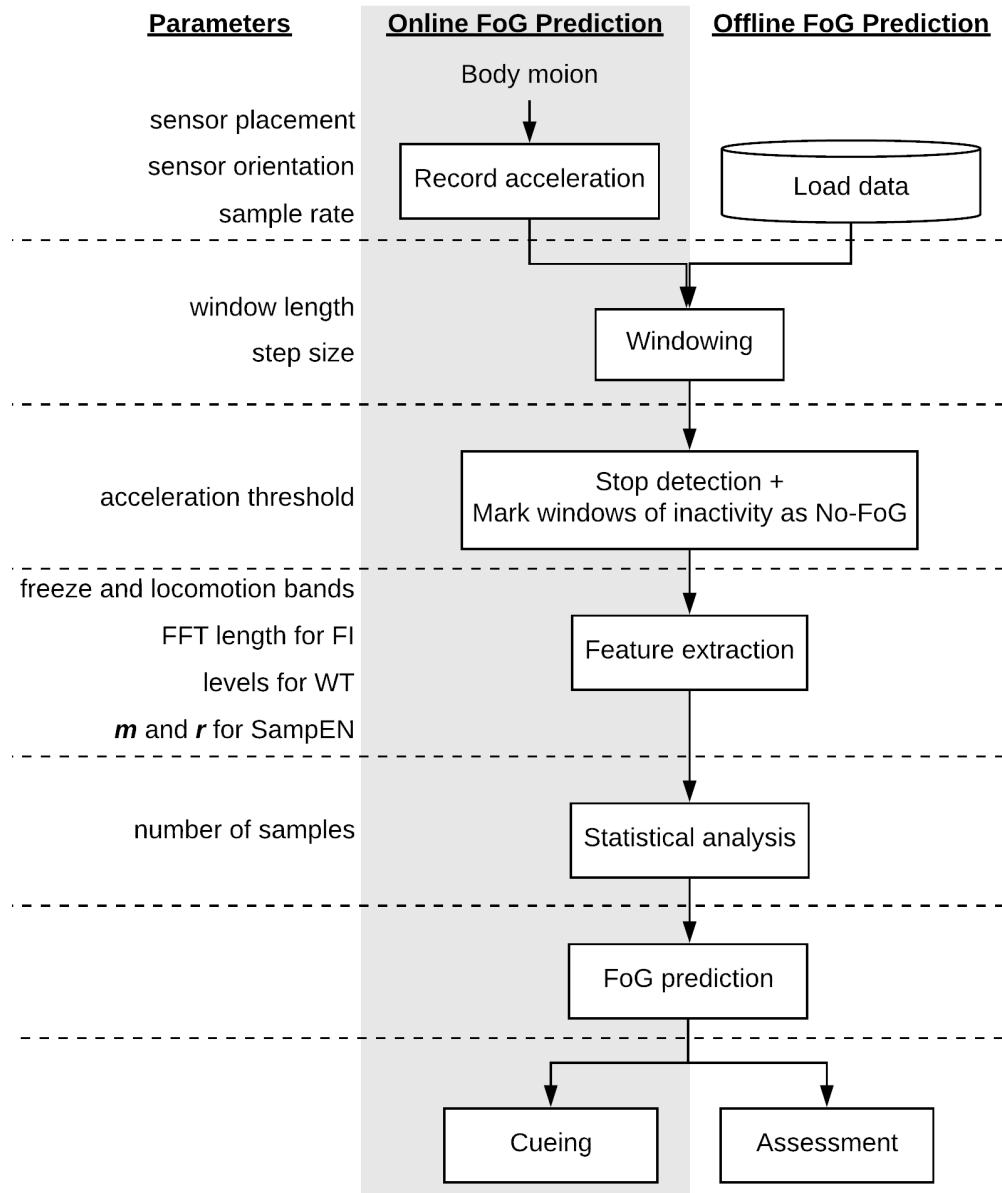


Figure 3.6: Flow chart comparing the online and offline FoG prediction algorithms including all parameters.

Table 3.1: Confusion matrix (NG: normal gait, FoG: freezing of gait).

		Predicted	
		NG	FoG
Actual	NG	True Negative (TN)	False Positive (FP)
	FoG	False Negative (FN)	True Positive (TP)

3.4 Results

We used 120 combinations of sensor locations (shank, thigh, waist, and majority voting), axes (F: perpendicular to frontal plane, V: perpendicular to transverse plane), window lengths, and features to find the best FoG prediction performance. The majority vote (MV) method identified FoG using all three accelerometers; at each window of data the combined result was set to “FoG” if at least two of the three sensors indicated a freeze event, and “NG” if only one or none of the sensors identified FoG. The preliminary results using the acceleration in the lateral axis were not comparable to the other axes, thus were excluded. Figure 3.7 shows the FoG identification results for all combinations of sensors, axes, window lengths, and features, averaged across participants. In the following, we refer to each combination using the array of `[sensor,axis>window.length,feature]`.

3.4.1 Predictivity

`[Thigh,V,4,FI]` and `[Shank,V,3,FI]` showed the highest rate of predictivity ($96.7 \pm 5.6\%$ and $95.7 \pm 3.9\%$, respectively). Additionally, using `[Waist,F,2.5,FI]`, $95.4 \pm 5.1\%$ of the recognized FoG events were predicted prior to occurring.

3.4.2 Sensitivity

`[Thigh,F,4,SampEn]` and `[Thigh,V,[2,2.5,3],FI]` showed the highest sensitivity ($99.7 \pm 0.7\%$ and $99.5 \pm 1.5\%$, respectively). The results also suggest that WT is the least sensitive feature to FoG events.

3.4.3 Specificity

The specificity of the algorithm to FoG events was inversely related to sampling window length. With a small window (2 or 2.5 seconds) most of the sensor/feature combinations exhibited specificity above 85%. Among all combinations, using majority voting of both horizontal forward and vertical axes showed highest specificity, with SampEn and window

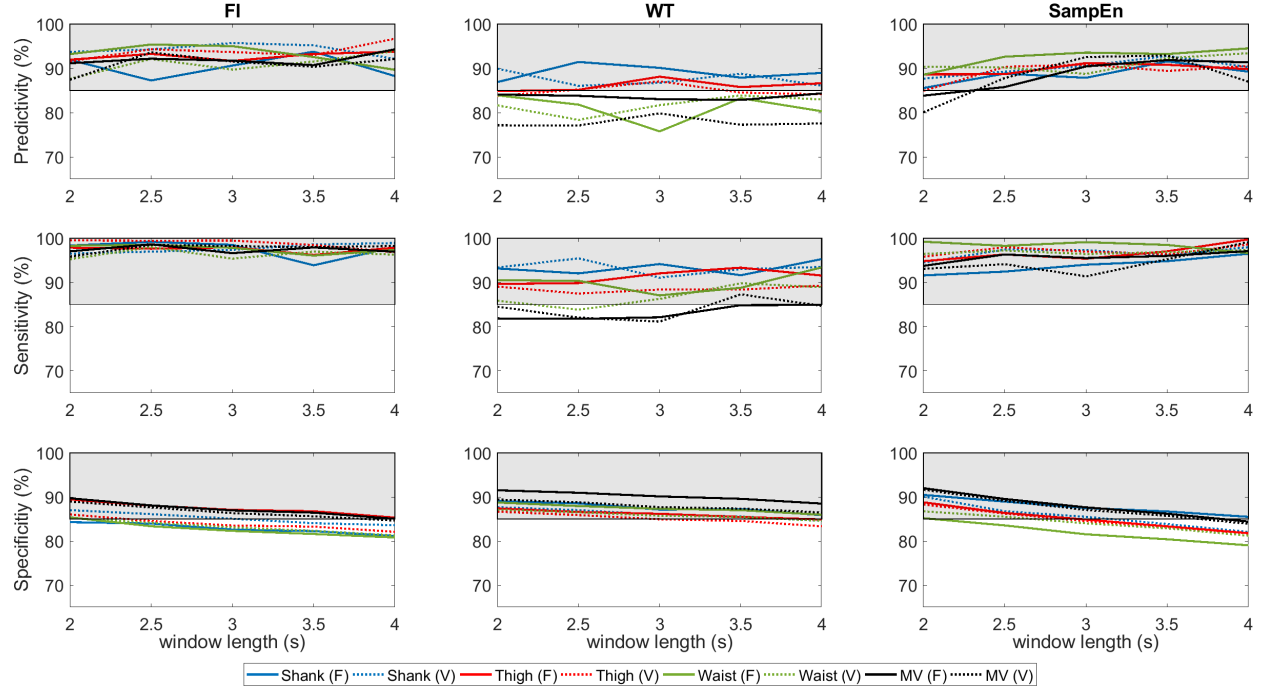


Figure 3.7: Predictivity, sensitivity, and specificity versus window length for the three selected features (MV refers to majority voting as using all the three sensors; (F) and (V) refer to forward and vertical axes of the sensors).

length of 2 seconds ($92.0 \pm 2.5\%$ and $91.6 \pm 2.1\%$ using [MV, F, 2, SampEn] and [MV, V, 2, SampEn], respectively).

The waist sensor was very sensitive to FoG using FI (96.0–98.5%) and SampEn (96.6–99.2%) with a robust predictivity (88.6–95.4%), but had the lowest specificity (<85%) among the selected combinations. Thus, waist is not a suitable sensor position for FoG detection applications.

All cases with specificity higher than 85% are listed in Table 3.2. Here, we selected 85% as a performance criterion which seems a relatively fair threshold compared with other studies.

3.5 Discussion

Using wearable inertial sensors, we found evidence that the pre-FoG periods can be separated from normal walking in patients with PD who suffer from freezing of gait. We tested the features that are usually used for FoG detection (i.e., after the FoG onset) to see how they can be used to predict FoG before its occurrence. At least four independent variables must be considered: the location and number of sensors, the sensor axes, the length of the sampling window, and the extracted feature. These choices represent tradeoffs between predictivity, sensitivity, and specificity; the longer the sampling window, the higher the predictivity and sensitivity at the expense of lower specificity (more false positives).

The best results were achieved when using majority voting (at least two sensors identifying FoG) from horizontal forward axis and FI feature with window length of 2 seconds. Two comparably performing combinations include the forward axis of the thigh and shank accelerometers, respectively, which suggests that the dysrhythmicity associated with FoG can be detected from forward motion of the thigh and shank. The performance for the other sensor positions, axes, window lengths and features are nevertheless quite good, suggesting that sensors can be placed at a variety of body locations without significant loss of performance.

As a recommendation, a single shank sensor makes for an ideal location as the accelerometer can be easily placed in the insole or on the lateral heel of a shoe to detect gait abnormalities [84, 85]. Among all combinations of the shank sensor listed in Table 3.2,

Table 3.2: Highest performing ($> 85\%$ specificity) combinations of sensor position, axis, window length, and feature sorted by descending specificity.

Sensor (axis)	Window Length [s]	Feature	Predictivity [%]	Sensitivity [%]	Specificity [%]
MV (F)	2.0	SampEn	83.8 ± 6.1	93.8 ± 4.6	92.0 ± 2.5
MV (F)	2.0	FI	91.2 ± 5.0	97.0 ± 4.4	89.7 ± 2.4
MV (F)	2.5	SampEn	85.8 ± 3.7	96.3 ± 4.3	89.6 ± 2.8
Thigh (F)	2.0	FI	92.0 ± 6.2	97.9 ± 2.0	89.5 ± 2.7
MV (V)	2.5	SampEn	87.8 ± 6.1	94.1 ± 4.5	89.1 ± 2.2
Shank (F)	2.5	SampEn	88.8 ± 7.5	92.5 ± 5.8	89.0 ± 4.1
Shank (F)	2.0	WT	86.9 ± 8.3	93.4 ± 3.0	89.0 ± 13.3
Thigh (F)	2.0	SampEn	88.7 ± 6.7	94.8 ± 3.4	88.8 ± 3.0
Shank (F)	2.5	WT	91.4 ± 5.5	92.1 ± 8.8	88.6 ± 3.2
Thigh (V)	2.0	SampEn	84.9 ± 12.9	95.8 ± 3.8	88.3 ± 2.9
MV (F)	2.5	FI	92.2 ± 4.0	98.6 ± 2.1	88.1 ± 3.0
Thigh (F)	2.5	FI	93.3 ± 4.7	97.6 ± 2.1	88.0 ± 3.2
Shank (V)	2.0	WT	89.9 ± 7.8	93.4 ± 3.2	87.7 ± 3.1
MV (V)	2.5	FI	93.6 ± 5.2	98.5 ± 1.8	87.6 ± 2.5
MV (F)	3.0	SampEn	90.4 ± 9.6	95.5 ± 5.0	87.6 ± 3.2
Shank (F)	3.0	WT	90.1 ± 6.4	94.2 ± 8.0	87.1 ± 3.6
Shank (V)	2.5	WT	86.1 ± 7.9	95.5 ± 5.3	87.0 ± 4.0
Shank (V)	2.5	SampEn	89.3 ± 6.7	97.4 ± 3.2	86.8 ± 3.3
Waist (V)	2.0	SampEn	90.4 ± 10.2	96.4 ± 3.5	86.8 ± 3.4
Waist (V)	2.5	SampEn	90.2 ± 6.9	97.2 ± 4.1	85.5 ± 3.8

[Shank,F,2.5,SampEn] demonstrated the highest mean and lowest standard deviation (less patient-dependency) in performance. Our results for the forward axis of the shank sensor (sensitivity of $92.5 \pm 5.8\%$, specificity of $89.0 \pm 4.1\%$) demonstrated improved performance as compared to [12] (sensitivity of 88.6%, specificity of 92.4%), which used the same dataset and optimized patient-dependent freeze index thresholds to detect FoG. Their method also included FoG prediction using the same time threshold (2 seconds before actual FoG occurrence) but no results were reported for predictivity. Moore *et al.* also used individually calibrated freeze thresholds and obtained sensitivity and specificity of 89% and 90%, respectively [9]. We believe that post-processing of the results can help reduce the false positive incidents and improve the specificity of the proposed FoG detection method.

Rezvanian and Lockhart computed summation of wavelet components in locomotor and freeze bands and defined an FoG index as the ratio of the two values [72]. They achieved 82.1% and 77.1% for the sensitivity and specificity, respectively, using a single shank sensor. Tripolitia *et al.* calculated entropy of signals from six accelerometers (on ankles, wrists, chest and waist) and two gyroscopes (chest and waist) and tested four classification algorithms to detect FoG [8]. Using Random Forests classification algorithm, FoG events were detected with 81.94% sensitivity and 98.74% specificity. These results show higher specificity and lower sensitivity compared to [MV,F,2,SampEn] combination in this study. However, the results for a single ankle sensor cannot be compared as was not reported in their study.

Pham *et al.* used anomaly score detector technique to identify patterns of gait that are not similar to previous windows of data. Their method achieved $94 \pm 23\%$ and $84 \pm 36\%$ sensitivity and specificity, respectively, using vertical acceleration of a sensor located at the ankle [86]. Their method resulted in higher sensitivity and lower specificity than that of this study. However, the standard deviation of both measures is an order of magnitude higher, which shows higher degrees of performance variability compared to our algorithm.

Palmerini *et al.* trained a linear discriminant analysis classifier to discriminate pre-FoG episodes from normal gait using a wearable multisensory setup [67]. They divided the data into 2-second windows of normal gait and pre-FoG, and discarded the windows overlapping with FoG events. This classifier was able to identify 83% of the pre-FoG episodes, i.e. 83% of the FoG event were predicted. Our selected combination of [Shank,F,2.5,SampEn] was

able to correctly predict 82.1% (product of predictivity and sensitivity) of total events, while [MV,F,2,SampEn] showed even higher rate of 88.5%.

The limitations of this study were the short data length for each participant (24.6 ± 6.6 minutes), the limited number of participants (10), the number who experienced FoG during the test (8), as well as lacking actual daily life environments. Furthermore, different stimuli trigger motor fluctuations in patients with PD (e.g., turns, walking in narrow spaces, and stops) [33, 87, 88]. These were included in the DAPHNet experiment but were not labeled [12]. Thus, we planned to conduct a similar experiment including activities of daily living and different real-world stimuli to examine the applicability of this technique for home-based rescue strategies and therapies.

For implementing the proposed method in an online FoG detection system, the complexity of the feature extraction method and computation time may be another element of interest. Future research needs to address the latency inherent to the feature extraction algorithms and statistical analyses with corresponding window size.

3.6 Conclusion

This work indicated prediction of FoG before its occurrence using freeze index, wavelet transform, and sample entropy features of lower limb acceleration is attainable by statistical analysis of successive windows of data. The analysis of 120 different combinations of sensor positions, sensor axis, sampling window length, and feature showed that all combinations of freeze index and sample entropy features with 2–2.5 seconds window length could be used for FoG prediction, although there were minor differences in performance. Since a single shank sensor interferes less with walking and can be integrated easily into a shoe, it is especially promising for real-time FoG prediction. A sensor at the waist could also be integrated into a belt. However, the waist sensor showed low specificity and, therefore, is less preferable for FoG cueing applications. The results suggested that the performance, especially the specificity, increased by using majority voting.

Chapter 4

Addressing the Class Imbalance Problem in FoG Identification

4.1 Introduction

The main issue regarding data-driven modeling of FoG detection may be raised by learning from an inherent imbalanced dataset, i.e., the duration of captured freezing of gait is almost always a small percentage of the entire dataset. Most of the conventional classification techniques assume balanced data distribution in data classes, which may lead to bias towards the majority class. To tackle the imbalance in FoG and normal gait instances, ensemble classifiers have been used to improve FoG detection [8, 23]. However, no study has been done on data oversampling methods that create artificial new sample of the minority class to improve classification performance. In this chapter we employed one of the most common data synthesis methods to create a balanced training dataset for different classifiers and compared the results with those of classifiers trained with the original dataset.

Previous studies achieved high classification accuracy for FoG detection, especially with patient-dependent models which reduces the effect of heterogeneity in data from different participants. However, these studies seldom reported the detection latency (i.e., time associated with classification after FoG onset) and prediction capability (i.e., time associated with classification prior to FoG onset). In this chapter we also investigated the ability of

the models to predict FoG events using wearable inertial sensor data, which ultimately may help patients prevent FoG through external preemptive cueing.

4.2 Background and Approach

Different methodologies have been suggested to face the imbalance challenge. Synthetic Minority Oversampling TEchnique (SMOTE) is a method that attempts to address the class-imbalance in the dataset by generating synthetic instances [89]. Using Euclidean distance to measure vicinity in feature space, the algorithm synthesizes artificial data among the minority data class. This oversampling method has also shown effectiveness when applied to data with low-dimensional feature sets [90]. The vanilla SMOTE algorithm randomly synthesizes data in the minority class without considering neighboring examples from other classes, which causes deviance in data variance and over-generalization. Different variants of the SMOTE algorithm have been proposed to combat this limitation. Among those, we have selected the ADaptive SYNthetic (ADASYN) sampling algorithm [91]. ADASYN has demonstrated performance improvements in various applications, including classification of physiological data. For instance, analysis of electroencephalogram (EEG) signals is a noninvasive brain imaging technique capable of automatically detecting various infrequent conditions. Applying ADASYN to augment the rare-class instances has been a promising alternative solution to address the possible biased performance of rare condition detectors such as automatic seizure detection [92], emotion detection [93], classifying sleep stages [94, 95], and driver behaviors [96]. This method has also been successfully evaluated on other modalities including uterine electromyogram (EMG) or electrohysterogram (EHG) to determine preterm birth and reduce the risk of preterm delivery [97, 98].

Recently, ADASYN has also been applied in the pre-processing pipeline for recognizing individuals with PD using variants of physiological signal modalities. Wu *et al.* investigated the capacity of voice-extracted features to manifest PD symptoms [99]. Mel-spectrogram as a nonlinear time-frequency representation of acoustics was consumed to train a convolutional neural network (CNN) model to automatically extract discriminative features of PD and healthy control participants. ADASYN was employed to augment PD data samples in the

CNN-synthesized feature space. Taleb *et al.* developed a clinical decision support system for early detection and monitoring the progression of PD stages using ADASYN [100]. Participants' handwriting portrayed the motor symptoms of PD in different stages, and oversampling with ADASYN demonstrated the best accuracy of PD diagnosis and prediction of PD stages when compared to other methods. In short, ADASYN has been successfully applied to enhance the detection of rare phenomena, particularly to detect people with PD impairments, but not yet to the detection of FoG onset.

Existing literature reported wide ranges of FoG detection and prediction accuracy for participants using primarily patient-dependent approaches. This is partly due to the fact that participants reacted differently to the FoG stimuli included in the experiments, which caused wide ranges of FoG to non-FoG data ratios. Oftentimes, high levels of class imbalance in data set aggravated model performance. Therefore, new techniques are needed to address this issue, particularly when patient-dependent models are to be developed and implemented in cueing intervention devices. This chapter compares the performance of different classifiers in detecting FoG using an imbalanced binary training dataset in tandem with ADASYN synthesized training sets. We also investigate how different data oversampling rates may jointly affect the performance of patient-dependent and patient-independent models. We applied three common approaches: using ensemble classifiers [101], adding new synthetic FoG samples to the training dataset to improve balance [91], and increasing misclassification cost for the minority class, i.e., FoG [102]. The selected classifier must have high performance in discriminating FoG from normal gait within an appropriate time period before or after FoG occurrence. The results of the study would serve to develop a therapeutic intervention tailored to the patients, a real-time FoG detection system to identify FoG episodes and provide stimuli to the patients to prevent and overcome freezing.

4.3 Methods

4.3.1 Experiment Protocol

Schaafsma *et al.* described five subtypes of freezing triggered by gait initiation, turning, walking through tight hallways, terminating gait, and open space hesitation [39]. In order to trigger FoG symptoms, our experiment protocol included turns within wide and narrow turning areas at either end of a hallway, varied path widths using shelves in the middle of the hallway (Object Area), expected stops before turns and unexpected stops instructed vocally when walking straight (Figure 4.1). The participants walked at a self-selected speed five times across the 38ft. long $\times$ 5ft. wide walkway. Acceleration data in two axes of horizontal forward (perpendicular to the frontal plane) and vertical (perpendicular to the transverse plane) were captured using two accelerometers (APDM Inc. Opal Sensors, www.apdm.com) placed on left and right ankles superior to the tibia/talus joint. Data were sampled at 128 Hz and stored on the sensors' internal memory. As the algorithm will be implemented in an online cueing device, we tested the feature extraction methods on an Android phone using a researcher developed application. The results suggested that, due to the limited computation power of cellphones compared with PCs, the data should be down-sampled to reduce the computation time for feature extraction. Thus, the sampling rate of the collected data was then reduced to 64 Hz using linear resampling.

4.3.2 Dataset

Eighteen participants with PD (12M/6F, 70.0 ± 8.7 years, Hoehn and Yahr score between 2 and 4) walked in a narrow hallway for 29.1 ± 8.2 minutes (524 minutes in total). 196 FoG episodes over 16.9 minutes (5.1 ± 5.5 seconds) were captured from nine participants. Figure 4.2 shows the distribution of FoG duration for all participants. 64% and 26% of the events took less than 5 and 10 seconds, respectively.

An experienced clinician determined FoG episodes and labeled acceleration signals using captured videos. The rest, included normal gait, stops and no activity periods. As for the pre-processing of the data, first, we excluded the data associated with stops and no activity

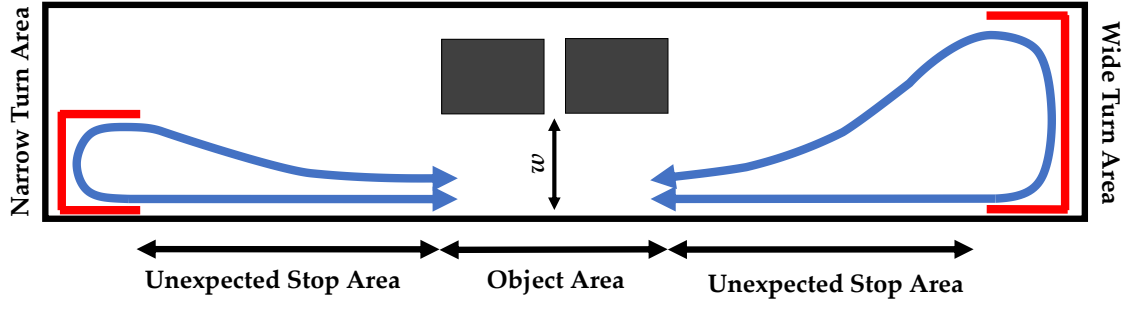


Figure 4.1: Experiment setting. Number of obstacles in the object area varied between 0, 1, and 2. The width of the walking path in the object area (w) varied between 150% and 100% of participant shoulder width.

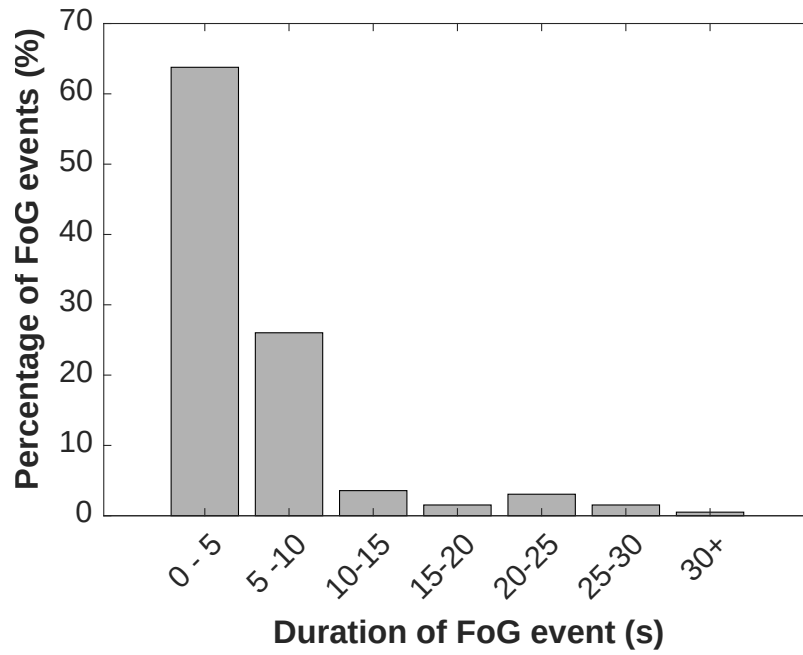


Figure 4.2: Distribution of FoG events duration.

periods to create a binary class dataset of normal gait and FoG. Data from one second pre-FoG periods were also included in the FoG class in order to enable the system to predict FoG before its actual occurrence.

In order to extract quantitative features from continuous sensor data, we used sliding windows of 2 seconds with 75% overlap to extract four features from acceleration data (Table 4.1). These four features indicated best performance on classification of gait using the collected data from the participants.

In Chapter 3 we showed that FoG is a dynamic process and using features from multiple successive windows can improve FoG detection. In this study, sample s_T in the dataset was formed by feature values from the current window ($t=T$) and five previous windows. Figure 4.3a demonstrates how an array of 24 elements is formed for vertical acceleration of left ankle (four features extracted from each of the six successive windows). For each sensor-axis combination, the array of 24 elements is created and concatenated to form a sample with 96 elements at time T (Figure 4.3b). For the next time step ($t=T+1$), the six windows are shifted forward one step (0.5 seconds) and the same process is followed to form sample s_{T+1} . The sample is labeled as FoG if at least one FoG labeled timestamp is included in the window nearest to the current time (the right most window). In Figure 4.3a, sample s_T is an example of normal gait, while s_{T+1} is an instance of FoG. Finally, the dataset was formed by samples from all participants and normalized so each feature set had zero mean and unit standard deviation. FoG labeled samples formed only 13.0% of the original dataset and the rest belonged to normal gait.

4.3.3 Data Synthesis

ADASYN uses a density distribution function Γ to determine the number of synthetic samples that need to be generated for each sample in the minority class. In order to create synthetic data sample for the minority class, first the number of synthetic data samples must be specified [103]:

$$G = (|S_{maj}| - |S_{min}|) \times \beta \quad (4.1)$$

Table 4.1: Feature set extracted from each window of data.

Feature	Description
Freeze Index	The power in freeze band (3–8 Hz) divided by the power in locomotion band (0.5–3 Hz) using FFT of the acceleration signal [9]
Sample Entropy	The negative logarithm of the ratio of conditional probability of data subsets of length m matching point-wise within a tolerance r , to conditional probability of data subsets of length $m + 1$ being within the same tolerance [78]
Power	Total power in the freeze and locomotion bands (0.5–8 Hz) of the signal
Standard Deviation	Mean deviation of data points compared to the average

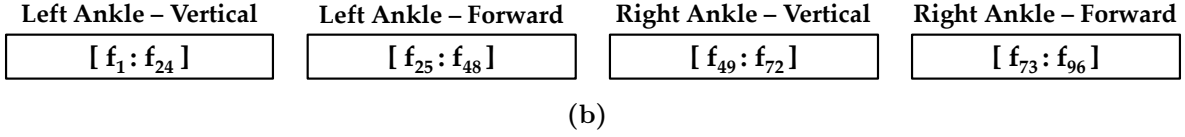
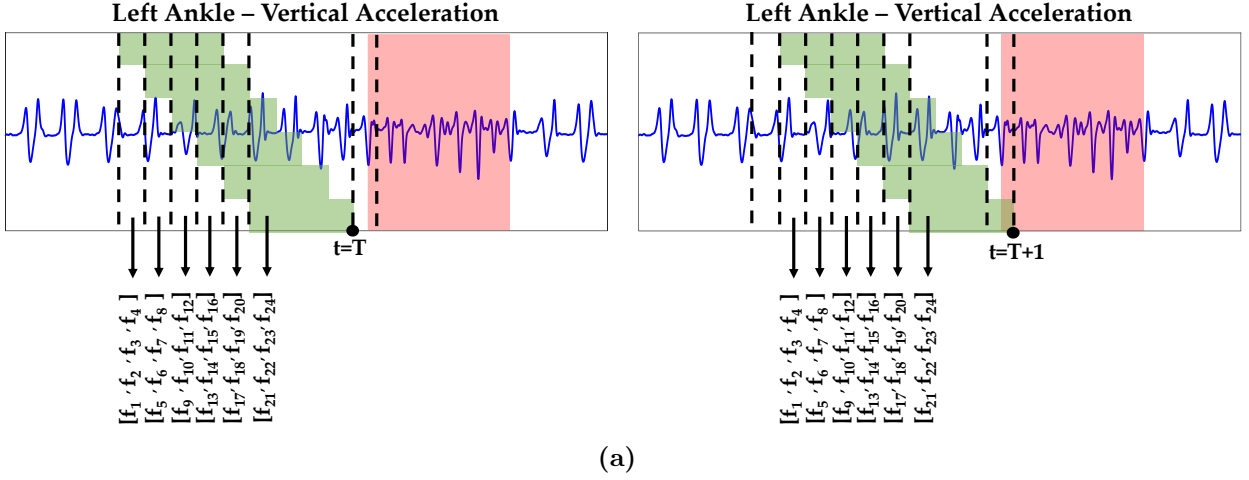


Figure 4.3: The process of creating samples from acceleration signals. (a) Extracting features from six successive windows at time $t=T$ (left) and the next time step, $t=T + 1$ (right). The red highlighted area shows the FoG labeled period using recorded videos and the green boxes show segments of data used to extract features from acceleration signal. (b) Aggregating arrays of features from different combinations of sensor-axis.

where $|S_{maj}|$ and $|S_{min}|$ are the number of samples in the majority and minority groups, respectively, and $\beta \in [0, 1]$ specifies the desired balance level after the synthetic data generation process. $\beta = 0$ means no new data will be created, and $\beta = 1$ creates a balanced dataset where both classes form 50% of the entire dataset.

Next, for each sample $x_i \in S_{min}$, the ratio of new synthetic data samples Γ_i in the vicinity of each sample x_i is calculated:

$$\Gamma_i = \frac{\Delta_i}{K_1} \quad (4.2)$$

where K_1 is the K-nearest neighbors according to Euclidean distance, and Δ_i is the number of samples from S_{maj} in the K-nearest neighbors of x_i .

Then, the number of synthetic data samples to be generated for each $x_i \in S_{min}$ is determined as:

$$g_i = \frac{\Gamma_i \times G}{Z} \quad (4.3)$$

where $Z = \sum_{i=1}^{|S_{min}|} \Gamma_i$.

Finally, new data samples for the minority class will be generated in the vicinity of $x_i \in S_{min}$:

$$x_{new} = x_i + (\hat{x}_i - x_i) \times \delta \quad (4.4)$$

where $\hat{x}_i$ is one of the K-nearest neighbors (K_2) for $x_i \in S_{min}$, and $\delta \in (0, 1]$ is a random number.

Figure 4.4 shows the t-distributed Stochastic Neighbor Embedding (t-SNE) visualization of the original data and synthesized data from all participants in this study. The algorithm projects original and synthesized data samples onto a 2-D map [104]. Unlike linear mapping methods such as principal component analysis (PCA) that uses the same linear mapping to all data samples, t-SNE learns embedded data characteristics in high dimensional feature space and preserve the distribution of data similarities in the 2D map. The similarities are measured with Euclidean distance, i.e., the same metric for generating synthesized data in ADASYN. The figure shows that FoG samples are scattered in the space, thus using SMOTE to generate random new instances for FoG class may not improve the FoG detection. On

the contrary, ADASYN creates samples on the boundaries which may potentially suggest a higher classification accuracy.

The value of β was changed from zero to one with steps of 0.2 to investigate the effect of dataset imbalance on classifiers performance. For the original dataset, all classifiers were trained and tested using 60% and 40% of all data, respectively. For the ADASYN synthesized dataset, the same training data were used to generate new synthetic samples which were then added to the training set. The same testing set was used to evaluate model performance, i.e. no synthetic samples was included in the testing set.

4.3.4 Classifiers

We compared the performance of three classifiers on identification of FoG events: Support Vector Machine (SVM), k-Nearest Neighbors (kNN) and Multilayer Perceptron (MLP):

- **SVM:** A C-SVC classifier using radial basis kernel function with $\gamma = 1/N$ was trained and tested, where N is the number of features (equal to 96).
- **kNN:** $k = 3$ showed the best performance on the original dataset and we used the same parameter to train the classifier using synthesized dataset. We used the Euclidean distance metric to find the nearest neighbors.
- **MLP:** We trained a network with two hidden layers with 20 and 15 nodes over 50 training epochs. These parameters showed best performance using the original dataset.

Ensemble analysis is a method which is commonly used in many data mining problems such as classification [105], clustering [106], and anomaly detection [107] in order to reduce the dependence of the model on the specific data set or data locality. This greatly increases the robustness of the data mining process. In this study, we also used boosting and bagging methods to train ensemble classifiers to distinguish between normal gait and freezing episodes (*ClsfBoosting* and *ClsfBagging*, respectively). Finally, we trained ensemble classifiers of decision trees: *AdaBoost*, *TreeBaggers*, and *RandomForest*.

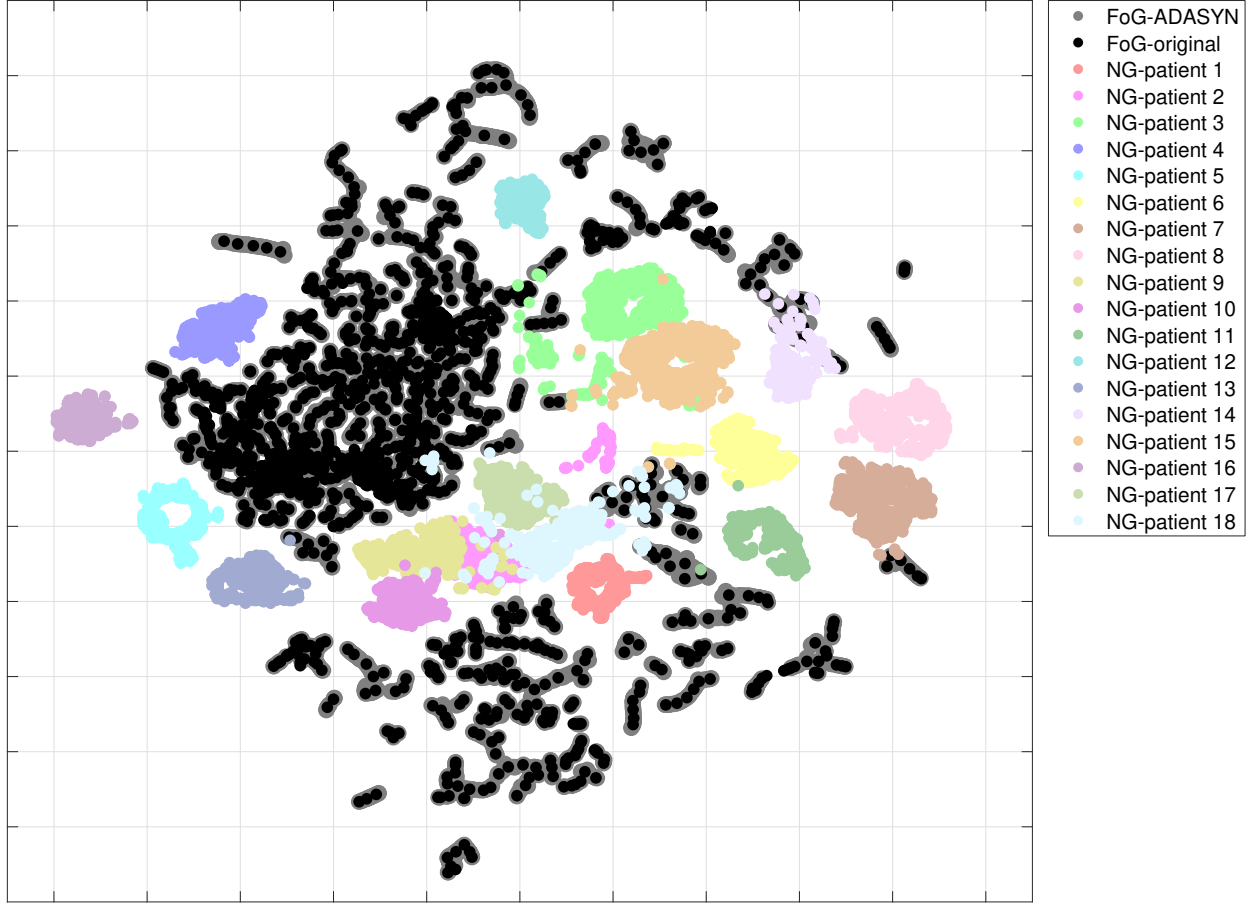


Figure 4.4: t-SNE visualization of the original dataset and new synthesized data using ADASYN ($\beta = 1$). *FoG-ADASYN*: new synthesized samples for FoG class, *FoG-original*: FoG samples from all participants, *NG*: normal gait.

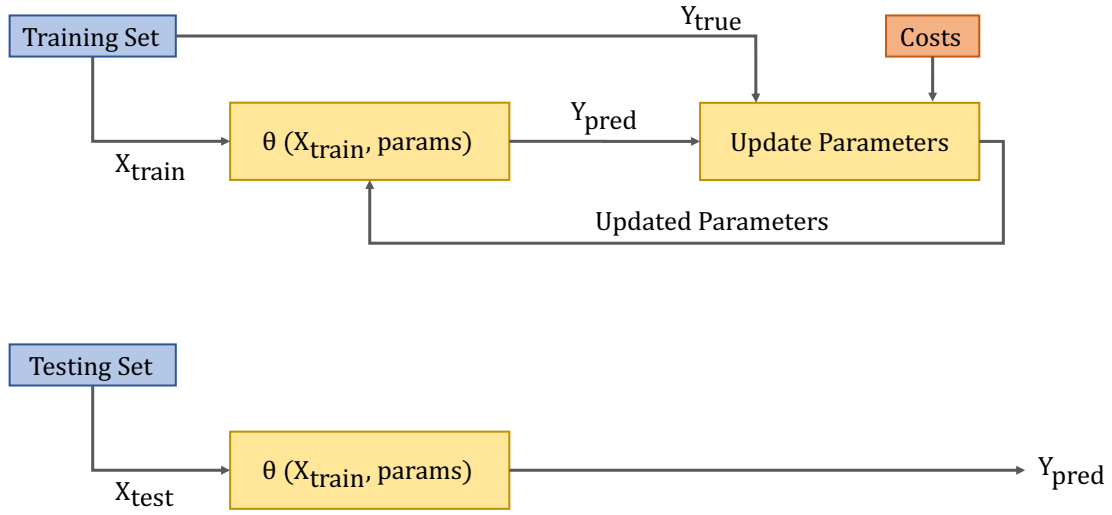
4.3.5 Cost Sensitive Classification

We investigated the misclassification cost, C , which quantifies the importance of incorrectly classifying samples from one class as the other. Cost-sensitive classification incorporates fixed and unequal misclassification costs between classes in decision-making and can be performed in two different ways: cost sensitive training and cost dependent classification. The first approach is to train a model with a parameter updating function that incorporates the costs (Figure 4.5a). In this case we need to provide the function with the costs associated with each of the four cases (True Negatives, False Negatives, False Positives, and True Positives) so that the model can learn to make optimal predictions accordingly. The second approach trains a regular model, but incorporates the costs when calculating the adjusted probabilities to make predictions for each sample (Figure 4.5b). This approach only works for models that predict a probability which can then be used to calculate the adjusted probabilities. Thus, we used cost sensitive training approach for training SVM and kNN models, and cost dependent classification for MLP models.

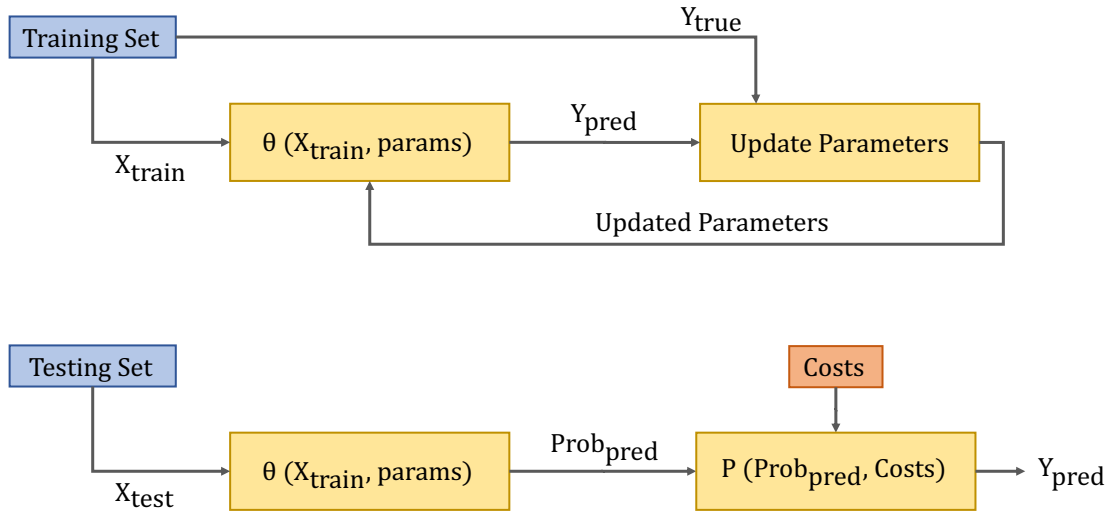
We set the cost associated with True Negative and True Positive cases equal to zero, meaning correct classification has no cost. However, as it is of essential importance to detect as many FoG episodes as possible, the misclassification cost of FoG class (False Negative cases) was set equal or greater than one ($C_{FoG} = 1, 2, 3$), while keeping the cost for normal gait class (False Positive cases) equal to one in all cases ($C_{NG} = 1$). $C_{FoG} > 3$ showed poor performance in preliminary results and was excluded from further analysis.

4.3.6 Patient-Dependent and Patient-Independent Settings

Although the FoG instances formed only 13% of the original dataset, using $\beta = 1$ to create new synthesized samples made a balanced dataset in which both normal gait and FoG classes formed 50% of the entire dataset. Figure 4.4 shows that the normal gait data corresponding to each participant formed a cluster that is separated from other participants. This could be due to the large variability in motor performance when performing the required tasks (i.e. walking, turning, initiation and termination), which may impair the generalizability of models to discriminate FoG from normal gait when tested on new participants. On the other



(a) Cost sensitive training



(b) Cost dependent classification

Figure 4.5: Cost sensitive classification approaches.

hand, the FoG samples from the entire population are dispersed and cannot be grouped in such distinguishable clusters as normal gait. It is partly due to the fact that FoG episodes happened during different tasks (i.e. walking in a straight line, turning and gait initiation) and participants had their own unique strategy to overcome freezing during any of these tasks. In order to investigate generalizability, we characterized the classifier performance on patient-dependent and patient-independent models using data from seven participants who froze more frequently during the experiments (22 ± 12.6 events). The other two participants who experienced FoG during the experiments, froze only once and, as the episodes were relatively short, were excluded from further analysis. In patient-dependent models, both training and testing data were from the same participant, while patient-independent models utilized leave-one-out cross validation (LOOCV).

In the next section, we presented the results for the three performance measures of selected classifiers for patient-dependent and -independent models. The sensitivity and specificity demonstrate the accuracy of the models in classifying FoG and normal gait classes, respectively. F-score, on the other hand, shows the overall performance of the models in classifying both classes and is used as the main performance measure to select the model with highest performance. Finally, we investigated FoG detection latency and prediction capability for the classifier with highest F-score in patient-dependent experiments.

4.4 Results

4.4.1 ADASYN Synthesized Datasets

Figure 4.6 shows results of SVM classifier trained with the ADASYN synthesized dataset and demonstrates how β , K_1 and K_2 can affect classifier performance. The results suggest that using higher K_1 resulted in higher sensitivity, but lower specificity and F-score. Also reducing K_2 would improve all three performance measures. For the following results we used $K_1 = K_2 = 5$ which showed best results in most cases.

Figure 4.7 compares the average performance of the SVM classifier in patient-dependent and patient-independent models for the original ($\beta = 0$) and synthesized datasets. The

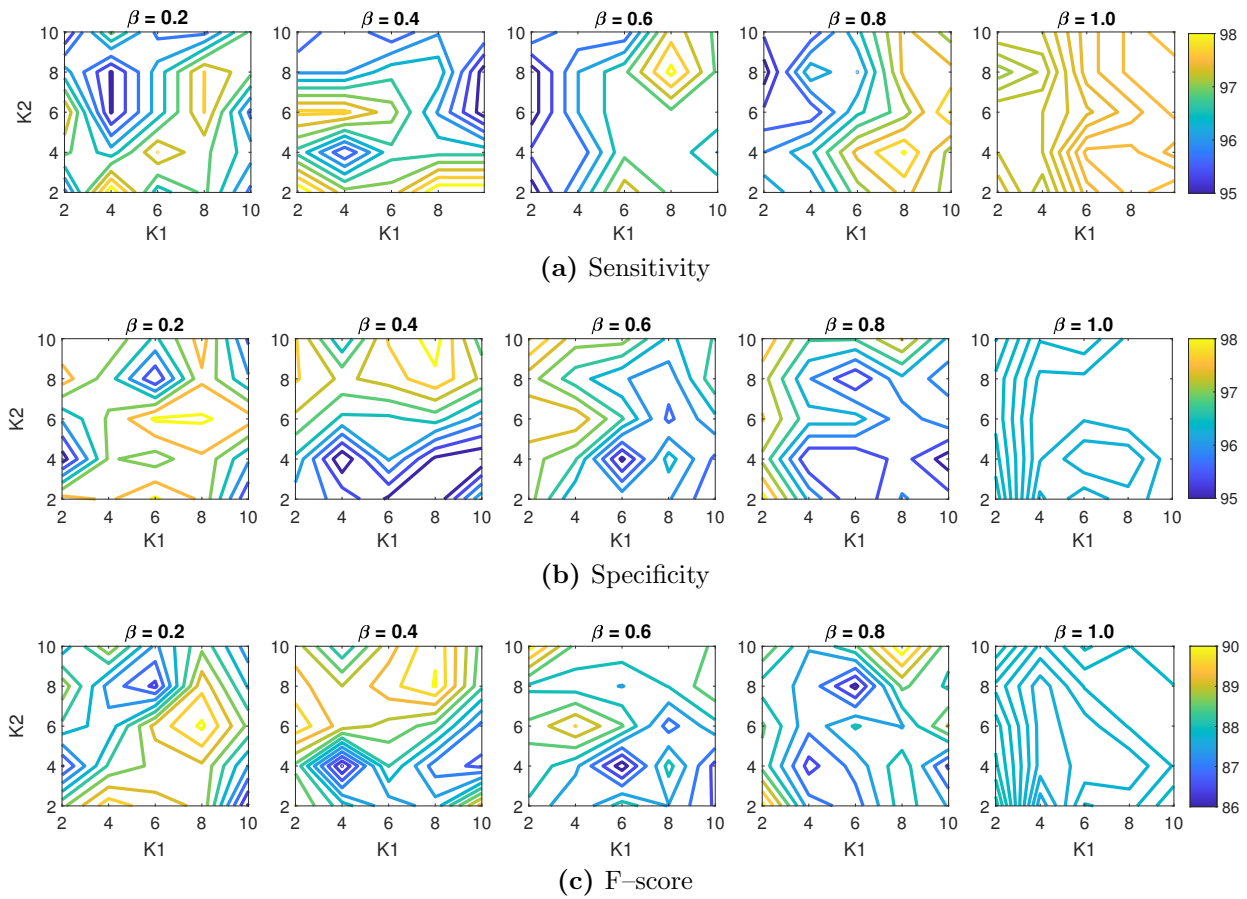


Figure 4.6: Sensitivity, specificity, and F-score of SVM classifier using synthesized dataset.

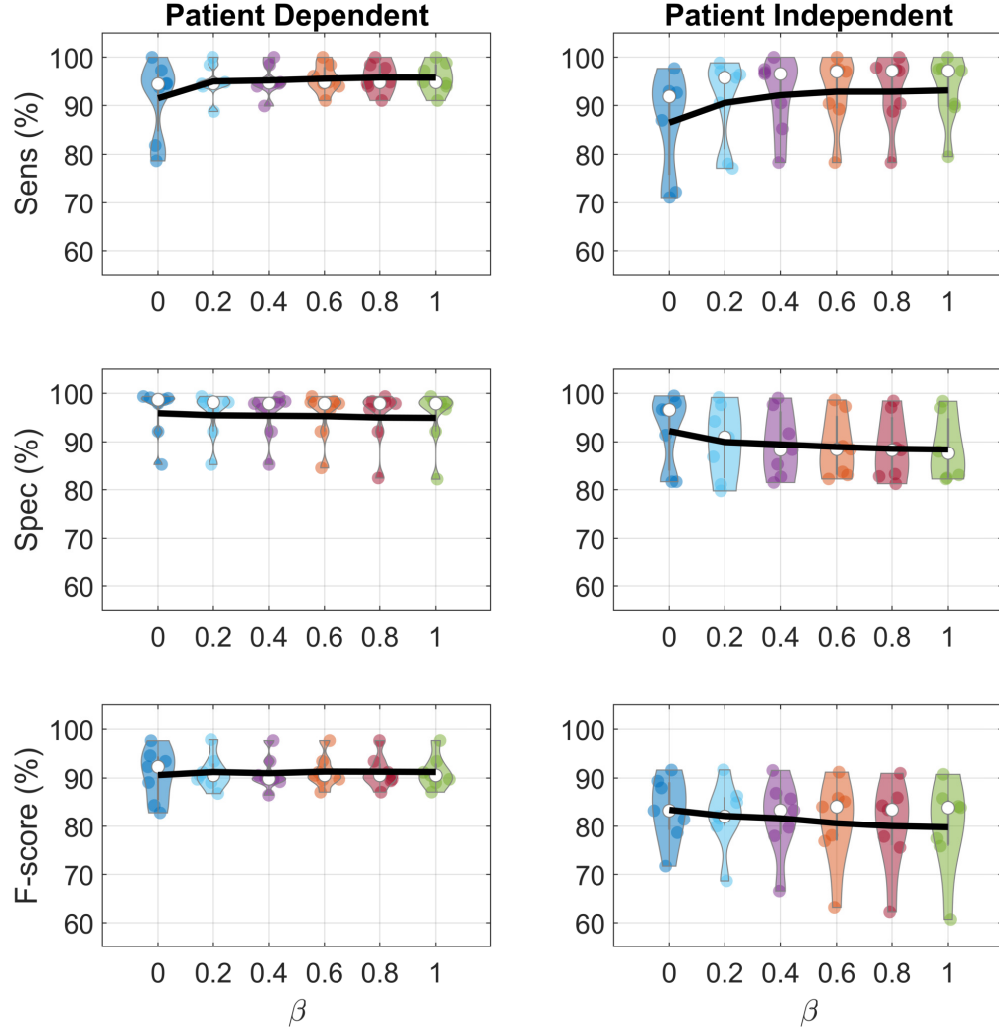


Figure 4.7: Violin plot of SVM classifier using original and synthesized datasets for patient–dependent and -independent models. Solid black lines and white circles show average and median of performance, respectively.

results suggest that adding synthesized samples to the minority class increases average F-score for patient-dependent models (from 90.6% for the original dataset to 91.2% for the balanced dataset) and diminishes average F-score for patient-independent models (from 83.4% for the original dataset to 79.8% for the balanced dataset). On the other hand, both approaches showed increased sensitivity using the balanced dataset (from 91.5% to 95.8% for patient-dependent, and 86.5% to 93.1% for patient-independent).

4.4.2 Patient-Dependent Models

Table 4.2 shows the average performance of patient-dependent models using the original imbalanced dataset. Among all the individual and ensemble classifiers, *SVM* showed highest Sensitivity (85.6%), and *ClsfBagging* demonstrated highest Specificity and F-score (95.8% and 87.7%, respectively). The results suggest that using ensemble classifiers led to improved FoG detection, measured by F-score as the overall performance measurement.

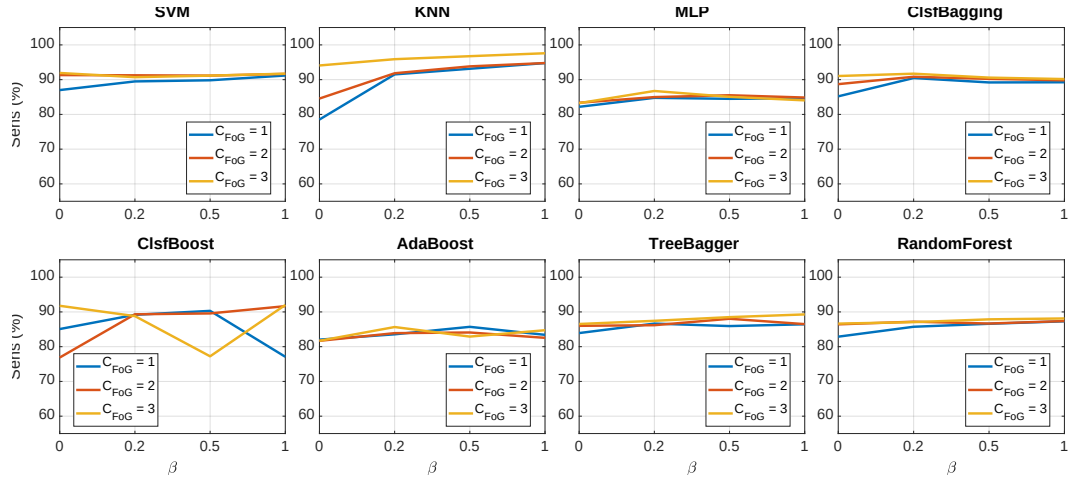
In order to see the impact of data imbalance on classifier performance, we changed the β value in ADASYN from zero (no synthetic data added to the training set) to one (creating a fully balanced training dataset). We also changed the ratio of cost between two classes (cost of FoG to normal gait) from one to three to investigate the effect of misclassification cost on performance of classifiers. Figure 4.8 shows how sensitivity, specificity, and F-score of classifiers changed by using different levels of imbalance and cost ratio. Table 4.3 shows the average performance of classifiers and their best results for the three performance measures in patient-dependent models. Among all classifiers and combinations of β and cost ratios, *kNN* showed highest sensitivity (97.6%) and specificity (96.2%), and *ClsfBagging* had the highest total performance measured by F-score (90.7%). The results show that using data synthesis and increased cost of misclassification, improved sensitivity of *ClsfBagging* from 85.2% to 90.8%, while keeping the specificity almost untouched, which resulted in improved F-score from 87.7% to 90.7%.

Table 4.2: Classifiers performance in patient-dependent models using the original imbalanced dataset and equal misclassification cost (results show average performance of all seven participants).

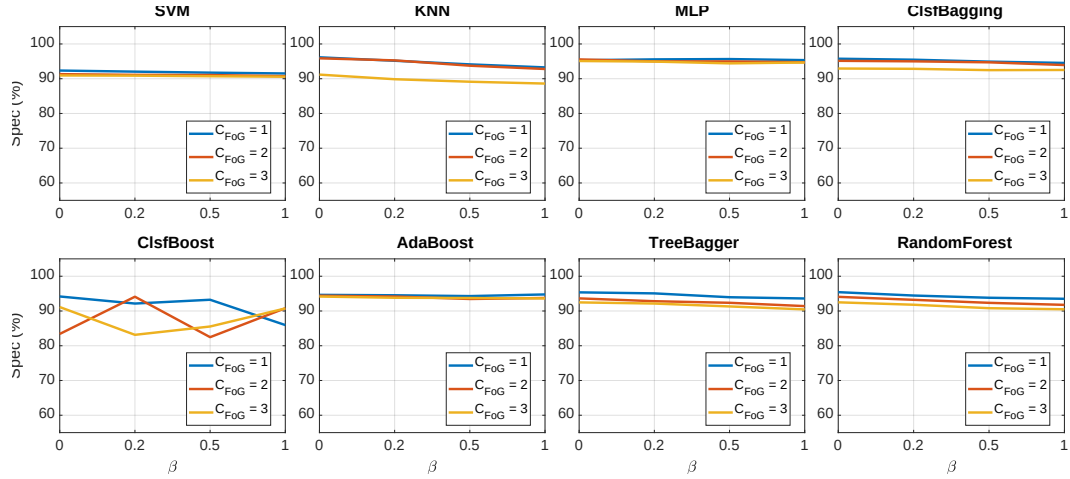
Classifier	Sensitivity (%)	Specificity (%)	F-score (%)
SVM	85.6	92.3	84.9
kNN	82.0	95.3	86.0
MLP	82.4	94.5	85.4
ClsfBagging	85.2	95.8	87.7
ClsfBoost	85.1	94.2	85.8
AdaBoost	82.0	94.6	82.8
TreeBagger	83.9	95.4	85.8
RandomForest	82.9	95.4	85.6

Table 4.3: Best performing classifiers in patient-dependent models (results show average performance of all seven participants).

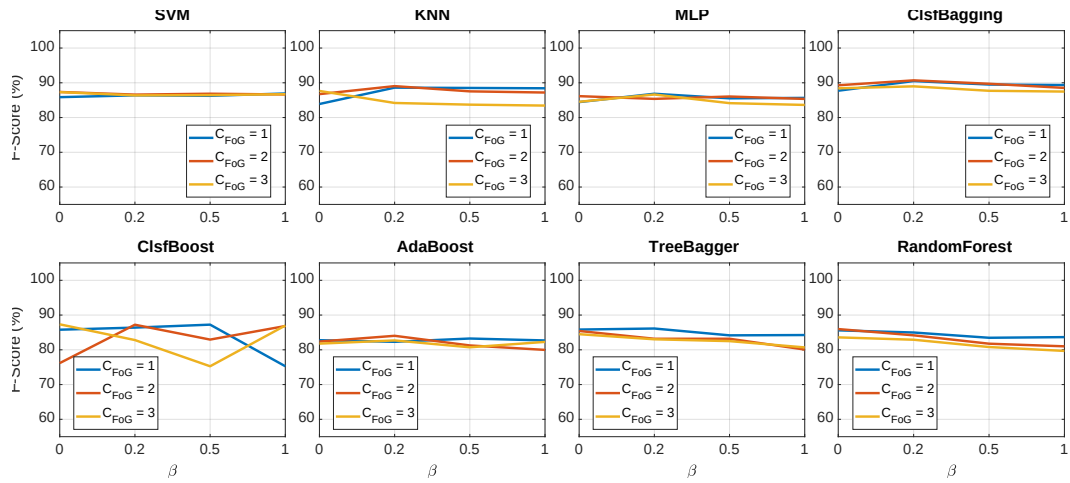
Classifier	C_{NG}	C_{FoG}	β	Sensitivity (%)	Specificity (%)	F-score (%)
kNN	1	3	1	97.6	88.6	83.4
kNN	1	1	0	82.0	95.3	83.9
ClsfBagging	1	2	0.2	90.8	95.0	90.7
ClsfBagging	1	1	0.2	90.5	95.5	90.5
ClsfBagging	1	2	0.5	90.2	94.7	89.7



(a) Sensitivity



(b) Specificity



(c) F-score

Figure 4.8: Performance measures of classifiers for patient-dependent models using synthetic data and misclassification cost.

4.4.3 Patient–Independent Models

Table 4.4 shows the average performance of patient–independent models using the original imbalanced dataset. Among all the individual and ensemble classifiers, *RandomForest* showed highest Sensitivity (77.0%), and *ClsfBagging* demonstrated highest Specificity (94.2%). *RandomForest* and *ClsfBagging* both demonstrated highest F–score (%74.5).

Figure 4.9 shows how the level of imbalance and ratio of cost affect sensitivity, specificity, and F–score of classifiers. Table 4.5 shows the average performance of classifiers for the three performance measures in patient–independent models. The *kNN* showed highest sensitivity (90.6%), and *ClsfBagging* had the highest specificity (94.2%) and F–score (76.3%). The results suggest that using data synthesis and increased cost of misclassification, improved sensitivity of *ClsfBagging* from 72.6% to 83.3% and impaired the specificity from 94.2% to 90.1%, which in sum resulted in improved F–score from 74.5% to 76.3%.

4.4.4 FoG Detection Latency

This section presents the results for FoG detection latency in patient–dependent models as it showed higher performance (see Tables 4.3 and 4.5). We refer to latency as the time between a FoG episode onset and the time when the system detects it. As the system runs the FoG detection algorithm every 0.5 second (step time in windowing the acceleration signal), the latency is also observed in steps of 0.5 second. A negative latency represents prediction of FoG (before its actual occurrence), whereas a positive latency represents detection (after its occurrence). We also assumed that delays caused by sensor data transmission are small and can be neglected.

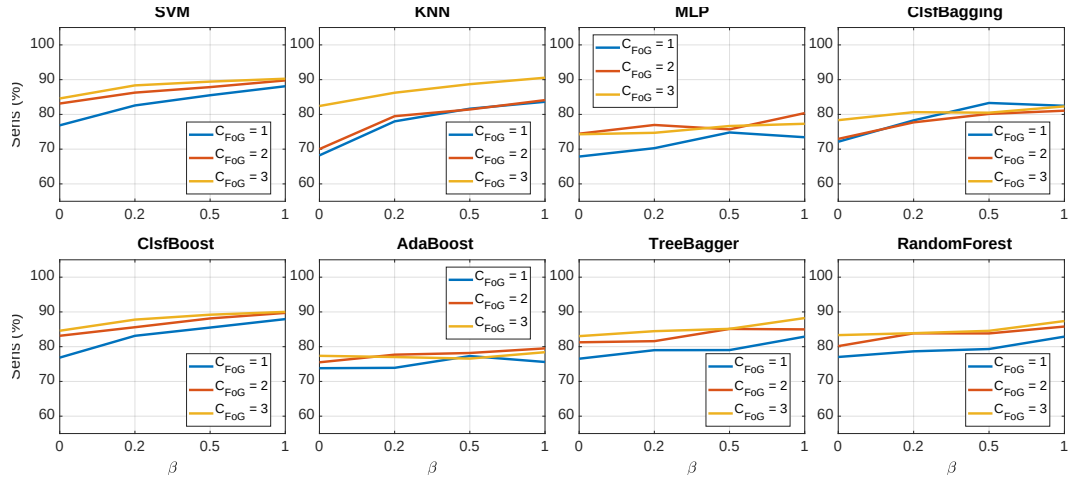
Here we only present the results for *ClsfBagging* ($\beta = 0.2$, $C_{FoG} = 2$) as it showed the highest F–score among classifiers in patient–dependent models. Figure 4.10 depicts the vertical and forward acceleration signal collected from the sensor placed on the right ankle of one participant as well as the labeled and identified events. Figure 4.11 shows the average detection latency of the *ClsfBagging* classifier for the seven selected participants. The classifier predicted 10.3%, 7.7%, 15.4% and 33.3% (66.7% in total) of the FoG episodes within 2, 1.5, 1, and 0.5 seconds before the actual FoG occurrence, respectively, and detected

Table 4.4: Classifiers performance in patient-independent models using the original imbalanced dataset and equal misclassification cost (results show average performance of all seven participants).

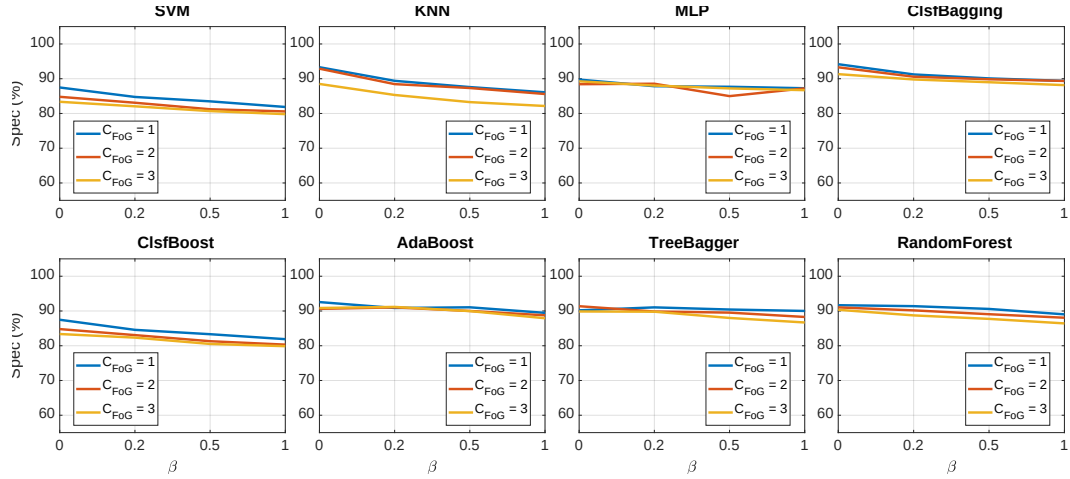
Classifier	Sensitivity (%)	Specificity (%)	F-score (%)
SVM	76.9	87.5	71.3
kNN	68.2	93.3	69.6
MLP	67.9	89.8	63.8
ClsfBagging	72.1	94.2	74.5
ClsfBoost	76.9	87.5	71.3
AdaBoost	73.8	92.6	72.2
TreeBagger	76.5	90.2	72.9
RandomForest	77.0	91.7	74.5

Table 4.5: Best performing classifiers in patient-independent models (results show average performance of all seven participants).

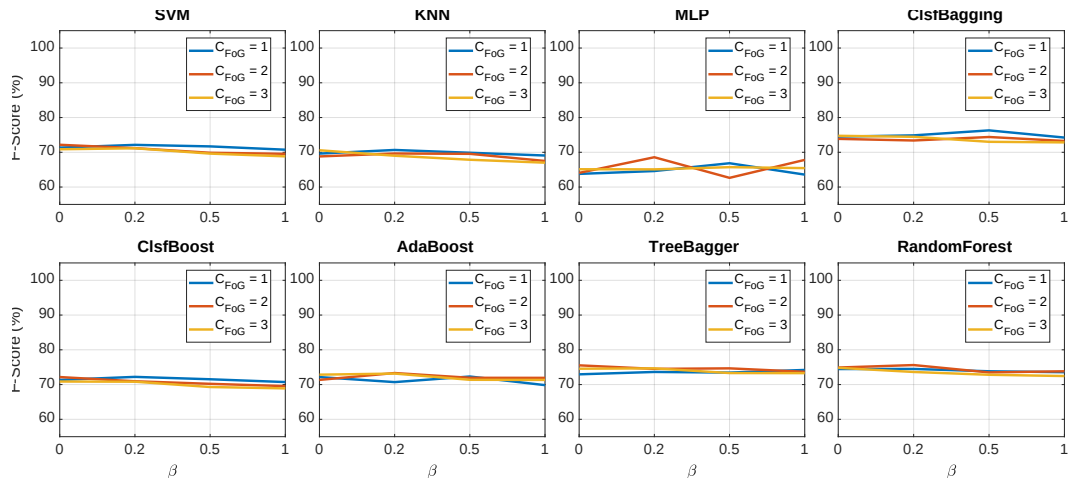
Classifier	C_{NG}	C_{FoG}	β	Sensitivity (%)	Specificity (%)	F-score (%)
kNN	1	3	1	90.6	82.2	67.0
ClfBagging	1	1	0	72.1	94.2	74.5
ClfBagging	1	1	0.5	83.3	90.1	76.3
RandomForest	1	2	0.2	83.8	90.2	75.6
TreeBagger	1	2	0	81.3	91.4	75.5



(a) Sensitivity



(b) Specificity



(c) F-score

Figure 4.9: Performance measures of classifiers for patient-independent models using synthetic data and misclassification cost.

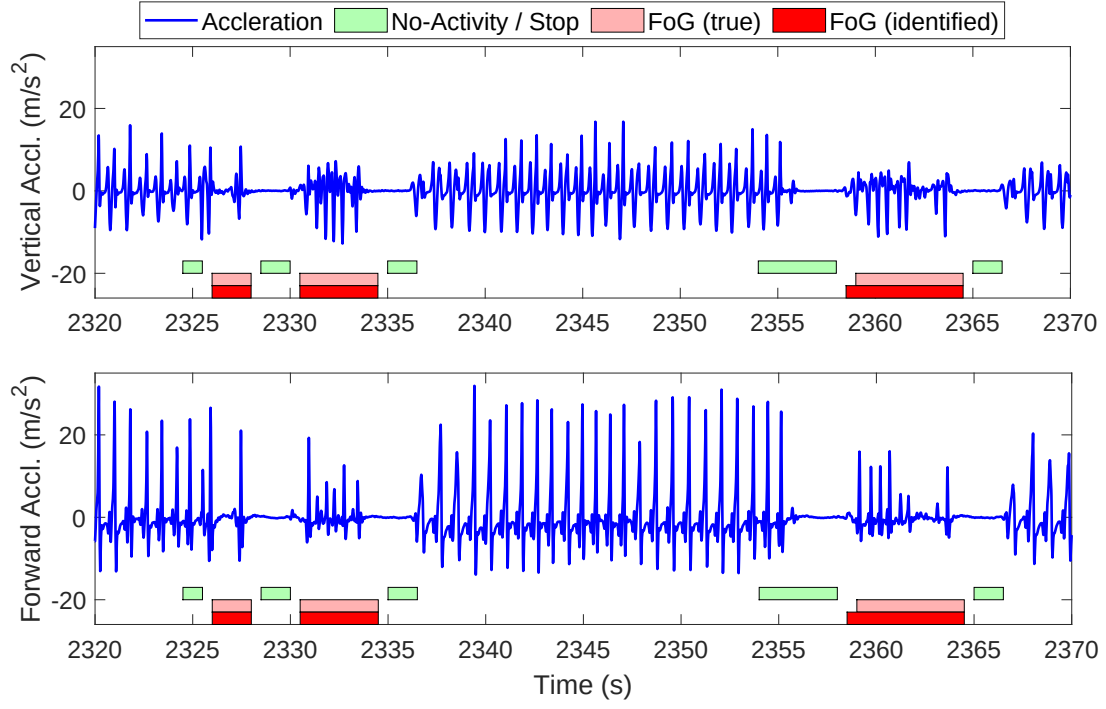


Figure 4.10: Acceleration signal collected from the right ankle sensor and the corresponding labeled and detected events using *ClsfBagging* with $\beta = 0.2$ and $C_{FoG} = 2$.

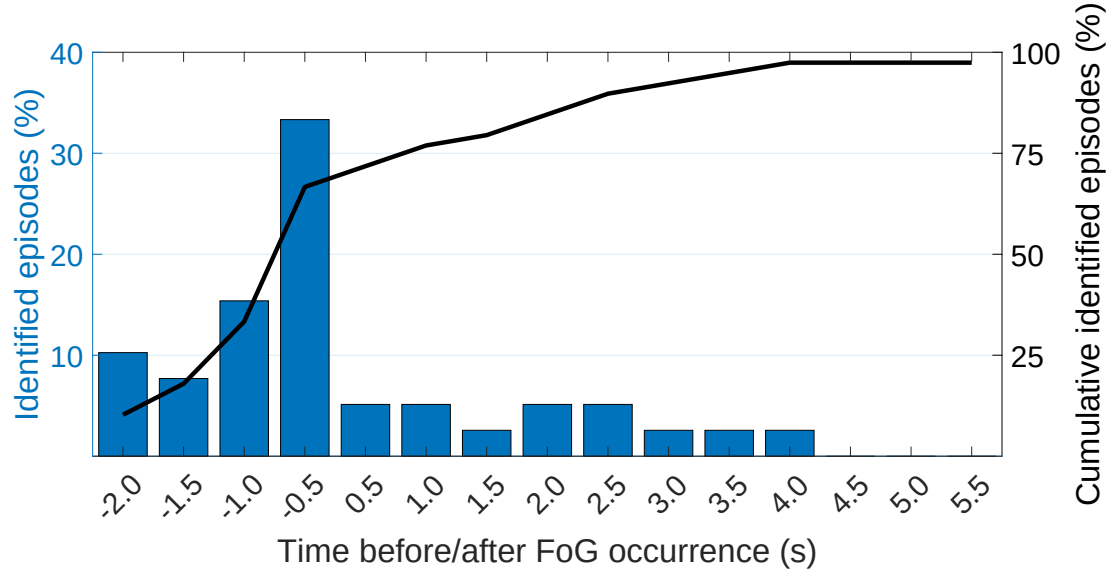


Figure 4.11: Average FoG detection latency of *ClfBagging* in patient-dependent models with $\beta = 0.2$ and $C_{FoG} = 2$. Negative values of time represent duration before FoG onset.

30.8% of the episodes within 4 seconds after FoG onset. The results shown in Figure 4.11 suggest that *ClsfBagging* is capable of detecting FoG events within 0.118 ± 1.587 second after FoG onset.

4.5 Discussion

This study investigated the impact of adding artificial new samples to the minority class using data synthesis methods and increased misclassification cost in favor of the minority class on an imbalanced dataset consisting of FoG and normal walking in patients with PD. The system takes time series signals as input and extracts features from 2-second windows; the features from six successive windows are then fed to the classifier. Comparing results from ADASYN synthesized dataset and the original dataset, it can be suggested that adding more synthetic samples to the training set (higher values of β) can shift classification bias in an imbalanced dataset toward the minority class and, thus, increase classifier sensitivity. On the other hand, more instances will be falsely recognized as FoG, and specificity will reduce. Hence, it depends on the patients' preference whether to receive feedback on more FoG events and, consequently, have more false signals, or to have a less sensitive and more specific model and probably miss few FoG events. The other factors in ADASYN, K_1 and K_2 , did not have a demonstrable impact on classification results. However, using a higher K_1 can help detect minority samples closer to the boundaries and improve accuracy of classification in the boundary regions. K_2 also determines how many points in the neighborhood of each sample must be used to create new samples, and using lower K_2 will limit the algorithm to samples closer to the initial data point. Finally, the new samples will be closer to the boundaries. The FoG class cost also affects the performance measures in the same way as new synthetic data ratio (i.e., higher cost classifies more instances as FoG resulting in increased sensitivity and reduced specificity and F-score).

In this study we labeled two windows prior to each freezing onset as FoG in order to enable the system to predict FoG before its actual occurrence. The system was able to predict 66.7% of FoG episode within 2 seconds before FoG onset. This method improved FoG detection latency (0.118 ± 1.587 s) compared with [23], another study using ensemble

classifiers (1.085 ± 0.731 s). Palmerini *et al.* also discriminated pre-FoG episodes from normal gait after removing data corresponding to FoG and with no sufficient motion [67]. Using patient-dependent training and testing, their method predicted 83% of the FoG episodes within 2 seconds before their onset. This comparison suggests that using different classifiers to discriminate FoG from prior normal gait or stop states may further improve detection accuracy.

The ensemble *ClfsBagging* classifier trained in this study outperforms the weighted SVM classifier in [25] that achieved 90% and 92% sensitivity and specificity, respectively. *ClfsBagging* is also capable of detecting more FoG episodes compared with [8] using Random Forests (90.8% vs 81.9% sensitivity), however the accuracy of classification for normal gait class is lower (95.0% vs 98.7% specificity). *ClfsBagging* also performs equally with the deep learning based model in [27] for detecting FoG episodes (90.8% vs. 91.9% sensitivity), and shows fewer false positives (95.0% vs. 89.5% specificity). However, labeling pre-FoG samples as FoG might have resulted in lower sensitivity and specificity when compared to [23] using ensemble classifiers (99.54% and 99.96%, respectively) and [28] using a deep convolutional neural network (99.64% and 99.99%, respectively).

In patient-independent models, *ClfsBagging* detected more FoG instances (83.3% vs. 66.3% sensitivity) but also showed more false positives (90.1% vs 95.4% specificity) compared with [23]. *ClfsBagging* also outperformed the deep convolutional neural network in [28] (74.43% and 90.60% sensitivity and specificity) which suggests that the feature set introduced in this study can yield a more generalized model to detect FoG. Although we expect higher predictability in patient-dependent compared to the patient-independent data, the comparable and significantly above chance-level predictability of patient-independent data is informative. Interestingly, the t-SNE illustration in Figure 4.4 portrays a heterogeneity of patients' non-freezing data. Such distribution in the patients extracted features may lead to a new avenue to wisely benefit from collective knowledge from previously recorded data. Instead of collecting about half an hour data from multiple sensors, a transfer learning scheme may enhance the training/calibration protocol of the FoG detector by reducing the data recording duration [108–110]. We may expand the study in future work by applying

different transfer learning approaches to enhance the model calibration process. This may ultimately help to increase the chance of user acceptance for the final device.

For clinical use, the proposed *ClsfBagging* model (an ensemble classifier formed by SVM, kNN and MLP individual classifiers and trained using bagging methods) will be integrated into an Android application which triggers vibrotactile cueing via a connected smartwatch. The system’s ability to detect FoG episodes, as well as the effects of cueing in ambient environments, will be investigated in future studies. However, considering that the most accurate performance was obtained using patient-dependent models, practical implementation will require data collection from each user in order to train an offline model using expert-labeled signals. A pre-trained model can also be updated using the new data and transfer learning techniques [68, 109].

There are two main limitations for the presented study: 1) Among the eighteen participants recruited, only nine froze during the test, two of whom with low number and duration of FoG episodes. This caused an imbalanced dataset with a small sample size to train and test classifiers. A larger pool of participants is required for further analysis of the results. 2) In a clinical setting and to prevent freezing, most of the participants were too focused on walking during the experiments, while they are normally distracted by other tasks in their homes. A set of data collected from participants performing dual-tasks in ambient environments would represent their behavior more realistically and a model trained on such data would be able to detect FoG episodes more effectively.

4.6 Conclusion

This work investigated the effect of synthesizing new samples and increased misclassification cost for the minority class on classification accuracy of a binary dataset using individual and ensemble classifiers. The results suggest that using ADASYN to create new samples for the FoG class and increasing its cost shift classification bias from majority class towards the distribution of the minority class data points located near the boundaries. This results in improved recognition of actual freezing instances (sensitivity) and increased incorrectly

identified FoG samples (false positive) and reduces specificity and F-score. The results also demonstrated that using ensemble classifiers improved performance of individual classifiers.

In addition to the accuracy of detecting FoG episodes, the latency of detection is also important for external cueing applications. The ideal system would be able to predict imminent FoG episodes and trigger preemptive cueing which can potentially prevent the episodes. The proposed model in this study was able to identify 97.4% of the FoG labeled samples within 2 seconds before to 4 seconds after FoG onset, 66.7% of which were predicted in the patient-dependent models.

The classifiers investigated here performed better on patient-dependent models (as compared to patient-independent models). This implies that the future studies on real-time FoG detection must implement an offline training process in which data will be collected and models will be trained as a pre-processing stage. Transfer learning techniques can also be used to update the already trained models using the data collected from a new patient. Future studies will be focused on the implementation of the trained classifiers in a real-time FoG prediction system to validate the results in ambient setting.

4.7 Future Works

This study demonstrated FoG prediction capability of the patient-dependent models. However, training specific models for each individual may not be practical as each individual freezes more frequently under certain conditions. This inter-subject variability results in varied class imbalance in the dataset and, consequently, performance of the patient-dependent models. To study the effect inter-subject and inter-trial variability, the models can be trained and tested on a dataset containing more participants performing more varied tasks.

Chapter 5

Addressing the One–Class Problem in FoG Identification

5.1 Introduction

In the previous chapter we studied how ensemble classifiers and oversampling methods can improve the FoG identification accuracy for patients who showed only a few FoG episodes during the experiments. For those who didn't experience FoG during the data collection sessions, we only had normal gait samples and we had to exclude them from further analysis. In this chapter, we aim to introduce a new method to identify FoG episodes in participants for whom no FoG data are available for model training. This method incorporates transfer learning to train a reference model using data from a group of participants for whom FoG and normal walking data are available. The reference model is then re-trained using only normal walking data from a new participant. Using only one class for re-training would result in all input samples being sorted into a single class. To address this issue, a deep one-class classifier is introduced to detect anomalies (i.e., FoG). With the encouraging results obtained, we expect this study to provide evidence of a novel algorithm for accurate identification of FoG onset using wearable sensor data. This may ultimately result in interventions to prevent FoG through the provision of preventive cues.

5.2 Background and Approach

Figure 4.4 suggests that normal gait samples for each participant can be clustered into one group while the FoG data is scattered in the space and can be distinguished from normal gait. Such marked disparity from normal behavior can be defined as “anomalies”. Anomaly detection (AD) is the task of identifying outliers (or “anomalous” samples) in data by training a model to accurately identify characteristics of “normal” samples. Deviations from normal behavior are then deemed to be anomalies. In AD problems it is assumed that the majority of the training dataset consists of normal samples. Thus, anomaly detectors are also known as one-class classifiers. Neural networks, one-class Support Vector Machine (OCSVM), and k-Nearest Neighbors (kNN) are the most common unsupervised algorithms used to detect anomalies in image [111–113] and time-series data [114–116]. AD algorithms have also been used to detect gait anomalies using wearable sensor data [117, 118].

While classical AD methods require feature engineering, novel methods use Deep Learning (DL) to learn relevant features automatically [119, 120]. A novel class of classification algorithms are Deep Belief Networks (DBNs) proposed as an unsupervised multi-class classifier and dimensionality reduction tool [121]. Each layer of a DBN is trained sequentially, i.e., the extracted features of the trained layers are used as inputs to the next layer. Erfani *et al.* developed a DBN-OCSVM hybrid model in which a DBN is trained to extract only generic underlying features in the input and reduce dimensionality to transform the higher dimensional data into a lower dimensional features space [119]. The OCSVM can then separate normal data from anomalies in the learned feature space more effectively.

In supervised AD, the objective of feature learning is to maximize the inter-class distance (between samples of different classes) and minimize the intra-class distance (between samples of the same class) [122]. One approach is to pre-train a model using a dataset with large number of classes to ensure feature “descriptiveness” (i.e., the features produce distinct representations for samples of different classes) [123]. Then, in order to differentiate between normal and anomalous classes, the model can be re-trained using data only from the normal class. However, this approach may learn a trivial solution as no penalty is assessed for misclassification (i.e., the learned features will be “compact” but not descriptive). Perera and

Patel proposed a one-class transfer learning method to simultaneously re-train a “reference” model using a multi-class dataset, and a “target” model (with shared layers with the reference model) using a single-class dataset [123]. The former minimized the descriptive loss, while the latter minimized the compactness loss of the learned features. A kNN classifier was then able to discriminate positive and negative classes using the features extracted from the fine-tuned target model.

Ruff *et al.* trained a Deep Support Vector Data Description (Deep SVDD) network to detect anomalies by minimizing the volume of a hypersphere enclosing the feature representation of the normal samples [124]. The volume of the sphere represents the variations in the features space and minimizing the volume means minimizing the intra-class variations (or improving the compactness) in the normal class. Then, any sample mapped to the area outside this sphere would be classified as anomaly.

These results demonstrate the potential applicability of AD and one-class classifiers to identification of infrequent gait deviation episodes. In the following sections, first, we will explain the data collection and data preparation steps. Then, we introduce our new deep gait anomaly detection algorithm to identify freezing of gait in patients with PD. Finally, we will discuss the results and select the best approach to train the models.

5.3 Methods

5.3.1 Dataset

In this study we used data from seven participants who experienced freezing during the clinical experiments (4M/3F, 74.3 ± 5.1 years, Hoehn and Yahr score between 2 and 4). Experiment protocols and data collection was discussed in Section 3.3. The signals were segmented into 2-second windows with steps of 0.25 second. The window size was chosen based on the minimum duration of freezing episodes and the sampling rate. Using a smaller step size results in more instances of training data, but also generates increased redundancy between adjacent segments. Windows were labeled FoG if at least 0.25 second (equivalent

to the step size of the sliding window) were marked as freezing. In this study, we refer to FoG-labeled samples as anomalous, and normal gait (NG) samples as normal.

In order to investigate the performance of the trained models in predicting FoG onset before its actual occurrence, we varied the number of pre-FoG windows included in the FoG samples of the training dataset. We also investigated the use of data augmentation on the accuracy of predictions (Table 5.1). Dataset **D0** is formed only by windows labeled FoG and NG. In **D2** (and **D4**), two (and four) windows before each FoG onset were included in FoG class (Figure 5.1); this was expected to facilitate the prediction of freezing onset. In **D0-A**, **D2-A**, and **D4-A**, signals from left and right ankles were mirrored (rotated 180° around the vertical axis) and added to the training set. This data augmentation compensated for instances wherein freezing happened only on one side of the body; the trained model would therefore be capable of detecting freezing episodes happening on right side if all training samples included left side freezing, and vice versa. We did not add pre-FoG nor augmented data in testing sets in order to keep the testing samples identical.

Freezing of gait is a relatively rare motor dysfunction in patients with PD, but the frequency of events increases with disease progression [125]. FoG is also more likely to happen towards the end of the effective medication dose periods [9]. The ratio of FoG to NG labeled windows varied between participants (Table 5.2). After excluding windows associated with stop and no activity, the remaining samples were split into training and testing sets. The first 70% of the samples were randomized and used to train the models, while the last 30% were kept with the same order to investigate the ability of the models to predict the occurrence of freezing episodes before their actual onset. The experiments started with easy tasks (walking back and forth in a hallway) and gradually shifted towards tasks requiring more attention and obstacle avoidance, resulting in an increased likelihood of freezing episodes at the end of each session, i.e. the testing datasets are likely to have more FoG labeled samples.

In the previous chapter we showed that improving FoG class ratio can improve the sensitivity of the models to detect more FoG samples in the testing set. However, in this study, we aimed to examine the performance of the deep one-class classifiers (anomaly

Table 5.1: Datasets used for model training (values in parenthesis show the number of pre-FoG windows added to the anomalous class of training set).

Dataset	Normal Class	Anomalous Class	Data Augmentation
D0	NG	FoG	-
D0-A	NG	FoG	yes
D2	NG	preFoG(2) + FoG	-
D2-A	NG	preFoG(2) + FoG	yes
D4	NG	preFoG(4) + FoG	-
D4-A	NG	preFoG(4) + FoG	yes

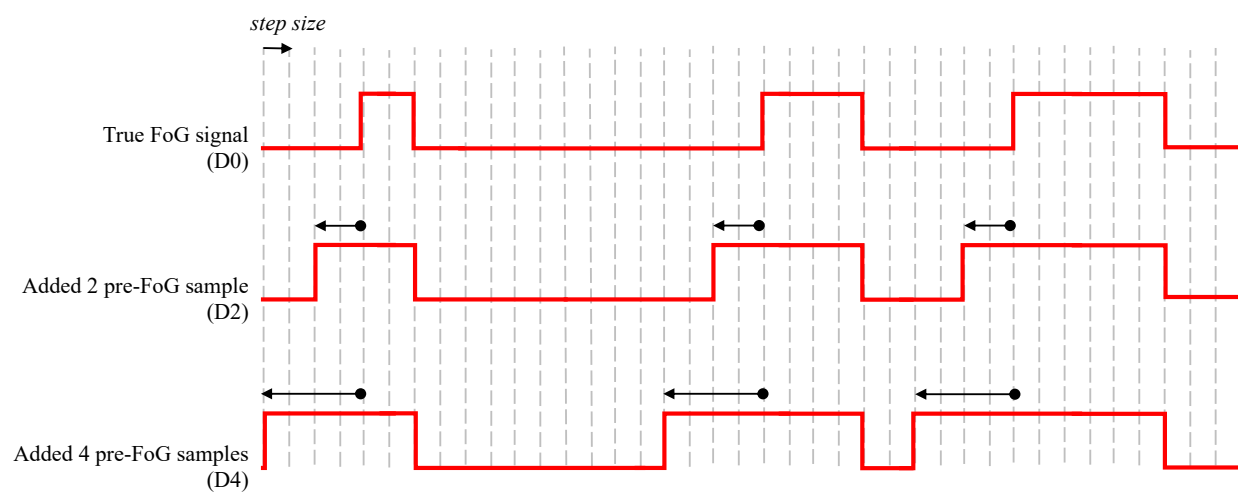


Figure 5.1: Adding pre-FoG samples to the FoG class in the training datasets.

Table 5.2: Distribution of labels in participants’ training and testing datasets.

Participant	Dataset	FoG Ratio (%)	Total Samples
1	Training	3.9	2880
	Testing	13.2	1234
2	Training	5.7	2076
	Testing	24.9	890
3	Training	24.0	2778
	Testing	44.2	1190
4	Training	45.6	1809
	Testing	56.3	776
5	Training	14.4	3156
	Testing	6.3	1353
6	Training	18.7	2053
	Testing	16.9	880
7	Training	30.6	2171
	Testing	34.5	930

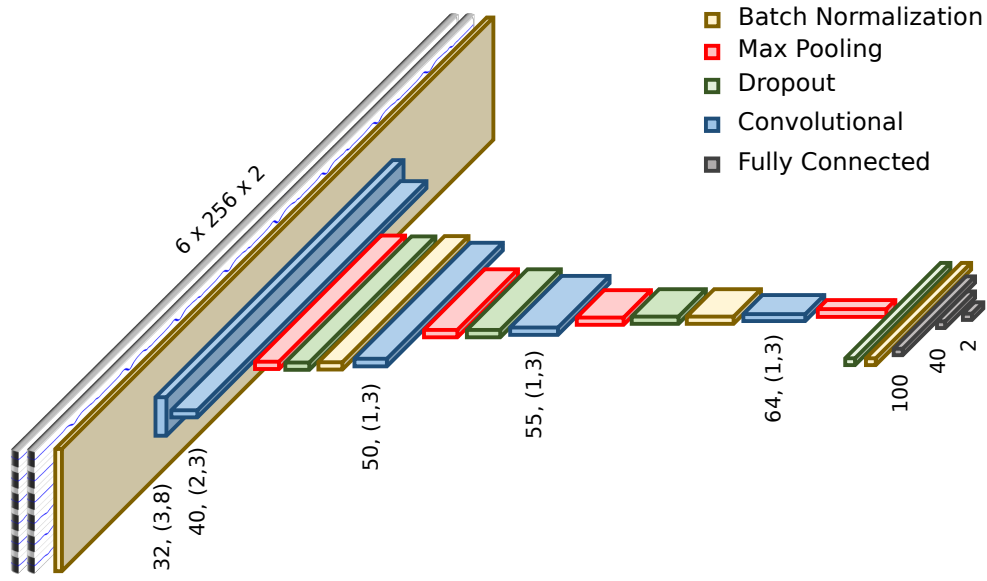
detectors) on the data with very few anomalous samples. Therefore, none of techniques in Chapter 4 were used to address the low FoG ratio.

5.3.2 Deep Gait Anomaly Detection (DGAD)

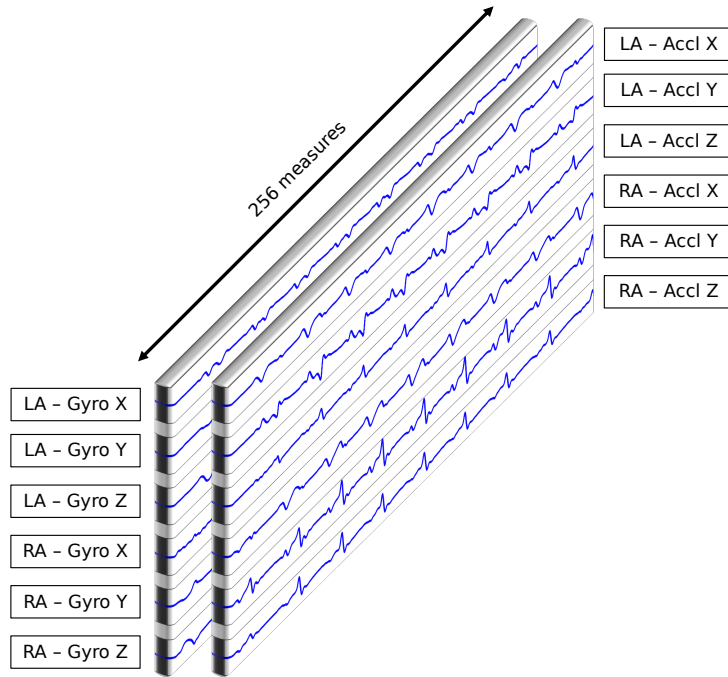
We propose a supervised hybrid convolutional neural network (CNN) architecture in which a series of convolutional layers are trained to extract relevant features, and three fully connected layers are trained to find the best representation of features in a 2-D space. A kNN classifier is then used to discriminate outliers (episodes of freezing of gait) from normal gait. The main objective of this study was to examine how to fine-tune a pre-trained deep anomaly detector for individual participants for whom only normal gait data are available. In order to investigate the accuracy of the models, we used LOOCV, i.e. normal and anomalous data from six participants were used to train a “reference” model which was then re-trained using only normal data from a new (“target”) participant.

Figure 5.2a illustrates the proposed structure of the reference and target models for the detection of gait abnormalities. The deep CNN architecture is composed of a cascade of five batch normalization, convolutional and max-pooling layers, followed by three fully connected. One of the main advantages of convolutional layers is that statistical features are automatically learned through supervised training. As the network gets deeper, the number of filters for the subsequent layer increase to capture more complex abstractions. The stride (representative of how each filter convolves around the input) for the first convolutional layer is set to (3,4) to capture features from all three dimensions of acceleration and rate gyro at a single step. The stride for the remaining convolutional and max-pooling layers is set to (1,1). Since the goal was to develop a reference model capable of extracting features from signals collected from different participants with different gait patterns, dropout layers were added to the network to improve model generalizability. The fully connected layers are then used to map the features into a two-dimensional space. In order to prevent the model from learning the trivial solution, the biases for the last two layers were fixed and set to zero [124].

The input layer of this model was fed using two channels of acceleration and angular velocity signals recorded from two ankle-worn IMUs (Figure 5.2b). The signals, sampled at 128 Hz, were first filtered using a Butterworth band-pass filter with cut-off frequencies of 0.5



(a)



(b)

Figure 5.2: (a) Structure of the reference and target models. The blocks show the output shape of each layer, and the values below show the number and size of kernels in convolutional layers or units in fully connected layers. (b) Input layer of the model (LA and RA correspond to left and right ankle, respectively)

and 8 Hz, and then segmented into 2-second windows (resulting in samples of 256 measures) with the steps of 32 measures.

The model was trained to map the filtered 2-second signals into a 2-D feature space. The samples from the normal class are mapped to the point $[1,1]$, whereas samples from the anomaly class are mapped to $[-1,-1]$ (see Figure 5.3). The Euclidean distance between the model output and the desired point was used as a loss function to update model parameters:

$$Loss = \frac{1}{N} \sum_{i=1}^N \sqrt{\sum_{j=1}^2 (\hat{y}_{ij} - y_{ij})^2} \quad (5.1)$$

where $\hat{y}_{ij}$ is the j -th output of the model for sample i , and y_{ij} is the j -th coordinate of the desired point for that sample (i.e., $y_{ij} = 1$ for the normal class, and $y_{ij} = -1$ for the anomaly class), and N is the number of samples.

5.3.3 Transfer Learning

The reference model is trained using data from a group of participants with samples from both NG and FoG classes and tested with data from a new participant; similar to the patient-independent paradigm in [23, 27, 28]. This is extremely advantageous given the rare occurrence of FoG in PD patients. The reference model is then trained only with NG samples of the target participant. This step differs from the conventional methods in which both NG and FoG samples are used for training the patient-dependent models. For the methods presented in [23, 27, 28], if FoG samples are not available for a participant they must be excluded from the analysis and patient-independent models are used instead, resulting in poor classification accuracy relative to patient-dependent models. In this study, we used transfer learning approach and a one-class classifier to develop a new method to retrain the reference models using only NG samples of the target participant.

After trained with data from six participants, the reference model was re-trained using only samples from the normal class in the first 70% of the data from the target participant. Training data were then randomly and equally divided into training and validation sets. During re-training, all layers of the reference model except the last four layers were frozen

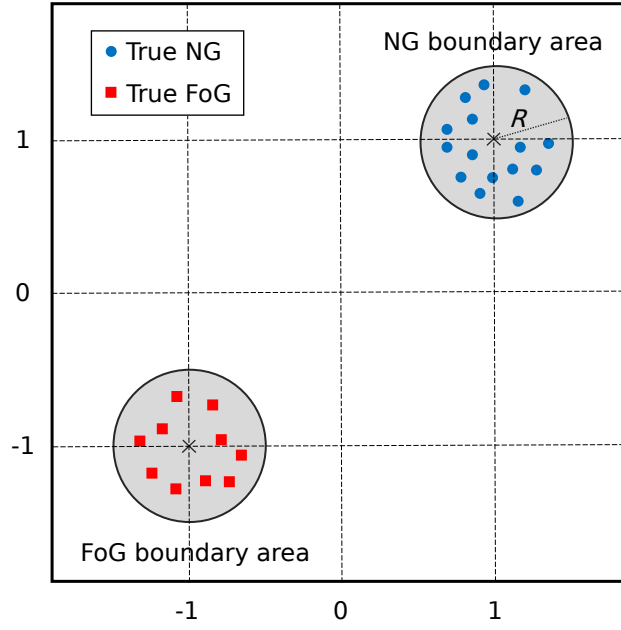


Figure 5.3: Ideal representation of model output. All NG and FoG samples are clustered in the NG and FoG boundary areas.

and the learning rate was reduced from 10^{-3} to 10^{-4} , as is commonly done in network fine-tuning. Re-training was stopped when the specificity (Equations 3.4) of samples in the validation set became equal or greater than a threshold (θ_{spec}). To find the specificity, model outputs for training and validation sets were stored as training and validation templates, respectively. A k-means clustering was used to find the centroid of the cluster ($\mathbf{C}$) in training templates. If the distance of a validation template from $\mathbf{C}$ was more than a predetermined value $\mathbf{R}$ (one-tenth of the Euclidean distances between $[-1,-1]$, $[1,1]$ points), it was classified as an anomaly. The training and validation templates were formed by NG samples. Then, in order to test the models, the samples from both classes in the next 30% of the data from the target participant were used to form testing templates. The same $\mathbf{C}$ and $\mathbf{R}$ from the training templates were used to classify the testing templates.

Automated cueing devices are designed to identify as many FoG episodes as possible and achieve high sensitivity. However, minimization of false positives (i.e., incorrectly identified FoG) is also important as users may adapt to cues or respond negatively to frequent stimuli, resulting in deteriorated intervention effectiveness. The threshold for specificity (θ_{spec}) of the validation set varied between 99%, 98%, and 97% corresponding to 1, 2, and 3 misclassification(s) per 100 instances of NG, respectively. An additional consideration is related to the ultimate application of this approach. DGAD will be implemented in a real-time FoG cueing application for participants who may have limited understanding of the underlying computational structure. Therefore, utility and a reduction of interference with typical activities are of maximal importance. Therefore, it is critical to minimize false positives to prevent participant adaptation to the cue.

5.3.4 Performance Measures

Sensitivity and specificity represent model classification accuracy of anomaly and normal classes, respectively (Equations 3.4). Because this algorithm is designed to be used in a real-time FoG detection application, the ratio of predicted and total identified (predicted or detected) FoG onsets to the total number of FoG onsets was also used as metrics. If the algorithm predicted the FoG onset within a 1-second epoch *prior* to actual FoG occurrence, the event was correctly *predicted*. If the algorithm detected the FoG onset within a 2-second

epoch *after* FoG onset, the event was correctly *detected*. The combined prediction and detection periods is then called *identification* period (Figure 5.4).

$$\text{Pred} = 100 \times \frac{\# \text{ of correctly predicted FoG onsets}}{\text{total } \# \text{ FoG onsets}} \quad (5.2a)$$

$$\text{Iden} = 100 \times \frac{\# \text{ correctly identified FoG onsets}}{\text{total } \# \text{ FoG onsets}} \quad (5.2b)$$

5.4 Results

5.4.1 Effect of Transfer Learning

In Chapter 4 we showed that normal gait data of each PD patient can be clustered and separated from the freezing data (see Figure 4.4). We also showed that patient-independent models (equivalent with reference models in this study) cannot detect FoG accurately in new patients due to gait variability. In this study, the reference models are trained with NG and FoG samples of six participants. The models are then retrained only with NG samples of the target participant, which improves the accuracy of target models in identification of the NG samples. Table 5.3 compares the sensitivity and specificity of the models trained with D4-A datasets.

Figure 5.5 shows how reference and target models map the NG and FoG samples in the testing sets of participants 3, 4, and 5. In the figure, the columns correspond to individual participants. The first and second rows show the projection of true-labeled NG and FoG samples, and the third shows classification results. The gray circles are centered at the centroid of the normal gait samples found using k-means clustering, and the radius depicts the boundary area for classification of the NG class.

The reference models are not able to discriminate between NG and FoG samples and the majority of the samples are mapped outside of the NG boundary area and, thus, incorrectly classified as FoG (i.e., high sensitivity and low specificity). The target models, however, are trained with NG samples of the target participant and mapped the NG samples into the NG boundary area and achieved higher specificity and lower sensitivity.

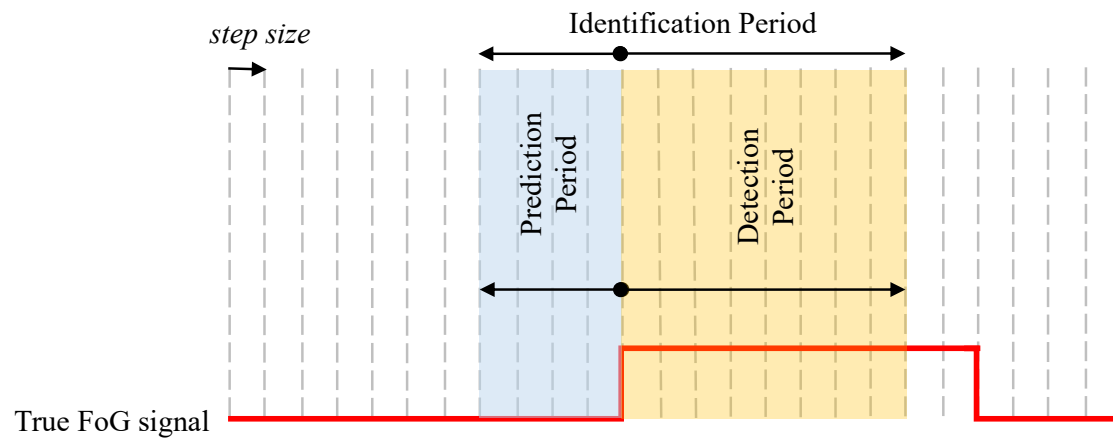
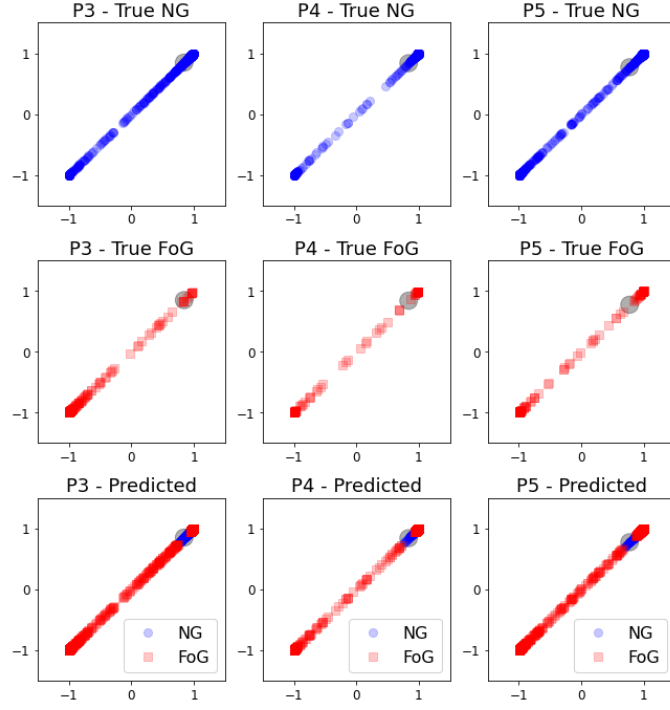


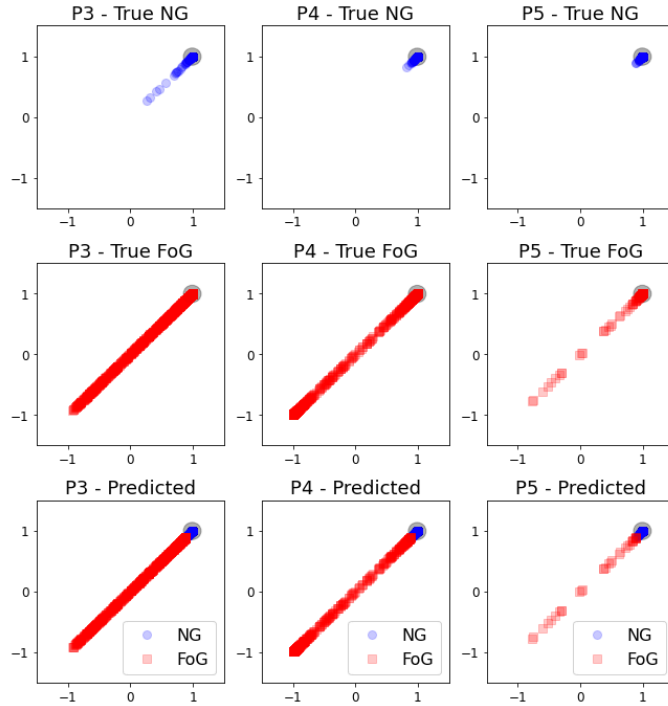
Figure 5.4: FoG onset prediction, detectio, and identification periods.

Table 5.3: Classification accuracy obtained using reference and target models trained with D4–A datasets.

Participant	Model			
	Reference		Target	
	Sens	Spec	Sens	Spec
1	73.2	94.4	64.8	98.3
2	97.3	97.6	65.5	100
3	99.6	10.5	69.6	96.8
4	99.1	3.6	73.5	97.6
5	97.7	1.9	41.7	99.5
6	91.9	27.9	74.5	99.0
7	100	1.3	51.5	99.1
Mean	94.1	33.9	63.0	98.6
std	9.6	43.4	12.2	1.1



(a) Reference Models



(b) Target Models

Figure 5.5: Mapped features from testing sets using reference and target models trained with D4-A datasets.

5.4.2 Effects of Data Augmentation and pre-FoG Windows

Augmenting the data doubled the size of the training set, which caused more data to be mapped to the centroid of the normal gait class, considering the higher ratio of NG instances compared with FoG. The results presented in Table 5.4 and Figure 5.6 show that using augmented training data improved Spec (accuracy of classification for NG samples) and reduced Sens (accuracy of classification for FoG samples). Table 5.5 also shows that data augmentation impaired Pred (prediction rate of the FoG onsets), although slightly improved Iden (ratio of predicted and detected FoG onsets). Figure 5.7 shows the number of FoG episodes predicted and detected for each participant using different datasets. The results suggest that although sensitivity deteriorates and smaller number of FoG windows can be identified using augmented data, more FoG episodes can be correctly identified within 2 seconds after the FoG onset, thus, cues can be provided shortly after the FoG occurrence.

Adding instances of pre-FoG (segments of signals prior to the FoG onset) to the FoG class raised the ratio of FoG class compare with NG in the training set, although the improvement was negligible due to the small number of FoG onsets. Therefore, Sens slightly improved using more windows of pre-FoG, while Spec remained almost unchanged (Table 5.4). However, the models learned to identify pre-FoG episodes and showed improved Pred and Iden compared with D0 and D0-A (Table 5.5). Thus, the models trained with D4-A are able to predict more FoG episodes and the following comparisons are based on the results obtained from these models. Figure 5.8 compares the output of the target models for three participants using D4 and D4-A approaches. For participants 2 and 3, training models using augmented data improved compactness of the features around the NG centroid (i.e. improved specificity), but with the cost of lower sensitivity. For participant 4, however, augmenting data pushed more samples outside the NG boundary area and led to higher sensitivity, but lower specificity.

5.4.3 Specificity Threshold (θ_{spec})

Using higher specificity thresholds on the validation sets to terminate the training of the target models, requires higher accuracy for classification of NG samples in the validation set. Thus, the models train for larger number of epochs and, consequently, model outputs become

Table 5.4: Classification accuracy ($\theta_{spec} = 99\%$).

P#	Dataset											
	D0		D0-A		D2		D2-A		D4		D4-A	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
1	57.4	97.3	63.6	97.6	63.6	97.5	63.6	98.1	63.0	97.4	64.8	98.3
2	81.2	100	82.5	99.9	83.9	99.9	79.8	100	88.8	99.3	65.5	100
3	72.7	97.2	69.2	97.1	78.5	96.2	73.7	96.2	82.2	95.6	69.6	96.8
4	66.6	96.5	65.7	99.1	67.5	96.5	79.6	98.2	63.6	97.9	73.5	97.6
5	36.9	100	34.5	99.9	34.5	100	44.0	99.8	42.9	99.7	41.7	99.5
6	85.2	98.9	70.5	99.5	83.2	98.8	79.2	98.8	84.6	98.2	74.5	99.0
7	40.9	98.0	40.6	99.0	35.9	98.5	35.3	99.0	41.5	98.0	51.1	99.1
Mean	63.0	98.3	60.9	98.9	63.9	98.2	65.0	98.6	66.7	98.0	63.0	98.6
std	18.9	1.4	17.2	1.1	21.0	1.5	18.4	1.3	19.5	1.3	12.2	1.1

Table 5.5: FoG onset prediction/detection ($\theta_{spec} = 99\%$).

P#	Dataset											
	D0		D0-A		D2		D2-A		D4		D4-A	
	Pred	Iden	Pred	Iden	Pred	Iden	Pred	Iden	Pred	Iden	Pred	Iden
1	0.0	57.1	0.0	85.7	0.0	85.7	0.0	85.7	14.3	71.4	0.0	85.7
2	0.0	100	12.5	100	0.0	100	0.0	100	25.0	100	0.0	100
3	38.5	61.5	38.5	61.5	46.2	76.9	46.2	69.2	53.8	100	46.2	69.2
4	14.3	100	0.0	85.7	14.3	85.7	0.0	85.7	42.9	85.7	14.3	85.7
5	0.0	71.4	0.0	71.4	0.0	71.4	0.0	71.4	0.0	85.7	0.0	71.4
6	57.1	100	42.9	100	57.1	100	42.9	100	71.4	100	42.9	100
7	37.5	87.5	25.0	100	37.5	62.5	25.0	100	37.5	50.0	50.0	100
Mean	21.1	82.5	17.0	86.3	22.2	83.2	16.3	87.4	35.0	84.7	21.9	87.4
std	21.5	17.5	17.3	14.1	22.6	13.0	19.8	12.4	22.4	17.3	21.8	12.4

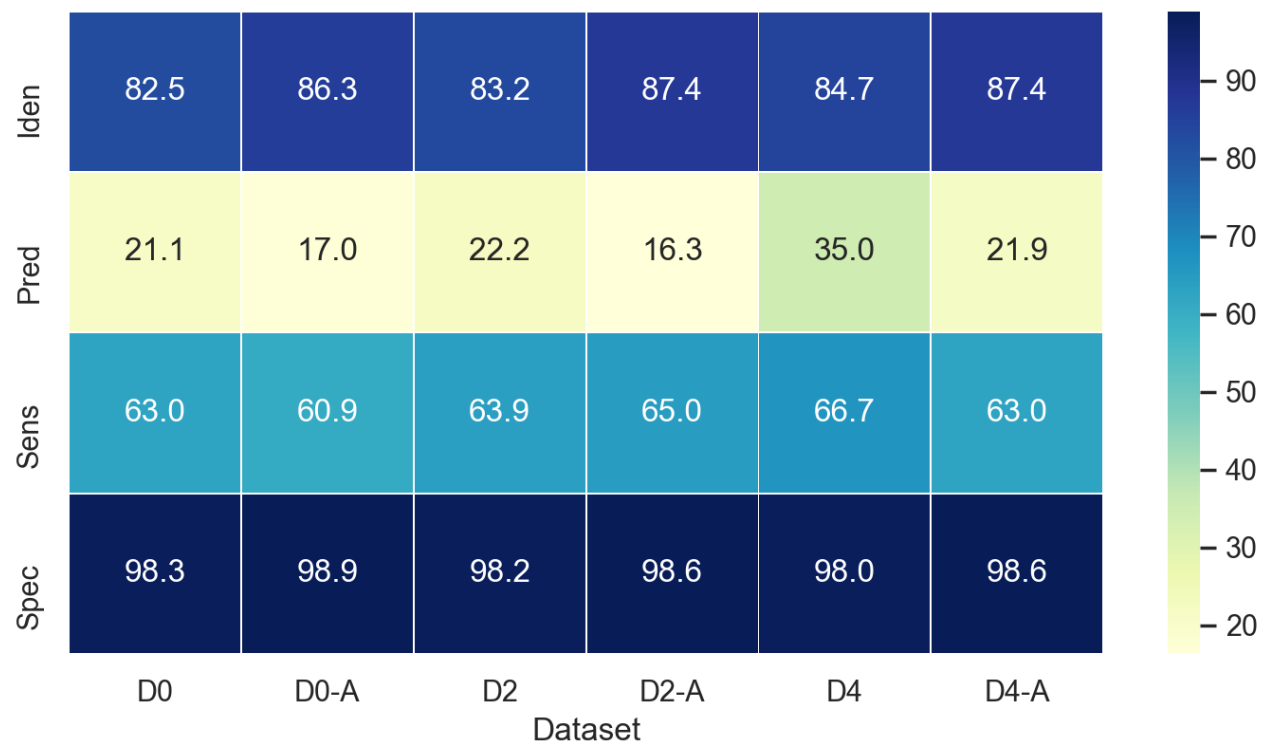


Figure 5.6: Average performance of the target models using different datasets for training ($\theta_{spec} = 99\%$).

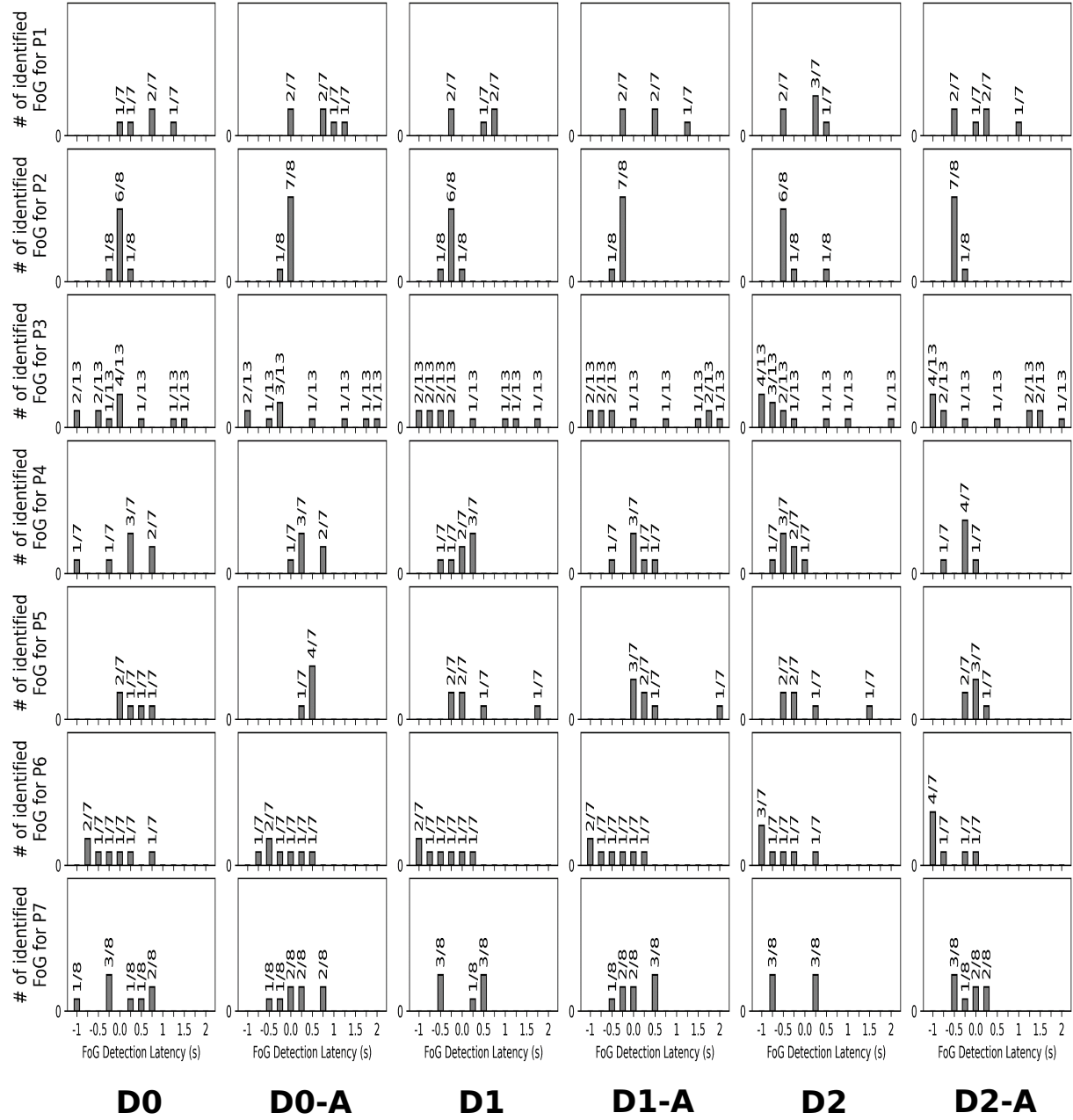


Figure 5.7: FoG detection latency ($\theta_{spec} = 99\%$). Negative and positive latency correspond to prediction and detection of FoG, respectively.

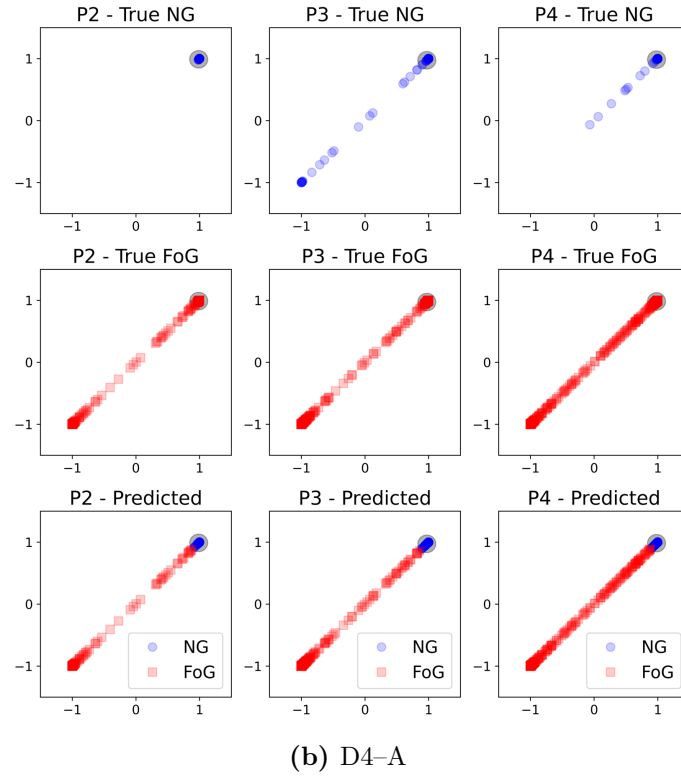
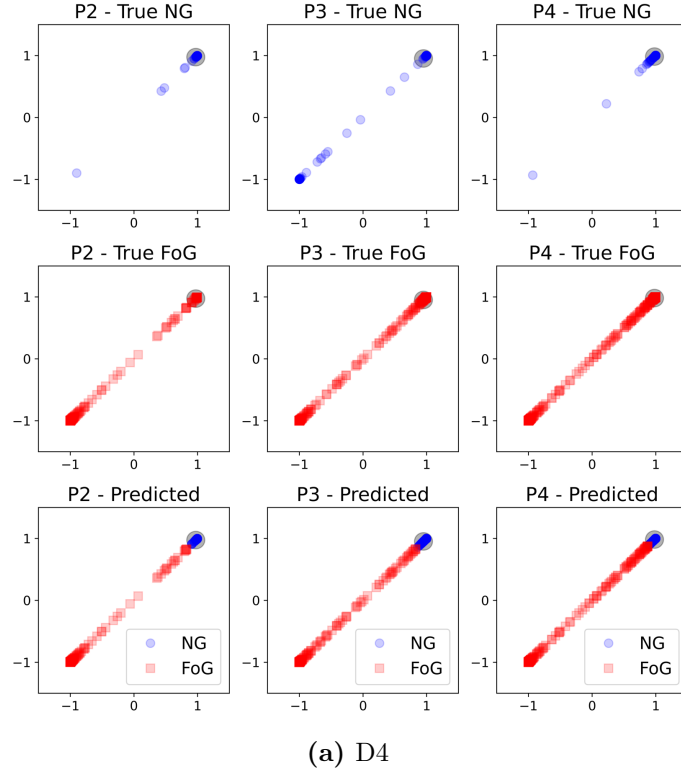


Figure 5.8: Testing templates and the classification boundary using $\theta_{spec} = 99\%$.

more compact around the NG centroid. Figure 5.9 illustrates how increasing θ_{spec} resulted in inclusion of more samples within the NG class boundary to reach the desired specificity. Table 5.6 compares the average FoG identification and latency results using D4–A datasets and different values of θ_{spec} . The results suggest that increasing the threshold from 97% to 99% improves Spec on the testing set, but cuts Sens by 16%. Higher thresholds also hamper the ability of the models to predict and identify the freezing onsets from 48.4% to 21.9% and 98.4% to 87.4%, respectively.

5.4.4 Number of Model Outputs

The last convolutional layer in Figure 5.2a outputs 128 features extracted from the raw signals. This layer is followed by three fully connected layers which are trained to map the features into a lower dimensional space and minimize the loss function. Increasing the number of output nodes will enable the model to classify data based on a larger number of extracted features, and potentially increase the accuracy; however, a higher output dimension increases the model complexity and increase the computation time, which will hamper the model performance for a real-time application on a mobile device with limited computation power.

Figure 5.3 depicts the desired model output when the last layer of the network consists of two nodes. We also compared the results for a varied number of outputs (from 1 to 10) for both reference and target models to identify the optimized model output size. The results presented in Table 5.7 and Figure 5.10 suggest that all models perform almost equally. However, for the sake of model simplicity we chose two outputs for the future analysis.

5.4.5 Normal Gait Boundary Area (R)

The parameter R defines the area around the normal gait centroid (C) inside which samples are classified as NG. Increasing R would result in more samples classified as NG and improve specificity, although hampering FoG identification. This relationship is visualized using the receiver operating characteristic (ROC) curve. Figure 5.11 shows the ROC curve for the target models trained with D4–A datasets and using $\theta_{spec} = 99\%$. The results suggest that

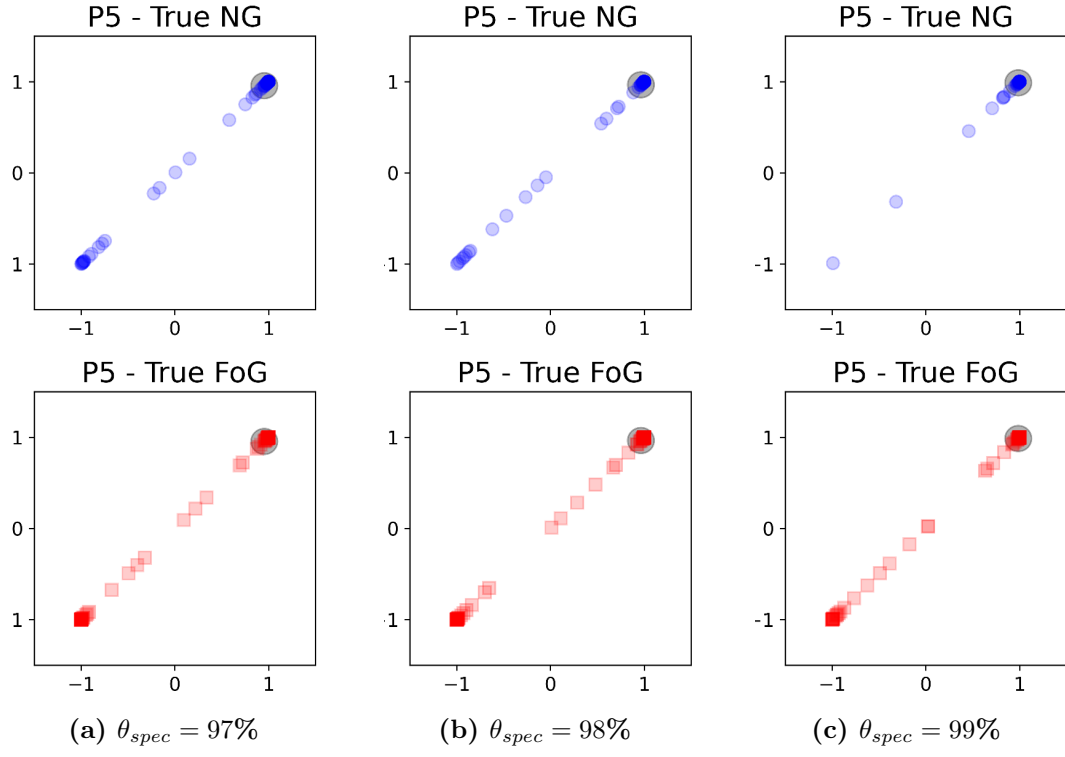


Figure 5.9: Effect of θ_{spec} on the testing sample classification using D4-A trained models.

Table 5.6: Average FoG identification performance using D2-A datasets for training and varied specificity thresholds.

θ_{spec}	Sens	Spec	Pred	Iden
97	78.9	94.9	48.4	98.4
98	73.2	96.5	37.5	90.6
99	63.0	98.6	21.9	87.4

Table 5.7: Average FoG identification performance using D4-A datasets and $\theta_{spec} = 99\%$ obtained with different number of model outputs.

Output Size	Sens	Spec	Pred	Iden
1	56.3	98.8	20.2	83.0
2	63.0	98.6	21.9	87.4
3	61.2	98.8	19.6	83.0
4	67.6	97.7	20.6	87.3
5	66.4	97.8	19.8	87.1
6	62.2	98.6	18.7	85.0
7	68.1	98.1	23.4	88.2
8	60.8	98.6	22.5	86.7
9	62.6	98.9	18.3	84.1
10	68.3	97.9	21.3	86.3

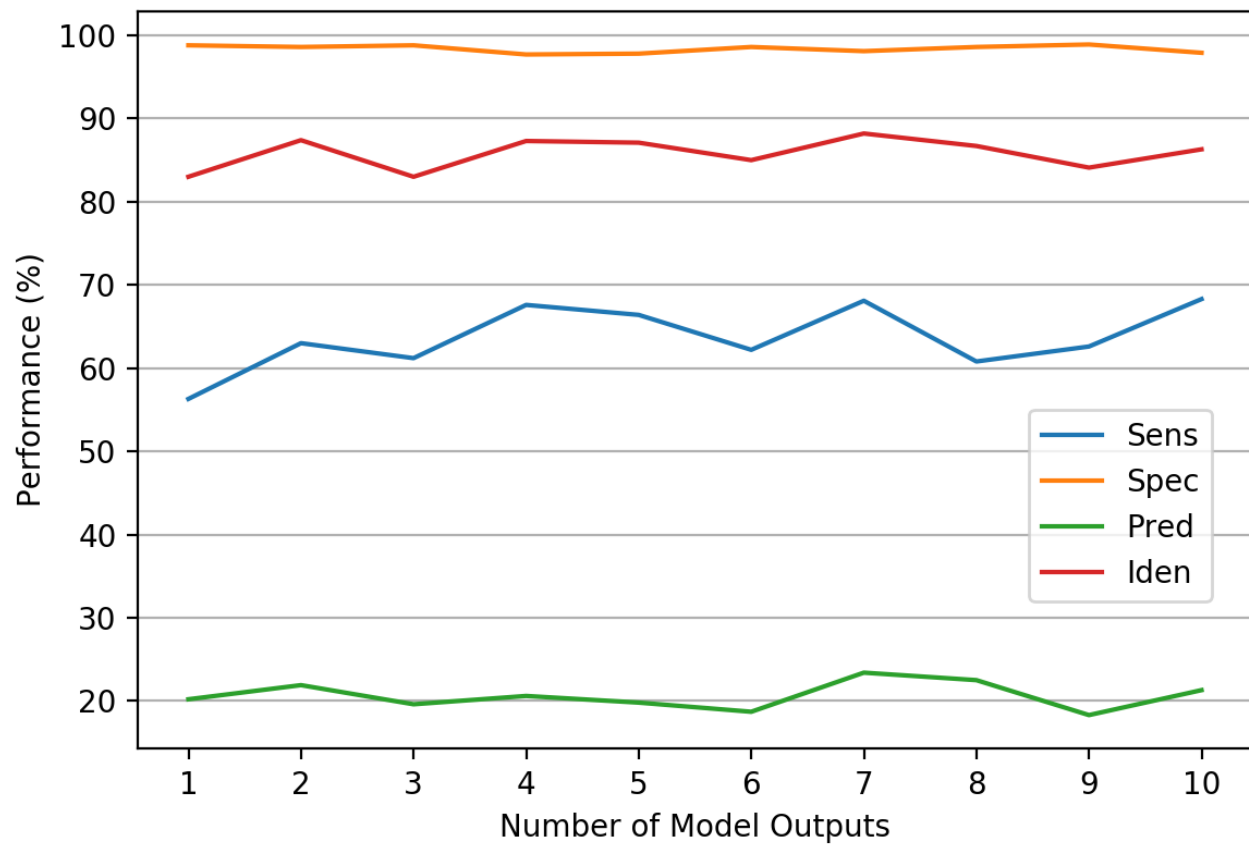


Figure 5.10: Average performance of the target models using different number of output nodes (models are trained using D4-A datasets and $\theta_{spec} = 99\%$).

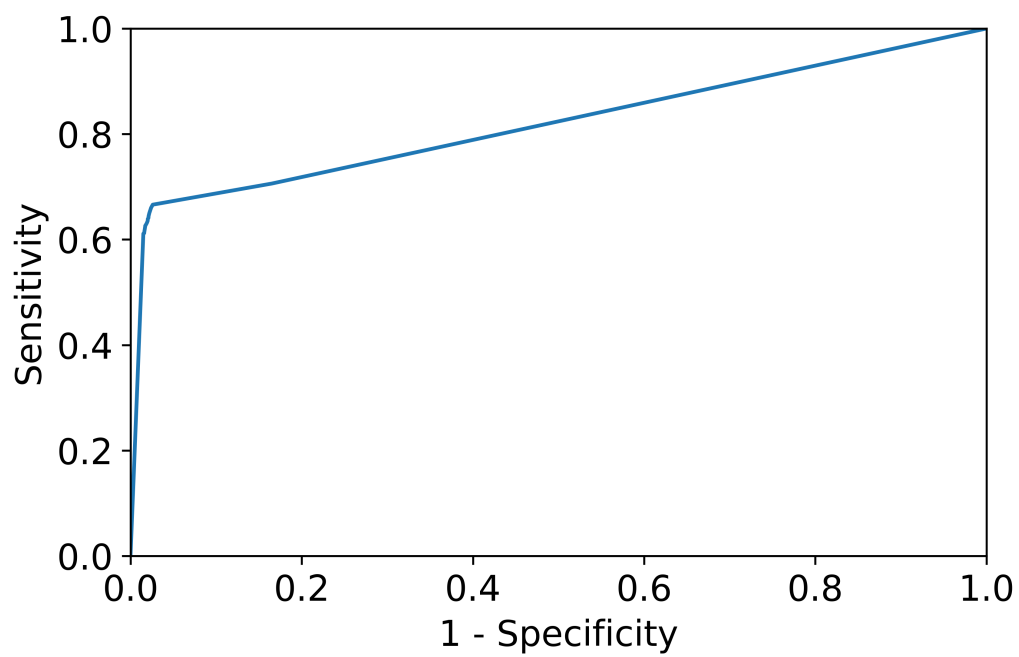


Figure 5.11: ROC curve using D4–A training datasets and $\theta_{spec} = 99\%$ (AUC = 0.82).

$R = 0.283$, equal to one-tenth of the Euclidean distances between $[-1,-1]$, $[1,1]$, shows the best performance combining all four metrics; Sens, Spec, Pred, and Iden.

5.4.6 Step Size Optimization

Figure 3.1 shows the sliding window method used to segment the signals in this study. The step size of the sliding windows (i.e., the time difference between the starting point of adjacent segments) defines the overlap between the adjacent segments and the number of training and testing instances available for each participant. Using lower step size values increases the number of samples but, on the other hand, may induce redundancy between adjacent windows. We previously used step size equal to 0.25 seconds (32 time points), while most of the studies in the literature used step size of only one time point. We compared the results using smaller step sizes to investigate the dependency of the model on this parameter. In order to prevent frequent false alarms in case of smaller step sizes, the output of the model can be median filtered, i.e. only trigger cues if at least a certain number of FoG events are identified within a predefined number of most recent samples.

Table 5.8 compare the results for different training and testing step sizes. The median filter (m/n) is defined by detection of at least m FoG windows among the most recent n windows. False positive (FP) rate is also calculated by the number of FPs per 5 minutes. The results suggest that training and testing with 0.125 and 0.5 second step sizes, respectively, 2/2 median filter on the model output improves the false positive rate to only 3.6 per 5 minutes while hampering prediction (4.2%) and identification rates (73.3%). The results also show that training and testing with 0.5 second step size without using a median filter enables 31.2% prediction and 91.8% identification rate with 16.2 false positives per 5 minutes.

5.4.7 Computational Complexity

All experiments were implemented with the TensorFlow library on a computer configured with Intel[®] Xeon[®] CPU @2.30GHz, Tesla P100-PCIE 16GB GPU, and 26GB RAM. The average training time for the reference models was 140.4 ± 4.4 s. The retraining and testing for the target models took 53.5 ± 17.7 s and 61 ± 10 ms, respectively.

Table 5.8: Model performace using different step sizes for training and testing. False Positive (FP) rate is calculated by the number of FPs per 5 minutes.

Step Size			Performance				
Training	Testing	Median Filter	Sens	Spec	Pred	Iden	FP Rate
0.5	0.5	1/1	71.5	97.3	31.2	91.8	16.2
0.5	0.5	2/2	62.7	98.6	6.4	80.7	8.4
0.5	0.5	2/3	72.1	97.7	6.4	80.7	13.8
0.25	0.25	1/1	63.0	98.6	21.9	87.4	16.8
0.25	0.25	2/2	60.9	99.1	15.4	85.4	10.8
0.25	0.25	2/3	66.1	98.8	15.4	85.4	14.4
0.25	0.5	1/1	66.1	97.6	21.1	91.7	14.4
0.25	0.5	2/2	57.2	98.7	6.3	80.7	7.8
0.25	0.5	2/3	66.9	98.0	6.3	80.7	12.0
0.125	0.125	1/1	59.6	98.5	27.3	79.4	36.0
0.125	0.125	2/2	57.0	98.9	15.5	79.4	26.4
0.125	0.125	2/3	59.8	98.7	15.5	79.4	31.2
0.125	0.125	3/3	54.8	99.1	13.5	77.3	21.6
0.125	0.125	3/4	57.5	98.9	13.5	77.3	26.4
0.125	0.125	4/4	52.6	99.3	7.4	76.2	16.8
0.125	0.25	1/1	59.8	98.4	23.2	79.4	19.2
0.125	0.25	2/2	55.4	99.0	11.5	79.4	12.0
0.125	0.25	2/3	60.2	98.7	13.5	79.4	15.6
0.125	0.5	1/1	59.8	98.5	15.3	82.2	9.0
0.125	0.5	2/2	51.3	99.4	4.2	73.3	3.6
0.125	0.5	2/3	60.3	98.9	4.2	73.3	6.6

5.5 Discussion

This study investigated a novel recognition algorithm for freezing of gait episodes in patients with Parkinson’s disease. Our contribution includes the use of transfer learning and addresses the challenge of training models for new patients who provide limited freezing data. The Deep Gait Anomaly Detection (DGAD) algorithm developed in this study uses a binary class dataset of normal gait (NG) and freezing of gait (FoG) from a group of participants to train a reference anomaly detector capable of differentiating infrequent FoG episodes from NG. The reference model was then fine-tuned using only NG data from a new participant to obtain a personalized target model for each new participant. We implemented a leave-one-out cross validation (LOOCV) approach to investigate the performance of different reference and target models. Although all seven participants in this study experienced FoG during data collection, we excluded FoG samples from training and validation sets during the retraining step to simulate the situation that the new participant had not shown FoG symptoms. However, we used both NG and FoG samples for testing the FoG identification accuracy. The results presented in Table 5.3 and Figure 5.5 show how retraining the models improves the false positive rate, which is essential for a real-time FoG detection algorithm.

This study is the first to propose a new method to train patient-dependent models using only normal gait data from the target patient. The conventional methods fail to train models for patients with only a single class (i.e., NG) and have to use patient-independent models, similar to the reference models in this study. Table 5.9 compares the performance of target models developed using DGAD algorithm with three patient-independent models in the literature. Although the sensitivity of DGAD model is lower than the others, the specificity is higher. Considering the higher ratio of the normal gait events compared with FoG episodes (Table 5.2) it is important to improve specificity to keep the false positive rate low and prevent adaptation to the cues. Thus, the results suggest superiority of DGAD algorithm for a real-time FoG cueing device for patients who do not show FoG symptom during the data collection experiments.

In developing this algorithm, we manipulated two independent variables: the number of pre-FoG windows included in the FoG class, and data augmentation. The results presented

in Table 5.4 suggest that augmenting data and adding four windows of pre-FoG (using D4-A dataset) can slightly improve the average specificity from 98.3% to 98.6%. The obtained sensitivity and specificity of the models are lower than those obtained in [23] using ensemble classifiers (99.54% and 99.96%), and [28] using a deep convolutional neural network (99.64% and 99.99%). Figure 5.12 compares the performance of patient-dependent models with those obtained using state-of-the-art methods. However, these studies used samples from both NG and FoG classes to train patient-specific models, which improves the ability of the models to discriminate between the two classes but fails to accurately discriminate FoG from NG in individuals who do not show the symptom during the data collection sessions. The DGAD algorithm outperforms the results achieved by Pham *et al.* using anomaly score detector techniques with adaptive thresholding to identify FoG episodes (89% and 94% for sensitivity and specificity, respectively) [86]. Although DGAD achieved lower sensitivity, the algorithm was still able to identify 87.4% of FoG onsets.

The results demonstrate that augmenting training data impairs FoG recognition accuracy (sensitivity), but improves classification of normal gait (specificity). The algorithm predicted and detected 21.9% and 65.5% of FoG episode within 1 second prior and 2 seconds post FoG onset, respectively, summing up to 87.4% total identification of FoG onsets. This method improved the average FoG detection latency (0.1 ± 0.1 second) compared with [23] (1.1 ± 0.7 second). Palmerini *et al.* also discriminated pre-FoG episodes from normal gait and achieved an 83% prediction rate for the FoG episodes within 2 seconds prior to onset [67]. This comparison suggests that adding pre-FoG segments to the FoG class enables models to discriminate pre-FoG from normal gait and improves FoG identification rate.

The results presented in Table 5.6 confirm that using a more conservative approach with a higher specificity threshold (θ_{spec}) to terminate re-training of the reference model results in higher specificity for the testing set. This specificity is essential in real-time FoG cueing applications to prevent adaptation to the stimuli. A higher θ_{spec} allows more epochs for model training; this causes features to become more compact around the normal gait centroid in the 2D space. Thus, the model outputs will demonstrate bias towards the normal gait class (improved specificity) and reduced ability to identify FoG episodes (lower Sens, Pred, and Iden).

Table 5.9: Average FoG identification performance using conventional patient-independent models and DGAD target models.

	Sens	Spec
Camps <i>et al.</i> [27]	91.9	81.5
Xia <i>et al.</i> [28]	74.4	90.6
Mazilu <i>et al.</i> [23]	66.3	95.4
DGAD	63.0	98.6

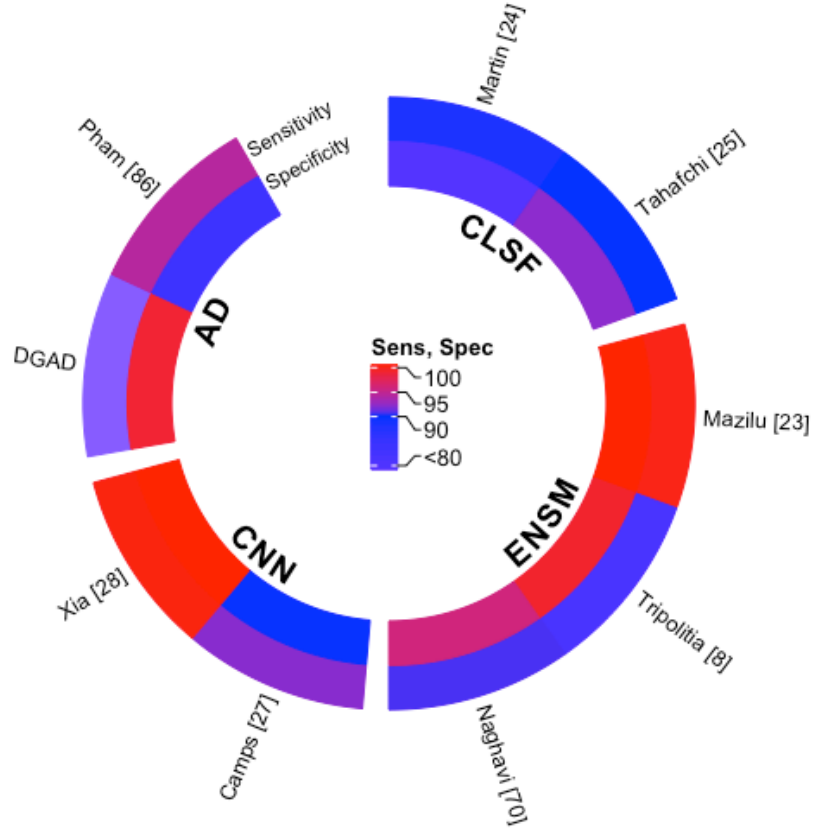


Figure 5.12: Circular heatmap of sensitivity and specificity obtained in FoG detection studies using patient-dependent models (CLSF: individual classifiers, ENSM: ensemble classifiers, CNN: convolutional neural networks, AD: anomaly detectors).

We also compared how the number of outputs can affect the model performance. The results did not reveal a fundamental improvement by increasing the number of outputs. Thus, we would suggest keeping the number of outputs as low as possible (equal to 2) to reduce the model parameters and improve the computation time on a mobile device with limited computation power.

We also investigated how a segmentation step size and using a median filter can improve model performance. The results suggest that using a smaller step size creates more samples for model training and slightly improves specificity, which is beneficial for a real-time FoG detection application. The median filter reduces the number of false positives and improves specificity while hampering the prediction and identification rate of the models. As we tend to limit the number of false positives to 10 per 5 minutes, we chose training and testing step sizes equal to 0.125 and 0.5 second, respectively, without using median filter on the output of the model (15.3% and 82.2% prediction and identification rates with 9.0 false alarms per 5 minutes). These values were used to train reference and target models for real-time FoG detection application described in Chapter 6.

5.6 Conclusions

The DGAD structure introduced in this study is composed of five layers of batch normalization, convolutional and max-pooling, followed by three fully connected layers. The deep convolutional network learns to extract relevant features from acceleration and angular velocity signals from ankle worn sensors during the training phase. The fully connected layers also learn how to map the data from the features space to a 2D space while reducing the intra-class and increasing the inter-class distances between centroids of the classes. Having such compact and distinct clusters of classes enables a kNN model to effectively classify the samples. The reference model requires data from both classes to learn discriminative features between normal and anomalous gait, but needs only normal gait data to refine the weights and biases for each individual participant.

This study also investigated the effect of augmenting (mirroring) signals collected from left and right ankles on the classification accuracy of gait data. The results suggested

that using augmented dataset for training the models resulted in reduced incorrect FoG identification (false positive) and improved specificity. The results also demonstrated that labeling segments prior to the FoG onset as FoG, improved correct FoG identification (true positive) and sensitivity.

This study demonstrated FoG prediction capability of the Deep Gait Anomaly Detection (DGAD) algorithm for new participant without having their anomalous gait data. However, the reference model requires data from both normal and anomalous gait states and having a larger dataset of more participants may improve the ability of the model to learn discriminative features and, ultimately, improve the accuracy of gait classification.

Chapter 6

Mobile Application Development

6.1 Introduction

In Chapters 3, 4, and 5 we described the development of algorithms to identify FoG episodes offline, i.e. as a post-process after data collection. We discussed how we improved the accuracy of the models in classification of both freezing and normal gait classes (sensitivity and specificity), and the ability of the model to predict FoG onset before its occurrence. In this chapter we describe the design requirements and development steps of the real-time FoG identification and cueing system. We implemented the FoG identification algorithm in an Android application which collects data from sensors, analyzes the received data, and triggers personalized cueing for the user.

6.2 Real-time PDFoG System Development

We have developed a novel system to address FoG in PD in ambient settings. The system collects data from patients with PD, analyzes data in real-time, and provides vibrotactile and auditory cueing through a smartwatch. The FoG detection algorithm implemented in an Android phone segments and classifies the collected data via IMUs worn on the users' body. If a segment is identified as FoG, the phone triggers user-specific cues on the watch. An author-developed application installed on the watch aggregates the cueing parameters (frequency, magnitude, and duration of the event) from the phone and triggers stimuli. The

users are instructed to synchronize their steps with the stimuli frequency. Magnitude and duration of the stimuli are customized to ensure participant sensitivity.

6.2.1 Design Requirements

Visual and auditory cueing interventions require patients to wear smart glasses or laser generators to trigger the feedback. Studies using glasses have shown that smart glasses may block the view of the lower limbs and cause an increase in FoG. In addition, blockage of the visual field has shown to decrease step length, velocity and cadence [126]. Vibrotactile feedback, on the other hand, can be provided via wristbands or smartwatches and easily adopted into routine daily activities. In addition, audio-visual processing shares many cognitive resources used for attention-dependent tasks and balance, while vibrotactile feedback may be processed subconsciously and, thus, faster [127]. Hence, vibrotactile could be a more efficient modality to overcome balance and gait dysfunction in PD patients by reducing the required executive control during complex motor tasks (i.e. dual-task walking) which require higher level of motor-cognitive integration and neuromuscular control. In this study, we selected vibrotactile through a smartwatch as the primary cueing modality. Users can also activate auditory cueing through the smartwatch as a secondary modality if they are not able to feel the vibrations effectively. We expected that using a cell-phone to analyze data and provision of cues through a watch reduces the burden for the patients to carry extra devices as everyone is carrying a phone and wearing a watch during daily routine activities nowadays. The usability and easiness of using the device will be evaluated at the end of the trials by asking the participants to fill a questionnaire.

In order to realize real-time FoG identification, we used wearable non-invasive Shimmer3 IMUs (www.shimmersensing.com) with integrated 9 DoF accelerometer, gyroscope and magnetometer, each with selectable range. The Shimmer3 IMUs support both logging data to SD card and streaming via Bluetooth, which is essential for real-time applications. The Android application first filters the received data from the sensors and logged the filtered signals as a text file on the phone's internal memory for future debugging. The application then analyzes the data and classifies gait into **Stop**, **NormalGait**, and **FoG**, and stores the results in a separate text file. The text files' names are in `yyyymmdd.hhmm` format so they

can be easily sorted by date/time and cross-referenced with the logs that participants were asked to keep in the provided notebooks.

The application for the Android watch needed to be as simple as possible with only one button on the screen as the screen is small and clicking on buttons would not be easy for the participants, especially for those with tremor symptoms. The application should also allow the user to check the status of the phone application (FoG detection running or stopped).

The phone and watch applications need to remain active in background if the user turned off the screen or left the application by pressing the home button on either of the devices. The Service component in Android enables the applications to perform a longer-running operation in background while not interacting with the user. Thus, both applications need to be written in a way that starts and stops the service when the user starts and stops the FoG detection component of the app, respectively.

6.2.2 Android Application Development

The phone application was formed of two main components: 1) **connect/disconnect** component allows the users to connect the devices to the phone when they want to start the application, and also disconnect the devices when they want to charge the devices. 2) **gait monitoring** component allows the users to start/stop the FoG detection.

The first version of the phone application was developed using the methods presented in Chapter 4. Patient-dependent classifiers were developed for each participant using the data collected during experiments in the clinic. We used Weka (www.cs.waikato.ac.nz/ml/weka) to train classifiers on a Windows operating computer and, then, imported the trained models into the Android application. We also used a publicly available filter library in Java (github.com/berndporr/iirj) to band-pass filter the received signals. The application was developed in IntelliJ IDEA (www.jetbrains.com/idea), an integrated development environment (IDE) written in Java.

6.2.3 User Interface

The buttons on the phone application menu are placed from top to down in the same order as the user needs to click and run each component of the application following the provided instructions. Figure 6.1 shows the phone menu when the user runs the application on the phone. More information about the software and the menu can be found in the instructions document provided for the participants (Appendix E). Table 6.1 lists the phone, watch and sensors models used in the PDFoG System.

Connect to Bluetooth Devices

The phone application requires participants to, first, connect IMUs and the provided smart-watch to the phone. Users can connect the Bluetooth device to the phone by pressing **CONNECT** button on the phone screen (Figure 6.1). The sensors automatically connect to the phone. In order to connect the watch to the phone, user needs to press the **CONNECT** button on the watch screen. Color-codes are used to show the connection status of each Bluetooth device on the phone screen (yellow: disconnected, green: connected). Red and green colors are also used on the watch screen to show the gait monitoring component status (red: FoG detection stopped, green: FoG detection running). The users can also activate cues manually using the **TEST WATCH** button on the phone screen.

Freezing of Gait detection

When all the sensors and the watch are connected, the user can start gait monitoring component (FoG detection) by pressing the **START** button. The data will then start to stream from the sensors to the phone and also logged on the sensors' internal memory. The FoG detection algorithm uses data from ankle-worn sensors to detect freezing. In the first version of the application, we used five sensors (on left and right ankles, left and right wrists, and waist) while the data from the wrist and waist sensors were collected only for future analyses (Figure 6.1).

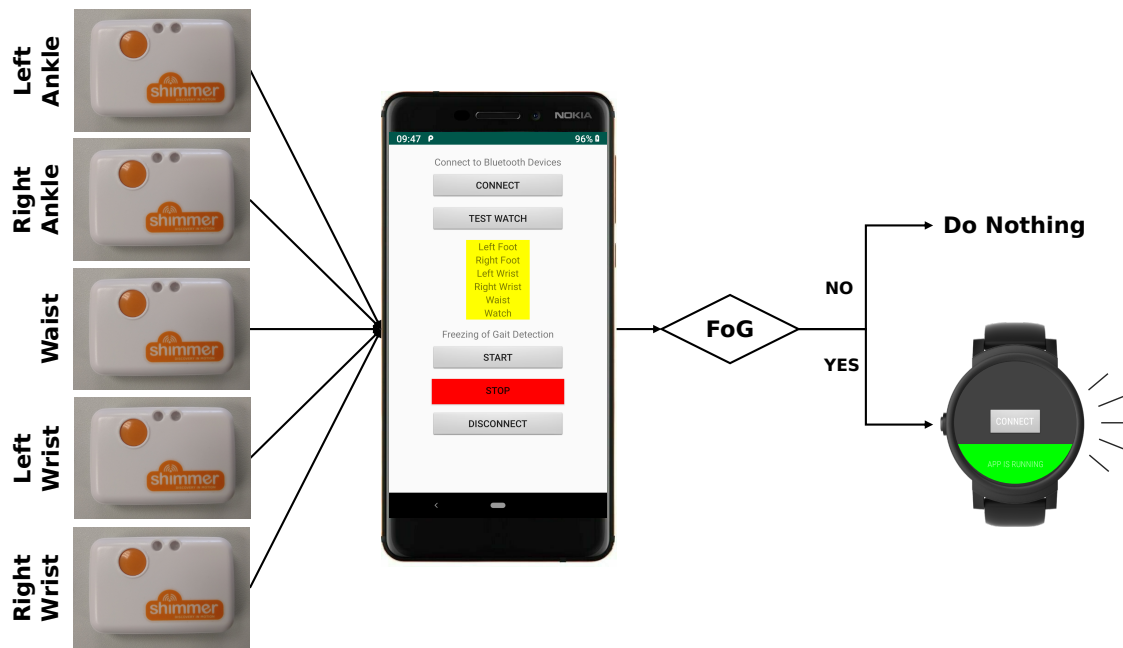


Figure 6.1: PDFoG system (version 1).

Table 6.1: Equipment model used to realize PDFoG system.

Device	Model
Phone	Nokia 6.1
Watch	Ticwatch E
IMU	Shimmer3 IMU

The **STOP** button terminates data streaming and logging, but keeps the sensors and the watch connected to the phone so that the users can start the application again without reconnecting the sensors. The participants are instructed to disconnect the phone (by pressing the **DISCONNECT** button) from the sensors and the watch when they want to charge the devices every night. The user manual (Appendix [E](#) explains the application menu and options).

6.3 Computational Complexity

The DGAD algorithm was implemented into the Android application running on a Nokia 6.1 cellphone configured with Octa-core Cortex-A53 CPU @2.20GHz, Qualcomm Adreno 508 GPU, and 3GB RAM. The average processing time for the model was 16 ± 5 ms to classify the state (NG or FoG) for each segment of the data.

6.4 Prototype Evaluation

We recruited two pilot participants for a 2-week investigation to evaluate the performance of the FoG detection models and the usability of the system. We found that one participant did not experience freezing during the tests; thus, training patient-dependent classifiers was not possible. Classical one-class classifiers like OCSVM also showed poor specificity and were not ideal for cueing application. For this participant, we used a patient-independent classifier in the Android application. This showed low specificity and resulted in user complaints regarding frequent false alarms. Another complaint from the participants was about the number of sensors they were required to wear, which discouraged one participant from continuing the in-home investigation. One participant also suggested adding auditory cueing as an option as the watch needed to be worn tightly to feel the vibrations. We added one switch key (**Audio**) to the menu so that the users can easily activate/deactivate auditory cues, i.e. beeps on the watch with the same pattern as the vibrations. They have access to this switch even when the gait monitoring component is running.

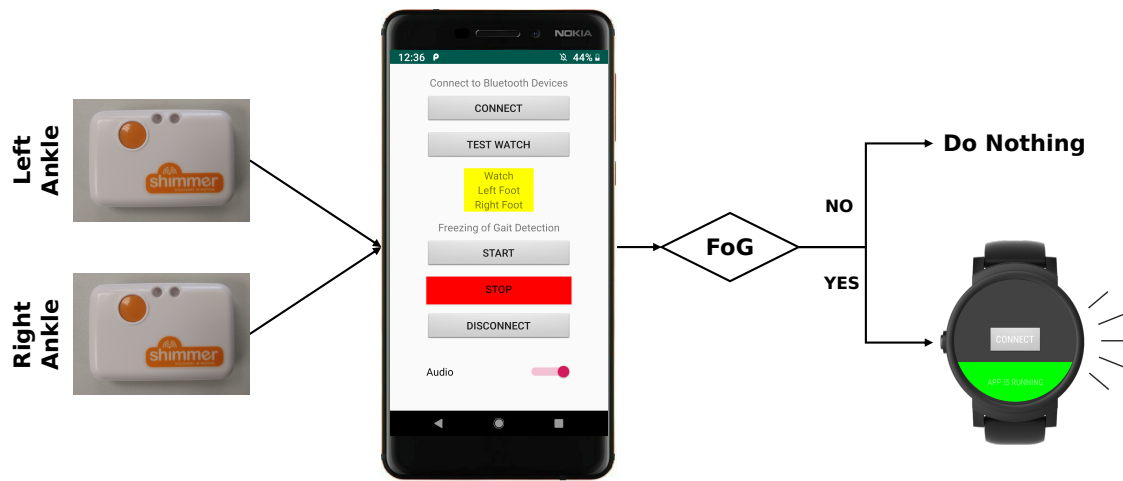


Figure 6.2: PDFoG system (version 2).

We modified the FoG detection algorithm using the methods described in Chapter 5. The second version of the application uses TensorFlow Lite (www.tensorflow.org/lite) an open source deep learning framework for on-device inferences, to analyze the received signals from accelerometer and gyroscope sensors and detect FoG in real-time. The current version also requires only two sensors worn on the ankles to reduce the burden for the participants (Figure 6.2).

6.5 Conclusion

In this chapter we explained the Android applications developed for real-time FoG prediction in ambient environments. The DGAD algorithm (introduced in Chapter 5) was implemented in an Android application to monitor gait patterns of PD patients in their homes. The computational complexity analysis showed that the phone is capable of analyzing the signals received from wearable sensors in 16 ± 5 ms which is sufficiently less than the step size in data segmentation algorithm, i.e. 500 ms, and ensures no delay in data analysis process. The participants will receive vibrotactile and auditory cues through a smartwatch (connected to the phone via Bluetooth) if an FoG episode is identified. The participants can enable/disable auditory cues based on their preference.

Chapter 7

Cueing System Evaluation: In-home Testing

7.1 Introduction

In Chapter 6 we discussed the Android application developed for real-time identification of FoG episodes. The software on phone collects data from two IMUs worn on the ankles and monitors the gait patterns. When the normal pattern is violated, the phone triggers vibrotactile and auditory cues on a connected Android watch. In this chapter we discuss the experiment design to evaluate the effectiveness of the designed cueing system in reducing the freezing frequency and severity in ambient environments. The system was validated by a 6-week in-home study, as FoG occurs more frequently at home than in clinical settings [35]. The results will demonstrate the extent to which an in-home intervention using automated, on-demand, contingent cueing can affect the frequency and duration of FoG episodes.

7.2 Study Design

The conducted research consisted of a longitudinal, within-subject intervention study. Each participant's involvement took seven weeks, which included: an initial assessment and gait data collection (day 0) at the University of Tennessee Brain and Spine Institute (UT-BSI);

a 6-week in-home intervention using the PDFoG system; and, an immediate follow up assessment. Figure 7.1 illustrates the study design.

7.2.1 Participants

A sample of 37 participants is required for a matched-pairs t-test with medium effect size ($f = 0.42$) on Freezing of Gait Questionnaire [52], 80% power and type I error rate of 5%. Past experience suggested a 25% attrition rate for such participants, so 46 community dwelling participants with PD were needed according to the standards of the UT Institutional Review Board (IRB). Along with Michelle Brewer, MD (neurologist), the research team oversaw the recruitment of participants with inclusion criteria: 1) a diagnosis of PD; 2) a Hoehn & Yahr (H&Y) ¹ Stage of 2–4 ; 3) FoG symptoms; and, 4) stable PD medication for the duration of the study. Exclusion criteria included: 1) individuals receiving deep brain stimulation; 2) sufficient pain to prevent task completion during the clinical data collection session; 3) participation in an ongoing exercise program; and, 4) clinical diagnosis of cognitive impairment as measured by Montreal Cognitive Assessment (MoCA score < 24, Appendices A and B).

7.2.2 Outcomes Measures

Gait and Fall Questionnaire (GFQ)

The primary outcome measure was the mean change in the total Gait and Fall Questionnaire (GFQ, Appendix C) pre versus post in-home cueing therapy. GFQ includes 13 questions about FoG symptom (frequency and severity of FoG symptom when initiating gait, turning, terminating gait, and walking through narrow spaces) and 3 questions about fall (when walking, standing, and freezing).

¹Named for its authors, H&Y is a commonly used scale for describing the severity of the symptoms in patients with Parkinson’s disease. It was originally published in 1967 and included stages 1 through 5. Since then, a modified H&Y scale was proposed with the addition of stages 1.5 and 2.5 to help describe the intermediate course of the disease.



Figure 7.1: In-home FoG cueing intervention study design.

Usability Questionnaire (UQ)

We also asked the participants to fill a usability questionnaire (Appendix D) at the end of the sixth week of the in-home study to evaluate the usability of the technology (hardware and software) in ambient environments. The usability questionnaire is designed based on the *Tele-healthcare Satisfaction Questionnaire – Wearable Technology* (TSQ-WT) which was developed to measure the satisfaction of users about using wearable technologies. TSQ-WT consists of six sections evaluating the benefit, usability, self-concept, privacy and loss of control, quality of life, and wearing comfort. Each section includes five items rated on a 5-point scale (0–4) with higher scores indicating more positive ratings. We added 16 more questions to the usability (4 questions) and wearing comfort (12 questions) sections to ask for more detail about each part of the system. We also provided the participants with a notebook and asked them to keep logs of the tests, record the start and end time of each session, and any major issue/improvement they (or their partners) could observe/suggest.

7.2.3 Participant Recruitment

Due to complications caused by COVID-19, we were able to recruit only ten (6M/4F, 72.6 ± 8.4 years, H&Y 3-4) participants. Table 7.1 lists the demographic information of the participants. Among the recruited participants:

- 4 participants (3M/1F, 69.3 ± 5.1 years, H&Y 3.5 ± 3.6 , and 13.8 ± 5.6 years of disease diagnosis) finished the 6-week in-home study and used the system for 4–6 hours every day (except the weekends and holidays).
- 1 participant had the system for six weeks but used it for only 4 days.
- 5 participants quit shortly after the beginning of the in-home tests as they couldn't use the software on the phone, or they were not able to use the system for the entire period of six weeks.

Table 7.1: Demographic information of the participants (bolded rows show participants who finished the in-home study).

Participant	Gender	Age (year)	H&Y	Disease Duration (year)
1	M	70	4	22
2	M	55	3	14
3	F	79	4	10
4	M	71	3	11
5	F	80	3	9
6	M	80	3	14
7	M	62	3	12
8	F	76	3	18
9	M	79	3	10
10	F	74	4	10

7.3 Results

7.3.1 Gait and Fall Questionnaire (GFQ)

In this section we discuss the pre-post analysis of the GFQ scores for the participants who finished the in-home study. As we were not able to recruit the required number of participants (46), we used a non-parametric Sign Test (an alternative to the matched-pairs t-test) to investigate the effectiveness of the intervention using the data collected from the GFQ. Average GFQ scores for pre and post in-home trials were 28.5 (21 to 37) and 25.5 (18 to 33), respectively, while the maximum possible score was 64. Average GFQ scores for gait related questions (1-11, 15, and 16) were 25 (20 to 31) and 22.5 (18 to 28) with 52 being the maximum possible score. Average GFQ scores for fall related questions (12-14) were 3.5 (1 to 6) and 3 (0 to 5), respectively. The Sign tests showed that there was not a statistically significant difference in GFQ score pre and post using the PDFoG system for six weeks, $p = 0.125$, despite the reduction in overall GFQ score for all participants. The Sign tests also showed no statistically significant difference in gait and fall related questions pre and post intervention, $p = 0.125$ and $p = 0.625$, respectively, despite the overall reduction in both categories of questions.

7.3.2 Usability Questionnaire (UQ)

In this section we discuss the comments we received from the participants about the sensors, phone, watch, and the software on the phone and watch. Table 7.2 compares the average scores for each section of the questionnaire. The average values are divided by the maximum possible score for each section to normalize the values between 0 (most unfavorable) and 4 (most favorable).

7.3.3 Observations

In this section we discuss our observations and suggestions on how future studies can be designed to improve participants motivation and knowledge of the technology.

Table 7.2: Scores for each section of the Usability Questionnaire. The scores show normalized scores (total score divided by the number of questions in each section).

Section	Average	Min	Max
Benefit	1.65	1.00	1.80
Usability	2.03	1.67	2.11
Self-concept	2.20	1.80	2.60
Privacy and Loss of Control	2.80	2.20	3.00
Quality of Life	2.05	1.20	2.60
Wearing Comfort	1.59	1.06	1.88

1. The first issue with our participant recruitment was caused by the spread of COVID-19 virus. From the list of 20 potential participants that we gathered with the help of Dr. Michelle Brewer, only 13 patients accepted to participate in the study and the others were reluctant to participate amid the coronavirus pandemic.
2. From the recruited participants (13), one was feeling sick at the time of clinical data collection and was sent to a hospital to test for COVID-19. He refused to come back later and was excluded from the study.
3. Two participants came for the initial assessment (Day 0), but quit before starting the in-home experiments as they were not experiencing FoG frequently and thought they may not benefit from their participation.
4. In the application user-interface design we used green and red color codes to show the connection status of the Bluetooth devices (watch and sensors). One participant suggested that might be a problem for color-blind users.
5. Two participants complained about false positives when they were walking up/down the stairs. Walking on stairs was not included in our clinical data collection tasks and, due to a different pattern with walking on a flat surface, may be detected as anomaly (i.e., freezing of gait).
6. The main reason for stopping the participation was the lack of knowledge of the cellphones and smartwatches. Although we explained every step in the manual document and recorded videos on the phones, a large portion of the participants lost their interest and motivation as they were not able to follow the instructions step by step. Although this was predictable in the PD population who suffer from cognitive impairment and are not used to using smartphones, the attrition rate (60%) was higher than our expectations.
7. In one case, the participant was not able to comprehend the instructions to run the experiments at home, but his wife, who used to work in IT sector, helped him throughout the in-home experiments and never faced a technical issue. In another

case, the participant had an engineering degree and was using the same type of wearable sensors in their studies on animals. He was the most enthusiastic participant in the study and finished the in-home experiments without any technical difficulties. These two cases suggest that the participants at this age and with impaired cognition require an assistant (partner, child, or caregiver) with some level of familiarity with smartphones and wearables to help them run the experiments independently and without researchers' supervision.

8. Patients were taking their medication before performing the clinical data collections; i.e., they were in their ON medication state. This would have caused less FoG episodes and more stable walking patterns. While using the system in their home, one participant who was showing different walking patterns when OFF medication or when walking in their home finding their way through multiple obstacles complained about the frequent false alarms (false positives). The issue was resolved by adding a median filter to the output of the model to reduce the false positive rate.

7.4 Discussion

7.4.1 Gait and Fall Questionnaire

The GFQ results shows that the 6-week in-home tests using the PDFoG system will improve gait and fall symptoms, and eventually improve the overall score. However, the therapy has not been significantly effective in treating the FoG symptom, nor the fall likelihood. A post-hoc analysis showed that the achieved power with the sample size of 4 participants is 30%. Thus, a larger population size is required to achieve the desired power.

7.4.2 Usability Questionnaire

The UQ results shows that the participants did not feel that they were benefiting from using this device (average score of 1.65 from 4). The average score for the Usability and Wearing Comfort (2.03 and 1.59, respectively) show that the participants were not facing difficulties using the device, except for wearing the watch and the sensors. This highlights the necessity

of a new design for the sensors as the first priority. One problem was reported to be the bands of sensors becoming loose after a period of wearing the sensors. The other common complaint from the participants was frequent disconnection of the sensors from the phone which eventually caused the application to stop. The disconnection could be due a hardware or software issue, or the participants leaving the phone in a distant location from the sensors. The UQ results also suggest that the participants were confident about their privacy and felt comfortable using a new technological product (2.80 and 2.20 average scores, respectively). The participants could not agree or disagree whether using this device can improve their quality of life (average score of 2.05).

7.4.3 Observations and Comments

The logs received from the participants show that the sensor–phone disconnection was a frequent complaint that happened almost every day for all of the participants. It was also reported that in many cases, the sensors failed to connect to the phone. In such cases, the producer company (Shimmer) suggested restarting the sensors would resolve the issue, however, it took more than one restarting in a few cases before the sensors could connect to the phone. The participants also complained about false alarms (i.e., false positives) when they were seating on a couch or standing still with limited movements of feet, although we had asked them to stop the application if they are not walking for a period of 5 minutes. This issue can be resolved in future developments of the system by adding an activity recognition algorithm before the FoG detection model. If recognized as walking, the data is fed to the FoG model, otherwise the FoG detection algorithm would be by–passed and the patient’s state would be non–FoG (see Figure 6.1). This suggestion will be discussed in more detail in Chapter 8.

One patient mentioned in his logs that the system has been helping him to prevent freezing in the spaces where he used to freeze (like walking through doorways), which highlights the capability of the device in predicting FoG and delivering cues before FoG occurrence. One other patient also stated that wearing the device has informed him about his posture and that he has been able to improve his balance using the device. This suggests that using such

devices can remind patients of their symptoms so they can adjust their motor behavior to improve gait and maintain balance.

Using the anomaly detection approach to identify FoG, characterizes the data segments as normal or anomaly (i.e., FoG). Accurate classification of signals requires inclusion of different tasks in the normal class, otherwise those incidents would be identified as anomalies and cause false alarms. We recommend including walking on stairs (going up and down), and walking with varied velocities (normal, lower, and higher) in the future studies.

7.5 Conclusion

In this chapter we explained the system evaluation study design to investigate the effectiveness of the designed on-demand cueing system on the FoG severity. We used standard clinical assessment (Gait and Fall Questionnaire) to identify the frequency and duration of the FoG symptoms as well as fall likelihood during a set of tasks performed routinely at home (e.g., walking through narrow spaces, initiating and terminating gait). we also used a Usability Questionnaire, designed specifically for wearable devices, to measure usability, wearability, and using comfort of the system (including phone, watch, and sensors).

We recruited 10 participants for a 6-week in-home study, although only 4 finished the study. The pre-post analysis of the GFQ results showed that in-home cueing therapy using the PDFoG system has not been significantly effective in treating the FoG symptom and fall likelihood, although the overall score was lower post intervention for all participants. We will discuss the suggestions for future studies in the next chapter.

Chapter 8

Conclusions and Future Works

8.1 Conclusions

Freezing of gait (FoG) is a common motor impairment in individuals with Parkinson's disease (PD). FoG impairs walking and is associated with increased fall risk. Although pharmacological interventions have shown promise during ON-medication periods, FoG remained difficult to treat during the medication OFF state and in advanced stages of the disease. External cueing systems are recent innovations providing stimuli that help patients overcome freezing. Continuously cued therapy has been effective in treating gait deviations, although it induced greater levels of fatigue in PD patients and, thus, is found unfavorable to the majority of the patients. Intelligent cueing systems, however, analyze gait signals through wearable sensors and deliver cues only at moments when specific gait alterations are detected. Such on-demand external feedback does not necessarily become part of the movement representation and may decrease the chance of development of cue-dependency.

Providing cues *after* detection of an FoG episode may result in cognitive overload, by superimposing an external stepping rhythm, and aggravate FoG. On other hand, predicting FoG *before* onset enables delivery of preemptive cueing and may prevent FoG. In this study we developed methods to predict FoG and improve FoG recognition accuracy. We recruited participants with PD and collected their gait data in a clinical environment. Using the collected data, we trained and tested algorithms for FoG prediction. We also implemented the algorithms into an Android application and provided the participants with a set of

sensors, watch, and phone to tests the effectiveness of the developed FoG cueing device in their homes. The details have been discussed in different chapters and the following highlights the main outcomes of each chapter.

In Chapter 3 we used an online available dataset to investigate the possibility of detecting FoG using statistical inference between multiple successive windows of signals. We extracted features from the acceleration sensors and hypothesized that the change of pattern in the feature values is a marker of FoG. We compared the results obtained using different sensor position, axis, sampling window length and features. The results suggested that using sample entropy of acceleration collected from a single sensor on the ankle enables 88.8% predictivity, 92.5% sensitivity, and 89.0% specificity.

In Chapter 4 we used machine learning algorithms to discriminate FoG episodes from normal gait. We used the extracted features from acceleration signals to train individual and ensemble classifiers in patient-dependent and patient-independent paradigms. We also used ADaptive SYNthetic (ADASYN) resampling algorithm and cost sensitive classification to improve the automatic detection of FoG episodes. The ensemble classifier using bagging techniques demonstrated highest performance in patient-dependent mode when synthetic FoG samples were added to the training set and FoG misclassification cost was set as twice that of normal gait. The model identified 97.4% of the FoG onsets, with 66.7% being predicted.

The detection and prediction still remained challenging particularly for individuals who did not exhibit sufficient instances of freezing episodes during the clinical trials for model training. To address this fundamental issue, a Deep Gait Anomaly Detector (DGAD) was introduced in Chapter 5. The algorithm implements a transfer learning-based approach to train a reference model using FoG and normal gait samples of participants who exhibited freezing during the experiments. The convolutional neural network is trained to extract features from acceleration and rate gyro signals and then reduce the dimensionality of the features in a way that FoG and normal gait data have longest Euclidean distance. The model is then retrained using only normal gait data from a new patient to form a patient-specific model. The effect of augmenting training data and adding pre-FoG segments (data occurring prior to the freezing onset) to the FoG training data was also investigated. The

DGAD algorithm demonstrated average sensitivity and specificity of 63.0% and 98.6% using augmented data and additional pre-FoG segments. The target models identified 87.4% of FoG onsets, with 21.9% being predicted on average for each participant. This study demonstrates our algorithm’s potential for accurate identification of anomalous gait episodes and delivery of preventive cueing for patients for whom no FoG data is available to train models.

In Chapter 6 we explained the Android applications developed for real-time FoG prediction. The DGAD algorithm was implemented in an Android application to monitor gait patterns of PD patients in their homes. The phone triggers vibrotactile and auditory cues through a smartwatch (connected to the phone via Bluetooth) if an FoG episode is identified.

A 6-week in-home study was designed in Chapter 7 to evaluate effectiveness of the cueing system on FoG severity in ambient environments. The study required 18 PD participants, but only 10 were recruited due to complications caused by the spread of COVID-19 virus. Among the recruited participants, only 4 were able to finish the 6-week study. The nonparametric statistical analysis revealed no significant improvement in FoG and fall at the end of the in-home experiments. However, a larger study is still required to reach the required statistical power to show the effectiveness of intervention. The participants’ responses to the usability questionnaires showed that the sensors caused frequent problems for the participants, which was the main reason for participants to quit in the middle of the study.

8.2 Future Works

8.2.1 Recurrent Neural Networks

In Chapter 5 we introduced a convolutional neural network (CNN) that was trained to extract features from the signals. The model also had three fully connected layers to reduce the dimensions of the features while reducing the intra-class and increasing the inter-class distances between samples of normal gait (NG) and freezing of gait (FoG). Using data augmentation methods, we were able to improve the FoG prediction rate. Another approach

to further improve FoG prediction would be to use recurrent neural networks. These networks have been widely used in text prediction and natural language processing (NLP) [128–132], and time-series forecasting [133–135]. Long-short-term-memory (LSTM) networks have also been used in detection of FoG episodes [68]. The suggested structure of an LSTM–DGAD FoG predictor is shown in Figure 8.1. The first row shows the current structure of the DGAD network and the training algorithm. The network is trained to reduce the Mean Squared Error (MSE) between model output and the desired output ([1,1] for NG, and [-1,-1] for FoG) for the current time point (t^*). The second and third rows show the LSTM layers training process. The convolutional layers are fixed (non-trainable) and only the LSTM layers are training. The features extracted from the last n time points ($t^* - n, t^* - n + 1, \dots, t^*$) are used as input to the first LSTM layer and the features from the next time step ($t^* + 1$) are used as the true output of the model. The LSTM layers will be trained to predict the features for the next time step using the previous n steps. The last row shows the final LSTM–DGAD structure in which all the layers are fixed. The network extracts features from the raw signals (using convolutional layers), predicts features for the next time step (using LSTM layers), and, finally, reduced the dimensions (using fully connected layers). A conventional classifier (SVM or kNN) will be used to classify the reduced features as NG or FoG. The LSTM layers can also predict features for multiple future steps, which enables activation of cues longer before FoG occurrence.

8.2.2 Activity Recognition

Another major complaint about the PDFoG software was about frequent false alarms while patients were seating or standing with only limited movement. Currently, an acceleration threshold is being used to filter out the low activity motions and identify stop periods. However, in order to better address the non-gait false alarm issue, we suggest using an activity recognition algorithm to see if there is enough walking related motion in the received signal. If not, the FoG detection algorithm will get bypassed, i.e., the model won't analyze the received data and cues won't be activated (Figure 8.2). Although, this approach may disable the system to accurately identify FoG at gait initiation, this may reduce false alarm rate and keep the participants less bothered and more motivated.

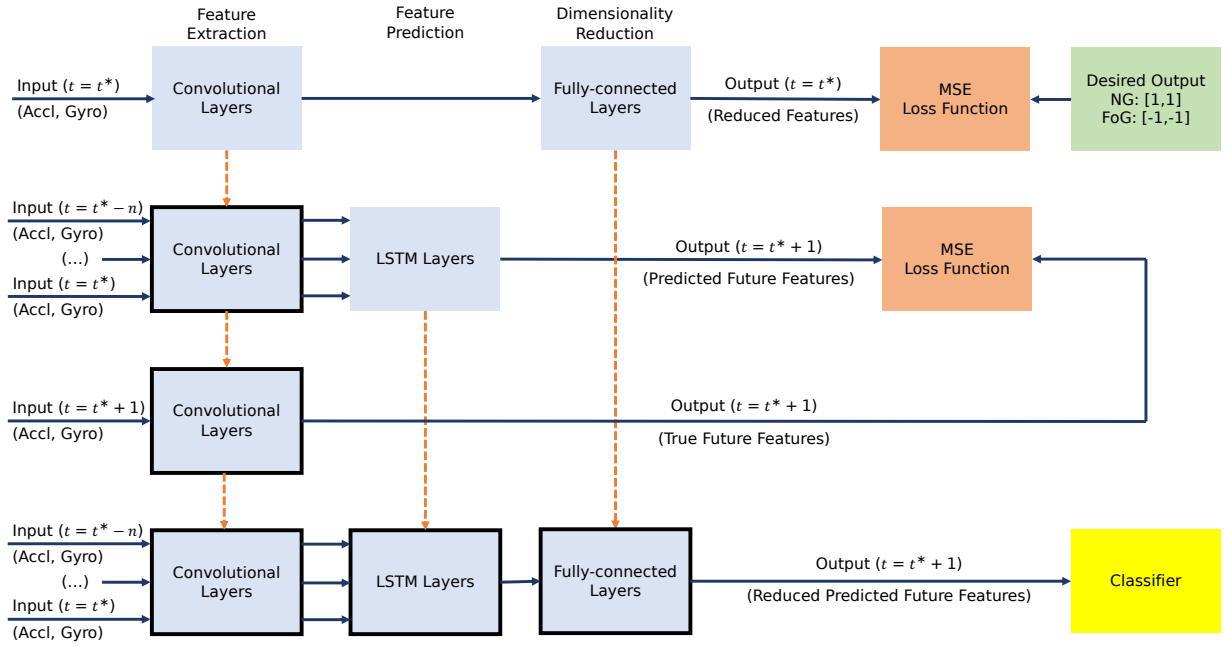


Figure 8.1: Structure of the suggested LSTM-DGAD network. First row shows training process for the current DGAD network. Second and third rows show training process for the LSTM layers. The fourth row shows the suggested network structure to predict FoG. Black solid outlines indicate that the layers are not trainable.

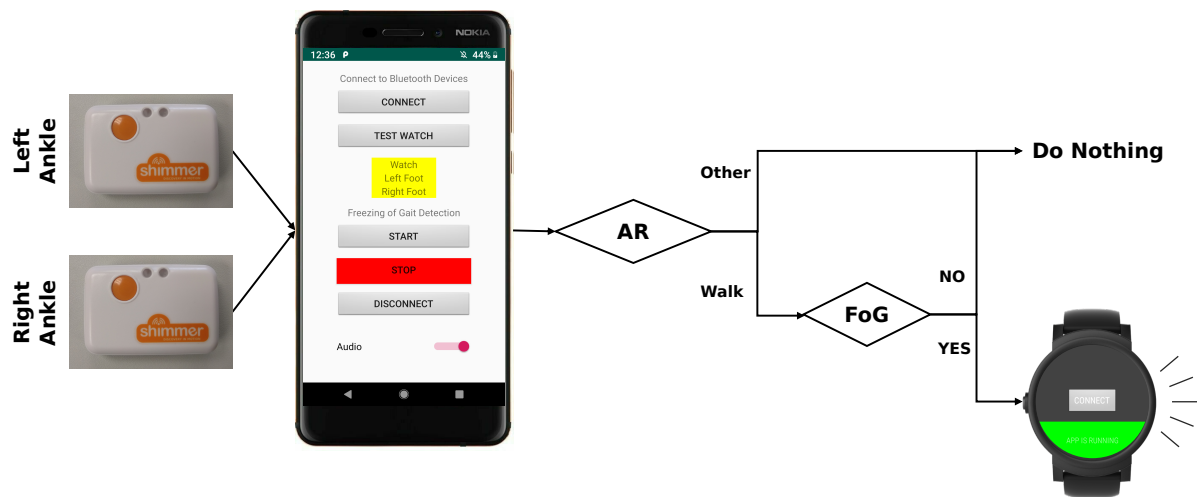


Figure 8.2: An activity recognition (AR) algorithm analyzes the received data to identify if the user is walking. The FoG identification algorithm analyzes the data only if the user is recognized as walking.

8.2.3 Larger Study

With the experiences obtained in this study, we aim to propose a larger study to further investigate the effectiveness PDFoG cueing system on the FoG symptom. We have been collaborating with a US based company that manufactures IMUs and develops software for fall prediction in patients with Parkinson’s disease. We hope that, using their sensors, we can address the complaints about the sensors (loose attachment, frequent disconnection, etc.). Also, with this company’s experience in software development for elderly people, we hope to reduce the required patients’ interaction with the software and improve user acceptance.

Bibliography

- [1] A. Nieuwboer, G. Kwakkel, L. Rochester, D. Jones, E. Van Wegen, A. M. Willems, F. Chavret, V. Hetherington, K. Baker, and I. Lim, “Cueing training in the home improves gait-related mobility in parkinson’s disease: The rescue trial,” *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 78, no. 2, pp. 134–140, 2007. [1](#), [7](#), [11](#), [15](#), [26](#)
- [2] S. Fahn, “Description of parkinson’s disease as a clinical syndrome,” *Annals of the New York Academy of Sciences*, vol. 991, no. 1, pp. 1–14, 2003. [Online]. Available: <http://dx.doi.org/10.1111/j.1749-6632.2003.tb07458.x> [1](#), [4](#)
- [3] G. DeMaagd and A. Philip, “Parkinson’s disease and its management: Part 1: Disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis,” *Pharmacy and Therapeutics*, vol. 40, no. 8, pp. 504–32, 2015.
- [4] A. Mirelman, H. Bernad-Elazari, A. Thaler, and et al., “Arm swing as a potential new prodromal marker of parkinson’s disease,” *Movement Disorders*, vol. 31, no. 10, 2016. [1](#), [4](#)
- [5] R. Pahwa and K. E. Lyons, “Levodopa-related wearing-off in parkinson’s disease: identification and management,” *Current Medical Research and Opinion*, vol. 25, no. 4, pp. 841–849, 2009, pMID: 19228103. [Online]. Available: <https://doi.org/10.1185/03007990902779319> [1](#), [5](#)
- [6] T. Xie, J. Vigil, E. MacCracken, A. Gasparaitis, J. Young, W. Kang, J. Bernard, P. Warnke, and U. J. Kang, “Low-frequency stimulation of stn-dbs reduces aspiration and freezing of gait in patients with pd,” *Neurology*, vol. 84, no. 4, pp. 415–420, 2015. [Online]. Available: <https://n.neurology.org/content/84/4/415> [7](#)

- [7] C. Huang, H. Chu, Y. Zhang, and X. Wang, “Deep brain stimulation to alleviate freezing of gait and cognitive dysfunction in parkinson’s disease: Update on current research and future perspectives,” *Frontiers in Neuroscience*, vol. 12, p. 29, 2018. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnins.2018.00029> 1, 5, 7
- [8] E. E. Tripoliti, A. T. Tzallas, M. G. Tsipouras, G. Rigas, P. Bougia, M. Leontiou, S. Konitsiotis, M. Chondrogiorgi, S. Tsouli, and D. I. Fotiadis, “Automatic detection of freezing of gait events in patients with parkinson’s disease,” *Computer Methods and Programs in Biomedicine*, vol. 110, no. 1, pp. 12–26, 2013. [Online]. Available: <http://dx.doi.org/10.1016/j.cmpb.2012.10.016> 1, 2, 9, 10, 20, 33, 35, 63
- [9] S. T. Moore, H. G. MacDougall, and W. G. Ondo, “Ambulatory monitoring of freezing of gait in parkinson’s disease,” *Journal of Neuroscience Methods*, vol. 167, no. 2, pp. 340 – 348, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0165027007004281> xii, 1, 4, 5, 6, 9, 14, 17, 19, 21, 33, 42, 69
- [10] N. Giladi, T. Treves, E. Simon, H. Shabtai, Y. Orlov, B. Kandinov, D. Paleacu, and A. Korczyn, “Freezing of gait in patients with advanced Parkinson’s disease,” *Journal of Neural Transmission*, vol. 108, no. 1. 1
- [11] S. T. Moore, D. A. Yungheer, T. R. Morris, V. Dilda, H. G. MacDougall, J. M. Shine, S. L. Naismith, and S. J. Lewis, “Autonomous identification of freezing of gait in parkinson’s disease from lower-body segmental accelerometry,” *Journal of NeuroEngineering and Rehabilitation*, vol. 10, no. 1, p. 19, Feb 2013. [Online]. Available: <https://doi.org/10.1186/1743-0003-10-19> 2, 21
- [12] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster, “Wearable assistant for parkinson’s disease patients with the freezing of gait symptom,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 436–446, March 2010. 2, 5, 9, 15, 16, 17, 21, 24, 33, 34
- [13] V. A. Goodwin, S. H. Richards, R. S. Taylor, A. H. Taylor, and J. L. Campbell, “The effectiveness of exercise interventions for people with parkinson’s disease: A systematic

- review and meta-analysis,” *Movement Disorders*, vol. 23, no. 5, pp. 631–640, 2008. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.21922> 2
- [14] C. L. Tomlinson, S. Patel, C. Meek, C. P. Herd, C. E. Clarke, R. Stowe, L. Shah, C. Sackley, K. H. O. Deane, K. Wheatley, and N. Ives, “Physiotherapy intervention in parkinson’s disease: systematic review and meta-analysis,” *BMJ*, vol. 345, 2012. [Online]. Available: <https://www.bmj.com/content/345/bmj.e5004> 2
- [15] G. M. Petzinger, B. E. Fisher, S. McEwen, J. A. Beeler, J. P. Walsh, and M. W. Jakowec, “Exercise-enhanced neuroplasticity targeting motor and cognitive circuitry in parkinson’s disease,” *The Lancet Neurology*, vol. 12, no. 7, pp. 716–726, 2013. 2
- [16] E. E. van Wegen, M. A. Hirsch, M. Huiskamp, and G. Kwakkel, “Harnessing cueing training for neuroplasticity in parkinson disease,” *Topics in geriatric rehabilitation*, vol. 30, no. 1, pp. 46–57, 2014. 2
- [17] L. Rochester, K. Baker, V. Hetherington, D. Jones, A.-M. Willems, G. Kwakkel, E. V. Wegen, I. Lim, and A. Nieuwboer, “Evidence for motor learning in Parkinson’s disease: Acquisition, automaticity and retention of cued gait performance after training with external rhythmical cues,” *Brain Research*, vol. 1319, pp. 103–111, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0006899310000041> 2
- [18] S. J. Spaulding, B. Barber, M. Colby, B. Cormack, T. Mick, and M. E. Jenkins, “Cueing and gait improvement among people with parkinson’s disease: A meta-analysis,” *Archives of Physical Medicine and Rehabilitation*, vol. 94, no. 3, pp. 562–570, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0003999312010842> 2
- [19] P. A. Rocha, G. M. Porfírio, H. B. Ferraz, and V. F. Trevisani, “Effects of external cues on gait parameters of Parkinson’s disease patients: A systematic review,” *Clinical Neurology and Neurosurgery*, vol. 124, pp. 127–134, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.clineuro.2014.06.026> 2

- [20] P. Ginis, A. Nieuwboer, M. Dorfman, A. Ferrari, E. Gazit, C. G. Canning, L. Rocchi, L. Chiari, J. M. Hausdorff, and A. Mirelman, “Feasibility and effects of home-based smartphone-delivered automated feedback training for gait in people with Parkinson’s disease: A pilot randomized controlled trial,” *Parkinsonism and Related Disorders*, vol. 22, pp. 28–34, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.parkreldis.2015.11.004> 2
- [21] E. Dorsey, R. Constantinescu, J. Thompson, K. Biglan, R. Holloway, K. Kiebertz, F. Marshall, B. Ravina, G. Schifitto, A. Siderowf *et al.*, “Projected number of people with parkinson disease in the most populous nations, 2005 through 2030,” *Neurology*, vol. 68, no. 5, pp. 384–386, 2007. 2, 4
- [22] J. L. Zimbelman, S. P. Juraschek, X. Zhang, and V. W.-H. Lin, “Physical therapy workforce in the united states: forecasting nationwide shortages,” *PM&R*, vol. 2, no. 11, pp. 1021–1029, 2010. 2
- [23] S. Mazilu, M. Hardegger, Z. Zhu, D. Roggen, G. Troester, M. Plotnik, and J. Hausdorff, “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. [Online]. Available: <http://eudl.eu/doi/10.4108/icst.pervasivehealth.2012.248680> 2, 10, 35, 62, 63, 75, 97, 98
- [24] D. Rodríguez-Martín, A. Samà, C. Pérez-López, A. Català, J. M. M. Arostegui, J. Cabestany, À. Bayés, S. Alcaine, B. Mestre, A. Prats, M. C. Crespo, T. J. Counihan, P. Browne, L. R. Quinlan, G. ÓLaighin, D. Sweeney, H. Lewy, J. Azuri, G. Vainstein, R. Annicchiarico, A. Costa, , and A. Rodríguez-Molinero, “Home detection of freezing of gait using Support Vector Machines through a single waist-worn triaxial accelerometer,” *PLoS ONE*, vol. 25, no. 3, pp. 1–12, 2017. [Online]. Available: <http://www.mdpi.com/1424-8220/14/10/18800><http://dx.doi.org/10.1016/j.neucli.2015.09.009><http://link.springer.com/>

[10.1007/s00415-017-8424-0](http://dx.doi.org/10.1016/j.parkreldis.2015.03.028)<http://dx.doi.org/10.1016/j.parkreldis.2015.03.028><http://www.dovepress.com/a-practical-method-for-the-detection-o> 10

- [25] P. Tahafchi, R. Molina, J. A. Roper, K. Sowalsky, C. J. Hass, A. Gunduz, M. S. Okun, and J. W. Judy, “Freezing-of-gait detection using temporal, spatial, and physiological features with a support-vector-machine classifier,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2017, pp. 2867–2870. 2, 10, 63
- [26] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, “Time series classification using multi-channels deep convolutional neural networks,” in *International Conference on Web-Age Information Management*. Springer, 2014, pp. 298–310. 2
- [27] J. Camps, A. Samà, M. Martín, D. Rodríguez-Martín, C. Pérez-López, J. M. M. Arostegui, J. Cabestany, A. Català, S. Alcaine, B. Mestre, A. Prats, M. C. Crespo-Maraver, T. J. Counihan, P. Browne, L. R. Quinlan, G. Ò. Laighin, D. Sweeney, H. Lewy, G. Vainstein, A. Costa, R. Annicchiarico, Àngels Bayés, and A. Rodríguez-Molinero, “Deep learning for freezing of gait detection in parkinson’s disease patients in their homes using a waist-worn inertial measurement unit,” *Knowledge-Based Systems*, vol. 139, pp. 119 – 131, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0950705117304859> 3, 10, 63, 75, 98
- [28] Y. Xia, J. Zhang, Q. Ye, N. Cheng, Y. Lu, and D. Zhang, “Evaluation of deep convolutional neural networks for detection of freezing of gait in parkinson’s disease patients,” *Biomedical Signal Processing and Control*, vol. 46, pp. 221 – 230, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S174680941830185X> 3, 11, 63, 75, 97, 98
- [29] L. Bertram and R. E. Tanzi, “The genetic epidemiology of neurodegenerative disease,” *The Journal of clinical investigation*, vol. 115, no. 6, pp. 1449–1457, 2005. 4

- [30] T. A. Boonstra, H. Van Der Kooij, M. Munneke, and B. R. Bloem, "Gait disorders and balance disturbances in parkinson's disease: clinical update and pathophysiology," *Current opinion in neurology*, vol. 21, no. 4, pp. 461–471, 2008. 4
- [31] J. G. Nutt, B. R. Bloem, N. Giladi, M. Hallett, F. B. Horak, and A. Nieuwboer, "Freezing of gait: moving forward on a mysterious clinical phenomenon," *The Lancet Neurology*, vol. 10, no. 8, pp. 734 – 744, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1474442211701430> 4
- [32] A. Elkouzi, E. N. Bit-Ivan, and R. J. Elble, "Pure akinesia with gait freezing: a clinicopathologic study," *Journal of clinical movement disorders*, vol. 4, no. 1, p. 15, 2017. 4
- [33] K. A. Ehgoetz Martens, C. G. Ellard, and Q. J. Almeida, "Does anxiety cause freezing of gait in parkinson's disease," *PLOS ONE*, vol. 9, no. 9, pp. 1–10, 09 2014. [Online]. Available: <https://doi.org/10.1371/journal.pone.0106561> 5, 34
- [34] J. Spildooren, S. Vercruysse, K. Desloovere, W. Vandenberghe, E. Kerckhofs, and A. Nieuwboer, "Freezing of gait in parkinson's disease: The impact of dual-tasking and turning," *Movement Disorders*, vol. 25, no. 15, pp. 2563–2570, 2010. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.23327> 5
- [35] M. Mancini, B. R. Bloem, F. B. Horak, S. J. Lewis, A. Nieuwboer, and J. Nonnekes, "Clinical and methodological challenges for assessing freezing of gait: Future perspectives," *Movement Disorders*, vol. 34, no. 6, pp. 783–790, 2019. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.27709> 5, 111
- [36] D. S. Peterson, L. A. King, R. G. Cohen, and F. B. Horak, "Cognitive contributions to freezing of gait in parkinson disease: implications for physical rehabilitation," *Physical therapy*, vol. 96, no. 5, pp. 659–670, 2016.
- [37] A. Nieuwboer, W. de Weerd, R. Dom, and E. Lesaffre, "A frequency and correlation analysis of motor deficits in parkinson patients," *Disability and*

- Rehabilitation*, vol. 20, no. 4, pp. 142–150, 1998, pMID: 9571381. [Online]. Available: <https://doi.org/10.3109/09638289809166074>
- [38] K. A. Ehgoetz Martens, F. Pieruccini-Faria, and Q. J. Almeida, “Could sensory mechanisms be a core factor that underlies freezing of gait in parkinson’s disease,” *PLOS ONE*, vol. 8, no. 5, pp. 1–7, 5 2013. [Online]. Available: <https://doi.org/10.1371/journal.pone.0062602> 5
- [39] J. D. Schaafsma, Y. Balash, T. Gurevich, A. L. Bartels, J. M. Hausdorff, and N. Giladi, “Characterization of freezing of gait subtypes and the response of each to levodopa in parkinson’s disease,” *European Journal of Neurology*, vol. 10, no. 4, pp. 391–398, 2003. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1468-1331.2003.00611.x> 5, 38
- [40] T. Ellis, J. Cavanaugh, G. Earhart, M. Ford, K. Foreman, and L. Dibble, “Which measures of physical function and motor impairment best predict quality of life in parkinson’s disease,” *Parkinsonism & Related Disorders*, vol. 17, no. 9, pp. 693 – 697, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1353802011002227> 5
- [41] A. Suppa, A. Kita, G. Leodori, A. Zampogna, E. Nicolini, P. Lorenzi, R. Rao, and F. Irrera, “l-dopa and freezing of gait in parkinson’s disease: Objective assessment through a wearable wireless system,” *Frontiers in Neurology*, vol. 8, p. 406, 2017. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fneur.2017.00406> 5
- [42] B. S. Connolly and A. E. Lang, “Pharmacological treatment of parkinson disease: A review,” *JAMA*, vol. 311, no. 16, pp. 1670–1683, 2014. [Online]. Available: [+http://dx.doi.org/10.1001/jama.2014.3654](http://dx.doi.org/10.1001/jama.2014.3654) 5
- [43] N. Giladi, “Medical treatment of freezing of gait,” *Movement Disorders*, vol. 23, no. S2, pp. S482–S488, 2008. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.21914> 5

- [44] B. Thanvi, N. Lo, and T. Robinson, “Levodopa-induced dyskinesia in parkinson’s disease: clinical features, pathogenesis, prevention and treatment,” *Postgraduate medical journal*, vol. 83, no. 980, pp. 384–388, 2007. 5
- [45] S. Ahn, Y. Chen, T. Bredow, C. Cheung, and F. Yu, “Effects of non-pharmacological treatments on quality of life in parkinson’s disease: A review,” *journal of Parkinson’s Disease and Alzheimer Disease*, vol. 4, no. 1, pp. 1–10, 2017. 5
- [46] J. Vandenbossche, N. Deroost, E. Soetens, D. Coomans, J. Spildooren, S. Vercruysse, A. Nieuwboer, and E. Kerckhofs, “Freezing of gait in parkinson’s disease: disturbances in automaticity and control,” *Frontiers in human neuroscience*, vol. 6, p. 356, 2013. 7
- [47] P. M. Fitts and M. I. Posner, “Learning and skilled performance in human performance,” *Belmont, CA: Brock-Cole*, 1967. 7
- [48] Z. Kadivar, D. M. Corcos, J. Foto, and J. M. Hondzinski, “Effect of step training and rhythmic auditory stimulation on functional performance in parkinson patients,” *Neurorehabilitation and Neural Repair*, vol. 25, no. 7, pp. 626–635, 2011. [Online]. Available: <https://doi.org/10.1177/1545968311401627> 7, 9
- [49] P. Ginis, E. Nackaerts, A. Nieuwboer, and E. Heremans, “Cueing for people with parkinson’s disease with freezing of gait: A narrative review of the state-of-the-art and novel perspectives,” *Annals of Physical and Rehabilitation Medicine*, vol. 61, no. 6, pp. 407 – 413, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877065717304049> 7, 9, 11
- [50] D. S. Peterson and K. Smulders, “Cues and attention in parkinsonian gait: Potential mechanisms and future directions,” *Frontiers in Neurology*, vol. 6, p. 255, 2015. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fneur.2015.00255> 7
- [51] F. Debaere, N. Wenderoth, S. Sunaert, P. Van Hecke, and S. P. Swinnen, “Internal vs external generation of movements: differential neural pathways involved in bimanual coordination performed in the presence or absence of augmented visual feedback,” *Neuroimage*, vol. 19, no. 3, pp. 764–776, 2003. 7

- [52] S. Donovan, C. Lim, N. Diaz, N. Browner, P. Rose, L. R. Sudarsky, D. Tarsy, S. Fahn, and D. K. Simon, “Laserlight cues for gait freezing in Parkinson’s disease: An open-label study,” *Parkinsonism and Related Disorders*, vol. 17, no. 4, pp. 240–245, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.parkreldis.2010.08.010> 7, 112
- [53] U. M. Fietzek, F. E. Schroeteler, K. Ziegler, J. Zwosta, and A. O. Ceballos-Baumann, “Randomized cross-over trial to investigate the efficacy of a two-week physiotherapy programme with repetitive exercises of cueing to reduce the severity of freezing of gait in patients with parkinson’s disease,” *Clinical Rehabilitation*, vol. 28, no. 9, pp. 902–911, 2014. [Online]. Available: <https://doi.org/10.1177/0269215514527299> 7
- [54] M. Plotnik, S. Shema, M. Dorfman, E. Gazit, M. Brozgol, N. Giladi, and J. M. Hausdorff, “A motor learning-based intervention to ameliorate freezing of gait in subjects with parkinson’s disease,” *Journal of neurology*, vol. 261, no. 7, pp. 1329–1339, 2014. 7
- [55] S. Ghai, I. Ghai, G. Schmitz, and A. O. Effenberg, “Effect of rhythmic auditory cueing on parkinsonian gait: A systematic review and meta-analysis,” *Aging and Disease*, vol. 9, no. 5, pp. 901–923, 2017.
- [56] I. Lim, E. van Wegen, C. de Goede, M. Deutekom, A. Nieuwboer, A. Willems, D. Jones, L. Rochester, and G. Kwakkel, “Effects of external rhythmical cueing on gait in patients with parkinson’s disease: a systematic review,” *Clinical Rehabilitation*, vol. 19, no. 7, pp. 695–713, 2005, pMID: 16250189. [Online]. Available: <https://doi.org/10.1191/0269215505cr906oa> 7
- [57] D. Sweeney, L. R. Quinlan, P. Browne, M. Richardson, P. Meskell, and G. ÓLaighin, “A technological review of wearable cueing devices addressing freezing of gait in parkinson’s disease,” *Sensors*, vol. 19, no. 6, p. 1277, 2019. xii, 8
- [58] G. Frazzitta, R. Maestri, D. Uccellini, G. Bertotti, and P. Abelli, “Rehabilitation treatment of gait in patients with parkinson’s disease with freezing: A comparison between two physical therapy protocols using visual and auditory cues with or

- without treadmill training,” *Movement Disorders*, vol. 24, no. 8, pp. 1139–1143, 2009. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.22491> 7
- [59] E. Cubo, S. Leurgans, and C. G. Goetz, “Short-term and practice effects of metronome pacing in parkinson’s disease patients with gait freezing while in the ‘on’ state: randomized single blind evaluation,” *Parkinsonism & Related Disorders*, vol. 10, no. 8, pp. 507 – 510, 2004. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1353802004000859> 9
- [60] S. Mazilu, A. Calatroni, E. Gazit, A. Mirelman, J. M. Hausdorff, and G. Tröster, “Prediction of freezing of gait in parkinson’s from physiological wearables: An exploratory study,” *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1843–1854, Nov 2015. 9
- [61] A. Nieuwboer, R. Dom, W. De Weerd, K. Desloovere, L. Janssens, and V. Stijn, “Electromyographic profiles of gait prior to onset of freezing episodes in patients with parkinson’s disease,” *Brain*, vol. 127, no. 7, pp. 1650–1660, 2004. [Online]. Available: <http://dx.doi.org/10.1093/brain/awh189> 9
- [62] B. T. Cole, S. H. Roy, and S. H. Nawab, “Detecting freezing-of-gait during unscripted and unconstrained activity,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2011, pp. 5649–5652. 9
- [63] S.-B. Koh, K.-W. Park, D.-H. Lee, S. J. Kim, and J.-S. Yoon, “Gait analysis in patients with parkinson’s disease: Relationship to clinical features and freezing,” *Journal of Movement Disorders*, vol. 1, no. 2, pp. 59–64, 2008. [Online]. Available: <http://www.e-jmd.org/journal/view.php?number=72> 9
- [64] A. Delval, A. H. Snijders, V. Weerdesteyn, J. E. Duysens, L. Defebvre, N. Giladi, and B. R. Bloem, “Objective detection of subtle freezing of gait episodes in parkinson’s disease,” *Movement Disorders*, vol. 25, no. 11, pp. 1684–1693, 2010. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.23159> 9

- [65] M. Plotnik, N. Giladi, Y. Balash, C. Peretz, and J. M. Hausdorff, “Is freezing of gait in parkinson’s disease related to asymmetric motor function,” *Annals of Neurology*, vol. 57, no. 5, pp. 656–663, 2005. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ana.20452> 9
- [66] J. M. Hausdorff, J. D. Schaafsma, Y. Balash, A. L. Bartels, T. Gurevich, and N. Giladi, “Impaired regulation of stride variability in parkinson’s disease subjects with freezing of gait,” *Experimental Brain Research*, vol. 149, no. 2, pp. 187–194, Mar 2003. [Online]. Available: <https://doi.org/10.1007/s00221-002-1354-8> 9
- [67] L. Palmerini, L. Rocchi, S. Mazilu, E. Gazit, J. M. Hausdorff, and L. Chiari, “Identification of characteristic motor patterns preceding freezing of gait in Parkinson’s disease using wearable sensors,” *Frontiers in Neurology*, vol. 8, no. AUG, pp. 1–12, 2017. 11, 24, 33, 63, 97
- [68] V. G. Torvi, A. Bhattacharya, and S. Chakraborty, “Deep domain adaptation to predict freezing of gait in patients with parkinson’s disease,” in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec 2018, pp. 1001–1006. 11, 64, 125
- [69] N. Naghavi and E. Wade, “Prediction of freezing of gait in parkinson’s disease using statistical inference and lower–limb acceleration data,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 947–955, May 2019. 12
- [70] N. Naghavi, A. Miller, and E. Wade, “Towards real-time prediction of freezing of gait in patients with parkinson’s disease: Addressing the class imbalance problem,” *Sensors*, vol. 19, no. 18, p. 3898, 2019. 12
- [71] N. Naghavi, S. Borhani, and E. Wade, “Improving machine learning based detection of freezing of gait using data synthesis methods,” in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, pp. 1432–1437. 12

- [72] S. Rezvanian and T. Lockhart, "Towards Real-Time Detection of Freezing of Gait Using Wavelet Transform on Wireless Accelerometer Data," *Sensors*, vol. 16, no. 4, p. 475, 2016. [Online]. Available: <http://www.mdpi.com/1424-8220/16/4/475> 15, 17, 20, 33
- [73] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000. [Online]. Available: <http://ajpheart.physiology.org/content/278/6/H2039> 15, 20
- [74] J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou, "The appropriate use of approximate entropy and sample entropy with short data sets," *Annals of Biomedical Engineering*, vol. 41, no. 2, pp. 349–365, 2013. 15
- [75] M. S. Baby, A. J. Saji, and C. S. Kumar, "Parkinson's disease classification using wavelet transform based feature extraction of gait data," in *2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, April 2017, pp. 1–6. 20
- [76] A.-E. B. and B. B., "Leg motion classification with artificial neural networks using wavelet-based features of gyroscope signals," *Sensors*, vol. 11, no. 2, pp. 1721–1743, 2011. 20
- [77] M. A. D. Brodie, M. J. M. Coppens, S. R. Lord, N. H. Lovell, Y. J. Gschwind, S. J. Redmond, M. B. Del Rosario, K. Wang, D. L. Sturnieks, M. Persiani, and K. Delbaere, "Wearable pendant device monitoring using new wavelet-based methods shows daily life and laboratory gaits are different," *Medical & Biological Engineering & Computing*, vol. 54, no. 4, pp. 663–674, Apr 2016. [Online]. Available: <https://doi.org/10.1007/s11517-015-1357-9> 20
- [78] Y. Tochigi, N. A. Segal, T. Vaseenon, and T. D. Brown, "Entropy analysis of tri-axial leg acceleration signal waveforms for measurement of decrease of physiological variability in human gait," *Journal of Orthopaedic Research*, vol. 30, no. 6, pp. 897–904, 2012. [Online]. Available: <http://dx.doi.org/10.1002/jor.22022> 20, 21, 42

- [79] H. Zach, A. M. Janssen, A. H. Snijders, A. Delval, M. U. Ferraye, E. Auff, V. Weerdesteyn, B. R. Bloem, and J. Nonnekes, “Identifying freezing of gait in parkinson’s disease during freezing provoking tasks using waist-mounted accelerometry,” *Parkinsonism & Related Disorders*, vol. 21, no. 11, pp. 1362 – 1366, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1353802015004423> 21
- [80] J. Han, H. Jeon, B. Jeon, and K. Park, “Gait detection from three dimensional acceleration signals of ankles for the patients with parkinson’s disease,” in *IEEE International Special Topic Conference on Information Technology in Biomedicine, Ioannina, Epirus, Greece*, 2006, pp. 1–4. 21
- [81] M. El-Gohary, S. Pearson, J. McNames, M. Mancini, F. Horak, S. Mellone, and L. Chiari, “Continuous monitoring of turning in patients with movement disability,” *Sensors (Switzerland)*, vol. 14, no. 1, pp. 356–369, 2014. 24
- [82] S. Mazilu, A. Calatroni, E. Gazit, D. Roggen, J. Hausdorff, and G. Tröster, “Feature learning for detection and prediction of freezing of gait in Parkinson’s disease,” *Proc. Workshop Math. Methods Biomed. Image Analysis*, vol. 7988 LNAI, p. 144, 2013. 24
- [83] Y. Jiang and K. E. Norman, “Effects of visual and auditory cues on gait initiation in people with Parkinson’s disease,” *Clinical Rehabilitation*, vol. 20, no. 1, pp. 36–45, jan 2006. [Online]. Available: <https://doi.org/10.1191/0269215506cr925oa> 26
- [84] M. Chen, B. Huang, and Y. Xu, “Intelligent shoes for abnormal gait detection,” in *2008 IEEE International Conference on Robotics and Automation*, May 2008, pp. 2019–2024. 31
- [85] J. Barth, J. Klucken, P. Kugler, T. Kammerer, R. Steidl, J. Winkler, J. Hornegger, and B. Eskofier, “Biometric and mobile gait analysis for early diagnosis and therapy monitoring in parkinson’s disease,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2011, pp. 868–871. 31

- [86] T. T. Pham, S. T. Moore, S. J. Lewis, D. N. Nguyen, E. Dutkiewicz, A. J. Fuglevand, A. L. McEwan, and P. H. Leong, “Freezing of Gait Detection in Parkinson’s Disease: A Subject-Independent Detector Using Anomaly Scores,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 11, pp. 1–1, 2017. [Online]. Available: <http://ieeexplore.ieee.org/document/7845616/> 33, 97
- [87] S. Mellone, M. Mancini, L. A. King, F. B. Horak, and L. Chiari, “The quality of turning in parkinson’s disease: A compensatory strategy to prevent postural instability,” *Journal of NeuroEngineering and Rehabilitation*, vol. 13, no. 1, pp. 1–9, 2016. 34
- [88] M. Bishop, D. Brunt, and J. Marjama-Lyons, “Do people with parkinson’s disease change strategy during unplanned gait termination,” *Neuroscience Letters*, vol. 397, no. 3, pp. 240 – 244, 2006. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0304394005014321> 34
- [89] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002. 36
- [90] R. Blagus and L. Lusa, “SMOTE for high-dimensional class-imbalanced data,” *BMC Bioinformatics*, vol. 14, no. 1, Mar. 2013. [Online]. Available: <https://doi.org/10.1186/1471-2105-14-106> 36
- [91] H. He, Y. Bai, E. A. Garcia, and S. Li, “Adasyn: Adaptive synthetic sampling approach for imbalanced learning,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, June 2008, pp. 1322–1328. 36, 37
- [92] M. Kaleem, A. Guergachi, and S. Krishnan, “Patient-specific seizure detection in long-term eeg using wavelet decomposition,” *Biomedical Signal Processing and Control*, vol. 46, pp. 157–165, 2018. 36
- [93] E. T. Pereira and H. M. Gomes, “The role of data balancing for emotion classification using EEG signals,” in *2016 IEEE International Conference*

- on *Digital Signal Processing (DSP)*. IEEE, Oct. 2016. [Online]. Available: <https://doi.org/10.1109/icdsp.2016.7868619> 36
- [94] M. Sharma, D. Goyal, P. Achuth, and U. R. Acharya, “An accurate sleep stages classification system using a new class of optimally time-frequency localized three-band wavelet filter bank,” *Computers in Biology and Medicine*, vol. 98, pp. 58–75, Jul. 2018. [Online]. Available: <https://doi.org/10.1016/j.combiomed.2018.04.025> 36
- [95] R. Shoorangiz, S. J. Weddell, and R. D. Jones, “Prediction of microsleeps from EEG: Preliminary results,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Aug. 2016. [Online]. Available: <https://doi.org/10.1109/embc.2016.7591764> 36
- [96] L. Yang, R. Ma, H. M. Zhang, W. Guan, and S. Jiang, “Driving behavior recognition using EEG data from a simulated car-following experiment,” *Accident Analysis & Prevention*, vol. 116, pp. 30–40, Jul. 2018. [Online]. Available: <https://doi.org/10.1016/j.aap.2017.11.010> 36
- [97] U. R. Acharya, V. K. Sudarshan, S. Q. Rong, Z. Tan, C. M. Lim, J. E. Koh, S. Nayak, and S. V. Bhandary, “Automated detection of premature delivery using empirical mode and wavelet packet decomposition techniques with uterine electromyogram signals,” *Computers in Biology and Medicine*, vol. 85, pp. 33–42, Jun. 2017. [Online]. Available: <https://doi.org/10.1016/j.combiomed.2017.04.013> 36
- [98] D. Despotovic, A. Zec, K. Mladenovic, N. Radin, and T. L. Turukalo, “A machine learning approach for an early prediction of preterm delivery,” in *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*. IEEE, Sep. 2018. [Online]. Available: <https://doi.org/10.1109/sisy.2018.8524818> 36
- [99] K. Wu, D. Zhang, G. Lu, and Z. Guo, “Learning acoustic features to detect parkinson’s disease,” *Neurocomputing*, vol. 318, pp. 102–108, Nov. 2018. [Online]. Available: <https://doi.org/10.1016/j.neucom.2018.08.036> 36

- [100] C. Taleb, M. Khachab, C. Mokbel, and L. Likforman-Sulem, “A reliable method to predict parkinson’s disease stage and progression based on handwriting and re-sampling approaches,” in *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)*. IEEE, Mar. 2018. [Online]. Available: <https://doi.org/10.1109/asar.2018.8480209> 37
- [101] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, “A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, July 2012. 37
- [102] S. Kotsiantis, D. Kanellopoulos, P. Pintelas *et al.*, “Handling imbalanced datasets: A review,” *GESTS International Transactions on Computer Science and Engineering*, vol. 30, no. 1, pp. 25–36, 2006. 37
- [103] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009. 41
- [104] L. v. d. Maaten and G. Hinton, “Visualizing data using t-sne,” *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008. 44
- [105] T. G. Dietterich, “Ensemble methods in machine learning,” in *Multiple Classifier Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1–15. 45
- [106] A. Topchy, A. K. Jain, and W. Punch, “Clustering ensembles: models of consensus and weak partitions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1866–1881, Dec 2005. 45
- [107] H. Sarvari, C. Domeniconi, and G. Stilo, “Graph-based selective outlier ensembles,” in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, ser. SAC ’19. New York, NY, USA: ACM, 2019, pp. 518–525. 45
- [108] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009. 63

- [109] S. Borhani, R. Abiri, X. Zhao, and Y. Jiang, “A transfer learning approach towards zero-training bci for eeg-based two dimensional cursor control,” in *Society for Neuroscience (SfN2017)*, 2017. 64
- [110] S. J. Pan, J. T. Kwok, Q. Yang *et al.*, “Transfer learning via dimensionality reduction,” in *AAAI*, vol. 8, 2008, pp. 677–682. 63
- [111] R. Chalapathy, A. K. Menon, and S. Chawla, “Anomaly detection using one-class neural networks,” *CoRR*, vol. abs/1802.06360, 2018. [Online]. Available: <http://arxiv.org/abs/1802.06360> 67
- [112] X. Hu, S. Hu, X. Zhang, H. Zhang, and L. Luo, “Anomaly detection based on local nearest neighbor distance descriptor in crowded scenes,” *The Scientific World Journal*, vol. 2014, 2014.
- [113] X. Chen and E. Konukoglu, “Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders,” *arXiv preprint arXiv:1806.04972*, 2018. 67
- [114] Y. Yu, Y. Zhu, S. Li, and D. Wan, “Time series outlier detection based on sliding window prediction,” *Mathematical problems in Engineering*, vol. 2014, 2014. 67
- [115] E. Ramirez, M. Wimmer, and M. Atzmueller, “A computational framework for interpretable anomaly detection and classification of multivariate time series with application to human gait data analysis,” in *Artificial Intelligence in Medicine: Knowledge Representation and Transparent and Explainable Systems*. Cham: Springer International Publishing, 2019, pp. 132–147.
- [116] M. T. Martinez, P. L. De Leon, and D. Keeley, “Anomaly detection for predicting falls risk using smartphone gait data.” Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), Tech. Rep., 2016. 67
- [117] T.-N. Nguyen, H.-H. Huynh, and J. Meunier, “Skeleton-based abnormal gait detection,” *Sensors*, vol. 16, no. 11, p. 1792, Oct 2016. [Online]. Available: <http://dx.doi.org/10.3390/s16111792> 67

- [118] G. Cola, M. Avvenuti, A. Vecchio, G. Yang, and B. Lo, “An on-node processing approach for anomaly detection in gait,” *IEEE Sensors Journal*, vol. 15, no. 11, pp. 6640–6649, 2015. [67](#)
- [119] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, “High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning,” *Pattern Recognition*, vol. 58, pp. 121 – 134, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320316300267> [67](#)
- [120] J. Chen, S. Sathe, C. Aggarwal, and D. Turaga, “Outlier detection with autoencoder ensembles,” in *Proceedings of the 17th SIAM International Conference on Data Mining, SDM 2017*, 2017. [67](#)
- [121] Y. Liu, S. Zhou, and Q. Chen, “Discriminative deep belief networks for visual data classification,” *Pattern Recognition*, vol. 44, no. 10, pp. 2287 – 2296, 2011, semi-Supervised Learning for Visual Content Analysis and Understanding. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320310005789> [67](#)
- [122] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, “A discriminative feature learning approach for deep face recognition,” in *European conference on computer vision*. Springer, 2016, pp. 499–515. [67](#)
- [123] P. Perera and V. M. Patel, “Learning Deep Features for One-Class Classification,” *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5450–5463, 2019. [67](#), [68](#)
- [124] L. Ruff, R. A. Vandermeulen, N. Görnitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, “Deep one-class classification,” *35th International Conference on Machine Learning, ICML 2018*, vol. 10, pp. 6981–6996, 2018. [68](#), [73](#)
- [125] D. Grabli, C. Karachi, M.-L. Welter, B. Lau, E. C. Hirsch, M. Vidailhet, and C. François, “Normal and pathological gait: what we learn from parkinson’s disease,” *J Neurol Neurosurg Psychiatry*, vol. 83, no. 10, pp. 979–985, 2012. [69](#)
- [126] S. Janssen, B. Bolte, J. Nonnekens, M. Bittner, B. R. Bloem, T. Heida, Y. Zhao, and R. J. van Wezel, “Usability of three-dimensional augmented visual cues delivered by

- smart glasses on (freezing of) gait in parkinson's disease," *Frontiers in neurology*, vol. 8, p. 279, 2017. 103
- [127] V. Ivkovic, S. Fisher, and W. H. Paloski, "Smartphone-based tactile cueing improves motor performance in parkinson's disease," *Parkinsonism & related disorders*, vol. 22, pp. 42–47, 2016. 103
- [128] A. Graves, "Generating sequences with recurrent neural networks," *arXiv preprint arXiv:1308.0850*, 2014. [Online]. Available: <https://arxiv.org/abs/1308.0850v5> 125
- [129] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *ICML*, 2011, pp. 1017–1024. [Online]. Available: https://icml.cc/2011/papers/524_icmlpaper.pdf
- [130] J. A. Pérez-Ortiz, J. Calera-Rubio, and M. L. Forcada, "Online text prediction with recurrent neural networks," *Neural processing letters*, vol. 14, no. 2, pp. 127–140, 2001.
- [131] Y. Goldberg, "Neural network methods for natural language processing," *Synthesis Lectures on Human Language Technologies*, vol. 10, no. 1, pp. 1–309, 2017.
- [132] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of cnn and rnn for natural language processing," *arXiv preprint arXiv:1702.01923*, 2017. [Online]. Available: <https://arxiv.org/abs/1702.01923v1> 125
- [133] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 240–254, 1994. 125
- [134] C. H. Aladag, E. Egrioglu, and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology," *Applied Mathematics Letters*, vol. 22, no. 9, pp. 1467–1470, 2009.
- [135] B. Oancea and Ş. C. Ciucu, "Time series forecasting using neural networks," *arXiv preprint arXiv:1401.1333*, 2014. [Online]. Available: <https://arxiv.org/abs/1401.1333v1> 125

Appendix

A Montreal Cognitive Assessment (MoCA)

MONTREAL COGNITIVE ASSESSMENT (MOCA)

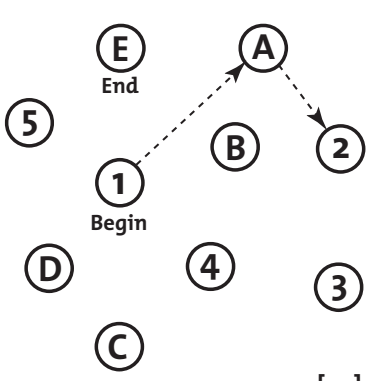
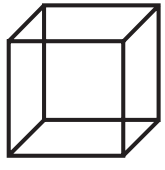
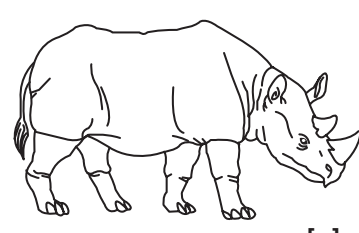
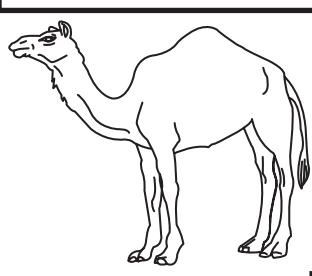
NAME :

Education :

Date of birth :

Sex :

DATE :

VISUOSPATIAL / EXECUTIVE		POINTS																			
 [] []	 Copy cube	Draw CLOCK (Ten past eleven) (3 points) [] [] [] Contour Numbers Hands	___/5																		
NAMING		POINTS																			
 []	 []	 []	___/3																		
MEMORY		No points																			
Read list of words, subject must repeat them. Do 2 trials. Do a recall after 5 minutes.		<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th></th> <th>FACE</th> <th>VELVET</th> <th>CHURCH</th> <th>DAISY</th> <th>RED</th> </tr> <tr> <td>1st trial</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>2nd trial</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>		FACE	VELVET	CHURCH	DAISY	RED	1st trial						2nd trial						
	FACE	VELVET	CHURCH	DAISY	RED																
1st trial																					
2nd trial																					
ATTENTION		POINTS																			
Read list of digits (1 digit/ sec.).		Subject has to repeat them in the forward order [] 2 1 8 5 4 Subject has to repeat them in the backward order [] 7 4 2	___/2																		
Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors [] F B A C M N A A J K L B A F A K D E A A A J A M O F A A B		___/1																			
Serial 7 subtraction starting at 100 [] 93 [] 86 [] 79 [] 72 [] 65 4 or 5 correct subtractions: 3 pts, 2 or 3 correct: 2 pts, 1 correct: 1 pt, 0 correct: 0 pt		___/3																			
LANGUAGE		POINTS																			
Repeat : I only know that John is the one to help today. [] The cat always hid under the couch when dogs were in the room. []		___/2																			
Fluency / Name maximum number of words in one minute that begin with the letter F [] _____ (N ≥ 11 words)		___/1																			
ABSTRACTION		POINTS																			
Similarity between e.g. banana - orange = fruit [] train - bicycle [] watch - ruler		___/2																			
DELAYED RECALL		POINTS																			
Has to recall words WITH NO CUE		<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th></th> <th>FACE</th> <th>VELVET</th> <th>CHURCH</th> <th>DAISY</th> <th>RED</th> </tr> <tr> <td>[]</td> <td>[]</td> <td>[]</td> <td>[]</td> <td>[]</td> <td>[]</td> </tr> </table>		FACE	VELVET	CHURCH	DAISY	RED	[]	[]	[]	[]	[]	[]							
	FACE	VELVET	CHURCH	DAISY	RED																
[]	[]	[]	[]	[]	[]																
Optional		<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th></th> <th>Category cue</th> <th>Multiple choice cue</th> </tr> <tr> <td>[]</td> <td></td> <td></td> </tr> </table>		Category cue	Multiple choice cue	[]															
	Category cue	Multiple choice cue																			
[]																					
ORIENTATION		[] Date [] Month [] Year [] Day [] Place [] City ___/6																			
© Z.Nasreddine MD Version November 7, 2004 www.mocatest.org		Normal ≥ 26 / 30 TOTAL ___/30 Add 1 point if ≤ 12 yr edu																			

B MoCA - Instructions

The Montreal Cognitive Assessment (MoCA) was designed as a rapid screening instrument for mild cognitive dysfunction. It assesses different cognitive domains: attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation. Time to administer the MoCA is approximately 10 minutes. The total possible score is 30 points; a score of 26 or above is considered normal.

1. **Alternating Trail Making:**

Administration: The examiner instructs the subject: *"Please draw a line, going from a number to a letter in ascending order. Begin here [point to (1)] and draw a line from 1 then to A then to 2 and so on. End here [point to (E)]."*

Scoring: Allocate one point if the subject successfully draws the following pattern: 1 –A- 2- B- 3- C- 4- D- 5- E, without drawing any lines that cross. Any error that is not immediately self-corrected earns a score of 0.

2. **Visuoconstructional Skills (Cube):**

Administration: The examiner gives the following instructions, pointing to the **cube**: *"Copy this drawing as accurately as you can, in the space below"*.

Scoring: One point is allocated for a correctly executed drawing.

- Drawing must be three-dimensional
- All lines are drawn
- No line is added
- Lines are relatively parallel and their length is similar (rectangular prisms are accepted)

A point is not assigned if any of the above-criteria are not met.

3. **Visuoconstructional Skills (Clock):**

Administration: Indicate the right third of the space and give the following instructions: *"Draw a **clock**. Put in all the numbers and set the time to 10 after 11"*.

Scoring: One point is allocated for each of the following three criteria:

- Contour (1 pt.): the clock face must be a circle with only minor distortion acceptable (e.g., slight imperfection on closing the circle);
- Numbers (1 pt.): all clock numbers must be present with no additional numbers; numbers must be in the correct order and placed in the approximate quadrants on the clock face; Roman numerals are acceptable; numbers can be placed outside the circle contour;
- Hands (1 pt.): there must be two hands jointly indicating the correct time; the hour hand must be clearly shorter than the minute hand; hands must be centred within the clock face with their junction close to the clock centre.

A point is not assigned for a given element if any of the above-criteria are not met.

4. **Naming:**

Administration: Beginning on the left, point to each figure and say: *“Tell me the name of this animal”*.

Scoring: One point each is given for the following responses: (1) camel or dromedary, (2) lion, (3) rhinoceros or rhino.

5. **Memory:**

Administration: The examiner reads a list of 5 words at a rate of one per second, giving the following instructions: *“This is a memory test. I am going to read a list of words that you will have to remember now and later on. Listen carefully. When I am through, tell me as many words as you can remember. It doesn’t matter in what order you say them”*. Mark a check in the allocated space for each word the subject produces on this first trial. When the subject indicates that (s)he has finished (has recalled all words), or can recall no more words, read the list a second time with the following instructions: *“I am going to read the same list for a second time. Try to remember and tell me as many words as you can, including words you said the first time.”* Put a check in the allocated space for each word the subject recalls after the second trial.

At the end of the second trial, inform the subject that (s)he will be asked to recall these words again by saying, *“I will ask you to recall those words again at the end of the test.”*

Scoring: No points are given for Trials One and Two.

6. **Attention:**

Forward Digit Span: Administration: Give the following instruction: *“I am going to say some numbers and when I am through, repeat them to me exactly as I said them”*. Read the five number sequence at a rate of one digit per second.

Backward Digit Span: Administration: Give the following instruction: *“Now I am going to say some more numbers, but when I am through you must repeat them to me in the backwards order.”* Read the three number sequence at a rate of one digit per second.

Scoring: Allocate one point for each sequence correctly repeated, (*N.B.*: the correct response for the backwards trial is 2-4-7).

Vigilance: Administration: The examiner reads the list of letters at a rate of one per second, after giving the following instruction: *“I am going to read a sequence of letters. Every time I say the letter A, tap your hand once. If I say a different letter, do not tap your hand”*.

Scoring: Give one point if there is zero to one errors (an error is a tap on a wrong letter or a failure to tap on letter A).

Serial 7s: Administration: The examiner gives the following instruction: “Now, I will ask you to count by subtracting seven from 100, and then, keep subtracting seven from your answer until I tell you to stop.” Give this instruction twice if necessary.

Scoring: This item is scored out of 3 points. Give no (0) points for no correct subtractions, 1 point for one correction subtraction, 2 points for two-to-three correct subtractions, and 3 points if the participant successfully makes four or five correct subtractions. Count each correct subtraction of 7 beginning at 100. Each subtraction is evaluated independently; that is, if the participant responds with an incorrect number but continues to correctly subtract 7 from it, give a point for each correct subtraction. For example, a participant may respond “92 – 85 – 78 – 71 – 64” where the “92” is incorrect, but all subsequent numbers are subtracted correctly. This is one error and the item would be given a score of 3.

7. **Sentence repetition:**

Administration: The examiner gives the following instructions: “I am going to read you a sentence. Repeat it after me, exactly as I say it [pause]: ***I only know that John is the one to help today.***” Following the response, say: “Now I am going to read you another sentence. Repeat it after me, exactly as I say it [pause]: ***The cat always hid under the couch when dogs were in the room.***”

Scoring: Allocate 1 point for each sentence correctly repeated. Repetition must be exact. Be alert for errors that are omissions (e.g., omitting “only”, “always”) and substitutions/additions (e.g., “John is the one who helped today;” substituting “hides” for “hid”, altering plurals, etc.).

8. **Verbal fluency:**

Administration: The examiner gives the following instruction: “Tell me as many words as you can think of that begin with a certain letter of the alphabet that I will tell you in a moment. You can say any kind of word you want, except for proper nouns (like Bob or Boston), numbers, or words that begin with the same sound but have a different suffix, for example, love, lover, loving. I will tell you to stop after one minute. Are you ready? [Pause] Now, tell me as many words as you can think of that begin with the letter F. [time for 60 sec]. Stop.”

Scoring: Allocate one point if the subject generates 11 words or more in 60 sec. Record the subject’s response in the bottom or side margins.

9. **Abstraction:**

Administration: The examiner asks the subject to explain what each pair of words has in common, starting with the example: “Tell me how an orange and a banana are alike”. If the subject answers in a concrete manner, then say only one additional time: “Tell me another way in which those items are alike”. If the subject does not give the appropriate response (*fruit*), say, “Yes, and they are also both fruit.” Do not give any additional instructions or clarification.

After the practice trial, say: “Now, tell me how a train and a bicycle are alike”. Following the response, administer the second trial, saying: “Now tell me how a ruler and a watch are alike”. Do not give any additional instructions or prompts.

Scoring: Only the last two item pairs are scored. Give 1 point to each item pair correctly answered. The following responses are acceptable:

Train-bicycle = means of transportation, means of travelling, you take trips in both;

Ruler-watch = measuring instruments, used to measure.

The following responses are **not** acceptable: Train-bicycle = they have wheels; Ruler-watch = they have numbers.

10. Delayed recall:

Administration: The examiner gives the following instruction: *“I read some words to you earlier, which I asked you to remember. Tell me as many of those words as you can remember. Make a check mark (✓) for each of the words correctly recalled spontaneously without any cues, in the allocated space.*

Scoring: **Allocate 1 point for each word recalled freely without any cues.**

Optional:

Following the delayed free recall trial, prompt the subject with the semantic category cue provided below for any word not recalled. Make a check mark (✓) in the allocated space if the subject remembered the word with the help of a category or multiple-choice cue. Prompt all non-recalled words in this manner. If the subject does not recall the word after the category cue, give him/her a multiple choice trial, using the following example instruction, *“Which of the following words do you think it was, NOSE, FACE, or HAND?”*

Use the following category and/or multiple-choice cues for each word, when appropriate:

FACE: category cue: part of the body multiple choice: nose, face, hand

VELVET: category cue: type of fabric multiple choice: denim, cotton, velvet

CHURCH: category cue: type of building multiple choice: church, school, hospital

DAISY: category cue: type of flower multiple choice: rose, daisy, tulip

RED: category cue: a colour multiple choice: red, blue, green

Scoring: **No points are allocated for words recalled with a cue.** A cue is used for clinical information purposes only and can give the test interpreter additional information about the type of memory disorder. For memory deficits due to retrieval failures, performance can be improved with a cue. For memory deficits due to encoding failures, performance does not improve with a cue.

11. Orientation:

Administration: The examiner gives the following instructions: *“Tell me the date today”.* If the subject does not give a complete answer, then prompt accordingly by saying: *“Tell me the [year, month, exact date, and day of the week].”* Then say: *“Now, tell me the name of this place, and which city it is in.”*

Scoring: Give one point for each item correctly answered. The subject must tell the exact date and the exact place (name of hospital, clinic, office). No points are allocated if subject makes an error of one day for the day and date.

TOTAL SCORE: Sum all subscores listed on the right-hand side. Add one point for an individual who has 12 years or fewer of formal education, for a possible maximum of 30 points. A final total score of 26 and above is considered normal.

C Gait and Fall Questionnaire (GFQ)

1. During your best state—do you walk:

- 0 Normally
- 1 Almost normally—somewhat slow
- 2 Slow but fully independent
- 3 Need assistance or walking aid
- 4 Unable to walk

2. During your worst state—do you walk

- 0 Normally
- 1 Almost normally—somewhat slow
- 2 Slow but fully independent
- 3 Need assistance or walking aid
- 4 Unable to walk

3. Are your gait difficulties affecting your daily activities and independence?

- 0 Not at all
- 1 Mildly
- 2 Moderately
- 3 Severely
- 4 Unable to walk

4. Do you feel that your feet get glued to the floor while walking, making a turn or when trying to initiate walking (freezing)?

- 0 Never
- 1 Very rarely—about once a month
- 2 Rarely—about once a week
- 3 Often—about once a day
- 4 Always—whenever walking

5. How long is your longest freezing episode?

- 0 Never happened
- 1 1–2 s
- 2 3–10 s
- 3 11–30 s
- 4 Unable to walk for more than 30 s

6. How long is your typical start hesitation episode (freezing when initiating the first step)?

- 0 None
- 1 Takes longer than 1 s to start walking
- 2 Takes longer than 3 s to start walking
- 3 Takes longer than 10 s to start walking
- 4 Takes longer than 30 s to start walking

7. How long is your typical turning hesitation (freezing when turning)

- 0 None
- 1 Resume turning in 1–2 s
- 2 Resume turning in 3–10 s
- 3 Resume turning in 11–30 s
- 4 Unable to resume turning for more than 30 s

8. How long is your typical destination hesitation (freezing when approaching the target, such as when stepping onto a scale or approaching a chair to sit down)?

- 0 None
- 1 Resume walking in 1–2 s
- 2 Resume walking in 3–10 s
- 3 Resume walking in 11–30 s
- 4 Unable to resume walking for more than 30 s

9. How long is your typical tight quarters hesitation (freezing when attempting to get through narrow space such as a doorway)?

- 0 None
- 1 Resume walking in 1–2 s
- 2 Resume walking in 3–10 s
- 3 Resume walking in 11–30 s
- 4 Unable to resume walking for more than 30 s

10. How long is your typical freezing episode while walking on straight?

- 0 None
- 1 Resume walking in 1–2 s
- 2 Resume walking in 3–10 s
- 3 Resume walking in 11–30 s
- 4 Unable to resume walking for more than 30 s

11. How long is your typical freezing episode during stressful time-demanding situations, such as when the telephone rings, at elevators or street crossing?

- 0 None
- 1 Resume walking in 1–2 s
- 2 Resume walking in 3–10 s
- 3 Resume walking in 11–30 s
- 4 Unable to resume walking for more than 30 s

12. How often do you fall?

- 0 Never
- 1 Very rarely—about once a year
- 2 Rarely—about once a month
- 3 Often—about once a week
- 4 Very often—once a day or more

13. How often do you fall when standing?

- 0 Never
- 1 It happened once or twice
- 2 It happened 3–12 times in the last 6 months
- 3 More than once a week
- 4 Whenever trying to walk unassisted

14. How often do you fall because of freezing episodes?

- 0 Never
- 1 It happened once or twice
- 2 It happened 3–12 times in the last 6 months
- 3 More than once a week
- 4 Whenever trying to walk unassisted

15. Do you experience festinating gait? (Festinating gait = accelerated, short steps, gait)

- 0 Never
- 1 Very rarely—about once a month
- 2 Rarely—about once a week
- 3 Often—about once a day
- 4 Whenever walking

16. How often do you fall because of festinating gait?

- 0 Never
- 1 It happened once or twice
- 2 It happened 3–12 times in the last 6 months
- 3 More than once a week
- 4 Whenever trying to walk unassisted

D Usability Questionnaire

Participant ID: _____

Date: _____

Use only 0, 1, 2, 3, 4 for answers, nothing else. Do not leave any blanks.

0 strongly disagree

1 mostly disagree

2 neither agree nor disagree

3 mostly agree

4 strongly agree

Area and Statement		Value
Benefit		
1.	I can benefit from this technology	
2.	The effort of using this technology/method is worthwhile for me	
3.	I am confident I am getting the most out of this technology/method	
4.	This technology/method is helping me to achieve my goals	
5.	I would recommend this technology/method to other people in my situation	
Usability		
6.	The use of this technology/method does not require effort	
7.	The technology/method is reliable according to my estimation and experience so far	
8.	This technology/method is easy to use	
9.	I feel safe when using this technology/method	
10.	I feel good while using this technology/method	
11.	I can easily sense the vibrations of the watch	
12.	I can easily hear the beeps of the watch	
13.	I am pleased with the look and design of the software on the phone	
14.	I am pleased with the look and design of the software on the watch	
Self-concept		
15.	The use of this technology/method is an interesting challenge for me	
16.	This technology/method helps me improve my independence	
17.	The use of this technology/method is making me feel younger than I am	
18.	I (would) feel comfortable using this technology/method visible around others	
19.	I like to use technologic products or systems like this technology/method	

Area and Statement		Value
Privacy and Loss of Control		
20.	The use of this technology/method does not require supervision	
21.	I use this technology/method by request of others (e.g., physician, therapist, relatives)	
22.	I am sure that my personal data are stored or processed in an appropriate way	
23.	I am sure that the use of this technology/method may not have any unpredictable negative consequences for me	
24.	This technology/method does not require me to disclose personal facts that I prefer to keep to myself	
Quality of Life		
25.	Using this technology/method improves my physical well-being	
26.	This technology/method does not evoke unpleasant feelings	
27.	This technology/method enhances my social contacts	
28.	This technology/method helps me to maintain or increase my independence (e.g., regarding mobility, communication, medication)	
29.	The use of this technology/method has a positive effect on me	
Wearing and Carrying Comfort		
30.	Wearing this device (parts of the device) is comfortable	
31.	I am pleased with the size of the device (parts of the device)	
32.	I am pleased with the look and design of the device (parts of the device)	
33.	I am pleased with the weight of the device (parts of the device)	
34.	The body-worn parts of the device are easy to adjust (fix, fasten)	
35.	Carrying the provided phone is comfortable	
36.	I am pleased with the size of the provided phone	
37.	I am pleased with the weight of the phone	
38.	I am pleased with the design of the provided phone	
39.	Wearing the provided watch is comfortable	
40.	I am pleased with the size of the provided watch	
41.	I am pleased with the weight of the provided watch	
42.	I am pleased with the design of the provided watch	
43.	Wearing the sensors is comfortable	
44.	I am pleased with the size of the sensors	
45.	I am pleased with the weight of the sensors	
46.	I am pleased with the look and design of the sensors	

E User Manual

,

1 Table of Contents

1	Beginning of the Day.....	2
1.1	Check the battery level on the Sensors.....	2
1.2	Check the battery level on the Watch.....	3
1.3	Check the battery level on the Phone.....	3
1.4	System Setup.....	4
1.5	“CONNECT”	6
1.6	Select “Audio”	7
1.7	“TEST WATCH”	7
1.8	“START”	8
2	End of the Day.....	11
2.1	“STOP”	11
2.2	“DISCONNECT”	12
2.3	Charging the Sensors.....	12
3	List of Items.....	14
4	Technical Support	14

Participant ID: P##_YYMMDD

1 Beginning of the Day

1.1 Check the battery level on the Sensors

1. Unplug the sensors from the charging station
2. Use the side button on the sensors to turn them on (see Figure 1).



Figure 1. Turn On/Off button on the sensors

3. Figure 2 shows LED-L (left) and LED-R (right) of the sensors.
LED-L shows the battery level, LED-R shows the connection status.
If the LED-L is blinking **red**, the battery level is low.
If the LED-L is blinking **yellow**, the battery level is medium.
If the LED-L is blinking **green**, the battery level is full.



Figure 2. LEDs on the sensors

4. If any of sensors are in low or medium level, connect them to the docking station and wait until all the sensors are fully charged and LED-L is blinking **green**.

1.2 Check the battery level on the Watch

1. Press the side button on the watch to turn on the screen.
2. The battery level can be seen on the screen as shown in Figure 3.



Figure 3. Battery level of the watch can be seen on the home screen (red box)

3. If the battery level is below 80%, connect the watch to the charger and wait until fully charged.

1.3 Check the battery level on the Phone

1. Press the side button on the Phone to turn on the screen.
2. Scroll up using your finger from the bottom edge towards the top edge to unlock the phone (Figure 4).

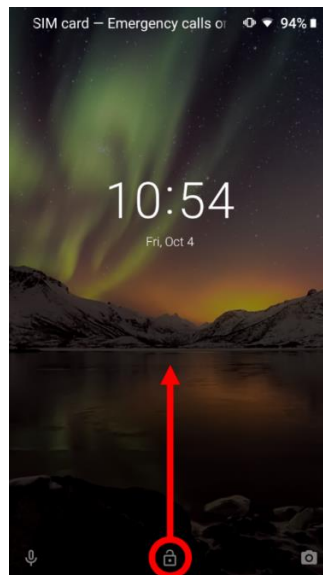


Figure 4. To unlock the phone, press the unlock icon on the bottom of the screen and scroll up while keeping your finger on the screen

3. The battery level is shown on the top right corner of the screen.
4. If the battery level is below 80%, connect the phone to the charger and wait until fully charged.

1.4 System Setup

1. When all the sensors are turned on and ready to use, wait for the LED-R (Right) on all the sensors to blink “blue” or “blue and green” (see Figure 5).

NOTE: If the LED-R is blinking “only green”, turn off the sensor and then turn it on after 3 seconds.



Figure 5. After turning on the sensors, LED-R will rapidly blink (for 0.1 second) “blue” or “blue and green”

2. When LED-R on all the sensors are blinking “blue” or “blue and green”, run the “FOG” application on the home screen of the phone (see Figure 6).



Figure 6. Open "FOG" app on the home screen of the phone

3. A screen may prompt asking your permission to turn on the Bluetooth. Press “ALLOW” to continue (see Figure 7).

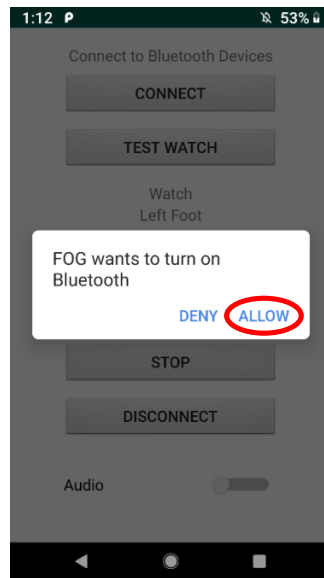


Figure 7. On the Bluetooth activation menu, press ALLOW to continue

4. Press the **UT** icon on the watch home screen to open the application (see Figure 8).
NOTE: If the watch screen was turned off, press the side button to turn it on again.



Figure 8. Press the **UT** icon on the watch home screen to open the application

1.5 “CONNECT”

1. On the watch screen, press “CONNECT” to connect to the phone (see Figure 9).

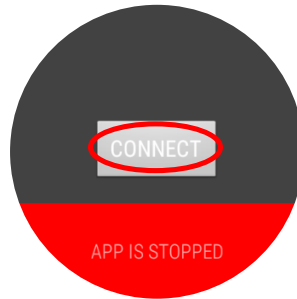


Figure 9. Press "CONNECT" to connect to the phone

2. On the phone, press “CONNECT” and **wait for 10 seconds** for the watch and sensors to connect.
 - a) Watch (see Figure 10):
 - If connected, it will be shown on the screen with **green** background color.
 - If disconnected, it will be shown on the screen with **yellow** background color.
 - b) Sensors (see Figure 10):
 - The connected sensors will be shown on the screen with **green** background color.
 - The disconnected sensors will be shown on the screen with **yellow** background color.

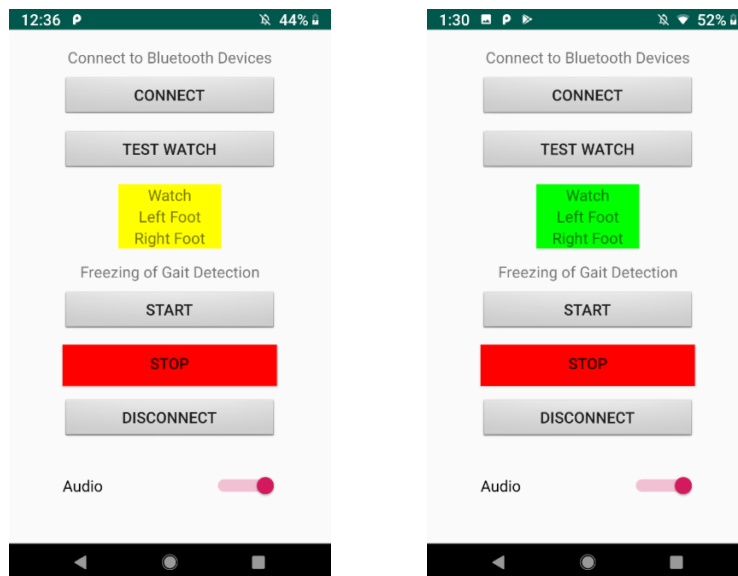


Figure 10. [left] The watch and sensors are **disconnected** (displayed by **yellow**),
[right] The watch and sensors are **connected** (displayed by **green**)

3. If the watch remained disconnected, press the “CONNECT” button on the watch (see Figure 9) and then press the “CONNECT” button on the phone again.

4. If any of the sensors remained disconnected:
 - a. Make sure all the disconnected sensors are turned on and the LED-R is blinking blue and green as shown in Figure 5.
 - b. Press the “CONNECT” button again and **wait for 10 seconds**.
 - c. If any of the sensors still remained disconnected, turn it off and wait for 2 seconds. Then turn the sensor on and press “CONNECT” on the phone screen.
 - d. If any of the sensors failed to connect, repeat from (b) until all the sensors are connected.

1.6 Select “Audio”

The watch will vibrate every time freezing is detected. You can also activate auditory feedbacks in the form of beeps using the “Audio” switch on the bottom of the phone screen (see Figure 11). When the switch is red, the auditory feedback is activated. You can turn this switch on and off at any time.

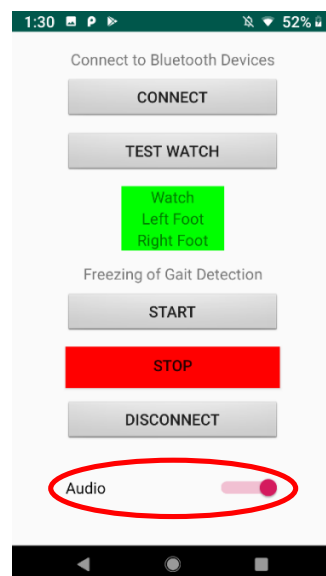


Figure 11 . When the “Audio” switch is red, the auditory feedback is activated

1.7 “TEST WATCH”

1. Follow the instructions in sections 1 on page 4 and to set up the devices, and section 0 on page 6 to connect the sensors and watch to the phone.
2. When the watch and all the sensors are connected, press the “TEST WATCH” button on the phone screen to trigger vibrations (and auditory beeps, if activated) on the watch (see Figure 12).
 - a. If the watch vibrates, the system is ready.
 - b. If the watch doesn’t vibrate, press the “CONNECT” button on the watch (see Figure 9) and then press the “CONNECT” button on the phone.
 - c. If the watch is shown on the phone screen with **green** background color but still doesn’t vibrate, check if the watch is charged above 80% (see section 1.2 on page 3). If the battery level is below 80%, connect the watch to charger and wait until it is charged.

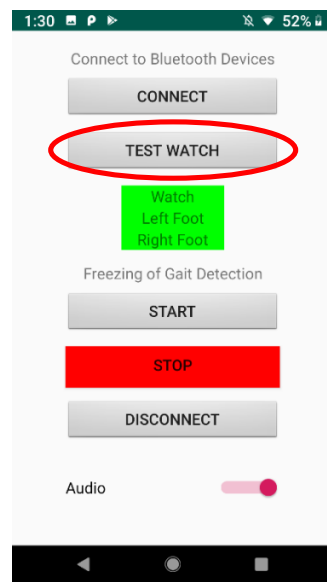


Figure 12. When the all devices are connected (shown green on the screen), press “TEST WATCH” button on the phone screen to trigger vibrations (and auditory beeps, if activated) on the watch

1.8 “START”

1. Wear the sensors on the ankles. The arrow on the labels of all sensors must be upward when standing normally (see Figure 13).

NOTE: Make sure that the sensors are fixed in their place and check their orientation every 30 minutes and adjust them if needed.



Figure 13. Foot sensor location and orientaion abobe the ankle

2. Wear the watch on your dominant hand snugly. In order to for you to feel the vibrations, the watch should stay in place on your wrist.
3. Check that the “STOP” button on the phone screen is **red** (see Figure 14), which means the automatic detection of freezing has not started yet. If it is not red, press “STOP” on the phone screen.

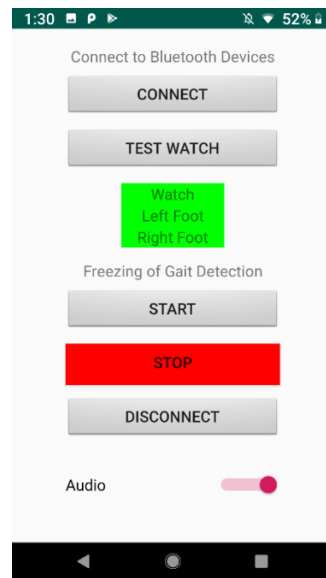


Figure 14. The “STOP” button must be **red** before starting the automatic detection of freezing

4. Press the “START” button on the screen to start automatic detection of freezing. This will change the color of “START” button to **green**, and the “STOP” button to **gray** (see Figure 15).

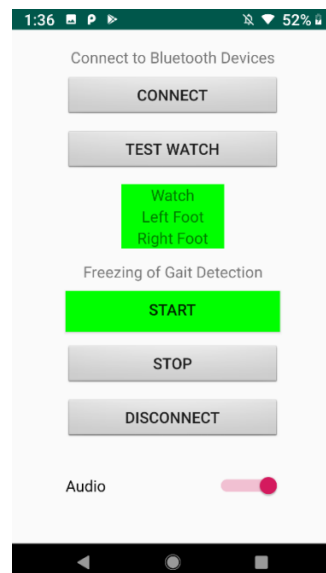


Figure 15. The “START” button will become green after starting the automatic detection of freezing

5. Turn the screen off on the phone using the side power button.
NOTE: When the screen is off, the application will remain active.
6. When the automatic detection of freezing is started, the notification on the watch will turn into **green**.

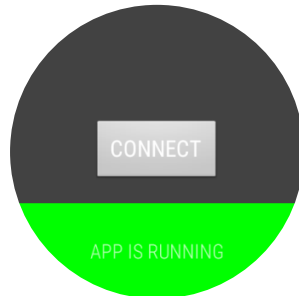


Figure 16 . When the application is started, the notification on the watch will turn **green**.

7. If the distance between the sensors and the phone become more that 6 ft., the sensors will disconnect, and the application will stop. You can check the application status using either of the following methods:
 - a. The “START” button on the phone screen will become **gray**, and the “STOP” button will turn **red**.
 - b. The notification on the watch screen will become **red**.
8. It is highly recommended to check the screen on the watch (or phone) every 15 minutes. If stopped, check the status of the watch and sensors on the phone screen.
9. If any of the watch or sensors become disconnected, follow the steps from section 0 on page 6, to reconnect.

2 End of the Day

2.1 “STOP”

1. Press the “STOP” button on the screen to stop automatic detection of freezing. This will:
 - a. change the color of “START” button to **gray**, and the “STOP” button to **red** (see Figure 17).
 - b. change the color of notification on the watch screen to **red** (see Figure 18).

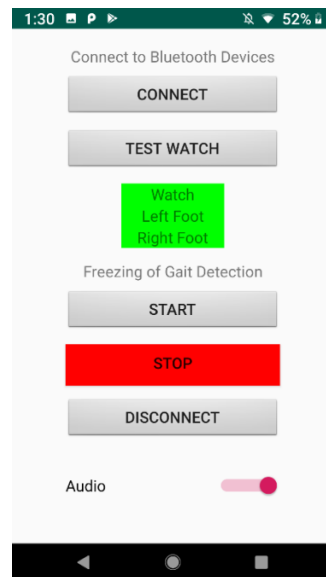


Figure 17. The “STOP” button on the phone will become red after stopping the automatic detection of freezing

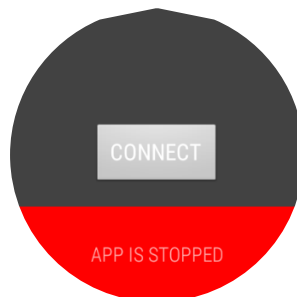


Figure 18. The notification on the watch will become red after stopping the automatic detection of freezing

2. To store and charge the sensors, follow the instructions in section 2.2 on page 12 to disconnect devices. Otherwise, if you want to use the device later on the same day, keep the watch and sensors connected to the phone.

NOTE: If you disconnect the watch and sensors, you need to configure them again. But if you keep them connected to the phone, they will be ready for the next session and you don’t have to configure them again.

NOTE: Because the devices consume battery when connected, we strongly recommend disconnecting and charging them if you are not going to use the system for the rest of the day.

2.2 “DISCONNECT”

1. Press “Disconnect” button on the phone’s screen. This will disconnect all the sensors and the watch and their color on the screen will turn to **yellow** (see Figure 19).

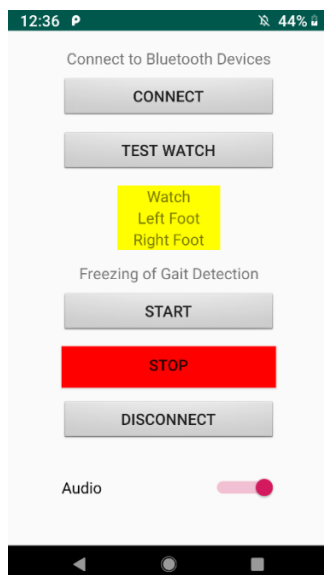


Figure 19. The background color of the watch and the sensors become **yellow** when disconnected

2. Charge the sensors (section 2.3)
3. Connect the watch to its charger.
4. Connect the phone to its charger.

NOTE: Disconnect the sensors, watch and phone from charger as soon as they are fully charged (see instructions in sections 1.1, 1.2 and 1.3 to check the battery levels) as overcharging will damage the battery. **Normally, two hours is enough for the sensors and watch. The phone will also be fully charged in three hours.**

2.3 Charging the Sensors

1. Turn off the sensors using the side button (see Figure 20).



Figure 20. Turn On/Off button on the sensors

2. Make sure the charging station is plugged in and the lights are on (see Figure 22)



Figure 21. Orange lights on the charging station will turn on when it is plugged in

3. Plug the sensors to the charging station and make sure the second lights on the docking station are on (see Figure 22). **Orange** means the sensors are being charged, and **green** shows that sensors are fully charged and can be disconnected.



Figure 22. Green lights on the charging station will turn on when the sensors are Fully charged

NOTE: Disconnect the sensors, watch and phone from charger as soon as they are fully charged (see instructions in sections 1.1, 1.2 and 1.3 to check the battery levels) as overcharging will damage the battery. **Normally, two hours is enough for the sensors and watch. The phone will also be fully charged in three hours.**

3 List of Items

Sensors		Phone		Watch	
Item	Quantity	Item	Quantity	Item	Quantity
Case	1	Box	1	Box	1
Sensors with bands	2	Phone	1	Watch	1
Charging Base	1	Adapter	1	Adapter	1
Adapter	1	Adapter Cable	1	Adapter Cable	1
Adapter Cable	1				

4 Technical Support

Nader Naghavi

Email: nnaghavi@utk.edu

Phone: (xxx) xxx – xxxx

Hours: 9 AM – 5 PM

Vita

Nader Naghavi received his B.Sc. and M.Sc. degrees in Mechanical Engineering from Sharif University of Technology, Iran, in 2011, and the University of Tehran, Iran, in 2014, respectively. He joined the Department of Mechanical, Biomedical, and Aerospace Engineering in the University Tennessee, Knoxville, in August 2017 as a research assistant pursuing a Ph.D. degree in Mechanical Engineering. His research interests include the applications of Machine Learning and Deep Learning in physiological signal processing. His current research is focused on developing mobile applications to detect and quantify motor impairment in patients with Parkinson's disease using data received from wearable devices.