Prediction and Detection of Freezing of Gait in Parkinson's Disease using Plantar Pressure Data

by

Gaurav Shalin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2021

AUTHOR'S DECLARATION

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Dr. Jonathan Kofman, as Gaurav Shalin's supervisor and Dr. Edward Lemaire and Dr. Julie Nantel as Gaurav Shalin's co-supervisors, and Scott Pardoel contributed in project conceptualization and data collection protocol development, and guidance on data analysis. Drs. Kofman, Lemaire, and Nantel also participated in reviewing all thesis chapters. Dr. Julie Nantel assisted in participant recruitment and data collection at University of Ottawa. Scott Pardoel and Gaurav Shalin performed the data collection described in Chapter 3.

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative disorder affecting movement and is characterized by symptoms such as tremor, rigidity, and Freezing of Gait (FOG). FOG is a walking disturbance seen in more advanced stages of PD. FOG is characterized by the feeling of feet being glued to the ground and has been associated with higher risks of falls. While falling can have great repercussions in individuals with PD, leading to restricted movement and independence, hip fracture, and fatal injury, even the disturbance of FOG alone can lead to decreased mobility, inactivity, and decreased quality of life. Determining methods to counter FOG can potentially lead to a better life for people with PD (PwPD).

Freezing episodes can be countered with the help of external intervention such as visual or auditory cues. Such intervention when administered during the freeze has been found to alleviate the freeze and thus prevent freeze-related falls. This sheds the importance of detecting or predicting a freeze event. Once a freeze is detected or predicted, an intervention can be administered to help prevent the freeze altogether (in case of prediction) or help resume normal walking (in case of detection).

Different wearable sensors have been used to collect data from participants to understand FOG and develop approaches to detect and predict it. Plantar pressure data has earlier been used in gait related studies; however, they have not been used for FOG detection or prediction. Based on the hypothesis that plantar pressure data can capture subtle weight shifts unique to FOG episodes, this research aimed to determine if plantar pressure data alone can be used to detect and predict FOG.

In this research, plantar (foot sole) pressure data were collected from shoe-insole sensors worn by 11 participants with PD as they walked a predefined freeze-provoking path while on their normal antiparkinsonian medication. The sensors included IMU, EMG, and plantar pressure foot insoles; however, for the research in this thesis, only plantar pressure data were used. The walking trials were also video recorded for labelling the data. A custom-built application was used to synchronize data from all sensors and label them. This was followed by feature extraction, dataset

balancing, and z-score normalization. The datasets generated were then classified using Long-short term memory (LSTM) networks.

The best model had an average 82.0% (SD 6.25%) sensitivity and 89.4% (SD 3.60%) specificity for one-freezer-held-out cross validation tests. For the participants who did not freeze during the walking trials, an average 87.7% specificity was achieved. Since, FOG detection is done with the aim to provide an intervention, a freeze episode analysis was completed, and it was found that the model could correctly detect 95% of freeze episodes. The misclassified freezes and false positives were analyzed with respect to active (walking and turning) and inactive states (standing). The model's specificity performance for one-freezer held out cross validation tests was found to improve to 93.3% when analyzing the model only on active states. FOG prediction was done afterwards, including data before FOG (labelled Pre-FOG) in the target class. The best FOG prediction method achieved an average 74.02% (SD 12.48%) sensitivity and 82.99% (SD 5.75%) specificity for one-freezer-held-out cross validation tests.

The research showed that plantar pressure data can be successfully used for FOG detection and prediction. Moving away from window-based model also helped the research in reducing the freeze detection latency. However, further research is required to improve the FOG prediction performance and a bigger sample size should be used in future research.

Acknowledgements

I would like to thank my supervisor Professor Jonathan Kofman for the continuous guidance and support that he provided during my master's degree. I would like to thank my co-supervisors Professor Edward Lemaire and Professor Julie Nantel for the guidance and invaluable insights they provided. I would like to thank my lab mate Scott Pardoel for his continuous support throughout my journey. I would like to express my gratitude toward my friends Arvind, Anmol, Harsh, Avi, Arnab, Devinder, and Manpreet for helping and guiding me through my personal life and lab-work.

I would like to thank my grandmother Late Pushplata Devi for helping me become the person I am today. Above all, I would like to thank my mother Dr. Shashibala Shiv Shankar for her incredible love and support.

I also wish to thank all the participants whose data was used in this work. I would also like to thank 'Compute Canada' for the computational resources provided.

This research was supported by Microsoft Canada; Waterloo Artificial Intelligence Institute, and Network for Aging Research, at University of Waterloo; and Natural Sciences and Engineering Research Council of Canada (NSERC).

Dedication

This is dedicated to my family.

Table of Contents

AUTHOR'S DECLARATION ii
Statement of Contributions iii
Abstract iv
Acknowledgements vi
Dedication vii
List of Figures x
List of Tables xi
List of Abbreviations xiii
Chapter 1 Introduction 1
1.1 Research Objectives 4
1.2 Research Contributions 4
1.3 Thesis Outline 5
Chapter 2 Background 6
2.1 Freezing of Gait and Functional Mobility 6
2.2 Freezing of Gait Detection 7
2.2.1 Gradient Descent Optimization Algorithms in Neural Networks 9
2.2.2 Convolutional Neural Networks 10
2.2.3 Recurrent Neural Networks 12
2.2.4 Long Short-Term Memory 13
2.3 Freezing of Gait Prediction 16
2.4 Summary 18
Chapter 3 Data Collection 20
3.1 Research Overview 20
3.2 Participants and Inclusion Criteria 21
3.3 Plantar Pressure Measurement 21
3.4 Clinical Assessment 22
3.5 Test Protocol 22
3.6 Results 23
Chapter 4 Freezing of Gait Detection 25
4.1 Data preprocessing and labeling 25
4.2 Experiment 1: FOG Detection with Raw Data 25

4.2.1 Methods 25
4.2.2 Results 26
4.3 Experiment 2: FOG Detection with Features 27
4.3.1 Feature Extraction 27
4.3.2 LSTM Model 28
4.3.3 Experiment 2a: LSTM Without Data Balancing 29
4.3.4 Experiment 2b: LSTM with Balanced Training Set 29
4.3.5 Experiment 2c: Determination of the Best LSTM Model 31
4.3.6 Cross Validation 32
4.4 Results 33
4.5 FOG Detection Latency 35
4.6 False Positive and Not-detected Analysis 36
4.7 Discussion 38
Chapter 5 Freezing of Gait Prediction 41
5.1 FOG prediction 41
5.2 Results 43
5.3 Discussion 44
Chapter 6 Conclusion 46
6.1 Conclusions Related to Objectives 46
6.2 Future work 48
References 50

List of Figures

Figure 2.1: Recurrent Neural Network unrolled in time. 13
Figure 2.2: LSTM unit representation depicting how activation, memory state, and output at current time step are obtained from activation, input, and memory state at previous time step. 15
Figure 3.1: F-Scan system: A plantar pressure insole sensor (left) and sensors worn in shoes (right). 21
Figure 3.2: Walking Path. Participants started from a chair, walked straight, navigating multiple cones, walked as far into a narrow hallway as possible, turned 180°, and walked back to the starting position. Participants stopped (voluntarily) once while walking back to the starting position. Participants also stopped at the end of each test protocol. 22
Figure 3.3: PD participant turning in a narrow hallway while holding a tray with a cup. Assistant follows for safety. 23
Figure 3.4: Plantar pressure array sample frame (kPa). Dark blue indicates zero pressure. 24
Figure 4.1: 2D CNN classifier architecture. 26
Figure 4.2: False positives and misclassified freezes distributed across active (walking, turning) and inactive (standing) states with the 2-layer LSTM model. Not defined refers to the beginning and end of a trial when a participant was not performing a specific activity. 36

List of Tables

Table 3.1: Data collection summary with freeze episodes, their mean and total duration for each participant. 24

Table 4.1: LSTM Network configurations tried: Number of LSTM layers, number of units in each LSTM layer, different learning rate schedule were tried to find the best network configuration. 31

Table 4.2: Best performing LSTM network configuration for FOG detection. 32

Table 4.3: FOG Detection: One-freezer-held-out cross validation for the 2-layer LSTM model. Each LSTM layer had 16 units. LSTM layers were followed by a time-distributed fully-connected layer and a Softmax activation. 34

Table 4.4: FOG Detection: All-non-freezer-held-out validation for the 2-layer LSTM model. Each LSTM layer had 16 units. LSTM layers were followed by a time-distributed fully-connected layer and a Softmax activation. 34

Table 4.5: FOG Detection: One-freezer-held-out cross validation for the 3-layer LSTM model. Each LSTM layer had 32 units. LSTM layers were followed by a time-distributed fully-connected layer and a Softmax activation. 34

Table 4.6: FOG Detection: All-non-freezer-held-out validation for the 3-layer LSTM model. Each LSTM layer had 32 units. LSTM layers were followed by a time-distributed fully-connected layer and a Softmax activation. 35

Table 4.7: Average FOG detection latency and standard deviation in one-freezer-held-out cross validation with the 2-layer LSTM model. A negative freeze detection latency means that the freeze was detected before the true freeze onset, as desired. 35

Table 4.8: Percentage of active and inactive states resulting in false positives with the 2-layer LSTM model. 37

Table 4.9: FOG Detection: One-freezer-held-out cross validation using the 2-layer LSTM model only on active states and on both active and inactive states. The active states include walking and turning during the trials and excludes standing. 37

Table 4.10: FOG Detection: All-non-freezer-held-out validation using the 2-layer LSTM model only on active states and on both active and inactive states. The active states include walking and turning during the trials and excludes standing. 37

Table 4.11: Computer memory requirement of different models in KB. 38

Table 5.1: Network configuration used for FOG prediction. The 2-layer LSTM network had 16 units in each layer while the 3-layer LSTM network had 32 units in each layer. LSTM layers were followed by a time-distributed fully-connected layer and Softmax activation..... 43
Table 5.2: FOG Prediction: One-freezer-held-out cross validation with the 2-layer LSTM model after 4 training epochs. 43
Table 5.3: FOG Prediction: All-non-freezer-held-out validation with the 2-layer LSTM model after 4 training epochs. 44
Table 5.4: FOG Prediction: One-freezer-held-out cross validation with the 3-layer LSTM model after 3 training epochs. 44
Table 5.5: FOG Prediction: All-non-freezer-held-out validation with the 3-layer LSTM model after 3 training epochs. 44

List of Abbreviations

AR: Autoregressive

ARMA: Autoregressive Moving Average

CNN: Convolutional Neural Network

COP: Center of pressure

ECG: Electrocardiography

EEG: Electroencephalogram

EMG: Electromyography

FI: Freeze Index

FOG: Freezing of gait

GRF: Ground Reaction Force

GRU: Gated Recurrent Unit

GUI: Graphic User Interface

IMU: Inertial Measurement Unit

LOPO: leave one participant out

LSTM: Long short-term memory

MGD: Multivariate Gaussian Distribution

PD: Parkinson's disease

PwPD: People with Parkinson’s disease

RBF: Radial Basis Function

ReLU: Rectified Linear Unit

RNN: Recurrent Neural Network

SCR: Skin Conductance Response

SD: Standard deviation

SVM: Support Vector Machine

Tanh: Hyperbolic Tangent

UPDRS: Unified Parkinson’s Disease Rating Scale

Chapter 1
Introduction

Parkinson's disease (PD) is a progressive neurodegenerative condition characterized by motor symptoms such as rest tremor (tremor occurring mostly at rest that lessens during physical activities), bradykinesia (slowness in physical action and difficulty in initiating movements) rigidity, loss of postural reflexes, and freezing of gait (FOG) [1]. FOG is a gait disorder, defined as the inability to step effectively and move forward despite the clear intention to do so, and occurs in up to 26% of people in the early stage of PD [2-4] . While FOG can occur in both early and advanced stages of PD, it is known to affect 20% to 60% of people in the more advanced stages of the disease [5].

Most FOG episodes are brief, typically lasting less than ten seconds [6], and can manifest as shuffling small steps, trembling in place, or total akinesia (when a person losses the ability to move their muscle) [7]. Episode severity is determined by duration, frequency, and intensity [8]. FOG manifests frequently at home when the participant is not being observed [9], but is difficult to evoke in a lab setting [10,11]. In some people, FOG episodes occur during medication "ON state" due to dopaminergic drugs [7]. However, in most people, FOG episodes occur while off medication ("OFF State", where medication has worn off) [2-4]. FOG episodes are also more severe in "OFF State" and rarely manifest as total akinesia during "ON State" [12].

FOG can cause falls [13], and even a brief FOG episode of less than 10 seconds can result in forward or lateral falls [7]. A forward fall happens when the feet are suddenly unable to move while walking. Lateral falls happen when a person freezes during a turn. Most falls due to FOG occur during the transition from walking to a freeze episode. While recurrent falls can increase a person with PD's mortality risk, reducing survival years by seven [14], even repeated freeze episodes can negatively affect the overall mobility of a person with PD, their level of activity and independence, and thus quality of life [15-17]. Therefore, research aimed toward reducing the occurrence of FOG in people with PD (PwPD) can have an important beneficial impact on their independence and quality of life.

Cueing can alleviate FOG by providing external stimuli (such as light or sound) that facilitate gait initiation and continuation [18]. Cueing during a freeze episode can help a person resume walking. However, continuous cueing may be distracting when a person is not walking, can lead to cue dependency [19] and induces greater fatigue [20] in PwPD. Furthermore, cueing with a pre-set rhythm that is not matched to the person's specific gait at each instant may induce FOG [21]. Cueing only during a freeze could lead to end or reduce the duration of a freeze. This would require a freeze detection system. Since many falls occur during the transition from walking to a freeze, FOG episodes should be predicted in order to prevent such falls. An FOG prediction system could prevent a FOG episode when used with cueing just before the transition to FOG. The development of methods for both freeze detection and prediction are thus important toward reducing freeze duration and occurrence.

FOG detection and prediction need physiological parameters that discriminate FOG from walking or standing. Wearable sensors such as electrocardiography (ECG), electromyography (EMG), and Inertial Measurement Unit (IMU: accelerometer, gyroscope, magnetometer) can capture physiological and biomechanical parameters relevant to FOG. ECG, EMG, and IMU have been used for detecting and predicting FOG [22]. Plantar pressure (i.e., pressure between the foot and the ground) is another gait related biomechanical parameter that has been used in several gait related studies such as fall-risk assessment [23,24], but not for FOG detection or prediction. Plantar pressure insoles can provide a 2D array of pressure readings from the foot's plantar (sole, bottom) surface. If plantar pressure manifests differently in a freeze episode than normal walking, this wearable sensor may be useful for FOG detection and prediction since the entire sensor and electronics can be easily used in footwear (i.e., easy to don and doff, unobtrusive, does not affect movement, etc.).

A FOG detection or prediction system needs a classifier to differentiate FOG, or episodes prior to a freeze, from normal walking and standing. Several classifiers including support vector machines (SVM), decision trees, and artificial neural networks have been used for FOG detection and prediction [22]. At each instant, plantar pressure data is a 2D matrix of pressure values. A Convolutional neural network (CNN) can be used to classify the raw 2D matrices of pressure values as belonging to a FOG or Non-FOG class. A CNN is an artificial neural network that can

learn hierarchies of patterns across fixed-size inputs through kernels and can be useful in FOG detection and prediction. Most machine learning classifiers would require features to be extracted from 3D plantar pressure data (row, column, pressure value) before training and implementing a model. A CNN can be used to extract features from the raw data, or handcrafted features may be used. The extracted features can then be fed into a machine-learning classifier for classification of the data.

Since wearable sensors provide a data stream (temporal sequence), a range of inputs should be considered while learning patterns from the sequence. Recurrent Neural Networks (RNN) are deep learning classifiers that learn from a temporal sequence. An RNN is an artificial neural network that can store previous information as an internal state (memory). Unlike a conventional feed-forward neural network or convolutional neural network (CNN), an RNN can have variable length inputs. Since freeze episodes vary in temporal length, RNNs could be suitable for FOG detection and prediction. Long short-term memory (LSTM) classifiers are a version of RNN that can learn from longer sequences [25]. Candidate values are calculated from the input data. LSTM units use a memory cell that is updated based on these candidates. LSTM classifiers have outperformed RNN for FOG prediction [26]. Thus, RNN based on LSTM units can be useful for FOG detection and prediction models.

For FOG detection or prediction using plantar pressure data, the raw plantar pressure data could be classified using a CNN, or alternatively, features that are extracted from the plantar pressure data can be fed into a classifier. Further, either handcrafted features could be extracted, or a CNN can be used to extract features. These extracted features can then be classified using machine learning classifiers or an LSTM network.

For FOG detection, one of the best results in a leave-one-participant-out (LOPO) cross validation have been achieved by a K-means clustering algorithm using IMU data [27]. The algorithm achieved 92.4% (SD 1.2%) sensitivity and 94.9% (SD 2.3%) specificity for ten PD participants. For FOG prediction in a participant-independent setting (i.e., a model was validated on a participant's data unseen by the model), one of the best results was obtained with a neural network using EEG data [28]. The neural network achieved 85.56% sensitivity and 86.60%

specificity on 5 held-out PD participants. The above results can be a benchmark for developing new FOG detection and prediction models based on plantar pressure data only.

1.1 Research Objectives

The research objectives of this thesis are as follows:

a) Determine if an LSTM network can detect FOG episodes with sensitivity > 92.4% (SD 1.2%) and specificity > 94.9% (SD 2.3%) [27] in a person-independent model using only plantar pressure data.

b) Determine if an LSTM network can predict FOG episodes with sensitivity > 85.86% and specificity > 80.25% [28] in a person-independent model using only plantar pressure data.

c) Determine if an LSTM network can achieve the benchmarks in a) and b) using custom features extracted from only plantar pressure data.

d) Determine if an LSTM network can achieve FOG detection and prediction, with similar specificity between participants who did not freeze during the walking trials and participants who froze during the trials.

e) Determine if only classifying FOG during active states (i.e., walking, moving, etc.) can reduce freeze episode misclassification and false positives by an LSTM network.

1.2 Research Contributions

The main contributions of this research are as follows:

a) This research developed the first viable FOG detection system that uses only in-shoe plantar pressure measurements. This can lead to an efficient wearable FOG mitigation system that is unobtrusive to a person with PD, thereby enhancing wearing compliance.

b) This research demonstrated that adding activity-state recognition to a wearable FOG mitigation system greatly reduces FOG-detection false positives. This could lead to improve future user experience by avoiding unneeded cueing that might lead to system abandonment.

c) This research identified a range of hyperparameters that can be used in LSTM network training for FOG detection and prediction. This can guide future research by providing an appropriate hyperparameter search space.

d) This research used data at each data-sampling time instant so that the model can be computationally efficient and mean FOG detection latency can be kept low in a wearable system, in comparison to an approach using data time windowing.

1.3 Thesis Outline

Chapter 2 discusses Parkinson's disease, FOG, and the FOG relationship to falling; reviews literature on FOG detection and prediction systems; and briefly reviews RNN and LSTM theory.

Chapter 3 presents the data collection process, including, participant recruitment, materials and sensors, and participant preparation during data collection and test protocol.

Chapter 4 presents FOG detection models; including, data setup, feature extraction and normalization; freeze episode misclassification, and false-positive analysis with respect to activities during the walking trials; and discussion.

Chapter 5 presents the FOG prediction method along with results and discussion. The thesis concludes in Chapter 6.

Chapter 2
Background

2.1 Freezing of Gait and Functional Mobility

Parkinson's disease (PD) is a slowly progressive neurodegenerative disorder characterized by symptoms such as bradykinesia, tremor, and rigidity [8], with the most distressing symptoms being FOG, postural instability, sleep disturbance, and difficulty in concentrating [8]. Trembling is the main complaint during early-stage PD, while FOG is rare in the early stage, manifesting only as gait hesitation [6]. As the disease progresses, balance impairment and FOG become more common, leading to falls [2,13]. Most falls in PD are caused by balance disorders rather than environmental causes (i.e., collisions, etc.) [29].

FOG manifests as total akinesia, trembling in place, and shuffling forward with small steps. FOG episodes generally last less than 10 seconds and rarely more than 30 seconds [6]. FOG events are more frequent in the OFF-medication state than in the ON state [13]. Most FOG episodes occur while turning or initiating walking, and while walking straight in a narrow pathway or trying to avoid obstacles [13]. Sudden changes in posture are attributed to falls in PD participants, particularly turning the trunk or multitasking while walking or balancing [13]. While most PD individuals can have a conversation while walking, they have difficulties performing other dual tasks during gait (such as carrying an empty or loaded tray, answering repetitive questions, or walking in reduced illumination), leading to FOG or balance loss, more so as these tasks become more complex [30]. Under complex conditions, individuals with PD can have poor gait and balance when they try to perform all tasks equally well [30].

Attention and vision can have both positive and negative effects on FOG occurrence [16]. While FOG episodes are more likely to occur in crowded places or as the person navigates narrow passages, focusing attention on each step through an external cue can alleviate FOG [31]. Auditory, visual, or tactile cues can help PwPD overcome freezing episodes and resume walking [32,33]. Auditory cues are the most common type for FOG and are generally administered using a computerized tone, such as a short beep, which are played in accordance with a participant's normal waking cadence [34]. Auditory cues can reduce FOG duration and frequency [5]. Laser

lines projected on the ground [35] have been used as a visual cue for alleviating FOG. These cues are often accompanied by a freeze detection system to automatically identify freezing episodes and activate the assistive cue only when needed.

2.2 Freezing of Gait Detection

The objective of a FOG detection system is to correctly identify all FOG occurrences and avoid falsely recognizing Non-FOG events as FOG occurrences. Further, a usable FOG detection system should be easily incorporated in a person's daily mobility activities. Data from the system can be captured using wearable sensors. A wearable physiological or biomechanical sensor that can be integrated with clothes or shoes would be desirable. Most research on FOG detection used 3D accelerometer data [36-39] or data from multiple IMU sensors [38,40]. IMU have been placed at the hip, above the knee, or on the ankle [36,41]. IMU located on the lower leg performed best, while hip-located IMU showed the highest standard deviation across participants and the thigh position was the most uncomfortable [41].

Plantar pressure data are also promising for FOG detection and prediction. Plantar pressure is based on the reaction force exerted by ground on participant's shoe insoles. Plantar pressure has been used for rehabilitating people with a spinal cord injury [42], post-traumatic rehabilitation [43], and stroke rehabilitation [44]. Since plantar pressure distribution can be useful in assessing lower limb function, this measure has been helpful in providing better rehabilitation strategies [45]. Apart from rehabilitation purposes, plantar pressure data have been used for fall-risk prediction [23] and classification [24] in older adults and to classify gait as belonging to a participant with PD or a healthy control [46].

Plantar pressure data can capture subtle weight shifts within a foot and between feet. Furthermore, the pressure distribution may vary distinctly between normal walking phases, during the transition from normal walking into a freeze, and during a freeze. Plantar pressures have been used in gait analyses for both PD and non-PD populations; however, they have not been used for FOG detection or prediction. The preliminary phase of this research using plantar pressure data for FOG detection and prediction has shown promise [47,48]; thus, using plantar pressure data may open new avenues in detecting or predicting FOG events.

Machine-learning algorithms have been used for classifying FOG [36,39,40,49]. Random forests have outperformed other machine learning classifiers such as pruned decision trees, K-nearest neighbor (KNN), and Naïve Bayes in a participant dependent setting [36] (i.e., each participant's data were in both training and test sets) and LOPO cross validation [40]. A random forest achieved 99.54% sensitivity and 99.96% specificity with 4 s windows but with the same participant data from the Daphnet dataset [50] in both training and testing groups [36]. Four time-domain features, signal energy, freeze index (power ratio in 3-8 Hz band and 0.5-3 Hz band), and signal power were computed from each acceleration data window. Use of the same participant data for training and testing created bias and reduced generalizability for this model [36]. Freeze detection latency (i.e., delay between FOG start and detection by the system) was 1.085 s (SD 0.731 s) for 4 s windows, decreasing to 0.235 s (SD 0.175 s) for 1 s windows with a pruned decision tree. The average FOG detection latency increased linearly with increasing window length, demonstrating a trade-off between detection performance and detection latency while choosing a window length.

Random forests achieved 89.30% (SD 9.47%) sensitivity and 79.15% (SD 7.48%) specificity for PD participants with FOG symptoms in LOPO cross validation [40]. Data were collected from 11 PD participants (only 5 showed FOG symptoms) with 93 FOG events using 6 accelerometers and 2 gyroscopes. Accelerometers were placed on both legs, both wrists, chest, and waist. Gyroscopes were located on the chest and waist. Manual data processing such as replacing missing values and filtering low frequency components were required before entropy was calculated as a feature from all sensor data. The model had a high standard deviation of 9.47% and was validated only on 5 PD participants, who showed FOG symptoms. Multiple sensors over different body locations would require more time to don and keep in place during the day, thereby being inconvenient to the user and perhaps reducing device compliance. Thus, the application of the FOG detection model based on accelerometers and gyroscopes is limited.

When a different sensor location and features were used than those in the above studies, SVM outperformed other machine learning classifiers [39,49]. Three-dimensional acceleration data from 15 participants with PD were collected from an IMU placed at participant's waist and 55 features were extracted from 3.2 s windows [49]. The SVM model on average had 6790 support

vectors, which required 1.49 MB computer memory, and achieved 74.7% sensitivity and 79% specificity with a LOPO validation. In a personalized setting, 50% FOG data and the corresponding non-FOG data from the target participant became part of the training set. Model performance in the personalized setting increased to 80.1% sensitivity and 88.1% specificity [49]. The personalized model with 9221 support vectors would need 2 MB computer memory.

In follow up research, features were reduced from 55 to 28 [39]. SVM with 1.6 s windows achieved 84.49 % sensitivity and 85.83 % specificity, in LOPO cross validation [39]. SVM with 1.6 s windows outperformed other "machine learning classifiers and window length" combinations. Using a stratified 10-fold cross validation when data were not split by participants (i.e, each participants' data were both in training and validation set), the performance increased to 91.7% sensitivity and 87.4% specificity. The SVM model with 7412 support vectors required 1.6 MB computer memory. Using a single IMU sensor allowed for a wearable system; however, multiple complex features were used, with preprocessing steps that could be computationally expensive for a wearable device.

2.2.1 Gradient Descent Optimization Algorithms in Neural Networks

In the standard gradient-descent learning algorithm in neural networks, gradients are calculated based on the true label and the output by the neural network. These gradients are then used to update the network parameters (i.e., weights and biases). However, gradient descent with momentum works faster than the standard gradient descent algorithm. In a gradient descent with momentum algorithm, the exponentially weighted average of gradients (called first moment estimate because the derivatives are directly used) is calculated and used to update the network parameters. Another useful optimization algorithm is RMSprop, which uses an exponential weighted-average of the squares of the derivative (called second moment estimate because of squares of derivatives) to update the parameters. The Adam optimizer combines both the gradient descent with momentum and gradient descent with RMSprop, and has been found to excel for many applications. The gradient-descent Adam optimizer speeds up the descent towards the minima and decreases the oscillation of the descent in other directions. The first moment estimations are denoted by m_t, while the second moment estimations are denoted by v_t . Both the

first and second moments are initialized as zero vectors before first iteration of the learning. At iteration *t*, the first and second moments are calculated using a moving average of the gradients (square of gradients in case of second moment).

Equation 2.1 and 2.2 show the first moment and second moment estimation, respectively, the first moment m_t is calculated as the exponential weighted moving average of gradients g_t, while the second moment v_t is calculated as the exponential weighted moving average of element-wise square of gradients.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{2.1}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{2.2}$$

$$\theta_{t+1} = \theta_t - \alpha m_t / (\sqrt{v_t} + \varepsilon)\,, \tag{2.3}$$

where β_1 and β_2 denote the exponential decay rate for the first and second moment estimation, respectively. The parameter θ at time *t-1* is updated based on Equation 2.3, where α denotes the learning rate. Learning rate (α), exponential weight decay rate for first and second moment estimates (β_1 and β_2), and a very small value $\varepsilon = 10^{-7}$ govern the learning process with Adam optimizer. ε is used to avoid the denominator from becoming zero. The learning rate in Adam optimizer needs to be tuned; however, the β_1 and β_2 parameters are set to default values of 0.9 and 0.999, respectively.

2.2.2 Convolutional Neural Networks

While machine learning models have been successful for FOG detection, deep learning methods have the potential to achieve better performance without feature extraction. A Convolutional Neural Network (CNN) is a deep learning algorithm that learns a hierarchy of features across an image or a windowed signal, starting from low level features such as edges in earlier layers to high level features such as a human face in final layers.

A 1D CNN was used for FOG detection [38,51,52], with a 4-layer 1D CNN model [52] achieving 74.43% (SD 9.79 %) sensitivity and 90.59 % (SD 6.4 %) specificity in a LOPO cross validation. Acceleration data were acquired from 10 participants' left shank, left thigh, and lower back (8 participants froze during trials). The model was trained with stochastic gradient descent

(0.9 momentum coefficient), 0.001 constant learning rate, and 128 batch size. These hyperparameters (parameters set to control a model's learning process) can be a good starting place to train a model. Batch-normalization and drop-out showed little influence on classification performance, and thus may not be considered a priority while configuring a network for FOG detection. The model suffered from a large variability in sensitivity (SD 9.79 %) across participants [52].

A 1D CNN model with 2.56 s windows achieved 88.6% sensitivity and 78% specificity on 4 held-out participants [38]. Data were collected from 15 participants with an IMU placed at their waist. Five convolutional layers (each with 16 kernels) and 2 fully connected layers (with 32 neurons) were used in the CNN model. Various hyperparameters were optimized and it was concluded that initializing network weights with 'Glorot' weight initialization, activating convolutional layer's output with 'ReLU' activation function (rectified linear unit defined as y=x for x>0), and training the network with 'Adam' optimizer work best. To account for training data imbalance, weighted hinge loss was used as the loss function [38]. These hyperparameters and loss function can be used in research on FOG detection when dealing with an imbalanced dataset.

In follow up research, 1D CNN model performance improved to 91.9% sensitivity and 89.5% specificity on 4 held out participants [51]. The new model used 9-channel IMU data from 21 PD participants. The number of convolutional layers was reduced to 4 and network weights were initialized using 'Xavier' initialization. Thus, adding data from more participants and including multiple IMU channels can improve a deep learning model's performance. The model needed 37,121 parameters (145 KB computer memory) to classify a sample, which was an improvement over the previous SVM model's 1.6 MB memory requirement [39]. These results demonstrated deep learning model's advantages for FOG detection when a larger dataset is available for training. However, better validations strategies (LOPO cross validation, etc.) may have produced lower results than a favorable train-test group split [38,51]. A stacking strategy was used to combine spectral domain information in current and previous windows, which increased the computational load [38,51]. However, a 16 batch size [38,51] or 128 batch size [52] was used to reduce time consumption in model training, and a 1 batch size has been recommended for the

best generalization error. (i.e., the difference in model's performance over training and test datasets) [53].

These deep learning methods reported good performance for FOG detection; however, the validation methods created uncertainty in model generalizability. More research is required to explore deep learning methods for FOG detection, especially when using sensor setups that are viable in a system that is worn all day by PwPD . Moreover, most FOG detection methods required signal preprocessing, which could increase the delay in a real time FOG detection system. More research is required to reduce or eliminate most time-consuming preprocessing steps.

2.2.3 Recurrent Neural Networks

A recurrent Neural Network is an artificial neural network that can model sequence data, taking a sequence as input and outputting a sequence. Unlike CNN, RNN input and output sizes are not fixed. RNN is used in speech recognition, music generation, sentiment classification, DNA sequence analysis, machine translation, video activity recognition, and other applications (i.e., where a standard neural network cannot be used). For these applications, each training and test sample varies in length, and features are learned across different positions/instances in the sequence.

RNN architectures include "many to many" (for machine translation), "many to one" (for sentiment classification), and "one to many" (for music generation). A "many to many" RNN has a sequence input and sequence output. RNN uses information only from values earlier in the sequence (i.e., earlier from the present position) and not from the sequence ahead. Bidirectional RNN learn from information ahead in the sequence; however, bidirectional RNNs are not suitable for real-time systems because a decision can be made only at the sequence's end. Bidirectional RNN were thus not explored in this thesis where the outcome must be applicable immediately from new streamed input in a wearable cueing control system.

RNN output prediction at a time instance is computed with input at that time but also from previous time steps. At each time step, an activation is calculated and passed on to the next time step for the network to use. Typically, a zero vector forms the activation needed for the first time-step.

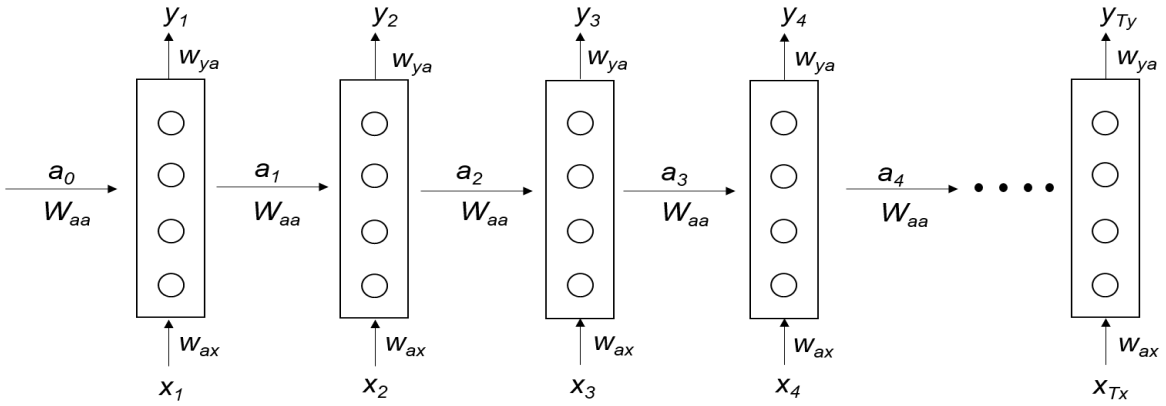

Figure 2.1: Recurrent Neural Network unrolled in time.

Equations 2.4 and 2.5 depict forward propagation in an RNN, w_a and $[a_{t-1}, x_t]$ are given by Equations 2.6 and 2.7, respectively.

$$a_t = g_1(w_a[a_{t-1}, x_t] + b_a) \tag{2.4}$$

$$y_t = g_2(w_{ya}\, a_t + b_y) \tag{2.5}$$

$$w_a = [w_{aa}\, w_{ax}] \tag{2.6}$$

$$[a_{t-1}, x_t] = [a_{t-1}\, x_t]^T. \tag{2.7}$$

In Equations 2.4 to 2.7, a_t is the activation at time *t* and x_t is the input at time *t*. In these equations, w_{aa}, w_{ax} and w_{ya} are the weights while b_a and b_y are the biases. Activation at time *t* is calculated using activation at time *t*-1, input at time *t*, weights (w_{aa}, w_{ax}) and bias b_a , which are passed through an activation function g_1. The activation function g_2 is generally hyperbolic tangent (tanh) or rectified linear unit (ReLU). The output at time *t* is calculated using activation at time *t*, which is multiplied by weight w_{ya} and then bias b_y is added before passing through activation function g_2 (sigmoid or Softmax function). The weights and biases are shared across all time steps. An RNN unrolled in time is shown as a "many to many" architecture in Figure 2.1.

2.2.4 Long Short-Term Memory

Long short-term memory (LSTM) units were introduced to solve the vanishing gradients problem and thus allow RNNs to learn from longer sequences [25]. The vanishing gradients are

caused when gradients calculated at the output layer decrease exponentially with increasing hidden layers and are not able to backpropagate to the first few layers to update their weights. For an RNN, the gradient associated with later time steps does not affect the computations in earlier time steps. Gradients in a neural network are calculated based on the error between the network output and the true label that backpropagates through the network to update the weights iteratively.

An LSTM unit calculates a candidate value based on the current input and the previous activation. Update and forget gates allow a memory cell to keep or forget the old candidate value while including a new candidate value. If the update and forget gates are set up correctly, an LSTM could retain the candidate's value for a longer duration than a standard RNN, thus capturing long-term dependencies.

Equations 2.8 to 2.14 show forward propagation in an LSTM unit, where $\hat{c}_t$ is the candidate for replacing memory cell, u_t is the update gate, f_t is the forget gate, o_t is the output gate, σ is the sigmoid function and $*$ denotes element wise multiplication. w_u, w_f, and w_o represent the update gate, forget gate, and output gate weights, respectively.

$$\hat{c}_t = tanh(w_c[a_{t-1}, x_t] + b_c) \tag{2.8}$$

$$u_t = \sigma(w_u[a_{t-1}, x_t] + b_u) \tag{2.9}$$

$$f_t = \sigma\big(w_f[a_{t-1}, x_t] + b_f\big) \tag{2.10}$$

$$o_t = \sigma(w_o[a_{t-1}, x_t] + b_o) \tag{2.11}$$

$$c_t = u_t * \hat{c}_t + f_t * c_{t-1} \tag{2.12}$$

$$a_t = o_t * tanh(c_t) \tag{2.13}$$

$$y_t = softmax\big(w_y a_t + b_y\big), \tag{2.14}$$

where b_u, b_f, and b_o represent the gates biases. All the weights can be expanded in the same way as in Equation 2.6, and $[a_{t-1}, x_t]$ is given by Equation 2.7. Equation 2.12 shows how the candidate for replacing a memory cell, along with the update and forget gates, are used to decide the current entry in the memory cell. An LSTM unit's pictorial representation is shown in Figure 2.2 [54].

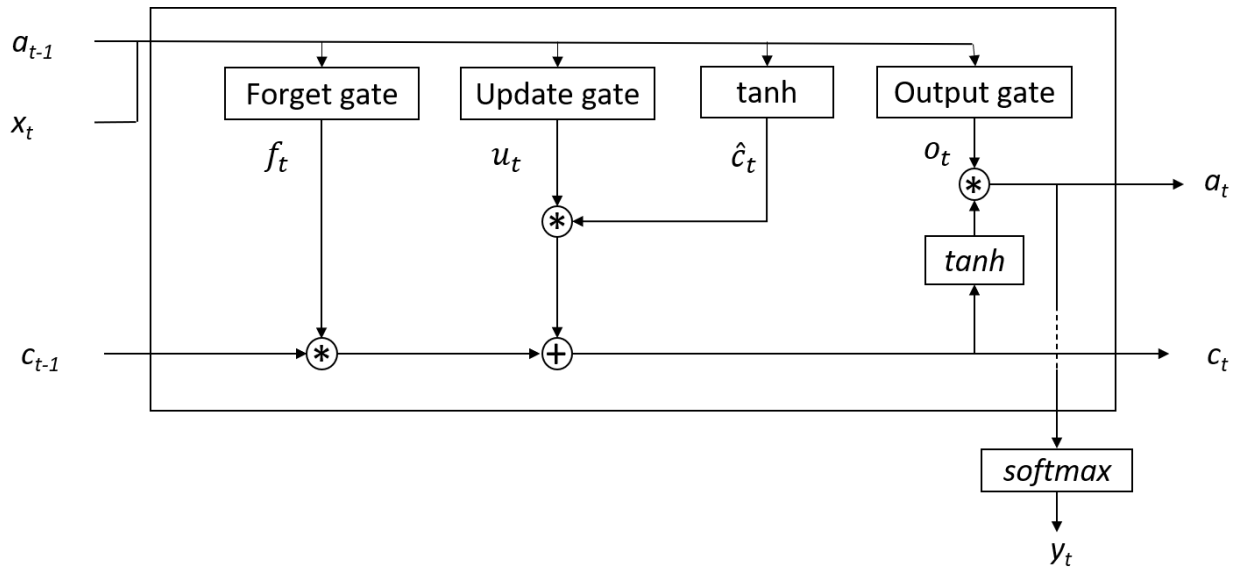

Figure 2.2: LSTM unit representation depicting how activation, memory state, and output at current time step are obtained from activation, input, and memory state at previous time step.

An LSTM network reported 83.38% (SD 10%) accuracy for FOG detection when trained with acceleration data acquired from 10 participants (8 froze) [37]. The 3D acceleration signals were acquired from participants' hip, knee, and ankle at 64 Hz frequency. The LSTM network was trained separately for each participant who froze, training on 70% the participant's data and testing on 30% of the data. The LSTM network used 100 units followed by 2 fully connected layers and was trained without feature extraction [37]. The research demonstrated that an LSTM network can be used for FOG detection. However, the model reported a 10% standard deviation across the 8 participants who froze [37]. Furthermore, accuracy is insufficient when validating on an imbalanced dataset. Since sensors are placed at multiple locations on participant's body, this approach may be inconvenient as a wearable system. Since there were only 8 participants who froze and the LSTM model was trained separately for each participant, the generalizability is limited [37].

2.3 Freezing of Gait Prediction

Since a FOG detection system provides a cue after the freeze has begun, falling risk due to freezing cannot be completely resolved. A preferred approach would be to predict the impending freeze and use a pre-emptive cue to help prevent the episode. Distinct FOG indicators can exist in gait parameters preceding a freeze episode [55]. Gait could deteriorate over a period of time as the person progresses into a freeze state (i.e., does not change abruptly). This period preceding a freeze is called pre-FOG [56] and has been predicted using IMU data [56], skin-conductance response (SCR) [5], and EEG data [57].

Data collected just before a freeze episode is considered as pre-FOG data (and labelled "Pre-FOG"). The Pre-FOG duration is determined empirically. Data collected before Pre-FOG is considered as non-FOG data (and labelled "Non-FOG"). A decision tree was used to classify Pre-FOG, FOG, and Non-FOG 3D shank acceleration data (1 s windows, 0.25 s overlap) from the Daphnet dataset [50]. After under sampling the Non-FOG data, a 10-fold cross validation was performed with all the data [56]. The Pre-FOG F1-measure (classification performance measure: harmonic mean of precision and recall) increased as the Pre-FOG duration increased from 1 s to 6 s [56]. On the other hand, the FOG and Non-FOG F1-measures decreased with increasing Pre-FOG duration. Pre-FOG durations of less than 4 s resulted in Pre-FOG F1-measures less than 0.4, while Pre-FOG durations greater than 4 s resulted in FOG F1-measures less than 0.65 [56]. These results suggest that Pre-FOG duration is a tradeoff between Pre-FOG and FOG classification performance.

A participant specific, anomaly-based model predicted 71.3% FOG episodes, 4.2 s in advance of a freeze [5]. Skin-conductance response (SCR) data were collected from 11 participants. Eight features including mean, median, standard deviation, minimum, maximum, difference between minimum and maximum, and number of local maxima and minima were calculated from SCR data. To predict FOG, Pre-FOG and FOG data were treated as anomalies and multivariate Gaussian distribution (MGD) was used for anomaly detection with a 3 s Pre-FOG length [5]. A FOG prediction model was built for each of the 11 participants. This method was adopted after participant dependent (i.e, a different algorithm threshold needed for each

participant) and statistically significant ($p < 0.01$) changes were found in features extracted from SCR just before FOG episodes, compared to normal walking [5].

FOG prediction using EEG data has shown promise [28,57]. A Wilcoxon Sum Rank test was used to select the best frequency and time domain features from EEG signals [57]. A multilayer perceptron neural network achieved 86.0% sensitivity and 74.4% specificity for FOG prediction using EEG data between 5 s and 1 s prior to FOG. In another study, EEG data was collected from 16 participants with FOG symptoms. EEG signals were filtered using a band-pass and band-stop filter after visually removing artifacts. A 3-layer backpropagation neural network using time-series analysis achieved 85.56% sensitivity and 86.60% specificity on 5 held-out participant's EEG data [28]. While EEG achieved good FOG prediction, the method required manual steps as well as preprocessing to remove noise [28]. Signals capturing different physiological or biomechanical parameters were used; however, further research is required to select a wearable sensor most suitable for FOG prediction.

A SVM model, based on classifying predicted features, achieved 93% (SD 4%) sensitivity and 87% (SD 7%) specificity for FOG prediction [58]. 3D acceleration data (64 Hz) from 10 PD participants' lower back, shank, and thigh were extracted from the Daphnet dataset [50]. Time and frequency-based features were calculated from each acceleration signal. Time series lag was calculated using autocorrelation. Future feature values were predicted using autoregressive models based on the time series lag [58]. A majority voting SVM fused 9 separately trained SVM classifiers to classify the predicted features as FOG or Non-FOG [58]. FOG was predicted within 1.72 s. The majority voting SVM had high computational cost, making the model unsuitable for a real-time system. The model gave a biased result since, in a participant dependent setting (i.e, prediction model for each participant), for each test window, there could be a training sample that overlapped the test window. More research is required before FOG prediction systems could reliably move from a Pre-FOG classification approach to a time series prediction-based approach.

Recently, deep learning has been used for FOG prediction. A 2-layer LSTM network was used (each layer with 50 neurons) for a 2-class model with Pre-FOG and FOG classes together in

the target class [26]. The network was trained with a 1000 batch size while using Adam optimizer with the cross-entropy loss function. The model used 50% data from all 10 participants for training and the remaining 50% data for testing [26]. The LSTM network achieved 87.54% accuracy for 1 s Pre-FOG duration, 85.54% accuracy for 3 s Pre-FOG duration, and 79.47% accuracy for 5 s Pre-FOG duration using acceleration signals [26]. A 2-layer LSTM network, when trained with Adam optimizer and cross-entropy loss function, can be useful for a FOG prediction model. Since the model used data from all participants in the training set, the model's application on a new participant is limited. Only accuracy was reported as a metric, which gives a biased analysis when learning from an imbalanced dataset.

2.4 Summary

Given the debilitating effects of FOG, such as reduced mobility, inactivity, and falls, the need for a FOG detection and prediction system (to be used with a device for intervention) is apparent. Previous research has provided a foundation for FOG detection and prediction; however, most research has focused on participant-dependent models (i.e., model was only validated on a participant's data, whose data was used for training the model) [5,26,36,37,58]. Since these models were not validated on an unseen participant's data, their applicability to a new participant is limited. Furthermore, participant dependent models gave a biased result due to a correlation between training and validation samples, which have been acknowledged, but not completely mitigated [36,58]. A few FOG detection [38,51] and FOG prediction [28] models have been validated on a set of held-out participant's data. However, the variation of the model's performance (i.e., standard deviation) across different held out participants was not reported. In deep learning models, a high batch size of 128 [52] or a high batch size of 1000 [26] was used, which led to poor generalization. A batch size of 1 has been recommended for the best generalization error [53]. There is a need for validating a FOG detection or prediction model with a leave-one-participant-out cross validation and a batch size of 1. This will allow the model to be evaluated on an unseen participant's data, and the performance variation across different participants can be known.

A FOG detection or FOG prediction model should aim to be wearable compliant. However, existing methods have used multiple sensors across different body locations making the sensors difficult to don and keep in place. Using plantar pressure insoles, which fit into participant's shoes, may be less obtrusive and more wearable compliant. A FOG detection or prediction model should require minimum computer memory and minimum preprocessing (including signal filtering) without any manual steps so that the model can be deployed for a real time application (such as in a microcontroller with limited computational resources). However, most research used noise filtering [28,40,57], manual steps [40], preprocessing such as combining windows in the spectral domain [38,51], or computationally intensive feature extraction [39,49] before FOG detection or prediction. A FOG detection model should aim to reduce freeze detection latency without compromising the freeze detection accuracy. Decreasing window length was found to decrease freeze detection latency at the cost of decreased freeze detection accuracy [36]. However, only a few studies have focused on finding and reducing freeze detection latency [36]. There is a need for a FOG detection and prediction system that requires fewer computational resources and decreases freeze detection latency without compromising accuracy.

Plantar pressure data detailed in Section 2.2 have been used for rehabilitation [42-44] and fall-risk analysis [23]. LSTM networks detailed in Section 2.2.4 have been used for FOG detection [37] and FOG prediction [26]. Using LSTM networks with plantar pressure data, the following chapters provide an analysis to assess the hypothesis that plantar pressure data can be used for FOG detection and prediction.

Chapter 3
Data Collection

3.1 Research Overview

The research described in this thesis aims to detect and predict FOG using plantar pressure data. Plantar pressure data from PD participants who show FOG symptoms are needed to train and validate a FOG detection or prediction model. Since no publicly available plantar pressure data from PD participants exist, the data needed to be collected. To collect data, PD participants were recruited and were asked to walk a Pre-defined freeze-provoking path while wearing plantar-pressure insole sensors. Data were labelled and a dataset was generated after data collection.

In preliminary modelling and testing in this research, classification algorithms performed poorly using raw plantar pressure data. This research was therefore continued with features extracted from plantar pressure data. The features were classified using an LSTM network for FOG detection and prediction. LSTM networks were trained on all but one participant's data who froze during the test protocol and were validated on the held-out participant's data. Hyperparameter tuning was done to find the best LSTM network architecture and training parameters for FOG detection. The best network was chosen based on the model's sensitivity and specificity results over one-held-out participant's data. The best LSTM network's performance for FOG detection was further validated with one-freezer-held-out cross validation. The model was also validated on held-out participants who did not freeze during the test protocol.

Since a FOG detection model should aim to detect freeze episodes with minimum latency, freeze detection latencies of the FOG detection model were calculated. FOG detection model performance was also investigated for different active (walking and turning) and inactive states (standing) during testing. Following this investigation, the FOG detection model was evaluated only on the active states, which led to an improvement in model specificity. For FOG prediction, data preceding freeze events (Pre-FOG) were relabelled to be included in the target class with FOG data. A binary classification was then done with the LSTM models for FOG prediction.

3.2 Participants and Inclusion Criteria

A convenience sample of 11 male participants was recruited from the Ottawa-Outaouais community with mean age 72 years (SD 5.5), mean height 1.77 m (SD 0.04), mean weight 80 kg (SD 11.37), and a mean 12 of years (SD 3.77) since PD diagnosis. Eligibility criteria were PD with FOG at least once a week, able to walk 25 m unassisted (without a cane or walking aid) and not having a lower limb injury or other comorbidities that impaired their ability to walk. The participants must not have undergone deep brain stimulation therapy or have other conditions that impaired balance and walking. Participants were on their normal antiparkinsonian medication schedule and dosage. Ethics clearance was obtained at both the University of Ottawa and the University of Waterloo. All participants provided written informed consent to participate.

3.3 Plantar Pressure Measurement

Plantar pressure data were recorded during multiple walking trials at 100 Hz using FScan pressure sensing insoles (Tekscan, Boston, MA; Figure 3.1). FScan insoles are thin (< 1 mm) plastic film sheets with 3.9 pressure sensing cells per cm^2 (25 cells per in^2). The insoles were equilibrated before participant arrival by applying uniform pressure to the entire sensor and adjusting the sensor constants to produce a uniform output [59]. Equilibration was performed at 138 kPa, 276 kPa, and 414 kPa.

Figure 3.1: F-Scan system: A plantar pressure insole sensor (left) and sensors worn in shoes (right).

3.4 Clinical Assessment

After providing informed consent, participants completed an information form that included disease history and medication schedule, New Freezing of Gait questionnaire, self-reported fall history questionnaire, and the Motor Examination section from the Unified Parkinson's Disease Rating Scale (UPDRS III).

3.5 Test Protocol

Data collection was performed in the Movement Performance Laboratory at the University of Ottawa. Participants were weighed and then the plantar pressure sensors were trimmed and fitted into their shoes. The plantar pressure sensors were connected to a computer and data streaming was verified. F-Scan step calibration was performed (i.e., standing on one leg and then quickly transferring weight to the other foot, to calibrate sensors to body weight). Participants were asked to walk a Pre-defined 25 m path that involved navigating multiple cones (requiring two 90° and two 180° turns); walking as far into a narrow, dead-end hallway as possible (2.1 m tall, 1.2 m wide, 2.4 m long), turning 180°, and walking back to the starting position (Figure 3.2). Participants were asked to stop once (voluntarily) while walking back to the starting position. Participants would also stop at the end of each test trial.

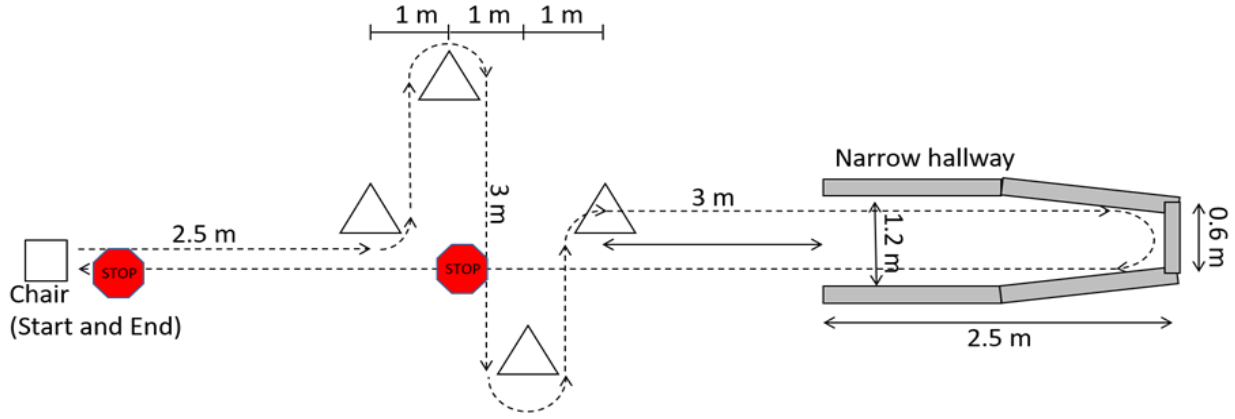

Figure 3.2: Walking Path. Participants started from a chair, walked straight, navigating multiple cones, walked as far into a narrow hallway as possible, turned 180°, and walked back to the starting position. Participants stopped (voluntarily) once while walking back to the starting position. Participants also stopped at the end of each test protocol.

Walking trials were recorded using a smartphone camera so that the trials could be analyzed after collection and freezing instances could be labelled. The camera closely followed the participant throughout the walking trial. For each walking trial, participants stood up from a sitting positing and did a foot stomp before starting to walk. The stomp was later used to synchronize plantar pressure data and the video. Participants completed up to 30 trials. The first five trials were baseline trials in which the participants completed the walking path without any additional tasks. After five baseline trials, additional tasks were added if the participant did not freeze (Figure 3.3). These included verbal (continuously speaking as many words as possible beginning with a specific letter) and motor tasks (holding a plastic meal tray with both hands, with objects on the tray). These tasks were performed individually or simultaneously to obtain more freezing episodes. Different difficulty levels were used when performing the motor task; for example, starting with three small wooden blocks on the tray and adding additional blocks as needed, to increase difficulty. Alternatively, the blocks were replaced with an empty paper coffee cup, a sealed water bottle, or the participant was asked to carry the tray with only one hand.

Figure 3.3: PD participant turning in a narrow hallway while holding a tray with a cup. Assistant follows for safety.

3.6 Results

A total of 241 minutes of walking data were collected, during which seven participants froze (Table 3.2). The data included 362 freeze episodes, with most freeze episodes corresponding to

Participant 7. A sample plantar pressure data frame from both feet of Participant 3 is shown in Figure 3.4. The blue regions indicate low pressure, dark blue regions indicate zero pressure, and yellow regions indicate higher pressure.

Table 3.1: Data collection summary with freeze episodes, with mean and total duration for each participant.

Participant	Most affected side	Number of FOG episodes	Mean (SD) FOG duration (s)	Total FOG duration (s)
1	Right	49	0.69 (0.26)	34.05
2	Left	35	2.64 (1.61)	92.35
3	Left	14	1.06 (0.53)	14.88
4	Left	0	-	-
5	Right	0	-	-
6	Left	10	4.23 (3.80)	42.29
7	Right	221	1.52 (1.48)	336.2
8	Right	24	1.51 (1.05)	36.16
9	Left	9	0.75 (0.35)	6.74
10	Left	0	-	-
11	Right	0	-	-

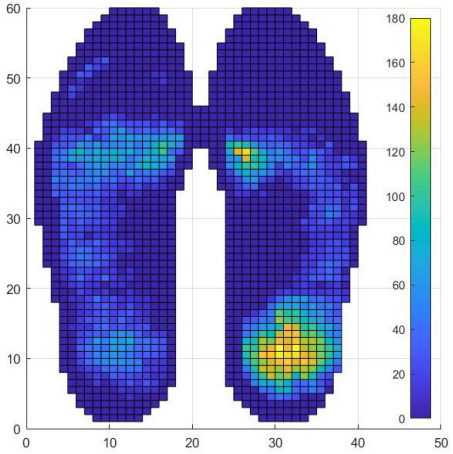

Figure 3.4: Plantar pressure array sample frame (kPa). Dark blue indicates zero pressure.

Chapter 4
Freezing of Gait Detection

4.1 Data preprocessing and labeling

Plantar pressure data were labelled using a custom-made GUI application in MATLAB R2019b (MathWorks, MA, USA). Trial video and plantar pressure frames were loaded and displayed in the GUI. Video and plantar pressure were synchronized using the right foot stomp of the participant at the start of each trial. The plantar pressure trial data start was set at the end of the spike in pressure. The video and plantar pressure data were viewed together, and each FOG episode start and end were marked in the video when detected. The program would then label all timestamps between the marked points as FOG and all other timestamps as Non-FOG.

Plantar pressure data from a foot at each instant was a 60x21 matrix of pressure values. Plantar pressure data collected at each time instant were stacked together as a time series. Data from each trial by each participant were kept as a separate time series, with a label for each timestamp. This preserved the time series information in the data which would be lost if data from any two trials were concatenated or mixed. Plantar pressure data from both feet were kept separate.

4.2 Experiment 1: FOG Detection with Raw Data

4.2.1 Methods

A 2D CNN network can classify a 60x21 matrix of pressure values (from each insole, at each time instance) into freeze events. However, a 2D CNN does not consider the data series and thus would fail to capture relevant time series information. Changing patterns over time is important in walking; thus, an LSTM network that captures time series relevant patterns could be better. A CNN could still be used to extract relevant features from each matrix of pressure values, before feeding these features to an LSTM network for classification.

To determine a 2D CNN's feature extraction capability with the plantar pressure dataset, a 2D CNN model was trained with data from five PD participants to classify FOG, Non-FOG, or Pre-FOG. Data acquired 2 s before all freeze episodes were relabelled as Pre-FOG and plantar

pressure data from both feet at each time instant were concatenated into a single 60x42 array. Since time series information was not needed for a CNN model, data from all trials with all participants were grouped together. The majority classes were undersampled, such that data in all three classes were equal in number. Data from all five participants were used for a 5-fold cross validation.

The 2D CNN model had two convolutional layers, the first layer with 8, 3x3 filters and the second with 16, 3x3 filters (Figure 4.1, graphic inspiration from [60]). The filter stride (pixels shifted by the filter over the input data) was set to 1 in both convolutional layers. A batch normalization layer and a ReLU nonlinearity unit followed each convolutional layer. Each convolutional layer was followed by a max-pooling layer with 2x2 filter size and a stride of 1. After the second pooling layer, two fully connected layers were used, the first with 10 neurons and the second with 3 neurons. A Softmax unit provided the final classifier output. A piecewise learning-rate schedule was used with a 0.02 initial learning rate and a 50% decrease every 10 epochs. The model was trained with Adam optimizer and 128 batch size for 45 epochs.

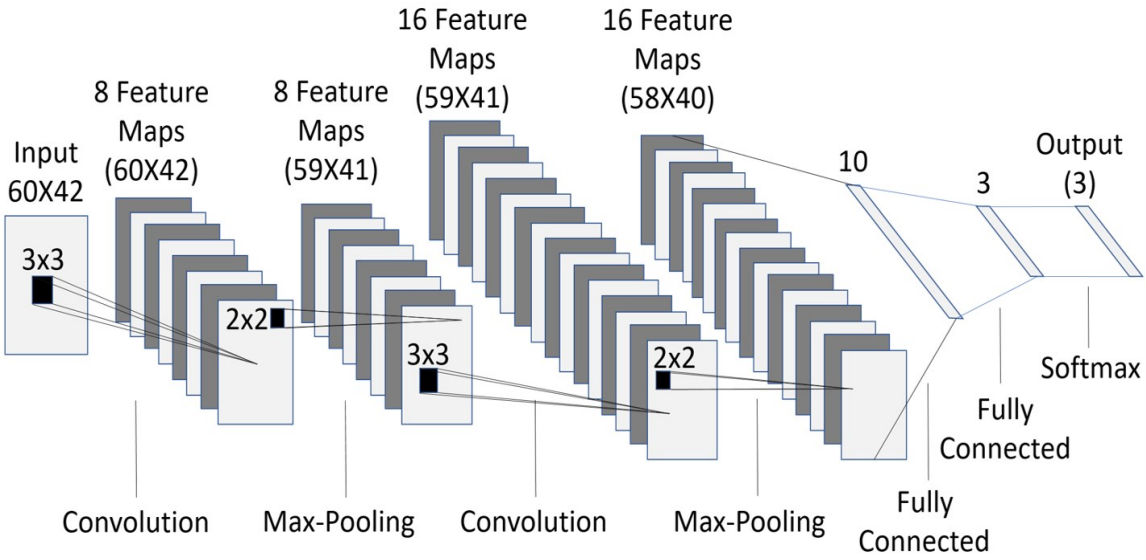

Figure 4.1: 2D CNN classifier architecture.

4.2.2 Results

In the three-class classification for a 2 s Pre-FOG duration, the 2D CNN model achieved 86.9% sensitivity and 95% specificity for Non-FOG class. For the FOG class, the model achieved

95.5% sensitivity and 96.5% specificity, while for the Pre-FOG class, the model achieved 86.9% sensitivity and 93.2% specificity. These results were obtained after balancing both training and validation sets. However, balancing the validation dataset led to a biased result. The validation dataset should represent a real-world test dataset, which is imbalanced, and thus should reflect the original data imbalance.

When the imbalanced dataset from all 11 participants was used to train and validate the model (5-fold cross validation), the model achieved 6.73% sensitivity for the Pre-FOG class, 43.68% specificity for the Non-FOG class, and 65.75% sensitivity for the FOG class. These results showed that the CNN model could not be used for FOG detection or prediction when the dataset is imbalanced. Apart from data imbalance, training the CNN model with raw plantar pressure data can lead to overfitting, increased noise sensitivity, and poor generalization. Since the raw data had 2520 (60x42) features, a CNN model would require a bigger dataset. Extracting handcrafted features from the plantar pressure data could reduce the number of features, thereby decreasing the population sample size needed for a network to learn patterns from the dataset. An approach based on handcrafted features was therefore tested, as described in Section 4.3.

4.3 Experiment 2: FOG Detection with Features

4.3.1 Feature Extraction

A feature is a data attribute or characteristic. A model trained with extracted features tends to be simpler and thus less prone to overfitting. A simple model generalizes better on unseen samples and is less sensitive to noise and outliers. A set of 16 features were extracted from plantar pressure data as follows:

- Centre of pressure coordinates (mm): Centre of pressure (COP) coordinates were calculated as weighted average positions, where the force magnitude at each pressure-sensor cell was multiplied by the distance of the sensor cell from the origin, the resulting products for all cells summed, and the sum was divided by the total ground reaction force. The COP coordinates were originally in cell numbers (a plantar pressure insole sensor has a 60x21 cell grid, each 5.08 mm apart) and then converted to distance from the origin by multiplying the

cell position (in x or y direction) with the distance between two cells (5.08 mm). COP coordinates in x (lateral) and y (medial) directions for both feet were calculated at each timestamp. The COP for a foot was set to zero to remove noise, if the total GRF from that foot was less than 5% of the total GRF for both feet.

- Centre of pressure velocity (cm/s): COP velocity was calculated by dividing the COP coordinates difference between two consecutive samples by the time difference between the samples. Since the plantar pressure data were collected at 100 Hz, the time difference between two consecutive samples is 0.01 s. COP velocity was calculated in both x and y axes for both feet.
- Centre of pressure acceleration (cm/s^2): COP acceleration was calculated by dividing the COP velocities difference between two consecutive samples by the time difference between the samples. COP acceleration was calculated in both x and y axes for both feet.
- Total ground reaction force (kPa): Total Ground Reaction Force (GRF) was obtained by adding pressure from all pressure cells in the plantar pressure sensor. Total GRF was determined for both feet at each timestamp.
- Fraction of total ground reaction force (unitless): Fraction of total GRF is the ratio of GRF from one foot divided by the total GRF for both feet. Fraction of total GRF was calculated for both feet at each timestamp.

These features were used in all feature-based experiments. For FOG detection with extracted features, data from each trial by each participant were kept as a separate time series and the plantar pressure data from both feet were kept separate. Data were split into training and validation set by participant, i.e., all the data from a participant were either in a training or validation set for an experiment.

4.3.2 LSTM Model

FOG detection was defined as a binary classification task, where samples were classified as FOG or Non-FOG. For all experiments in this thesis, LSTM networks were setup using a multiple-input multiple-output architecture. Each LSTM layer returned the full sequence to the model's

next layer. This allowed the model to classify each timestamp. LSTM layers used a hyperbolic tangent (tanh) activation function, followed by a time-distributed fully-connected layer (i.e., output at each time step passes through the fully connected layer) with 2 units and Softmax activation. Models were trained with Adam optimizer, using 0.9 decay rate for first and 0.999 decay rate for the second moment estimates and a cross entropy loss function.

Most deep learning frameworks (e.g., Tensorflow) require that all sequences in the same batch have the same length for vectorization. Vectorization uses network weights and inputs as vectors, allowing vector multiplication rather than repetitive element-wise multiplications. Thus, deep learning frameworks are built to use vectorization, allowing them to speed up the network training process. If two samples in a batch vary in length, then the computations needed for both samples are different (based on the sample length) and thus both cannot be computed simultaneously. Sequences of different lengths can be handled by a 1-batch size. All LSTM models were evaluated using the model's specificity and sensitivity on the validation set.

4.3.3 Experiment 2a: LSTM Without Data Balancing

A 2-layer LSTM model with 16 units in both layers was used for FOG detection. Data from all but one participant who froze during the trials formed the training set. Data from the held-out participant formed the validation set. The 2-layer LSTM model was trained with 0.01 constant learning rate for 30 epochs. The model failed to learn and classified every timestamp as Non-FOG. This was due to dataset imbalance, with most samples belonging to the Non-FOG class.

4.3.4 Experiment 2b: LSTM with Balanced Training Set

One way for a model to learn from imbalanced data is to randomly reduce the majority class samples in the training set while leaving the validation set untouched.

4.3.4.1 Data Balancing Approach 1

In a first data-balancing approach, the training set was built from:

- Only FOG labelled data from all trials with a freeze.

- Trials without a freeze were randomly chosen and data were added to the training set until the number of FOG instances was equal to or greater than the number of Non-FOG instances.

A 2-layer LSTM network, with 16 units, followed by a time-distributed fully-connected layer with 2 units and Softmax activation was trained for 30 epochs with 0.01 constant learning rate. This experiment was completed with data from all 11 participants. The one held-out validation results had very high sensitivity and 0 specificity (i.e., model was classifying most data points as belonging to FOG class).

4.3.4.2 Data Balancing Approach 2

In a second data-balancing approach, the model was again trained on FOG data extracted from trials with a freeze. The model was also trained on complete trials with no freeze. However, this time, the number of trials in the training set with no FOG was equal to the number of freeze episodes. The one held-out validation results were 0 sensitivity and very high specificity (i.e., model classified most data points as the Non-FOG class).

In both data-balancing methods above, the model was trained with instance labels in groups (i.e., all FOG labels in a freeze then all Non-FOG labels in a non-freeze period). Hence, the model classified data points together as a group, failing to properly classify transitions between Non-FOG and FOG. The network was not trained on any time series where a transition between both classes occur. A model should be able to learn that:

- both FOG and Non-FOG class data points occur in a group.
- generally, the FOG data-points group should occur between Non-FOG data-points groups. Thus, the model should learn the transition from Non-FOG to FOG data points and the transition from FOG to Non-FOG data points.

4.3.4.3 Data Balancing Approach 3

Based on the insights from the above three methods to balance the data, a different method of balancing the dataset was tried. At first, the length of each freeze episode was calculated. If the number of Non-FOG data points both before and after the freeze episode were at least half the

number of data points in the freeze episode, then Non-FOG data points equal to half the number of FOG data points both before and after the freeze episode, along with the freeze episode were extracted as a Non-FOG, FOG, Non-FOG (data points with labels in that order) time series from the trial. This series of extracted data points formed a training sample of the model. Such training samples were balanced.

If for a freeze episode, the number of data points before or after the freeze episode were less than half the number of data points in the freeze episode, then more data points were taken from one side (before or after, as available) to compensate for the less availability of data points on the other side and make the extracted time series balanced. However, in a few cases, there were not enough Non-FOG data points before and after a FOG event (this happened when multiple FOG episodes occurred close to each other). In such cases, the FOG episode along with the available Non-FOG data between freezes was extracted. Such time series (and thus the training sample and training batch) was slightly imbalanced, with more FOG data points than Non-FOG. This data organization performed well on one-held-out-freezer participant's data and was used for all experiments on FOG detection.

4.3.5 Experiment 2c: Determination of the Best LSTM Model

The best performing data organization from Experiment 2b was used for training with several network architectures and learning rate combinations while using Adam optimizer, cross entropy loss function, and a batch size of 1 (Table 4.1). All the models had a time-distributed fully connected layer with 2 neurons and a Softmax activation after the LSTM layers.

Table 4.1: LSTM Network configurations tried: Number of LSTM layers, number of units in each LSTM layer, different learning rate schedule were tried to find the best network configuration.

Hyperparameter	Values tried
Number of LSTM layers	1, 2, 3, 4, 5
Number of units in each LSTM layer	16, 32, 64
Constant learning rate	0.1, 0.01, 0.001, 0.0001
Learning rate decay with a decay rate (decay rate, initial learning rate)	(0.5, 0.005), (0.75, 0.001)
Learning rate decreases in discrete steps (initial learning rate)	Decreases to half every 5 epochs (0.01)

The 2-layer LSTM network was trained with 16, 32, and 64 units in both layers for 30 epochs with a 0.01 constant learning rate. Using 16 units in each layer worked best. The 2-layer LSTM network's performance did not improve beyond 30 epochs. Thus, in all subsequent experiments, the network was trained only until 30 epochs. With 16 units in each layer, networks with 1, 2, 3, 4, and 5 LSTM layers were trained. Networks with 1, 4, or 5 LSTM layers performed poorly. Thus, only networks with 2 LSTM layers (each with 16 units) and 3 LSTM layers (each with 32 units) were used for subsequent experiments.

Different learning rates were explored for the 2- and 3-layer LSTM network; 0.1, 0.01, 0.001, and 0.0001 constant learning rates were used. Learning rate decay with 0.5 decay rate (0.005 initial learning rate) and 0.75 decay rate (0.001 initial learning rate) were tried. All the different learning rate schedules were outperformed by the learning rate schedule for which the learning rate was reduced to half every 5 epochs after starting from a 0.01 initial learning rate. The best performing models were used for FOG detection (Table 4.2).

Table 4.2: Best performing LSTM network configuration for FOG detection.

Network or training parameter	Values / Options
LSTM layers (units in each LSTM layer)	2 layers (16 units) and 3 layers (32 units)
Initial learning rate	0.01
Learning rate decay	Decreases to half, every 5 epochs
Optimizer	Adam optimizer
Loss function	Cross entropy loss function
Batch size	1
Training epochs	30

4.3.6 Cross Validation

The aim of this research was to develop FOG detection and prediction models that could be applied on a new participant (a participant whose data the model has not been trained on). With k-fold cross validation, where FOG and non-FOG data would be pooled from all participants and then separated into k folds, a participant's FOG and non-FOG data would be used in both training and validation sets. Performance of the model on a new participant could thus not be determined. An alternative k-fold validation by participant would have too little freeze data in the training set, since only 7 participants froze. Therefore, participant held out cross validation was performed instead of k-fold cross validation.

Models were trained and evaluated with cross validation as follows:

- One-freezer-held-out-cross-validation: The model was trained on data from all participants who froze during trials except one participant (who froze), on whose data the model was validated. This was repeated for each freezer, such that each freezer was in the validation set once (7 folds).
- All-non-freezer-held-out validation: The model was trained on data from all participants who froze and was validated on all participants who did not freeze during the trials. This facilitated false positive assessment in situations where a participant does not freeze.

Each feature was normalized using z-score normalization by subtracting the feature value by the feature's mean and dividing by the feature's standard deviation. The mean and standard deviation calculated on the training set was used for normalizing both the training and the validation sets. The process was repeated for each cross-validation test, where a new mean and standard deviation were calculated from the cross-validation's training set. Z-score normalization is useful for removing outliers and bringing all features to a similar scale. One-freezer-held-out-cross-validation and all-non-freezer-held-out validation were performed with the 2-layer and 3-layer LSTM models after z-score normalization.

4.4 Results

The 2-layer LSTM model achieved 82.06% (SD 6.25%) sensitivity and 89.46% (SD 3.60%) specificity in one-freezer-held-out cross validation (Table 4.3). The model achieved 81.65% specificity in all-non-freezer-held-out validation (Table 4.4). The 3-layer LSTM model achieved a slightly improved 83.44% (SD 6.65%) sensitivity but a slightly lower 87.36% (SD 5.42%) specificity than the 2-layer LSTM model in one-freezer-held-out cross validation (Table 4.5). In all-non-freezer-held-out validation, using the 3-layer model improved the specificity to 87.70% (Table 4.6).

Table 4.3: FOG Detection: One-freezer-held-out cross validation for the 2-layer LSTM model. Each LSTM layer had 16 units. LSTM layers were followed by a time-distributed fully-connected layer and a Softmax activation.

Participant Held Out	Sensitivity	Specificity
Participant 1	83.03%	92.50%
Participant 2	77.17%	90.25%
Participant 3	72.50%	92.76%
Participant 6	85.90%	90.24%
Participant 7	77.42%	89.53%
Participant 8	86.18%	81.15%
Participant 9	92.24%	89.76%
Average	82.06% ± 6.25%	89.46% ± 3.60%

Table 4.4: FOG Detection: All-non-freezer-held-out validation for the 2-layer LSTM model. Each LSTM layer had 16 units. LSTM layers were followed by a time-distributed fully-connected layer and a Softmax activation.

Participant Held Out	Sensitivity	Specificity
Participant 4,5,10,11	-	81.65%

Table 4.5: FOG Detection: One-freezer-held-out cross validation for the 3-layer LSTM model. Each LSTM layer had 32 units. LSTM layers were followed by a time-distributed fully-connected layer and a Softmax activation.

Participant Held Out	Sensitivity	Specificity
Participant 1	83.64%	86.26%
Participant 2	83.45%	90.08%
Participant 3	71.70%	91.98%
Participant 6	86.93%	90.71%
Participant 7	77.13%	89.17%
Participant 8	87.63%	74.72%
Participant 9	93.56%	88.57%
Average	83.44% ± 6.65%	87.36% ± 5.42%

Table 4.6: FOG Detection: All-non-freezer-held-out validation for the 3-layer LSTM model. Each LSTM layer had 32 units. LSTM layers were followed by a time-distributed fully-connected layer and a Softmax activation.

Participant Held Out	Sensitivity	Specificity
Participant 4,5,10,11	-	87.70%

4.5 FOG Detection Latency

FOG detection latency is the time difference between freeze onset and the time when the model detects the freeze. A freeze episode was detected correctly if the model classified a freeze during the true freeze period. If multiple freeze classifications exist within the true freeze episode, the earliest classification was used to calculate freeze detection latency.

With a 2-layer LSTM model (16 units) in one-freezer-held-out cross validation, 95% of freeze episodes were detected correctly (Table 4.7). Seventeen freeze episodes from Participant 7 and one freeze episode from Participant 8 were not detected by the 2-layer LSTM model. The model achieved a maximum 0.1 s (SD 0.32 s) average freeze detection latency for Participant 7. However, standard deviations were larger, up to 0.85 s (for Participant 8). A negative freeze detection latency means that, on average, freeze episodes were detected before the freeze started, as desired.

Table 4.7: Average FOG detection latency and standard deviation in one-freezer-held-out cross validation with the 2-layer LSTM model. A negative freeze detection latency means that the freeze was detected before the true freeze onset, as desired.

Participant Held Out	Freezes - not detected	Freezes correctly detected	Average FOG detection latency (s)
Participant 1	0	49	-0.23 ± 0.55
Participant 2	0	35	0.02 ± 0.17
Participant 3	0	14	0.08 ± 0.25
Participant 6	0	9	-0.04 ± 0.36
Participant 7	17	204	0.10 ± 0.32
Participant 8	1	23	-0.55 ± 0.85
Participant 9	0	9	-0.47 ± 0.74
Total	**18**	**343**	**-**

4.6 False Positive and Not-detected Analysis

Freeze episodes not detected and false positive classifications by the 2-layer LSTM model, across active (walking and turning) and inactive (standing) states are shown in Figure 4.2. For the 2-layer LSTM model, 35.13% of false positives were during walking, 27.91% of false positives were during turning, 34.74% of false positives were during standing, and 2.22% of the data were not defined. Not defined refers to the beginning and the end of a trial when no specific activity was being performed. For the 2-layer LSTM model, 58.67% of misclassified freezes were during turning while 46.12% of misclassified freezes were during walking.

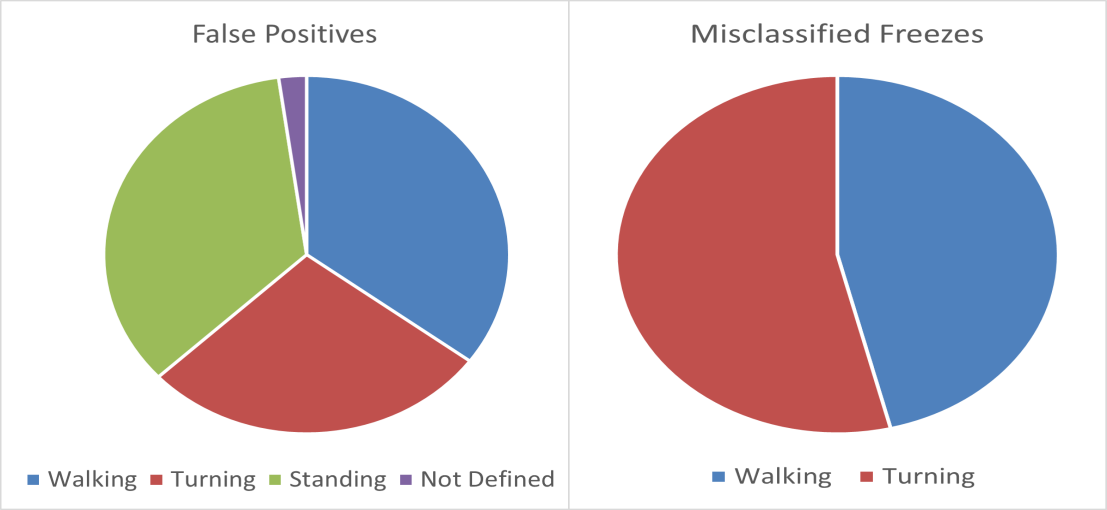

Figure 4.2: False positives and misclassified freezes distributed across active (walking, turning) and inactive (standing) states with the 2-layer LSTM model. Not defined refers to the beginning and end of a trial when a participant was not performing a specific activity.

Most standing states were misclassified as freeze by the model (Table 4.8). Almost no standing data were included in the final training set due to the data setup method (i.e., only freeze episodes, and data before and after the freeze episodes were included in the training set). Considering that standing state data could be classified using an activity recognition algorithm (not developed or implemented in this research) before FOG detection, the 2-layer LSTM model was also evaluated only on active states (turning and walking) and excluding standing. For the 2-

layer LSTM model, in one-freezer-held-out cross validation, the mean specificity increased by 3.8%, from 89.46 % (SD 3.6%) to 93.27% (SD 3.96%) when evaluating only on active states compared to classifying both active and inactive states, while the sensitivity decreased almost negligibly by 0.48% (Table 4.9). For all-non-freezer-held-out validation, the specificity increased from 81.65% to 88.39% (6.7% increase) (Table 4.10).

Table 4.8: Percentage of active and inactive states resulting in false positives with the 2-layer LSTM model.

Activity	Percentage of activity resulting in false positives
Standing	65.32%
Walking	3.72%
Turning	7.09%

Table 4.9: FOG Detection: One-freezer-held-out cross validation using the 2-layer LSTM model only on active states and on both active and inactive states. The active states include walking and turning during the trials and excludes standing.

Participant held out	Only active states		Both active and inactive states	
	Sensitivity	Specificity	Sensitivity	Specificity
1	79.24%	95.79%	83.03%	92.50%
2	77.16%	96.76%	77.17%	90.25%
3	72.50%	93.64%	72.50%	92.76%
6	85.89%	94.79%	85.90%	90.24%
7	77.87%	92.33%	77.42%	89.53%
8	86.18%	84.15%	86.18%	81.15%
9	92.24%	95.43%	92.24%	89.76%
Average	81.58% ± 6.26%	93.27% ± 3.96%	82.06% ± 6.25%	89.46% ± 3.60%

Table 4.10: FOG Detection: All-non-freezer-held-out validation using the 2-layer LSTM model only on active states and on both active and inactive states. The active states include walking and turning during the trials and excludes standing.

Participant held out	Only active states		Both active and inactive states	
	Sensitivity	Specificity	Sensitivity	Specificity
4,5,10,11	-	88.39%	-	81.65%

4.7 Discussion

The new method for FOG detection using plantar pressure data detected 95% of freeze episodes correctly (Table 4.7). The new method was the first attempt to use an LSTM network for FOG detection in a LOPO cross validation. The results could be considered reliable due to the use of participant independent LOPO cross validation; however, more participant data should be added to improve reliability. The 3-layer LSTM model achieved a consistent specificity for participants who did not freeze during the trials with the participant who froze. In the literature, LSTM networks for FOG detection were not validated on data from an unseen participant (i.e, a participant, whose data the network was not trained with) [37]. The 2-layer LSTM model showed an improvement in specificity when classifying only the actives states. The model can be viable since only 16 features from plantar pressure data have been used after z-score normalization without signal filtering. The 2-layer LSTM model needs 51 KB computer memory and can be stored on a microcontroller. Compared to SVM models [39,49] and a 1D CNN model [51], the new 2-layer LSTM model needs less computer memory (Table 4.11). The 2-layer LSTM model achieved a better freeze detection latency than a random forest with 4 s windows [36]. The random forest had achieved 1.08 s (SD 0.73 s) freeze detection latency compared to the new 2-layer LSTM model's 0.1 s (SD 0.3 s) maximum freeze detection latency.

Table 4.11: Computer memory requirement of different models in KB.

Model	Computer memory requirement (KB)
SVM [39]	1600
SVM [49]	1490
1D CNN [51]	145
New 2-layer LSTM	51

Compared to a SVM model [39] which achieved 84.49% sensitivity and 85.83% specificity on 15 participants in LOPO cross validation, the new 2-layer LSTM model in this research has slightly better specificity (89.46% (SD 3.6%)) but slightly worse sensitivity (82.06% (SD 6.25%)). The SVM model had used 28 features from 1.6 s windows of IMU data after low pass filtering and required 1.6 MB computer memory compared to the LSTM model's 16 features used with plantar pressure data without signal filtering. The SVM model's performance variations

across the 15 participants (as standard deviation) had not been reported [39]. On the other hand, the SVM model had a bigger sample size of 15 PD participants who froze, compared to the new LSTM model in this research, which had 7 PD participants who froze.

When classifying only the active states, the new 2-layer LSTM model achieved a better sensitivity and specificity (in both mean and standard deviation) than a 4-layer 1D CNN model [52]. The 4-layer 1D CNN model achieved 74.43% (SD 9.79 %) sensitivity and 90.59 % (SD 6.4%) specificity (only on active states) in LOPO cross validation on 8 participants. The 4-layer 1D CNN was trained with a 128-batch size and would have worse generalization compared to the new 2-layer LSTM model, which uses 1-batch size. With 1 batch size, however, the network will take more time to converge to the global optimum. A batch size of 1 can also perform poorly in many applications, and thus a larger batch size should be used in those applications. With the 2- and 3-layer LSTM models, however, 1-batch size has shown to perform well in FOG detection using the plantar pressure dataset. Compared to the use of plantar pressure insoles in this research, the 4-layer 1D CNN in the literature used acceleration data from participant's ankle, knee, and hip, which may be more obtrusive [52]. While wires and cuffs used with the plantar pressure insoles can also be obtrusive, more wireless technology is available.

A 4-layer 1D CNN model using 9-channel IMU data from 21 participants achieved better sensitivity (91.9%) and similar specificity (89.5%) on 4 held out participants [51]. However, the 1D CNN model required 145 KB computer memory. The 1D CNN model used 16 batch size leading to a slightly poorer generalization than the 2-layer LSTM model. The 1D CNN model had combined information from 2 consecutive 2.56 s windows in the frequency domain before classification, which would increase the computational cost [51]. The 4-layer 1D CNN's performance improved from previous work by the research group where a similar model was used with a smaller dataset (15 participants) [38]. This shows that deep learning models for FOG detection, including the new 2-layer LSTM model, can benefit from adding more participants to the dataset.

A k-means clustering algorithm [27] achieved a better sensitivity (92.4%) and specificity (94.9%) in a LOPO cross validation setting than the new LSTM model in this research. Entropy

was extracted from 1 s sliding windows (with 0.5 s overlap) of raw acceleration data from 10 PD participants. In a k-means clustering algorithm; however, outliers should be removed in advance. Further research on a larger dataset is required to validate the utility of unsupervised learning algorithms (i.e., learning algorithms which use unlabelled data) such as k-means clustering for FOG detection.

Chapter 5
Freezing of Gait Prediction

For FOG prediction, data collected just before a freeze episode was labelled as Pre-FOG. FOG is usually a 3-class classification problem, with the classes being Non-FOG, Pre-FOG, and FOG. FOG prediction has earlier been done as a binary classification [26], where data just before a freeze episode (Pre-FOG) and the freeze episodes were in the target class (i.e., the class which the model aims to detect) and the remaining Non-FOG data were in the non-target class. The same binary classification setup was used for FOG prediction in this research. The data length before a freeze (Pre-FOG duration) was chosen depending on the corresponding freeze episode's length. For short freeze episodes, the gait deterioration (from normal walking into a freeze) was assumed to be short compared to longer freeze episodes. Based on this assumption, for all freeze episodes that were 2 s or longer, a 2 s Pre-FOG duration was chosen. For all other freeze episodes shorter than 2 s, a Pre-FOG duration equal to the corresponding freeze episode length was chosen.

5.1 FOG prediction

Based on a freeze episode length, Non-FOG timestamps before FOG were relabelled to be included in the target class with FOG as Pre-FOG. For FOG detection with extracted features, data from each trial by each participant were kept as a separate time series and the plantar pressure data from both feet were kept separate. The 16 features extracted from plantar pressure data for FOG detection were used for FOG prediction. Data were split into training and validation set by participant, i.e., all the data from a participant were either in the training or validation set for an experiment. Data from all but one participant who froze during the trials formed the training set. Data from the held out participant formed the validation set.

For the model to learn from the imbalanced data, the majority class samples in the training set were undersampled, while leaving the validation set untouched. To balance the training dataset, the length of each target class episode (which includes both Pre-FOG and FOG) was calculated. If the number of Non-FOG data points both before and after the target class episode was at least half the number of data points in the target episode, then Non-FOG data points equal to half the number

of target class data points both before and after the freeze, along with the freeze episode are extracted from the trial. This series of extracted data points formed a training sample of the model. This training sample (and a batch) was balanced.

If the number of data points before or after the target class episode was less than half the number of data points in the target class, then more data points were taken from one side (before or after, as available) to compensate for the less availability of data points on the other side and to make the extracted time series balanced. However, in a few cases, there were not enough Non-FOG data points before and after a target class. In such cases, the target class episode along with the available Non-FOG data were extracted. Such time series (and thus the training samples) were slightly imbalanced, with more data points of the target class than of the non-target (Non-FOG) class.

The model architecture and training parameters that performed best for FOG detection were used for FOG prediction (Table 5.1). A 2-layer LSTM model with 16 units in each LSTM layer and a 3-layer LSTM model with 32 units in each LSTM layer were used for FOG prediction. The LSTM layers in both of these models were followed by a time-distributed fully connected layer with 2 units and Softmax activation. Both the 2-layer and the 3-layer LSTM models were trained with the Adam optimizer, a cross entropy loss function, and 1-batch size. A learning rate schedule was used with a 0.01 initial learning rate, which was decreased by half every 5 epochs.

One-freezer-held-out cross validations and all-non-freezers-held-out validation were performed. In one-freezer-held-out cross validations, the model was trained on data from all but one participant who froze and was validated on the held-out participant. The one-freezer-held-out cross validations were repeated for each participant who froze during the trials (7 folds). The all-non-freezers-held-out validation was performed by training the model on data from all participants who froze and validating on all participants who did not freeze during the trials.

Table 5.1: Network configuration used for FOG prediction. The 2-layer LSTM network had 16 units in each layer while the 3-layer LSTM network had 32 units in each layer. LSTM layers were followed by a time-distributed fully-connected layer and Softmax activation.

Hyperparameter	Value / Options
LSTM layers (units in each LSTM layer)	2 layers (16 units) and 3 layers (32 units)
Initial learning rate	0.01
Learning rate decay	Learning rate decreases by half every 5 epochs
Optimizer	Adam optimizer
Loss function	Cross entropy loss function
Batch size	1
Training epochs	30

5.2 Results

The best mean sensitivity and specificity were obtained after 4 training epochs for the 2-layer LSTM model and after 3 training epochs for the 3-layer LSTM model. The 2-layer LTSM model achieved 75.8% (SD 12.59%) average sensitivity and 76.9% (SD 7.15%) average specificity in one-freezer-held-out cross validation (Table 5.2). The same model achieved a higher 84.5% specificity for participants who did not freeze during the walking trials (i.e., Participants 4,5,10, 11) (Table 5.3). For the 3-layer LSTM model, average sensitivity decreased to 74% (SD 12.48 %) and specificity increased to 82.9% (SD 5.75%) (Table 5.4). However, the 3-layer model's specificity on participants who did not freeze decreased to 69.4% (Table 5.5).

Table 5.2: FOG Prediction: One-freezer-held-out cross validation with the 2-layer LSTM model after 4 training epochs.

Participant held out	Sensitivity	Specificity
Participant 1	66.96%	86.21%
Participant 2	97.24%	74.11%
Participant 3	90.16%	62.05%
Participant 6	69.13%	83.21%
Participant 7	61.83%	78.28%
Participant 8	65.50%	77.96%
Participant 9	80.31%	76.98%
Average	75.88% ± 12.59%	76.97% ± 7.15%

Table 5.3: FOG Prediction: All-non-freezer-held-out validation with the 2-layer LSTM model after 4 training epochs.

Participant held out	Sensitivity	Specificity
Participant 4,5,10,11	-	84.5%

Table 5.4: FOG Prediction: One-freezer-held-out cross validation with the 3-layer LSTM model after 3 training epochs**.**

Participant held out	Sensitivity	Specificity
Participant 1	78.83%	84.88%
Participant 2	94.04%	79.11%
Participant 3	82.20%	71.17%
Participant 6	71.51%	83.26%
Participant 7	51.54%	90.36%
Participant 8	64.68%	86.09%
Participant 9	75.33%	86.05%
Average	74.02% ± 12.48 %	82.99% ± 5.75%

Table 5.5: FOG Prediction: All-non-freezer-held-out validation with the 3-layer LSTM model after 3 training epochs.

Participant held out	Sensitivity	Specificity
Participant 4,5,10,11	-	69.4%

5.3 Discussion

The new 3-layer LSTM model for FOG prediction using features extracted from plantar pressure data achieved good specificity; however, improvement in sensitivity is desired. The new 3-layer LSTM model, with 74.02 % (SD 12.48%) mean sensitivity and 82.99% (SD 5.75%) mean specificity performed better than a multilayer perceptron neural network that did FOG prediction using EEG data [57]. The multilayer perceptron neural network achieved 73.19% mean sensitivity and 80.16 % mean specificity in classifying data between 5 s and 1 s before freeze onset for 5 held out participants using wavelet energy [57].

A backpropagation neural network using EEG signals from 16 participants and time series analysis predicted FOG with 85.56% sensitivity and 80.25% specificity on 5 held-out participants [28]. This result had better sensitivity but worse specificity than the new 3-layer LSTM model. Since EEG signal artifacts were removed with visual inspection and the signal needed to be filtered [28], the noise and complex pre-processing make the EEG approach challenging for use in a wearable system.

Most FOG prediction models in the literature have been set in a participant dependent setting (i.e., all participant's data used in both training and test set). For example, a 2-layer LSTM network was trained using all participant's 50% data and was tested on the remaining 50% [26]. While the model achieved 87.54% accuracy in FOG prediction with 1 s Pre-FOG duration and 85.54% accuracy with 2 s Pre-FOG duration [26], the results cannot be generalized on new participants because all participant's data were used both in the training and validation set. Furthermore, the 2-layer LSTM network in [26] will lead to poor generalization, since it was trained using a large 1000-batch size compared to the new 3-layer LSTM model's 1-batch size.

In another participant dependent setting (i.e, modeled and trained on the same participant's data), autoregressive (AR) models were used to predict feature values that were classified using a majority voting Support Vector Machine (SVM) [58]. A separate predictive model was built for each participant, achieving 93% (SD 4%) mean sensitivity and 87% (SD 7%) mean specificity. The majority voting SVM fused the results from 9 separately trained SVM classifiers, which made the model computationally expensive. Since the same participant's data were used both in the training and validation set, the results were biased due to the correlation between overlapping windows in the training and validation set. Compared to the new 3-layer LSTM model that used plantar pressure insoles in this research, the SVM model had used IMUs placed at participant's lower back, shank, and thigh, which would be more obtrusive [58]. Since a separate predictive model was built for each participant, the SVM model's validity could not be generalized to a new participant.

Chapter 6
Conclusion

Freezing of Gait is an incapacitating problem for people with advanced stage Parkinson's disease. While many research outcomes have been reported in the literature, FOG is still a problem. The research in this thesis was undertaken to determine if plantar pressure data could be effective for FOG detection. After data collection and preprocessing, features were extracted from plantar pressure data and were used to train LSTM models. For FOG detection, a 2-layer LSTM model achieved comparable results with existing literature. The approach of using plantar pressure insole sensors rather than IMU can be useful for FOG detection and prediction, since the entire sensor and electronics can be easily worn in footwear and would thus be unobtrusive and not affect movement.

6.1 Conclusions Related to Objectives

6.1.1 Determine if an LSTM network can detect FOG episodes with sensitivity > 92.4% (SD 1.2%) and specificity > 94.9% (SD 2.3%) [27] in a person-independent model using only plantar pressure data.

The new LSTM model for FOG detection using plantar pressure data achieved a lower 82.06% (SD 6.25%) sensitivity and 89.46% (SD 3.6%) specificity than the benchmark results that were obtained using the K-means clustering algorithm (an unsupervised learning algorithm) and IMU data [27]. When classifying only active states, the new LSTM model with plantar pressure data achieved a comparable 93.27% (SD 3.96%) specificity with the K-means clustering algorithm with IMU data. LSTM models have an advantage of not being highly susceptible to outliers, unlike the K-means clustering algorithm.

6.1.2 Determine if an LSTM network can predict FOG episodes with sensitivity > 85.86% and specificity > 80.25% [28] in a person-independent model using only plantar pressure data.

The new 3-layer LSTM model achieved a greater 82.99% (SD 5.75%) specificity but a lower 74.02% (SD 12.48%) sensitivity than the benchmark results. However, the results by the new

LSTM model in this research are more reliable since the results were obtained in a participant-held-out cross-validation setting, compared to the benchmark results, where the model was validated on 5 held out participants without a cross validation. The new LSTM models used plantar pressure data without any signal filtering, while the benchmark results were obtained used EEG signals which needed manual steps and preprocessing to remove noise. Furthermore, plantar pressure insoles are easier to use in a wearable system compared to EEG sensors.

6.1.3 Determine if an LSTM network can achieve the benchmarks in a) and b) using custom features extracted from only plantar pressure data.

The LSTM network using the 16 features extracted from plantar pressure data could not achieve the benchmark results for FOG detection. For FOG prediction, the LSTM network using the extracted features performed better than the benchmark in terms of specificity but could not achieve the sensitivity benchmarks. The new FOG detection and prediction models, however, used data from plantar pressure insoles which can be less obtrusive than EEG and IMU sensors in a wearable device. The current model of plantar pressure used wires and cuffs, which can also be obtrusive in a wearable system; however, a wireless system is available [61].

6.1.4 Determine if an LSTM network can achieve FOG detection and prediction, with similar specificity between participants who did not freeze during the walking trials and participants who froze during the trials.

For FOG detection, the 2-layer LSTM model achieved a lower specificity (81.65%) for participants who did not freeze compared to 89.46% (SD 3.6%) specificity for participants who froze. The 3-layer LSTM model achieved a better specificity (87.70%) for participants who did not freeze. For FOG prediction, the 3-layer LSTM model achieved a lower specificity (69.4%) for participants who did not freeze compared to 82.99% (SD 5.75%) specificity for participants who froze. The 2-layer LSTM model achieved a better specificity (84.5%) for participants who did not freeze for FOG prediction. The models that performed best for participants who froze performed slightly worse for participants who did not freeze during the trials, since a higher false positive rate occurred for participants who did not freeze than for those who froze. In a practical application

of the freeze detection or prediction system, it may be more important to achieve fewer false negatives to avoid missing a freeze and tolerate false positives as unnecessary cues.

6.1.5 Determine if only classifying FOG during active states (i.e., walking, moving, etc.) can reduce freeze episode misclassification and false positives by an LSTM network.

For FOG detection, most data during an inactive state (standing) were misclassified by the model. When classifying only the active states, the performance of the 2-layer LSTM model improved. The mean specificity increased by 3.8% while the sensitivity decreased by 0.48%. The specificity by the model had increased to 93.27% (SD 3.96%) which is comparable to the objective of achieving specificity greater than 94.9% (SD 2.3%). Thus, coupling the FOG detection system with an activity recognition system can decrease the false positives by the detection model, while the correctly detected FOG remains unaffected.

6.2 Future work

This research showed that features extracted from plantar pressure data can be reliably used for FOG detection with LSTM models. The thesis also showed that FOG detection models could move away from the window-based approach, which could save critical time in real-time implementation without compromising on freeze detection performance. To improve the new 2-layer LSTM model's performance, more features could be explored while including only the best features selected by a feature-selection technique. The training dataset used for the new LSTM models did not contain standing activity data, thus the standing data (inactive state) in the validation set were misclassified as FOG by the model. An energy threshold-based approach could be implemented to classify data corresponding to inactive states before FOG detection.

Model performance for FOG prediction should be improved before implementation, and a different approach may be needed for this task. Models based on time series prediction can be explored for FOG prediction. The network configurations that performed best for FOG detection were also used for FOG prediction. Hyperparameter optimizations should be done again for the FOG prediction task, which may lead to an optimal network architecture and training parameters for FOG prediction.

A personalized model could be made with transfer learning. The final few layer's weights in a participant independent deep learning model could be trained with the target participant's data. Apart from training the model with a larger dataset, future work could implement the LSTM model in a microcontroller for real-time FOG detection and prediction.

References

1. Jankovic J. Parkinson’s disease: clinical features and diagnosis. Journal of Neurology, Neurosurgery & Psychiatry. 2008 Apr 1;79(4):368-376.
2. Giladi N, Treves TA, Simon ES, Shabtai H, Orlov Y, Kandinov B, Paleacu D, Korczyn AD. Freezing of gait in patients with advanced Parkinson's disease. Journal of Neural Transmission. 2001 Jan 1;108(1):53-61.
3. Giladi N, McMahon D, Przedborski S, Flaster E, Guillory S, Kostic V, Fahn S. Motor blocks in Parkinson's disease. Neurology. 1992 Feb 1;42(2):333-339.
4. Giladi N, McDermott MP, Fahn S, Przedborski S, Jankovic J, Stern M, Tanner C, Parkinson Study Group. Freezing of gait in PD: prospective assessment in the DATATOP cohort. Neurology. 2001 Jun 26;56(12):1712-1721.
5. Mazilu S, Calatroni A, Gazit E, Mirelman A, Hausdorff JM, Tröster G. Prediction of freezing of gait in Parkinson's from physiological wearables: an exploratory study. IEEE Journal of Biomedical and Health Informatics. 2015 Aug 5;19(6):1843-1854.
6. Schaafsma JD, Balash Y, Gurevich T, Bartels AL, Hausdorff JM, Giladi N. Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson's disease. European Journal of Neurology. 2003 Jul;10(4):391-398.
7. Snijders AH, Nijkrake MJ, Bakker M, Munneke M, Wind C, Bloem BR. Clinimetrics of freezing of gait. Movement Disorders. 2008;23(S2):S468-474.
8. Backer JH. The symptom experience of patients with Parkinson's disease. Journal of Neuroscience Nursing. 2006 Feb 1;38(1):51-57.
9. Giladi N, Tal J, Azulay T, Rascol O, Brooks DJ, Melamed E, Oertel W, Poewe WH, Stocchi F, Tolosa E. Validation of the freezing of gait questionnaire in patients with Parkinson's disease. Movement Disorders. 2009 Apr 15;24(5):655-661.
10. Nieuwboer A, Weerdt WD, Dom R, Lesaffre E. A frequency and correlation analysis of motor deficits in Parkinson patients. Disability and Rehabilitation. 1998 Jan 1;20(4):142-150.

11. Nieuwboer A, Dom R, De Weerdt W, Desloovere K, Fieuws S, Broens-Kaucsik E. Abnormalities of the spatiotemporal characteristics of gait at the onset of freezing in Parkinson's disease. Movement Disorders. 2001 Nov;16(6):1066-1075.

12. Schrag A, Jahanshahi M, Quinn NP. What contributes to depression in Parkinson's disease? Psychological Medicine. 2001;31(1):65-73.

13. Bloem BR, Hausdorff JM, Visser JE, Giladi N. Falls and freezing of gait in Parkinson's disease: a review of two interconnected, episodic phenomena. Movement Disorders. 2004 Aug;19(8):871-884.

14. Wenning GK, Litvan I, Jankovic J, Granata R, Mangone CA, McKee A, Poewe W, Jellinger K, Chaudhuri KR, D'olhaberriague L, Pearce RK. Natural history and survival of 14 patients with corticobasal degeneration confirmed at postmortem examination. Journal of Neurology, Neurosurgery & Psychiatry. 1998 Feb 1;64(2):184-189.

15. De Boer AG, Wijker W, Speelman JD, De Haes JC. Quality of life in patients with Parkinson's disease: development of a questionnaire. Journal of Neurology, Neurosurgery & Psychiatry. 1996 Jul 1;61(1):70-74.

16. Martinez-Martin P. An introduction to the concept of "quality of life in Parkinson's disease". Journal of Neurology. 1998 Apr 1;245(1):S2-6.

17. Giladi N, Nieuwboer A. Understanding and treating freezing of gait in parkinsonism, proposed working definition, and setting the stage. Movement Disorders. 2008;23(SUPPL.2):423-425.

18. Nieuwboer A, Kwakkel G, Rochester L, Jones D, van Wegen E, Willems AM, Chavret F, Hetherington V, Baker K, Lim I. Cueing training in the home improves gait-related mobility in Parkinson's disease: the RESCUE trial. Journal of Neurology, Neurosurgery & Psychiatry. 2007 Feb 1;78(2):134-140.

19. Spildooren J, Vercruysse S, Meyns P, Vandenbossche J, Heremans E, Desloovere K, Vandenberghe W, Nieuwboer A. Turning and unilateral cueing in Parkinson's disease patients with and without freezing of gait. Neuroscience. 2012 Apr 5;207:298-306.

20. Ginis P, Heremans E, Ferrari A, Dockx K, Canning CG, Nieuwboer A. Prolonged walking with a wearable system providing intelligent auditory input in people with Parkinson's disease. Frontiers in Neurology. 2017 Apr 6;8:1-9.

21. Moreau C, Defebvre L, Bleuse S, Blatt JL, Duhamel A, Bloem BR, Destée A, Krystkowiak P. Externally provoked freezing of gait in open runways in advanced Parkinson's disease results from motor and mental collapse. Journal of neural transmission. 2008 Oct 1;115(10):1431-1436.

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

23. Howcroft J, Kofman J, Lemaire ED. Prospective fall-risk prediction models for older adults based on wearable sensors. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2017;25(10):1812-1820.

24. Howcroft J, Lemaire ED, Kofman J. Wearable-sensor-based classification models of faller status in older adults. PLoS One 2016;11(4):e0153240.

25. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation. 1997 Nov 15;9(8):1735-1780.

26. Torvi VG, Bhattacharya A, Chakraborty S. Deep domain adaptation to predict freezing of gait in patients with Parkinson's disease. 17th IEEE International Conference on Machine Learning and Applications (ICMLA) 2018:1001-1006.

27. Li B, Zhang Y, Tang L, Gao C, Gu D. Automatic detection system for freezing of gait in Parkinson's disease based on the clustering algorithm. 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) 2018:1640-1649.

28. Handojoseno AMA, Naik GR, Gilat M, Shine JM, Nguyen TN, Quynh TLY, et al. Prediction of freezing of gait in patients with Parkinson's disease using EEG signals. In Studies in Health Technology and Informatics. IOS Press; 2018: 124–131.

29. Bloem BR, Valkenburg VV, Slabbekoorn M, Willemsen MD. The Multiple Tasks Test: development and normal strategies. Gait & Posture. 2001;14(3):191-202.

30. Bloem BR, Valkenburg VV, Slabbekoorn M, van Dijk JG. The multiple tasks test. Strategies in Parkinson's disease. Experimental Brain Research. 2001;137(3-4):478-486.

31. Rubinstein TC, Giladi N, Hausdorff JM. The power of cueing to circumvent dopamine deficits: a review of physical therapy treatment of gait disturbances in Parkinson's disease. Movement Disorders. 2002;17(6):1148-1160.

32. Ginis P, Nackaerts E, Nieuwboer A, Heremans E. Cueing for people with Parkinson's disease with freezing of gait: a narrative review of the state-of-the-art and novel perspectives. Annals of Physical and Rehabilitation Medicine. 2018;61(6):407-413.

33. Nieuwboer A. Cueing for freezing of gait in patients with Parkinson's disease: a rehabilitation perspective. Movement Disorders. 2008;23(S2):S475-81.

34. Picelli A, Camin M, Tinazzi M, Vangelista A, Cosentino A, Fiaschi A, Smania N. Three-dimensional motion analysis of the effects of auditory cueing on gait pattern in patients with Parkinson's disease: a preliminary investigation. Neurological Sciences. 2010;31(4):423-430.

35. Barthel C, Nonnekes J, Van Helvert M, Haan R, Janssen A, Delval A, Weerdesteyn V, Debû B, Van Wezel R, Bloem BR, Ferraye MU. The laser shoes: a new ambulatory device to alleviate freezing of gait in Parkinson disease. Neurology. 2018 J;90(2):e164-171.

36. Mazilu S, Hardegger M, Zhu Z, Roggen D, Tröster G, Plotnik M, Hausdorff JM. Online detection of freezing of gait with smartphones and machine learning techniques. 6th International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health) and Workshops 2012: 123-130.

37. Ashour AS, El-Attar A, Dey N, Abd El-Kader H, Abd El-Naby MM. Long short-term memory based patient-dependent model for FOG detection in Parkinson's disease. Pattern Recognition Letters 2020;131:23-29.

38. Camps J, Samà A, Martín M, Rodríguez-Martín D, Pérez-López C, Alcaine S, Mestre B, Prats A, Crespo MC, Cabestany J, Bayés À. Deep learning for detecting freezing of gait episodes in

Parkinson’s disease based on accelerometers. In International Work-Conference on Artificial Neural Networks 2017:344-355.

39. Samà A, Rodríguez-Martín D, Pérez-López C, Català A, Alcaine S, Mestre B, Prats A, Crespo MC, Bayés À. Determining the optimal features in freezing of gait detection through a single waist accelerometer in home environments. Pattern Recognition Letters 2018;105:135-143.

40. Tripoliti EE, Tzallas AT, Tsipouras MG, Rigas G, Bougia P, Leontiou M, Konitsiotis S, Chondrogiorgi M, Tsouli S, Fotiadis DI. Automatic detection of freezing of gait events in patients with Parkinson's disease. Computer Methods and Programs in Biomedicine 2013;110(1):12-26.

41. Bächlin M, Hausdorff JM, Roggen D, Giladi N, Plotnik M, Tröster G. Online detection of freezing of gait in Parkinson's disease patients: a performance characterization. In Proceedings of the Fourth International Conference on Body Area Networks 2009;1-8.

42. Crosbie WJ, Nicol AC. Reciprocal aided gait in paraplegia. Spinal Cord 1990;28(6):353-363.

43. Neaga F, Moga D, Petreus D, Munteanu M, Stroia N. A wireless system for monitoring the progressive loading of lower limb in post-traumatic rehabilitation. In International Conference on Advancements of Medicine and Health Care through Technology 2011: 54-59.

44. Edgar SR, Swyka T, Fulk G, Sazonov ES. Wearable shoe-based device for rehabilitation of stroke patients. Annual International Conference of the IEEE Engineering in Medicine and Biology 2010:3772-3775.

45. Abdul Razak AH, Zayegh A, Begg RK, Wahab Y. Foot plantar pressure measurement system: A review. Sensors 2012;12(7):9884-9912.

46. Jeon HS, Han J, Yi WJ, Jeon B, Park KS. Classification of Parkinson gait and normal gait using spatial-temporal image of plantar pressure. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2008:4672-4675.

47. Pardoel S, Shalin G, Nantel J, Lemaire ED, Kofman J. Selection of Plantar-Pressure and Ankle-Acceleration Features for Freezing of Gait Detection in Parkinson's Disease using Minimum-

Redundancy Maximum-Relevance. 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society 2020: 4034-4037.

48. Shalin G, Pardoel S, Nantel J, Lemaire ED, Kofman J. Prediction of Freezing of Gait in Parkinson's Disease from Foot Plantar-Pressure Arrays using a Convolutional Neural Network. 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society 2020: 244-247.

49. Rodríguez-Martín D, Samà A, Pérez-López C, Català A, Moreno Arostegui JM, Cabestany J, Bayés À, Alcaine S, Mestre B, Prats A, Crespo MC. Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer. PloS One 2017;12(2):e0171764.

50. Bachlin M, Plotnik M, Roggen D, Maidan I, Hausdorff JM, Giladi N, Troster G. Wearable assistant for Parkinson's disease patients with the freezing of gait symptom. IEEE Transactions on Information Technology in Biomedicine 2009;14(2):436-446.

51. Camps J, Sama A, Martin M, Rodriguez-Martin D, Perez-Lopez C, Arostegui JM, Cabestany J, Catala A, Alcaine S, Mestre B, Prats A. 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 2018;139:119-131.

52. Xia Y, Zhang J, Ye Q, Cheng N, Lu Y, Zhang D. Evaluation of deep convolutional neural networks for detection of freezing of gait in Parkinson's disease patients. Biomedical Signal Processing and Control 2018;46:221-230.

53. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. Cambridge: MIT press; 2016 Nov 18.

54. Olah C. Understanding LSTM Networks. Available at: https://colah.github.io/posts/2015-08-Understanding-LSTMs/. Accessed 14/11, 2020.

55. Nieuwboer A, Dom R, De Weerdt W, Desloovere K, Janssens L, Stijn V. Electromyographic profiles of gait prior to onset of freezing episodes in patients with Parkinson's disease. Brain. 2004;127(7):1650-1660.

56. Mazilu S, Calatroni A, Gazit E, Roggen D, Hausdorff JM, Tröster G. Feature learning for detection and prediction of freezing of gait in Parkinson's disease. International Workshop on Machine Learning and Data Mining in Pattern Recognition 2013:144-158

57. Handojoseno AA, Shine JM, Nguyen TN, Tran Y, Lewis SJ, Nguyen HT. Analysis and prediction of the freezing of gait using EEG brain dynamics. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2014;23(5):887-896.

58. Arami A, Poulakakis-Daktylidis A, Tai YF, Burdet E. Prediction of gait freezing in Parkinsonian patients: a binary classification augmented with time series prediction. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2019;27(9):1909-1919.

59. Tekscan. F-Scan User Manual v.6.3x: Bipedial in-shoe pressure/force measurement system. Boston, MA: Tekscan, Inc.;2011.

60. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998;86(11):2278-324.

61. F-Scan64 | Tekscan. Available at: https://www.tekscan.com/products-solutions/systems/f-scan64. Accessed 4/11, 2021.