Development of a Bayesian Decision Tree Model for Task Classification and its Validation on Parkinson's Disease Patients

Melina Dekker

Development of a Bayesian Decision Tree Model for Task Classification and its Validation on Parkinson's Disease Patients

by Melina Dekker
Student number:
4304209

To obtain the degree of
Master of Science
in Biomedical Engineering

at Delft University of Technology
Faculty of Mechanical, Maritime and Materials Engineering

to be defended on 19 June 2020

Thesis Committee

Chair	Dr. ir. A.C. Schouten (Chair) TU Delft Biomechanical Engineering
First Supervisor	Dr. ir. W. Mugge (First supervisor) TU Delft Biomechanical Engineering
Daily Supervisor	Ir. Sander Minnoye (Daily supervisor) Cue2Walk
	Dr. ir. Daan M. Pool TU Delft Control & Simulation

Preface

This thesis project was a collaboration with Cue2Walk, a company based in the Hague. I have enjoyed (and sometimes struggled with) combining the research aspect required for the TU with the development aspect called for by Cue2Walk. On the one side, every decision I made needed to be justified in a good research article, while on the other side I mostly wanted to deliver a useful "end product". I'd like to think that my project ended up combining both of these aspects.

My graduation has for the most part been quite a smooth ride. Sometimes so much so that I wondered when something would go wrong. Then came along Covid-19. The time I put into figuring out my sensors, planning experiments, writing protocols, discussing with the Basalt Revalidatie about the location, applying a research ethics proposal- was suddenly fruitless. In the end, (as expected,) everything did not go according to my plan. I was fortunate in the way I could adjust my project proposal and get back on track.

In this report, I will introduce you to the background of this thesis project and the company Cue2Walk. The essential part of this report is the research article, which mainly concerns the results of the project. The report introduction provides more insight into the process of getting there. Of course, you are free to skip directly to the research article, where you will have the opportunity to read a condensed account of the most important components of my project, which is already pretty interesting on the whole! Enjoy.

Acknowledgements

A report such as this one, I say in all humility, is not written overnight. What's more, I doubt that this report would have been written at all by now, if it wasn't for the many people that I would like to thank.

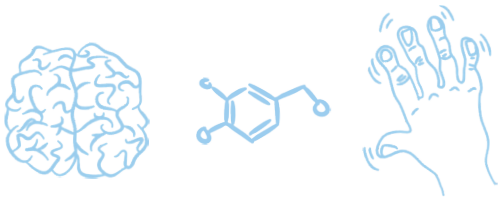
Starting with Cue2Walk. I met Martijn and Sander over a year ago, you could say by coincidence, because Cue2Walk was one of the few companies that replied to my e-mails regarding graduation projects. And as usual, in hindsight, I could not imagine having found a nicer place to graduate. I´m glad that I got to work on this project, and I'm curious to see what it will bring in the future. I also want to thank Jorrit at Basalt Revalidatie, for the experiments that could have been.

I want to thank my flatmates, who had to live with me during this period of thesis stress (kidding), and who are almost like family since the recent turn of events and the time we spent at home. Then there's my family! Thank you to infinity and back for being my family.

Thanks to the squad, of course to Hugo, Gailey and Johan, for over a thousand (I'm pretty sure) coffee and lunch breaks we took. Thanks to proofreaders Laura, Michaël, Jorn, Benjamin, Umit, Johan, Gailey and especially Lilja and Wouter. Thanks to Gaia, for teaching me about the world of academics. A shout out to Jorik for using his home office and for synchronizing his graduation schedule with mine, it has been fun.

Special thanks to my committee, Alfred and Winfred, for the meetings we had, and for your honesty. Thanks to Sander, worlds' best thesis supervisor, I've been able to learn a great deal from you.

Finally, thanks to the friends I met while studying in Delft, and many thanks to Life, for getting me where and who I am at the moment.



Introduction

Background

Medical records show that approximately 50.000 patients were diagnosed with Parkinson's Disease (PD) in the Netherlands in 2016. Of these patients, most patients were among the elderly population [1]. PD is a neurodegenerative disorder, a disorder of the central nervous system tampering with muscle control [2]. One of the main causes of PD is the loss of dopaminergic neurons in the brain. Dopamine acts as a messenger molecule and is essential for movement control and coordination [2]. Existing medication for PD patients is focussed on reducing the dopamine deficit in the brain [3].

One of the motor symptoms of PD is Freezing of Gait (FoG). Approximately half of the PD patients diagnosed over five years experience FoG, a short episodic inability to keep moving forward despite the intent to do so [4,5]. There is no existing cure for FoG, but symptom reduction is offered with a technique called cueing. Cueing stimulates the continuation of gait by giving rhythmic feedback in the form of auditory cues or visual stimuli [6].

The mechanism that underlies the success of cueing is not completely known. A possible explanation for its benefits is that cues shift the attention of patients towards the movement of their legs, leading them take steps consciously, thereby reducing the chance of experiencing FoG [7]. Another suggestion is that cues bypass a defective part of the PD brain, the basal ganglia, and thereby acts as a surrogate for the defective internal timing of gait [8].

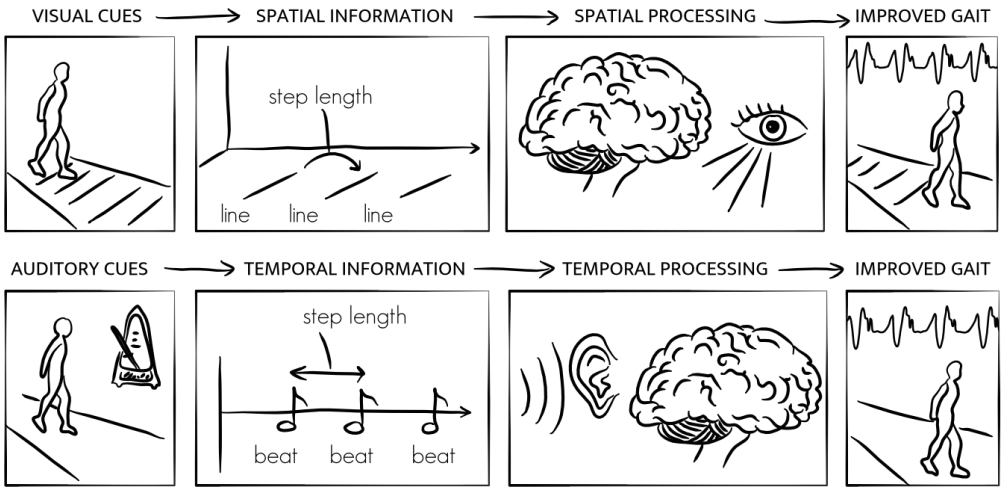


Figure I: Cueing in the form of visual or auditory stimuli can improve gait in PD patients - figure redrawn from [6]

Cueing is the most effective if cues are received as soon as a FoG episode starts instead of continuously [9]. For this reason, algorithms have been developed that identify when a FoG episode commences, to be used in a wearable device that gives cues to PD patients on demand [6]. These FoG detection algorithms use kinematic data (predominantly acceleration data) obtained at the lower limb of a PD patient to decide whether the patient is walking normally (cues are not required) or is approaching or experiencing a FoG episode (cues are required) [6]. FoG is often paired with complex rhythmic oscillations in frequency domain [10]. The gold standard for FoG detection is the "Freeze Index" (FI), developed by Moore et al., that compares vertical acceleration data in frequency domain [11]. The FI compares energy of an acceleration window in the "locomotor band" (0.5-3 Hz) to energy in the "freeze band" (3-8 Hz). If this ratio exceeds a personal threshold, the window is labelled as a FoG event [11].

The FI method has since been used by other researchers. Bachlin et al. improved the method by adding a threshold on the total energy in a window, to avoid false positive FoG detection [12]. Jovanov et al. build upon the method of Moore et al. by computing correlation of the FI to the total energy per window [13].
Coste et al. developed a method unrelated to the FI [14]. Their focus was on detecting festination, an increased cadence coupled with decreased stride length, that often occurs before a FoG episode. Cadence was calculated from forward acceleration data, stride length was estimated by integration, after which they were compared to predefined thresholds [14]. These examples illustrate the diversity in methods used for FoG detection and the desire to improve the existing FoG detection methods.

Implementing FoG detection methods on a wearable cueing device could be very beneficial to PD patients: when a patient experiences FoG, the device could initiate cues, and the patient can resume walking [6]. However, in practice, this is not as simple as it sounds. A lot of FoG detection research was done in a laboratory environment, and involved only walking in a straight line. As I will clarify in the research article, current research is not satisfactory for FoG detection in the home environment. This is an issue that Cue2Walk (a company developing a cueing wearable for PD patients) experiences at the moment. Together with Cue2Walk, I set out to find a solution to this predicament.

Company: Cue2Walk

Medical records show that over 52.000 patients were diagnosed with Parkinson's Disease (PD) in the Netherlands in 2018. Of these patients, most patients were among the elderly population [1]. PD is a neurodegenerative disorder, a disorder of the central nervous system tampering with muscle control [2]. One of the main causes of PD is the loss of dopaminergic neurons in the brain. Dopamine acts as a messenger molecule and is essential for movement control and coordination [2]. Existing medication for PD patients is focussed on reducing the dopamine deficit in the brain and can be used for a limited period only [3]. It is therefore essential to find long-term solutions to symptoms caused by PD.

One of the motor symptoms of PD is Freezing of Gait (FoG). Approximately half of the PD patients diagnosed over five years experience FoG, a short episodic inability to keep moving forward despite the intent to do so [4] [5]. There is no existing cure for FoG, but symptom reduction is offered with a technique called cueing. Cueing stimulates the continuation of gait by giving rhythmic feedback in the form of auditory cues, haptic feedback or visual stimuli [6].

The mechanism that underlies the success of cueing is not completely known. A possible explanation for its benefits is that cues shift the attention of patients towards the movement of their legs, leading them to take steps consciously, thereby reducing the chance of experiencing FoG [7]. Another suggestion is that cues bypass a defective part of the PD brain, the basal ganglia, and thereby acts as a surrogate for the defective internal timing of gait [8].

Related Products

PD is a progressive disease without a cure. The PD patient population is growing, and the market of wearable assistive devices expanded over the last years. Several wearable products related to PD were developed recently.

- The Parkinson Buddy is a device that resonates a continuous walking rhythm using a metronome [15]. The limitation of this device is that it is not "smart". The device gives a continuous signal, that causes habituation after some time (as one gets used to a ticking clock). Cue2Walk focuses on cueing on demand, only when FoG is detected.
- The Parkinson Smartwatch with a build-in movement sensor can monitor tremor in the home-environment [16]. This device can also keep track of mobility and sleep patterns. Statistics are stored in a smartphone app. The smartwatch is aimed at registration and interpretation of PD patient mobility and symptoms, but does not detect FoG due to the position at the wrist.
- Great Lakes NeuroTechnologies (GLNT) developed a device that measures the movement intensity of PD patients, and measures tremor, dyskinesia and activity intensity [17]. This device does not include FoG detection or monitoring.
- Medigait is a virtual reality device that projects a walking pattern on a virtual screen of glasses [18]. This (virtual and visual) method of cueing is only applicable in a limited number of situations indoors. Furthermore, this device does not include a FoG detection algorithm.

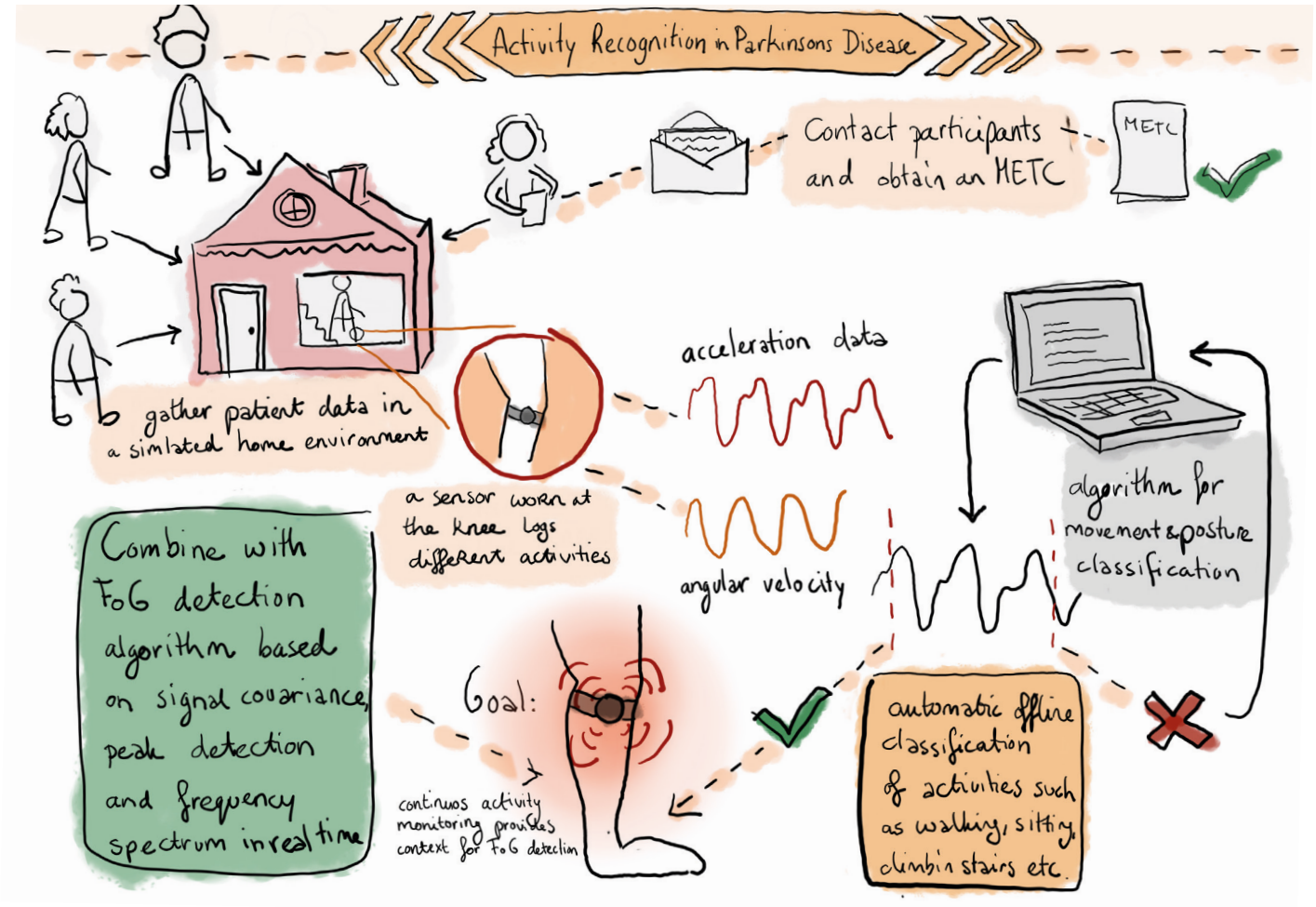
This list of related research and products is not exhaustive. However, it indicates the relevance of the field and the need for the device that Cue2Walk is developing

Thesis Project

Objectives

The goal of the thesis collaboration with Cue2Walk is to reduce the initiation of cues when they are not required by developing a method to improve FoG detection by reducing false positive detection. For the Cue2Walk cueing wearable, it is important to know when a patient is in ambulatory motion (FoG may occur) or when a patient is in a non-ambulatory motion or static position (FoG does not occur). Furthermore, Cue2Walk is planning to incorporate an activity logging feature to the Cue2Walk app. With this activity log, patients can keep track of their daily mobility and of the tasks that they have completed during the day. This information can be shared with doctors and physiotherapists, to evaluate disease progress, to make treatment plans and to set goals for a more active lifestyle. Keeping active is especially important for Parkinson's Disease patients to slow down the progression of the disease [19, 20].

Figure II: Initial sketch of the thesis outline. The plan was to do measurements on PD patients. However, Medical Ethical Review Committee (METC) application is a long process that was not achievable during the time span of the thesis



Experimental Plans

The initial plans for this thesis project included experiments with healthy subjects. Unfortunately, the experiments had to be canceled. Nevertheless, I will shortly describe the original protocol. Approximately ten healthy subjects were required to take part in these experiments. These experiments would take place at Basalt Revalidatie, The Hague.

CE certified MetaMotionR sensors from Mbientlab were ordered to record three-axial acceleration data [21]. Three sensors would be worn in bands and clips at the right leg:

- Hip: Against upper part of the pelvis, attached to the rightmost part of a belt or edge of trousers.
- Thigh: Above the knee in a band around the thigh at the rightmost part.
- Lower leg: Below the knee in a band around the lower leg at the rightmost part.

The following tasks would each be completed twice in randomized order. Except for tasks 3, 4 and 5, all tasks would be carried out for one minute at the time.

1. Straight lined walking: The participant is asked to walk in a straight line, turning around (naturally) when the end of a hall is approaching.
2. Walking in a zigzag: The participant is asked to walk around indicated objects in the room, while not trying to speed up this process (it is not an obstacle course race)
3. Walking the stairs: The participant is asked to walk up and down the stairs 6 times, stopping for at least 1 second before the first and after the last step.
4. Turning in place. The participant is asked to turn in place at his or her own pace. The participant can turn 4 times 15 seconds, to prevent dizziness.
5. Standing still. The participant is asked to stand on the spot, not forcibly trying to keep still.
6. Sitting on a kitchen chair.
7. Lying in bed. The participant is asked to lie down on his or her back.

Alterations

Due to the Covid-19 crisis, the experiments had to be canceled. Fortunately, I could access a publicly available dataset containing activity data of healthy subjects. Most of the tasks that I was planning to include in my experiments, were included in this dataset. Furthermore, cycling and vacuuming were added to the lists of tasks to be classified. I considered this an improvement to my research project.

Research Article

Development of a Bayesian Decision Tree Model for Task Classification and its Validation on Parkinson's Disease Patients

Melina Dekker - Delft University of Technology - Faculty of Mechanical, Maritime and Materials Engineering

***Abstract*—Freezing of Gait (FoG) is a debilitating walking problem affecting over 50% of Parkinson's Disease (PD) patients. Rhythmic cues during FoG can help patients to resume walking. FoG can be detected from lower limb acceleration data, but current detection algorithms lack context, resulting in false positive detection and initiation of cues when these are not required. Cues are only required during ambulatory motion. Adding a task classification model to an FoG detection device can help to increase specificity. Furthermore, using this model, an activity log of PD patients can be kept in the home environment, thereby facilitating disease evaluation.**

A decision tree model using Bayesian classifiers was developed for task classification. Predictors were extracted from raw data of an ankle-worn triaxial accelerometer and the best predictors were determined at each decision node using feature selection methods. The proposed decision tree model (trained using healthy subject data) was tested on PD patient data to evaluate transferability. Furthermore, to compare performance and computational speed, our Bayesian decision node classifiers were compared to Naive Bayes classifiers.

The proposed model was over 30 times faster, compared to the computational speed of using Naive Bayes classifiers at the decision nodes. Ambulatory windows were identified with a sensitivity over 91% for both healthy subjects and PD patients, showing that the Bayesian decision tree model developed in this study can provide context for an FoG detection wearable, to enable effective cueing. This will help patients to walk more confidently and to keep active.

Accelerometry, Task Classification, Freezing of Gait, Parkinson's Disease, Decision Tree Model, Bayesian

I. Introduction

A. Background

Keeping patients active is of great importance to slow down the progression of Parkinson's Disease (PD) and to maintain quality of life [1], [2], [3]. PD is a disorder of the central nervous system that interferes with the control of muscles and causes symptoms such as tremor, bradykinesia (slow movements) and walking problems [4], [5], [6]. Freezing of Gait (FoG) is a debilitating walking problem that affects over 50% of PD patients [7], [8]. It is an episodic inability to move forward despite the intent to do so, often described as the feeling of the feet being glued to the ground [7]. Approximately half of the FoG episodes last over 10 seconds [9]. FoG poses difficulties in performing everyday tasks and can contribute to dangerous situations such as falls [10], [11]. Walking difficulties and fear of falling caused by FoG can conduce an inactive lifestyle, which in turn does not help to slow down the progression of Parkinson's Disease [2].

A popular topic of research on FoG reduction is *cueing* [12], [13], [14]. This technique stimulates the initiation or continuation of movement by exposing patients to rhythmic auditory feedback, thus reducing FoG and improving gait [12], [15]. Cueing is the most effective if cues are received on-demand instead of continuously, as continuous cues can cause hindrance in daily life and can lead to habituation and lower response to cues [16], [17]. Automatic cueing on a wearable device could alleviate FoG symptoms in daily life [17]. FoG episodes need to be identified for cues to start automatically [18]. *FoG detection algorithms* were recently developed for this purpose [17]. These detection algorithms use accelerometer and gyroscope data obtained at the lower limb to identify FoG episodes during walking tasks, obtaining sensitivities of up to 92% [18], [19], [20], [21].

B. PD patient monitoring in the home environment

A major shortcoming in studies on FoG detection is that the clinical setting used in these studies does not reflect *real-life situations* [22]. Algorithms were built to identify FoG during walking but not during activities of daily living. False positives are likely to occur when using an FoG detection wearable in a home environment due to increased variability in movement signals [22]. To avoid habituation and hindrance during the long-term use of an automatic cueing device, false positives should be minimized. FoG episodes occur during ambulatory motion, most likely while walking through narrow spaces, turning and walking while dual-tasking [23]. Cues are not required during

many other activities in daily life, such as sitting or lying down. We propose that *task classification* can provide a context-aware framework for FoG detection, reducing false-positive FoG identification. This will contribute to the user-friendliness of an automatic cueing wearable.

Analogous to this claim, Takac et al. stated that the specificity of FoG detection is dependent on the context of a user, such as the current location or performed task [24]. Knowing the context might reduce FoG misclassification. Takac et al. provided context for FoG detection by indoor position tracking using a camera system [24]. In our study, we will focus on providing context for FoG detection by classifying common tasks of daily living.

Another limitation of PD patient monitoring in the home environment, unrelated to FoG detection, is that current means of *evaluating PD patient mobility* are limited, while the evaluation is of importance to analyze disease progression [25]. A common method to evaluate mobility is patient journaling, where patients are asked to periodically indicate their activities [26]. This method is both inconvenient to patients and subjective. A wearable device, continuously classifying common tasks, can be used to create an *activity log* for the user. Continuous tracking of patient mobility and creating an activity log can contribute to disease management. Dynamic adjustment of treatment could potentially be facilitated [22], [25].

C. Task classification models

Task classification research commonly starts with sensor placement [27]. FoG detection takes place mainly at the lower limb, the most convenient place to record rapid knee-trembling present during FoG [28], [7], [9]. In task classification studies, the waist or the lower leg are popular places to classify simple tasks [29], [27]. After sensor placement, task classification studies persist with data acquisition, pre-processing, predictor extraction, predictor selection, model learning and performance evaluation [27]. *Predictors* are measurable properties of an observed circumstance, class or category. Predictors used for task classification include the mean value, the minimum value, the frequency peak location and the total energy of acceleration data windows [27].

The focus of this study will be on decision tree models for task classification. Decision tree models have a hierarchical structure with multiple decision nodes, each containing a *classifier*: an algorithm labeling an input (data window) as belonging to a certain category or class [30], [31]. Skotte et al. developed a decision tree model for task classification in healthy subjects, obtaining sensitivities of 95%-100% [32]. As decision node classifiers, thresholds on SD (SD) and accelerometer angle were used. In another task classification study, Khan et al. point out that the identification of multiple activities results in complex decision boundaries in the predictor space, which are difficult for a single classifier to solve [33]. This promotes using a decision tree model, enabling decision-making in several steps using multiple classifiers.

In our study, a Bayesian classifier was developed for the decision nodes, as opposed to thresholds, with the following reasoning: when using thresholds for classification, a clear decision boundary is required. However, PD patients show deteriorated gait regularity, bradykinesia and a limited range of motion [34], [35]. Furthermore, the power spectrum obtained from PD patients is wider and less distinct than that of healthy subjects [36]. We hypothesize that predictors obtained from PD patients are affected by changes observed in their movements and that threshold-based classifiers may therefore not provide enough information for classification.

D. Requirements

For developing the classification model, the following set of requirements was extracted:

- **Obtrusiveness.** If a telecare system is obtrusive, for example in the number of sensors used, it interferes with patients' daily routines and will not be used [4]. A single sensor model will therefore be developed. Single sensor task classification models for healthy subjects were developed in recent studies, as presented in a review by Cheung et al. [37].
- **Energy use.** A wearable device for long-term use needs to operate energy efficiently. An accelerometer will thus be used. Accelerometers are common in task classification research and have the advantage of low power consumption [38].
- **Latency.** Tasks need to be classified fast to provide context to FoG detection and cueing. Cues should start with a low delay time (below 2 seconds after FoG starts) to induce prompt resumption of gait [39], [40]. Window size affects the overall lag of classification systems, so a small window size is required [4], [41]. Furthermore, a low sampling rate (but at least 20 Hz [42]) should be used to cut back on the computational speed of the model.
- **Subject independence.** The model needs to be subject-independent to classify data of a subject on which it was not trained. Hence, leave-one-out cross-validation will be used throughout this study [43], [44].

E. Objectives

From a developmental perspective, the objective of this study was twofold. Firstly, the misclassification of FoG on a home cueing wearable can be reduced by providing context in the form of task classification. FoG occurs during ambulatory tasks, therefore FoG detection is required in ambulatory windows (AWs) and not in non-ambulatory windows (nAWs). The first objective was thus to classify AWs and nAWs. Secondly, an activity log can provide insight into mobility and can provide a base for a patient's treatment plan. We aimed to enable continuous and automatic task classification on a wearable device, thereby facilitating disease evaluation.

A decision tree model was developed for task classification and Bayesian classifiers were constructed for classification at the decision nodes. Predictors were extracted from raw data of an ankle-worn tri-axial accelerometer and data obtained from healthy subjects was used to build the decision tree model. PD patient data were then used to assess the transferability of the model. Furthermore, the proposed Bayesian decision node classifiers were replaced with common Naive Bayes classifiers (as developed by The Mathworks Inc [45]) to compare performance and computational speed. The proposed decision tree model was over 30 times faster than when using Naive Bayes classifiers at the decision nodes. AWs were identified with sensitivity over 91% sensitivity for both healthy subjects and PD patients. However, transferability to PD patients was inadequate for task-specific classification. Therefore, training data from PD patients is required in future studies.

II. Method

A. Data

The process overview of building, improving, comparing and validating the proposed decision tree model is shown in Figure 1. Two publicly available datasets containing tri-axial acceleration data were used during this study. Dataset 1 was used to build the model. Dataset 2 was used to evaluate the transferability of the model to PD patients.

Dataset 1 - Healthy subjects - building the model The PAMAP2 physical activity monitoring dataset was created by Reiss et al. in 2012 [46]. Acceleration data were obtained from an inertial measurement unit (IMU) worn at the ankle. A Colibri wireless inertial measurement unit was used, sampling at 100Hz. The participants executed several tasks for approximately 200 seconds per task. The following tasks were included in the protocol; walking, vacuuming, ascending stairs, descending stairs, standing, sitting and lying.

Dataset 2 - PD Patients - validating the model The Daphnet Freezing of Gait Dataset was collected by Bächlin et al. [40]. Acceleration data were obtained from an ankle-worn IMU, sampling at 64 Hz. The dataset contains straight-lined walking and standing.

We will refer to the x-axis as directed horizontally forward, the y-axis as directed vertically upward and the z-axis as directed horizontally outward, using a right-handed coordinate system. Neither dataset specifically contained task transitions.

B. Processing and splitting data

Data were resampled to 50 Hz (The Mathworks, Inc *resample*). This sampling frequency was suggested in studies on task classification and in studies on FoG detection [47], [48], [49]. As low sampling frequencies can fail to pick up activity details and a high sampling frequencies result in higher computational load, 50 Hz was a suitable compromise [27].

After resampling, data were labeled (as containing a certain task) and split with a windowing technique called sliding window approach, an approach well suited for real-time applications [27]. Per class, data was split into consecutive windows that overlap by 50%, as applied in earlier research on task classification [32], [50], [51], [52]. The optimal size of acceleration data windows is 0.8-1.4 seconds for task classification research on healthy subjects [42]. PD patients generally exhibit bradykinesia [53], [54]. This study used a window size of 1.4 seconds to ensure that complete movement cycles were included, enabling derivation of roughly stationary averages of parameters in the interval. The total number of windows for each task after splitting is listed in Table I.

Table I: Number of windows for each task after splitting dataset 1 and dataset 2. A window size of 1.4 seconds was used and windows overlapped by 50%.

	Task	Dataset 1: Healthy subjects [windows]	Dataset 2: PD patients [windows]
1	Walking	3244	2012
2	Upright active	2318	0
3	Stairs (up and down)	2592	0
4	Cycling	2114	0
5	Standing	2608	2828
6	Sitting	2176	0
7	Lying	2416	0

There is no gold standard for filtering and denoising raw data before use [55]. Both filtered data and raw data were used in task classification studies [27], [29], [32]. Millecamps et al. suggested that denoising data before use does not improve predictor quality [55]. Furthermore, the computation of filtering raw data

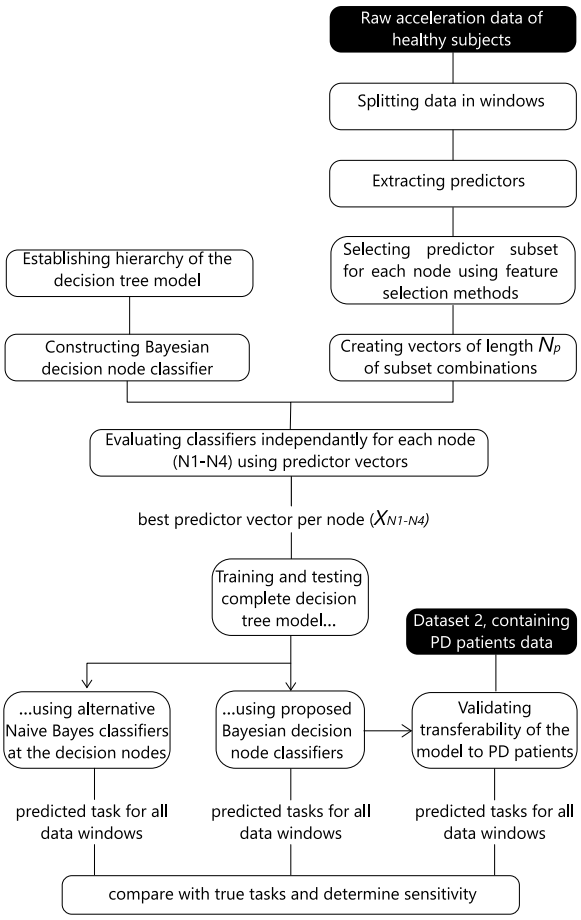


Figure 1: Process overview of building, improving, comparing and validating the proposed decision tree model. Raw acceleration data were labeled, split in windows (of 1.4 seconds) and predictors were extracted. Afterward, a subset of predictors was selected for each decision node classifier based on feature selection scores. All possible combinations of the predictor subset were used to train and test the decision node classifiers independently. Predictor vectors resulting in the highest sensitivity for decision node classifiers were used in the complete decision tree model. To compare classification results and computational speed, the same predictor vectors were used to test and train an alternative decision tree model with more the common Naive Bayes classifiers [45]. Lastly, to evaluate transferability, the proposed decision tree model was trained on dataset 1 (healthy subjects) and tested on dataset 2 (PD patients).

adds latency to a real-time system. Therefore, we extracted predictors from the raw data.

1) Extracting predictors: Forty-four predictors were extracted for every data window, as listed in Table II. These include five tri-axial (3D) predictors and 13 single-axial predictors. Single-axial predictors were included once for every (x,y, z-)axis. Seven of the 1D predictors were time-domain descriptive statistics of an acceleration window: the minimum value [56], the maximum value [56], the range [57], the interquartile range [57] , the arithmetic mean [57], the entropy [56] and the SD [50]. A new single-axial predictor (the number of samples below mean value) was determined as the number of samples within a data window with values below zero for the x- and z-axis and below 9.81 for the y-axis. Data were transferred to the frequency domain using the fast Fourier transform (FFT), subsequently removing the first (DC) value. Five predictors were determined per acceleration windows in the frequency domain: the maximum value [56], the SD [56], the entropy [50], the spectral centroid [56] and the total energy within the data window [56]. The tri-axial predictors were defined as follows: the (3D) interquartile range was determined as the sum of the interquartile ranges over all three axes. The tilt angle was calculated as defined by Lugade et al. [58]. Covariance was defined as the largest eigenvalue of the covariance matrix between the x- and y-axis, as suggested by Capela et al. [57]. Mean Amplitude Deviation (MAD) was determined as suggested by Ypya et al. [52]. Derivative was defined as the sum of the approximate derivatives for all axes [50].

Table II: Fourty-four predictors were extracted for every data window. These include five tri-axial predictors and 13 single-axial predictors. The single-axial predictors were included once for every (x,y,z-)axis.

1D Predictor name	3D Predictor name
Minimum	IQR 3D
Maximum	Tilt angle
Range	Covariance matrix eig.
IQR	Mean amplitude deviation (MAD)
Mean	Derivative
Entropy	
Standard deviation	
Number of samples below mean	
FFT-maximum	
FFT-SD	
FFT-entropy	
FFT-spectral centroid	
FFT-total energy	

C. Decision tree model

Seven tasks were classified in this study: walking, upright active (vacuuming), taking the stairs, cycling, standing, sitting and lying. Our model classifies the tasks at 4 decision nodes (N1-N4), as seen in Figure 2. A decision tree model served the purpose of this

study because the hierarchy of the decision tree could be organized to classify AWs and nAWs at the first two nodes (N1 and N2). This can provide context to a wearable FoG detection and cueing device for PD patients, as cueing is only required during ambulatory tasks. Task-specific classification took place at subsequent nodes.

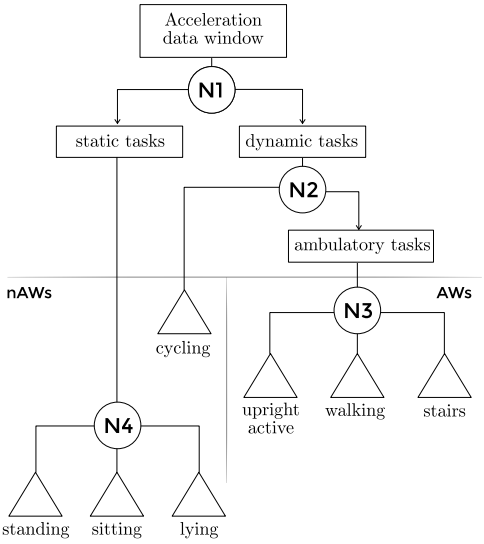


Figure 2: Decision tree model with a single acceleration data window as input and specific task labels as output. N1-N4 represent decision nodes, where classifiers are used to categorize incoming data. N1 and N2 distinguish between AWs and nAWs. Context for FoG detection and cueing is provided at these nodes. At N3 and N4, ambulatory tasks and static tasks are categorized respectively.

D. Bayesian decision node classifiers

We developed classifiers for the decision nodes (N1-N4), based on Bayes' rule from probability theory. Likelihoods $p(x_i|c)$ were computed from labeled training data, where x_i is a certain predictor and c is the class of a data window. The posterior probability of an instance belonging to a class was then calculated as follows:

$$p(c|x_i) = \frac{p(a)p(x_i|c)}{p(x_i)} \tag{1}$$

The a priori probability of the classes $p(a)$ was assumed to be uniform. The probability density function $p(x_i|c)$ was computed as normalized 50 bin histograms. The a-priori probability $p(x_i)$ could be neglected because of the normalization. For each data window, posterior probabilities were then calculated per class as a product of posterior probabilities of all Np predictors x_i in predictor vector $X_N1 - N4$, as:

$$p(c|X_N1 - N4) = \prod_{i=1}^{Np} p(x_i|c) \tag{2}$$

Windows were labeled as the class with the maximum posterior probability.

Validation and performance evaluation: The validation method used to obtain subject independent results were leave-one-out cross-validation [43], [44]. Data of one subject was retained as test data while the model was trained using data from all other subjects. This was repeated 8 times, once for every subject.

Sensitivity was determined to evaluate classification performance. Sensitivity for a certain task is independent of the number of windows for other tasks, making it a useful measure for imbalanced datasets. The number of windows for each task in both datasets was unbalanced, as indicated in Table I. Furthermore, false positives and true negatives could not be calculated for all tasks in dataset 2 because the dataset contained only two tasks (walking and standing). Sensitivity could be determined for both dataset 1 and dataset 2 (using true positives and false negatives) and was thus a suitable measure for comparison.

E. Selecting predictors

The process of evaluating all predictors and selecting final predictor vectors is depicted in Figure 3. Feature Selection (FS) methods were used to select a subset of 14 predictors, which were most predictive for the corresponding class. Dimensionality of the predictor space was thereby reduced: optimizing decision node classifiers using all predictors would not be realizable in terms of computation time. The number 14 was chosen as a trade-off between a high computation time and including as much relevant information as possible. The following three FS methods were used in previous task classification studies to rank predictors based on predictive power [50], [57], [59], [60]:

1) Minimum Redundancy Maximum Relevance (MRMR) algorithm. This method aims to find maximum dependency between a predictor and its assigned class and to exclude redundant predictors [59]. Mutual information between predictors and classes is defined using probability density functions and joint probabilities.
   We used the MRMR algorithm because it is a probability-based FS method that could provide relevant information for the probability-based decision node classifiers developed in this study.

2) Relief F is a popular FS method to rank features by relevance [50], [57]. This algorithm finds the nearest hits (data point belong to the same class) and the nearest misses (data point belong to different classes) for every predictor data point. We evaluated 30 nearest data points. The weight of every predictor starts at zero and updates according to Formula 3.

$$\omega_i = \sum_{j=1}^{M} (x_i^j - nearmiss(x^j)_i)^2 - (x_i^j - nearhit(x^j)_i)^2 \quad (3)$$

Where ω_i represents the weight of the i^{th} predictor, x_j is the value of the i^{th} predictor for data point j, M is the total number of data points, *nearhit* and *nearmiss* represent the nearest data point from the same and different class respectively [50]. The Relief F method is used in this study because it was suggested for datasets with strong interdependencies between predictors [57].

3) Variance ratio (VarRat), as suggested by Moore et al., is a simple method that compares within-class variance $W_c(x_i)$ to between-class variance $B_c(x_i)$ [60].

$$V(x_i) = \frac{B_c(x_i)}{W_c(x_i)} \quad (4)$$

A higher VarRat $V(x_i)$ for a predictor means that it is predictive for the corresponding class. A higher VarRat also decreases the probability that data points of different predictors are close to one another, making the variance ratio a useful tool to identify relevant predictors for Bayesian classification.

Scores from all three FS methods were normalized and added for each predictor. FS methods were repeated for N1-N4. Four lists containing predictor subsets (X_{S_N1-N4}) sorted on descending FS scores were obtained, indicating predictive power for each decision node.

F. Predictor vectors for decision node classifiers

The relation between classification sensitivity and the number of predictors used in the predictor vector for each decision node classifier (N_p) was evaluated. Optimal N_p for every node was assessed as follows: for every node, predictor vectors were created of all possible combinations of the predictor subset X_{S_N1-N4}, using an N_p of one up until eight. The predictor vectors were used to train and test all decision node classifiers independently. The sensitivity for N1-N4 was determined for every predictor vector. Feasibility of minimizing N_p per classifier (for a simpler and faster model) or maximizing N_p (for higher sensitivity) was evaluated and the optimum was established for N1-N4. Predictor vectors (X_{N1-N4}) resulting in the highest classification sensitivity for each node were subsequently used to train and test the complete decision tree model.

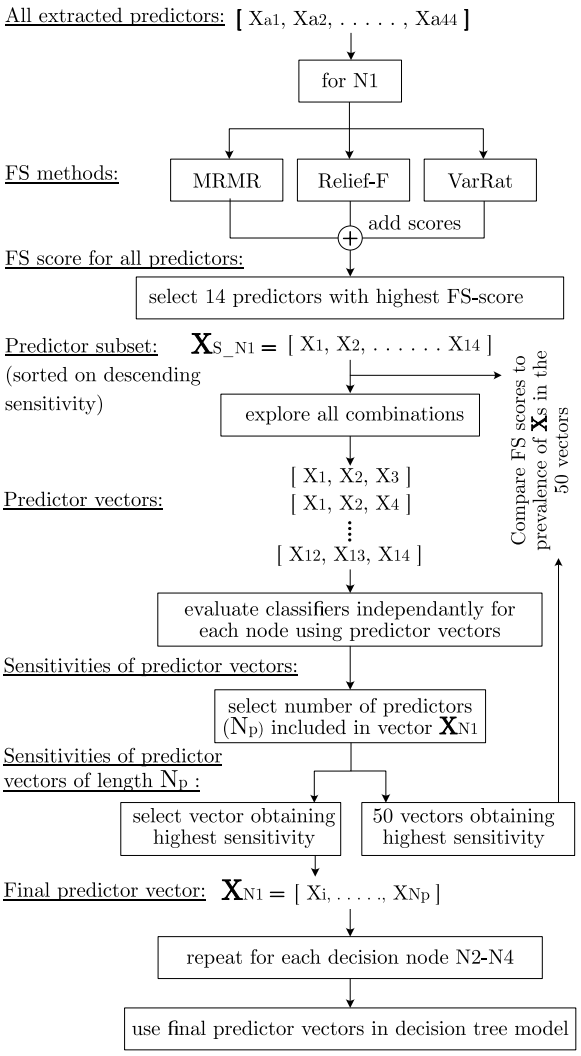


Figure 3: Schematic representation of the predictor selection and comparison process. The process was repeated four times, for decision nodes N1-N4, starting with all 44 extracted predictors and acquiring the final predictor vectors. Predictor subsets (selected as the 14 predictors obtaining the highest FS scores) were compared to their prevalence in the 50 predictor vectors obtaining the highest sensitivity for the corresponding decision node. By this analysis, we gain insight into the suitability of predictor types for each decision node, and we evaluate the correlation of FS scores to prevalence in the top 50 predictor vector. This would indicate the reliability of the FS methods. The number 50 was chosen to obtain an extensive overview of predictor prevalence.

G. Naive Bayes classifier

To compare classification results and speed of the proposed decision tree model, we used an alternative classifier at the decision nodes. The Naive Bayes classifier was used in previous studies on task classification [37], [57], [61] [62], and was reported to outperform other classification methods [27]. Analogous to our Bayesian decision node classifiers, Naive Bayes estimates densities of predictors within each class and models posterior probabilities according to Bayes' rule [63], [64]. The alternative decision tree model was trained and tested using X_{S_N1-N4}.

Comparing computational speed: The decision tree model was tested using both the proposed Bayesian decision node classifiers and the Naive Bayes classifier, processing identical data, to compare the respective computational speed. Time was recorded between receiving the input window with the corresponding predictor vector and assigning a specific task to the window, for every window in dataset 1.

H. Transferability to PD data

Processing data, splitting data and extracting predictors was executed as previously described for dataset 1. The proposed model was trained on dataset 1 (healthy subjects) and tested on dataset 2 (PD patients). PD patients show bradykinesia and move and in a less symmetrical way compared to healthy subjects [34], [53]. Therefore, it was uncertain whether the predictors of tasks performed by healthy subjects would be descriptive for tasks performed by PD patients. We thus decided to evaluate the five best predictor vectors previously acquired for healthy subjects (listed in Appendix B) to test each decision node classifier independently. This includes (X_{N1-N4}) and the following four vectors obtaining the highest sensitivities for each node. Performance per node was evaluated and the preferred predictor vectors were determined. These vectors were then used to test the complete decision tree model, to evaluate the transferability of the model from healthy subjects to PD patients.

III. Results

A Bayesian decision tree model for task classification was developed in this study. Predictors were extracted from raw data and the most suitable predictors (X_{S_N1-N4}) were determined for N1-N4 using FS methods. The subsets of 14 best predictors were combined in predictor vectors of length N_p to train and test each decision node classifier separately. The vectors resulting in the highest sensitivity per node (X_{N1-N4}) were used in the complete decision tree model. The model (trained on healthy subject data) was subsequently tested on PD patient data to evaluate transferability. Furthermore, the proposed Bayesian decision node classifiers were compared to Naive Bayes classifiers to compare performance and speed.

A. Selecting Predictors

Using FS methods, predictor subsets X_{S_N1-N4} were found for N1-N4, which were considered the most predictive for the classes included at the corresponding nodes (Table III). We evaluated the characteristics (x-,y- or z-axis, time-domain versus frequency domain) of the predictor subsets. Z-axis predictors were encountered least among the highest FS scores, although they did obtain high FS scores at N1 (separating static and dynamic tasks) and N2 (classification of static tasks. Y-axis predictors obtained high FS scores at N3 (classification of ambulatory tasks) and N4 (classification of static tasks). Frequency domain predictors were among the highest-scoring predictors especially at N1 but were not often selected at N4. In general, there was no single predictor characteristic that most frequently obtained high FS scores. All acceleration axes provided useful information for task classification, as did both time domain and frequency domain predictors. However, not all predictors were equally suited for N1-N4 of the task classification process, as demonstrated by the variation of predictors in subset X_{S_N1-N4}.

B. Predictor vectors for decision node classifiers

To find the optimal N_p for each Bayesian decision node classifier, the sensitivity of the classifiers was determined for an N_p of one up until eight. The sensitivity of each decision node classifier was determined for all combinations of X_{S_N1-N4}. It was hypothesized that using Bayesian classifiers would be advantageous as opposed to using thresholds on single predictors. Results are shown in Figure 4. Using a higher N_p resulted in higher sensitivity than using a single predictor (threshold) for each node. The optimal number of predictors was N_p = 3 at N1 and N_p = 4 at N3 and N4. At N2, sensitivity increased with an increasing N_p. The optimal N_p at N2 was selected as four, as using more predictors adds latency to the model, and as the increase in sensitivity was minimal when adding more predictors. In general, increasing N_p for a Bayesian classifier should result in higher sensitivity because more information is available to the classifier. Contrary to expectations, this was only the case for N2.

The predictors in subset X_{S_N1-N4} (as listed in Table III) were combined in vectors to train and test the four Bayesian decision node classifiers at N1-N4

Table III: Subset of predictors (X_{S_N1-N4}) with the highest FS scores (sorted on descending FS score) for N1-N4 and their prevalence in 50 predictor vectors that resulted in the highest sensitivities of corresponding decision node classifiers. In general, all acceleration axes as well as time domain and frequency domain predictors were suitable for task classification. However, distinct predictor types were suited for distinct steps of the task classification process, as demonstrated by the variation of predictors selected for each decision node. Next to each selected predictor, its prevalence (Pr) in the top 50 predictor vectors is shown in percentage. High FS scores were not directly related to high sensitivity when the predictors were used in the Bayesian decision node classifiers. This decreases reliability of the FS methods.

Predictor subset N1	Pr [%]	Predictor subset 2	Pr [%]	Predictor subset 3	Pr [%]	Predictor subset 4	Pr [%]
IQR 3D	0	FFT Entropy x	0	Min y	70	Angle 3D	48
FFT totalEnergy z	14	MAD 3D	25	IQR y	34	Mean y	38
FFT SD x	2	Max z	28	Range y	42	Min y	58
Range z	0	Range z	22	FFT totalEnergy y	38	BelowMean y	38
FFT Max x	2	FFT totalEnergy x	26	SD y	0	Max y	26
IQR x	4	Entropy x	50	FFT SD y	0	Entropy z	28
SD z	22	Range x	0	FFT Entropy x	20	Entropy y	24
FFT SD z	18	FFT totalEnergy z	0	Mean y	46	Entropy x	26
FFT SpectralCentroid x	36	Entropy z	60	FFT Max y	0	Mean x	24
BelowMean x	36	SD x	24	Max y	26	Max x	10
FFT Entropy x	44	SD z	22	FFT totalEnergy x	36	FFT SpectralCentroid x	70
BelowMean y	46	Max x	0	MAD 3D	0	FFT SpectralCentroid y	10
IQR z	24	Entropy y	86	BelowMean y	58	CovarianceEig 3D	0
FFT Entropy y	52	FFT Entropy z	30	Range x	30	Min x	0

and sensitivity was determined. The vectors contained three predictors at N1 and four predictors at N2-N4. Vectors obtaining the highest sensitivity per node were examined to see if predictors with the highest FS scores contributed to the highest decision node classifier sensitivity. This would demonstrate the reliability of the FS methods. The prevalence (Pr) of the predictors in X_{S_N1-N4} in the top 50 combinations per node is found in Table III next to the respective predictor. At N3 and N4, as expected, high FS scores often coincided with high prevalence. For example, minimum acceleration of the y-axis scored highest using FS methods for N3 and was present in 70% of the top 50 predictor vectors for N3. However, at N1 and N2, the correlation between FS scores and high prevalence in the top 50 vectors was not present. The predictors with the highest FS scores were not found in the top 50 predictor vectors for these nodes. High FS scores of a predictor were thus not directly related to a high sensitivity when used in the Bayesian decision node classifiers, decreasing the reliability of the FS methods.

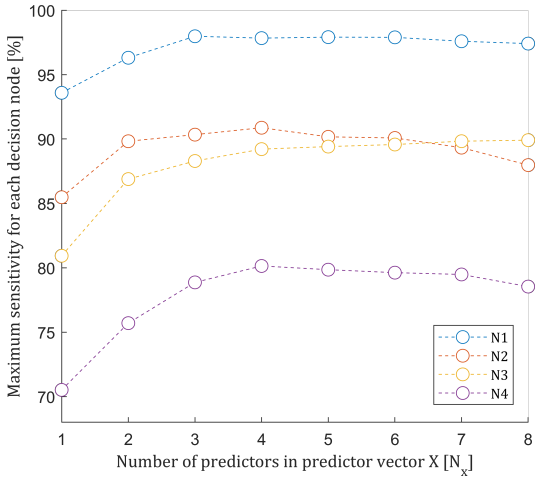


Figure 4: Relation between number of predictors N_p used per node and the maximum obtained sensitivity. A general trend was observed where sensitivity increases with an increasing N_p (up until four) was used. The optimal N_p was selected as N_p = 3 at N1 and N_p = 4 at N2-N4.

C. Complete decision tree model

Predictor vectors X_{N1-N4}, listed in Table IV, obtained the highest sensitivity and were subsequently used at their respective nodes to train and test the proposed decision tree model (see Figure 2). The confusion matrix of the classification results is shown in Figure 5. Walking, cycling and lying were classified correctly in most instances. The most common misclassifications were upright active classified as cycling, taking the stairs classified as walking and sitting classified as standing. The top right corner and the bottom left corner of the confusion matrix show low occupancy, meaning that dynamic tasks were not often classified as static tasks and vice versa.

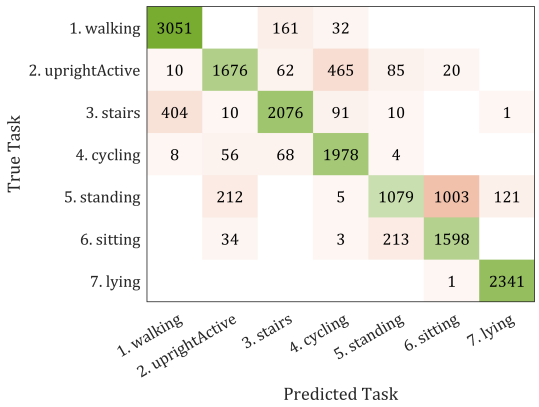


Figure 5: Confusion matrix of classifying dataset 1 using the proposed decision tree model. The rows show the true class of each window that was classified and columns show the class assigned to each instance. Walking, cycling and lying were classified correctly in most cases. Standing was often misclassified as sitting.

Table IV: Predictor Vectors X_{N1-N4} that resulted in the highest classification sensitivity per node. The predictor combinations were used at their respective nodes to train and test the complete decision tree model.

Predictor vector X_{N1}	Predictor vector X_{N2}
FFT SpectralCentroid x	MAD 3D
IQR z	Entropy y
FFT Entropy y	FFT Entropy z
	Entropy z
Predictor vector X_{N3}	**Predictor vector X_{N4}**
Range y	Min y
FFT totalEnergy x	BelowMean y
BelowMean y	Max y
IQR y	FFT SpectralCentroid x

1) Alternative validation method: Monte Carlo cross-validation: X_{N1-N4} were used again for the proposed decision tree model, this time training and testing the model using Monte-Carlo (MC) cross-validation. The confusion matrix is included in Appendix C. We investigated whether leave-one-out cross-validation, the default validation method for developing a subject independent model, could perform as well as a model trained with data from all subjects. Task-specific sensitivities are presented in Table V for both validation methods. The average sensitivity for leave-one-out cross-validation and MC cross-validation was comparable (81.9% compared to 82.9%). Test data was thus classified with high sensitivity, even when test data consisted of tasks performed by an unfamiliar subject.

2) Alternative decision node classifier: Naive Bayes: As an alternative to the Bayesian decision node classifiers, the decision tree model was trained and tested (with X_{N1-N4}) using the Naive Bayes classifier at the decision nodes. Functionality was compared to the proposed Bayesian decision node classifier. The confusion matrix is shown in Appendix C. The most common misclassifications were sitting classified as standing (and vice versa), taking the stairs classified as walking, upright active classified as cycling and standing classified as upright active.

Task-specific sensitivity is shown in Table V. Average sensitivity obtained with the Naive Bayes classifier is lower than for the proposed Bayesian decision node classifier (77.7% compared to 81.9%) and SD for the Naive Bayes classifier was higher than for the proposed Bayesian decision node classifier (11.9% compared to 10.4%). These results demonstrate the preference of the proposed Bayesian decision node classifiers as opposed to the Naive Bayes classifiers frequently used in task classification studies.

D. Computational speed

The time between receiving the input window with its corresponding predictors and assigning a specific task was recorded for every window in dataset 1. The proposed decision tree model took on average 1.9 μs (SD = 0.44 μs). The alternative decision tree model with Naive Bayes classifiers took on average 57 μs (SD = 25 μs). The proposed decision tree model was over 30 times faster.

E. Transferability to PD data

1) Predictor values: Predictors with high predictive power for healthy subjects, as found in Table III, were compared to the same predictors for PD patient data. It was hypothesized that predictor values may diverge, as the movement of PD patients is often slower and less symmetrical compared to healthy subjects [36]. The predictor distributions for certain tasks are shifted or wider compared to those obtained from healthy subjects, as seen in Figure 6. This confirms the presumption that predictor distributions differ between healthy subjects and PD patients. Consequently, certain predictors that were highly predictive for certain tasks in healthy subjects were not useful when transferring the model to PD data and we had to reconsider the predictor vectors used to classify dataset 2.

2) Testing classifiers with top 5 predictor combinations: Because not every predictor could be used to transfer the model to PD data, we evaluated the Bayesian decision node classifiers using the top 5 predictor vectors for healthy subjects (listed in Appendix B) to find the combinations that were most transferable to PD patients. Correctly classified instances per node are shown in Figure 7. The

Table V: Performance of the proposed decision tree model using leave-one-out cross-validation, using Monte-Carlo cross-validation and using PD patient data to validate the transferability of the model. Performance is additionally shown for the alternative decision tree model that used Naive Bayes classifiers. Sensitivity was high for the proposed decision tree model, indicating its ability to classify tasks subject-independently. Sensitivity in PD patients was low, meaning that model transferability from healthy subjects to PD patients was inadequate.

Task	walking	upright active	stairs	cycling	standing	sitting	lying	Average
Proposed decision tree model:								
Leave-one-out cross-validation sensitivity [%]	93.8	72.3	80.2	93.6	47.0	86.3	100.0	81.9
Standard Deviation [%]	8.2	7.4	6.5	3.8	26.0	20.7	0.1	10.4
Monte-Carlo cross-validation sensitivity [%]	96.7	73.3	80.3	93.0	47.3	90.3	99.4	82.9
Standard Deviation [%]	1.3	1.7	1.7	1.9	1.3	1.2	0.6	1.4
PD patients sensitivity [%]	55.3				51.5			53.4
Standard Deviation [%]	33.3				23.3			28.3
Alternative model with Naive Bayes classifier:								
Leave-one-out cross-validation sensitivity [%]	87.4	75.4	80.8	91.0	31.2	78.1	100.0	77.7
Standard Deviation [%]	15.3	8.9	4.7	3.5	24.8	25.7	0.0	11.9

sensitivity obtained by the top 5 predictor vectors varies considerably. Once more, these results indicate that not all predictors were suited for transferring the model to PD patients. The predictor vectors that resulted in the highest sensitivity for PD data were subsequently used in the proposed decision tree model to classify dataset 2, and are listed in Appendix B.

The proposed decision tree model was trained using healthy subject data and tested on PD patient data. The confusion matrix is included in Appendix C. Walking instances were often classified as taking the stairs and numerous standing instances were classified as sitting. Task-specific sensitivity is shown in Table V. Sensitivity for walking instances was 55.1% and sensitivity for standing instances was 51.5% (*SD* = 5.8%). The low sensitivities and high SD values signify a low transferability of the model from healthy subjects to PD patients.

F. Classifying ambulatory tasks

Task windows were either ambulatory windows (AWs) for walking, upright active or taking the stairs, or non-ambulatory windows (nAWs) for all other tasks.

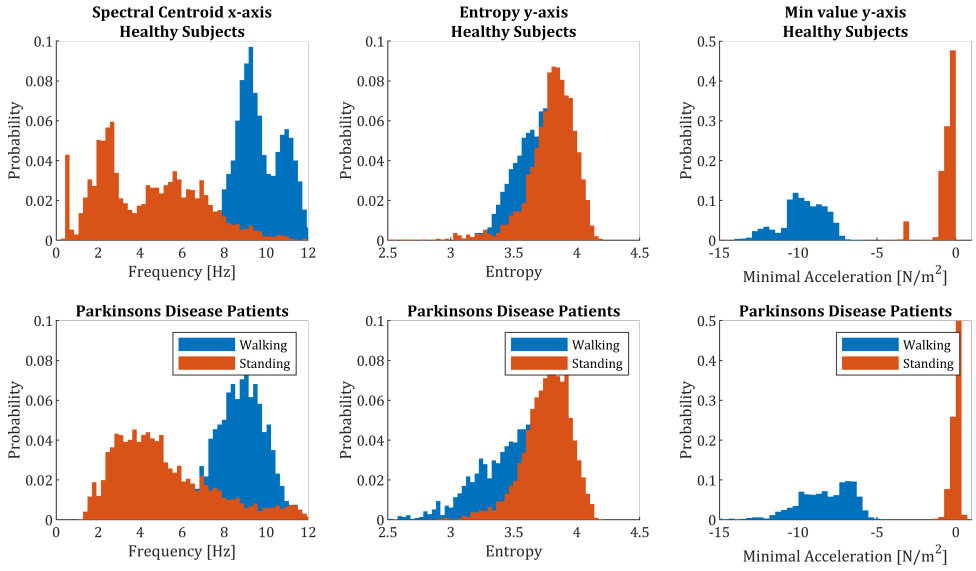


Figure 6: Probability distributions of relevant predictors for healthy subjects (Table III) compared to the same predictors for PD patient data. For the spectral centroid of the x-axis and the minimum acceleration of the y-axis, the predictor distribution in PD patients was shifted compared to that of healthy subjects. For the entropy of y-axis acceleration, the distribution in PD patients was wider compared to that of healthy subjects. These results illustrate that some predictors suitable to classify certain tasks in healthy subjects might not be suitable to transfer the model to PD patients.

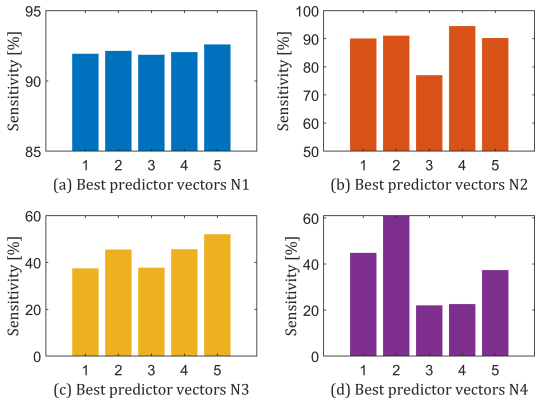


Figure 7: Classification per node using the top 5 predictor vectors for healthy subjects. Especially at N2 and at N4, sensitivity obtained by the top 5 predictor vectors varies considerably (over 10% at N2 and over 30% at N4). Thus, not every predictor vector was suited for transferring the model from healthy subjects to PD patients.

AWs were separated from nAWs at N1 and N2. When correctly identified, context can be provided for FoG detection and cueing. Classification sensitivity of N1 and N2 can be seen in Table VI for both healthy subjects and PD patients.

Table VI: Sensitivity of the Bayesian decision node classifiers at N1 and N2 for both healthy subjects and PD patients. The results support the feasibility of using the proposed decision tree model for the first objective of this study: to improve the misclassification of FoG on a home cueing wearable for PD patients by providing context in the form of AW and nAW classification.

Ambulatory/non-Ambulatory Windows	AWs	nAWs
Healthy subjects sensitivity [%]	91.3	95.8
Standard Deviation [%]	2.8	2.7
PD patients sensitivity [%]	94.9	91.6
Standard Deviation [%]	6.2	9.2

AWs were classified with a sensitivity of 91.3% and 94.9% in healthy subjects and PD patients respectively, and nAWs with a sensitivity of 95.8% and 91.6% in healthy subjects and PD patients respectively. These results support the feasibility of using the proposed decision tree model for the first objective of this study: to improve the misclassification of FoG on a home cueing wearable for PD patients by providing context in the form of AWs and nAWs.

IV. Discussion

The proposed decision tree model using Bayesian decision node classifiers has a computational speed over 30 times higher than when common Naive Bayes classifiers were used at the decision nodes [45]. Transferability from healthy subjects to PD patients was inadequate for task-specific classification, as predictors obtained from healthy subject data were dissimilar to those obtained from PD data, due to differences in gait characteristics. However, classifying ambulatory tasks was attainable by transferring the model to PD patients: AWs and nAWs were identified with over 91% sensitivity for both healthy subjects and PD patients.

A. Limitations on the datasets

The *sensor placement* at the ankle posed certain restrictions to this study. In other task classification studies, locations around the lower limb and the waist were explored [27], [32], [42], [57]. Skotte et al. obtained promising results for classification at the thigh by using the inclination angle and tilt angle of the sensor [32]. Karantonis et al. placed a sensor at the hip and obtained high accuracy using the tilt angle of the sensor [42]. The angle of the ankle-worn sensor provided little relevant information during our study. With data obtained at the ankle, it was challenging separating standing and sitting windows from one another. It should be kept in mind that using other sensor locations can allow the use of other predictors which could facilitate the classification of certain tasks.

Another limitation was the *available data of PD patients*. Dataset 2, used to validate the transferability of the proposed decision tree model, contained only walking and standing data. More comprehension of applying the proposed model to PD patients could be obtained by analyzing data of the five remaining tasks. Furthermore, the decision tree model could only be trained with healthy subject data, that turned out to be dissimilar to PD patient data. In future research, training the model with PD patient data may contribute to more accurate task classification.

Feature selection methods are a valuable tool for selecting a subset of predictors, helping to reduce dimensionality in predictor space. However, high FS scores did not guarantee that predictors would perform well in classification. Predictors with high FS scores were not always common in the most successful predictor vectors at the corresponding node. It could even be the case that we failed to include certain predictors with high predictive power in the predictor subsets X_{S_N1-N4}: at N1 and N2, there was *no apparent correlation between the FS score and prevalence* in the best predictor vectors. Possible solutions would be (1) to increase the number of predictors N_p used in the vectors to train and test each decision node classifier,

(2) to examine possibilities of using other FS methods or (3) to select predictors that score highest for the most FS methods, instead of adding the normalized scores from each method. The last solution (robustness criteria) was used by Moore et al. to select features for FoG detection [60]. Furthermore, we suggest that FS should only be used for dimensionality reduction and not for selecting final predictor vectors used in a classification model.

To find the optimal number of predictors (N_p) in the predictor vector for N1-N4, sensitivity was determined for each decision node classifier, using an N_p of one to eight. In general, increasing N_p for a Bayesian classifier should result in higher sensitivity because more information is available to the classifier. This was only true for N2. A reason for the unexpected behavior at the other nodes could be the fact that Bayes' rule is based on the assumption that *predictors are independent of one another*. In the case of task classification, however, predictors are often highly correlated [57]. For example, if the range of a data window is large, the interquartile range is likely to be large and so are the minimum and maximum value within the data window. Therefore, using a high N_p might not necessarily add new relevant information. Especially considering the possibility that we did not include all predictors with the highest predictive power in the subset X_{S_N1-N4} for each node. However, in this study, using a low N_p is advantageous, as numerous predictors in each predictor vector would increase the computational load of the model [65], making it unsuited for real-time use. The decrease in sensitivity when using a high N_p was thus not an issue in this study.

B. Limitations of classification

The proposed decision tree model classifies multiple tasks in healthy subjects with a high average sensitivity of 81.9%. In the introduction of our study, we mentioned that Skotte et al. obtained task-specific sensitivities of 95-100% using a decision tree model [32]. There are a few possible reasons why these results surpass the results of our study. Firstly, Skotte et al. filtered the data before extracting predictors (using a low-pass filter with a 5 Hz, 4th order Butterworth filter) and divided data in windows of 2 seconds. This processing method possibly contributes to the derivation of more consistent predictor values. Secondly, the sensor was placed at the thigh instead of at the ankle. Tilt angles were used to create a clear decision boundary between sitting and standing and between cycling and ambulatory tasks. Thirdly, only two static tasks were included in the research of Skotte et al. (sitting and standing) and running was included instead of upright active (in our study). The difference in task intensity results in more apparent decision boundaries. *The type and number of classified tasks* included in a study can greatly affect average sensitivity, while not necessarily establishing the superiority of the applied methods.

AWs and nAWs were classified with high sensitivity (over 91%) for both healthy subjects and PD patients. Classification of ambulatory tasks resulted in higher sensitivity in the data of PD patients than in healthy subjects (94.9% as opposed to 91.3%). A possible explanation is that the ambulatory task that was most often classified as non-ambulatory (upright active, see Figure 5) was not included in the PD data. Sensitivity would likely drop if upright active would be included. The non-ambulatory task that was most often classified as ambulatory was standing (Figure 5). Since this task was present in the PD data, we consider the sensitivity of nAWs in PD patients as reliable.

Task-specific classification on PD patients was inadequate in our study: walking data were recognized with a sensitivity of 55.3% for PD patients, as opposed to 83.8% for healthy subjects. PD walking instances were frequently classified as taking the stairs, a misclassification that was not entirely unexpected: PD patients show *reduced walking speed and decreased symmetry* in gait, which could be interpreted as characteristics of taking the stairs [34], [36], [66]. Classification for this task is likely to improve when training the model with data obtained from PD patients, as predictors would then include these characteristics.

We desired *high transferability* of our model because obtaining data from healthy subjects is simpler than obtaining PD patient data, and datasets of healthy subjects are (publicly) available [46]. Training a model with this data and transferring it directly to PD patients would a straightforward method of development. Transferability was compared to the transferability of other studies transferring a classification model from healthy subjects to PD patients. Nguyen et al. reported a sensitivity of 97.6% when transferring a classification model developed for elderly subjects to PD patients [35], demonstrating the transferability of their model that classified standing, turning, walking and sitting. A plausible explanation is that movement patterns of elderly subjects is more similar to that of PD patients. A further explanation is that 17 sensors were worn by the subjects, a number that would not be achievable in the home environment. In another study evaluating transferability, Albert et al. used a single mobile phone acceleration sensor to classify walking, standing and sitting in both healthy subjects and PD patients, obtaining average accuracy of 86.0% for healthy subjects [25]. However, when applying the

classification parameters derived from healthy subjects to PD patients, accuracy dropped to 60.3% [25]. Analogous to the conclusion of our study, transferring a model directly from an able-bodied to a PD population was not feasible. In future research, we will thus have to obtain PD patient data to finalize the proposed model for application in a cueing wearable.

C. Contributions

For comparison, the proposed decision tree model was trained and tested using MC cross-validation with the hypothesizes that this method would result in higher sensitivity than leave-one-out cross-validation. With MC cross-validation, training data are more similar to test data, with the assumption that predictors for a certain task are more similar within-subject than between subjects. However, the difference in obtained average sensitivity was small (81.9% for leave-one-out cross-validation and 82.9% for MC cross-validation). We can compare this to the study of Albert et al., where leave-one-out cross-validation resulted in 10% more misclassifications compared to using MC cross-validation. This comparison emphasizes the subject-independent aspect of the proposed decision tree model.

In this study, we demonstrated that *specific predictors types* are suited for specific nodes in the decision tree. Frequency domain predictors are suited to separate static tasks from dynamic tasks. This was in accordance with the fact that frequency analysis provides information about the periodicity of a signal: static tasks do not involve periodic movements. At N3 and N4, classifiers should detect subtle differences between similar tasks (such as walking and taking the stairs, standing and sitting). At these nodes, y-axis predictors obtained high FS scores (Table III), suggesting that y-axis acceleration data was most suitable to detect subtle differences between specific tasks. We can draw a parallel with FoG detection algorithms, where y-axis predictors are often used to detect the subtle differences between regular walking and steps at a higher cadence (festination) that precedes FoG [9], [67], [68]. The fact that predictors from certain axes and predictors from time or frequency domain are suitable for different steps of the classification process, underlines the advantage of using a decision tree model for task classification: in a decision tree, specific predictors types can be used for a suitable decision node classifier.

Using thresholds on single predictors at the decision nodes resulted in lower sensitivities compared to the proposed Bayesian classifiers, according to our expectations. Furthermore, we replaced the proposed Bayesian classifiers at decision nodes by the more common *Naive Bayes classifiers* for comparison. Naive Bayes classifiers are based on the same principles as the Bayesian classifiers and were therefore expected to yield similar classification results. The sensitivity obtained using the proposed Bayesian classifiers was on average lower than for the Naive Bayes classifiers (81.9% compared to 77.7%). Moreover, the computational speed of the decision tree model when using the proposed Bayesian classifiers was over 30 times faster than for the common Naive Bayes classifiers. Gao et al. investigated the computational speed of several classifiers commonly used for task classification: Support Vector Machine, Artificial Neural Networks, K-Nearest-Neighbour and Naive Bayes [69]. Compared to the Naive Bayes classifier, the other classifiers were up to three times faster, apart from K-Nearest-Neighbour, which was slower. We conclude that our Bayesian classifiers show sufficient computational speed, over 30 times higher than that of Naive Bayes. This promotes the benefits of using the proposed decision tree model, which is both fast and effective.

D. Application

The proposed Bayesian decision tree model is computationally fast and therefore promising for *real-time use*. A recommendation for real-time use is to increase the overlap between consecutive windows. An overlap of 50% was chosen to build the model as a bigger overlap could result in many data windows containing very similar information. For real-time use, however, a larger overlap results in a shorter update time, decreasing the latency between changing tasks and classification of a new task.

The proposed model type, a decision tree, was selected to best support the objectives of this study: the *hierarchy of the decision tree* specifically classifies AWs at the first two nodes, allowing for effective integration of an FoG detection scheme. As soon as AWs are detected, FoG detection should be started. Parallel to N3, an extra node can be added, containing an FoG detection algorithm.

Future research: Correct classification of AWs specifically can be of importance for FoG detection and to initiate cueing at the right time. We expect that an AW is less likely to be classified as a nAW when the user is approaching an FoG episode. FoG is often preceded by festination and is often paired with trembling of the legs, resulting in higher energy per window [7], [9]. Data approaching or containing an FoG episode would be more similar to walking or taking the stairs, as these tasks contain the most

energy per window. These data may thus be correctly classified as AWs. This expectation will be approved or disproved during future research with PD patients.

During this study, a criterion was used that maximizes the average sensitivity per class. However, for FoG detection, it may be more important to maximize sensitivity for AWs. Other predictor vectors might, in this case, be preferred at N1 and N2.

Further research is needed to optimize our model for PD patients. Task data from PD patients is required to train the decision tree model. Furthermore, FS methods used in this study need to be reevaluated to ensure that suitable predictors are identified. FoG detection in combination with the proposed decision tree model, using the context provided by N1 and N2, is promising for implementation on a wearable cueing device, thereby decreasing FoG misclassification.

V. Conclusion

A fast and simple Bayesian decision tree model for task classification has been developed in this study. The proposed model, classifying seven tasks using a single tri-axial accelerometer, has a high computational speed: over 30 times faster than when Naive Bayes classifiers were used at the decision nodes. One objective of this study was to develop a model for keeping an activity log by continuous task classification on a wearable device, thereby facilitating the mobility evaluation of PD patients. Promising task classification results were obtained using healthy subject data (81.9% sensitivity on average), but the model was not directly transferable from healthy subjects to PD patients. To use the model on a wearable device, it should be trained with PD patient data in the future. The second objective of this study was to provide context for a wearable cueing device. AWs and nAWs were identified with sensitivities over 91% for both healthy subjects and PD patients. Incorporating the proposed decision tree model in an FoG detection and cueing wearable reduces the problem of unrequired cues when FoG is falsely identified during non-ambulatory activities. User-friendliness and effectiveness of such a device can hence be improved. Effective cueing helps PD patients to walk more confidently and to ultimately enjoy a more active lifestyle, which could in turn be an effective way to delay the progression of Parkinson's Disease [2].

References

[1] T. Ellis and L. Rochester, "Mobilizing Parkinson's disease: The future of exercise," pp. S95–S100, 2018.

[2] T. Paillard, Y. Rolland, and P. S. de Barreto, "Protective effects of physical exercise in Alzheimer's disease and Parkinson's disease: A narrative review," pp. 212–219, jul 2015.

[3] A. D. Speelman, B. P. Van De Warrenburg, M. Van Nimwegen, G. M. Petzinger, M. Munneke, and B. R. Bloem, "How might physical activity benefit patients with Parkinson disease?" pp. 528–534, sep 2011.

[4] F. M. Khan, M. Barnathan, M. Montgomery, S. Myers, L. Cote, and S. Loftus, "A Wearable Accelerometer System for Unobtrusive Monitoring of Parkinson's Diease Motor Symptoms," in *2014 IEEE International Conference on Bioinformatics and Bioengineering*. IEEE, nov, pp. 120–125.

[5] S. V. Perumal and R. Sankar, "Gait and tremor assessment for patients with Parkinson's disease using wearable sensors," *ICT Express*, vol. 2, no. 4, pp. 168–174, dec 2016.

[6] S.-H. Lee and J. S. Lim, "Parkinson's disease classification using gait characteristics and wavelet-based feature extraction," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7338–7344, jun 2012.

[7] J. G. Nutt, B. R. Bloem, N. Giladi, M. Hallett, F. B. Horak, and A. Nieuwboer, "Freezing of gait: moving forward on a mysterious clinical phenomenon," *The Lancet Neurology*, vol. 10, no. 8, pp. 734–744, aug 2011.

[8] D. M. Tan, J. L. McGinley, M. E. Danoudis, R. Iansek, and M. E. Morris, "Freezing of Gait and Activity Limitations in People With Parkinson's Disease," *Archives of Physical Medicine and Rehabilitation*, vol. 92, no. 7, pp. 1159–1165, jul 2011.

[9] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in Parkinson's disease," *Journal of Neuroscience Methods*, vol. 167, no. 2, pp. 340–348, jan 2008.

[10] S. Rahman, H. J. Griffin, N. P. Quinn, and M. Jahanshahi, "Quality of life in Parkinson's disease: The relative importance of the symptoms," *Movement Disorders*, vol. 23, no. 10, pp. 1428–1434, jul 2008.

[11] B. R. Bloem, J. M. Hausdorff, J. E. Visser, and N. Giladi, "Falls and freezing of gait in Parkinson's disease: A review of two interconnected, episodic phenomena," *Movement Disorders*, vol. 19, no. 8, pp. 871–884, aug 2004.

[12] D. Sweeney, L. Quinlan, P. Browne, M. Richardson, P. Meskell, and G. ÓLaighin, "A Technological Review of Wearable Cueing Devices Addressing Freezing of Gait in Parkinson's Disease," *Sensors*, vol. 19, no. 6, p. 1277, mar 2019.

[13] I. Lim, E. van Wegen, C. de Goede, M. Deutekom, A. Nieuwboer, A. Willems, D. Jones, L. Rochester, and G. Kwakkel, "Effects of external rhythmical cueing on gait in patients with Parkinson's disease: a systematic review," *Clinical Rehabilitation*, vol. 19, no. 7, pp. 695–713, nov 2005.

[14] C. Nombela, L. E. Hughes, A. M. Owen, and J. A. Grahn, "Into the groove: Can rhythm influence Parkinson's disease?" *Neuroscience & Biobehavioral Reviews*, vol. 37, no. 10, pp. 2564–2570, dec 2013.

[15] E. van Wegen, C. de Goede, I. Lim, M. Rietberg, A. Nieuwboer, A. Willems, D. Jones, L. Rochester, V. Hetherington, H. Berendse, J. Zijlmans, E. Wolters, and G. Kwakkel, "The effect of rhythmic somatosensory cueing on gait in patients with Parkinson's disease," *Journal of the Neurological Sciences*, vol. 248, no. 1-2, pp. 210–214, oct 2006.

[16] P. J. McCandless, B. J. Evans, J. Janssen, J. Selfe, A. Churchill, and J. Richards, "Effect of three cueing devices for people with Parkinson's disease with gait initiation difficulties," *Gait & Posture*, vol. 44, pp. 7–11, feb 2016.

[17] S. Del Din, A. Godfrey, C. Mazzà, S. Lord, and L. Rochester, "Free-living monitoring of Parkinson's disease: Lessons from the field," *Movement Disorders*, vol. 31, no. 9, pp. 1293–1313, sep 2016.

[18] C. A. Coste, B. Sijobert, R. Pissard-Gibollet, M. Pasquier, B. Espiau, and C. Geny, "Detection of freezing of gait in Parkinson disease: preliminary results." *Sensors (Basel, Switzerland)*, vol. 14, no. 4, pp. 6819–27, apr 2014.

[19] P. Tahafchi, R. Molina, J. A. Roper, K. Sowalsky, C. J. Hass, A. Gunduz, M. S. Okun, and J. W. Judy, "Freezing-of-Gait

detection using temporal, spatial, and physiological features with a support-vector-machine classifier," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2017. IEEE, jul 2017, pp. 2867–2870.

[20] N. Naghavi and E. Wade, "Prediction of Freezing of Gait in Parkinson's Disease Using Statistical Inference and Lower–Limb Acceleration Data," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 947–955, may 2019.

[21] S. Rezvanian and T. Lockhart, "Towards Real-Time Detection of Freezing of Gait Using Wavelet Transform on Wireless Accelerometer Data," *Sensors*, vol. 16, no. 4, p. 475, apr 2016.

[22] D. Rodríguez-Martín, A. Samà, C. Pérez-López, A. Català, J. M. Moreno Arostegui, J. Cabestany, À. Bayés, S. Alcaine, B. Mestre, A. Prats, M. C. Crespo, T. J. Counihan, P. Browne, L. R. Quinlan, G. ÓLaighin, D. Sweeney, H. Lewy, J. Azuri, G. Vainstein, R. Annicchiarico, A. Costa, and A. Rodríguez-Molinero, "Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer." *PloS one*, vol. 12, no. 2, p. e0171764, 2017.

[23] W. R. Young, L. Shreve, E. J. Quinn, C. Craig, and H. Bronte-Stewart, "Auditory cueing in Parkinson's patients with freezing of gait. What matters most: Action-relevance or cue-continuity?" *Neuropsychologia*, vol. 87, pp. 54–62, jul 2016.

[24] B. Takač, A. Català, D. Rodríguez Martín, N. Van Der Aa, W. Chen, and M. Rauterberg, "Position and orientation tracking in a ubiquitous monitoring system for Parkinson disease patients with freezing of gait symptom," *Journal of Medical Internet Research*, vol. 15, no. 7, p. e14, jul 2013.

[25] M. V. Albert, S. Toledo, M. Shapiro, and K. Kording, "Using Mobile Phones for Activity Recognition in Parkinson's Patients," *Frontiers in Neurology*, vol. 3, p. 158, nov 2012.

[26] A. Fisher, M. Memo, F. Stocchi, and I. Hanin, *Advances in Alzheimer's and Parkinson's Disease*, ser. Advances in Behavioral Biology. Boston, MA: Springer US, 2008, vol. 57.

[27] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical Human Activity Recognition Using Wearable Sensors," *Sensors*, vol. 15, no. 12, pp. 31 314–31 338, dec 2015.

[28] A. L. Silva de Lima, L. J. W. Evers, T. Hahn, L. Bataille, J. L. Hamilton, M. A. Little, Y. Okuma, B. R. Bloem, and M. J. Faber, "Freezing of gait and fall detection in Parkinson's disease using wearable sensors: a systematic review," *Journal of Neurology*, vol. 264, no. 8, pp. 1642–1654, aug 2017.

[29] N. Kern, B. Schiele, and A. Schmidt, "Multi-sensor Activity Context Detection for Wearable Computing," in *Ambient Intelligence*. Springer, Berlin, Heidelberg, 2003, pp. 220–232.

[30] D. Janssens, G. Wets, T. Brijs, K. Vanhoof, T. Arentze, and H. Timmermans, "Integrating Bayesian networks and decision trees in a sequential rule-based transportation model," *European Journal of Operational Research*, vol. 175, no. 1, pp. 16–34, nov 2006.

[31] M. J. Mathie, B. G. Celler, N. H. Lovell, and A. C. F. Coster, "Classification of basic daily movements using a triaxial accelerometer," *Medical & Biological Engineering & Computing*, vol. 42, no. 5, pp. 679–687, sep 2004.

[32] J. Skotte, M. Korshøj, J. Kristiansen, C. Hanisch, and A. Holtermann, "Detection of Physical Activity Types Using Triaxial Accelerometers," *Journal of Physical Activity and Health*, vol. 11, no. 1, pp. 76–84, jan 2014.

[33] A. M. Khan, Y. K. Lee, S. Y. Lee, and T. S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1166–1172, sep 2010.

[34] M. Demonceau, A.-F. Donneau, J.-L. Croisier, E. Skawiniak, M. Boutaayamou, D. Maquet, and G. Garraux, "Contribution of a Trunk Accelerometer System to the Characterization of Gait in Patients With Mild-to-Moderate Parkinson's Disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1803–1808, nov 2015.

[35] H. Nguyen, K. Lebel, P. Boissy, S. Bogard, E. Goubault, and C. Duval, "Auto detection and segmentation of daily living activities during a Timed Up and Go task in people with Parkinson's disease using multiple inertial sensors," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 1, p. 26, dec 2017.

[36] J. Cancela, M. Pastorino, M. T. Arredondo, M. Pansera, L. Pastor-Sanz, F. Villagra, M. A. Pastor, and A. P. Gonzalez, "Gait assessment in Parkinson's disease patients through a network of wearable accelerometers in unsupervised environments," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, aug 2011, pp. 2233–2236.

[37] V. H. Cheung, L. Gray, and M. Karunanithi, "Review of Accelerometry for Determining Daily Activity Among Elderly Patients," *Archives of Physical Medicine and Rehabilitation*, vol. 92, no. 6, pp. 998–1014, jun 2011.

[38] M. Suzuki, H. Mitoma, and M. Yoneyama, "Quantitative Analysis of Motor Status in Parkinson's Disease Using Wearable Devices: From Methodological Considerations to Problems in Clinical Applications." *Parkinson's disease*, vol. 2017, p. 6139716, 2017.

[39] C. Punin, B. Barzallo, R. Clotet, A. Bermeo, M. Bravo, J. Bermeo, and C. Llumiguano, "A Non-Invasive Medical Device for Parkinson's Patients with Episodes of Freezing of Gait," *Sensors*, vol. 19, no. 3, p. 737, feb 2019.

[40] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. Hausdorff, N. Giladi, and G. Troster, "Wearable Assistant for Parkinson's Disease Patients With the Freezing of Gait Symptom," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 436–446, mar 2010.

[41] S. Mazilu, M. Hardegger, Z. Zhu, D. Roggen, G. Troester, M. Plotnik, and J. Hausdorff, "Online Detection of Freezing of Gait with Smartphones and Machine Learning Techniques," in *Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare*. IEEE, 2012.

[42] D. M. Karantonis, S. Member, M. R. Narayanan, M. Mathie, N. H. Lovell, S. Member, and B. G. Celler, "Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring," *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, vol. 10, no. 1, 2006.

[43] M. Kuhn and K. Johnson, *Applied predictive modeling*. Springer New York, jan 2013.

[44] D. Rodríguez-Martín, A. Samà, C. Pérez-López, A. Català, J. M. Moreno Arostegui, J. Cabestany, À. Bayés, S. Alcaine, B. Mestre, A. Prats, M. C. Crespo, T. J. Counihan, P. Browne, L. R. Quinlan, G. ÓLaighin, D. Sweeney, H. Lewy, J. Azuri, G. Vainstein, R. Annicchiarico, A. Costa, and A. Rodríguez-Molinero, "Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer," *PLOS ONE*, vol. 12, no. 2, p. e0171764, feb 2017.

[45] "Naive Bayes classification - MATLAB - MathWorks Benelux." [Online]. Available: https://nl.mathworks.com/help/stats/classificationnaivebayes-class.html

[46] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *Proceedings - International Symposium on Wearable Computers, ISWC*, 2012, pp. 108–109.

[47] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-Aware Human Activity Recognition Using Smartphones," *Neurocomputing*, vol. 171, pp. 754–767, jan 2016.

[48] O. Baños, M. Damas, H. Pomares, I. Rojas, M. A. Tóth, and O. Amft, "A benchmark dataset to evaluate sensor displacement in activity recognition," in *Proceedings of the 2012 ACM*

Conference on Ubiquitous Computing - UbiComp '12. New York, New York, USA: ACM Press, 2012, p. 1026.

[49] A. Nieuwboer, F. Chavret, A.-M. Willems, and K. Desloovere, "Does freezing in Parkinson's disease change limb coordination?" *Journal of Neurology*, vol. 254, no. 9, pp. 1268–1277, sep 2007.

[50] M. Zhang, "A feature selection-based framework for human activity recognition using wearable multimodal sensors — Proceedings of the 6th International Conference on Body Area Networks," in *A feature selection-based framework for human activity recognition using wearable multimodal sensors*, 2011.

[51] A. Bayat, M. Pomplun, and D. A. Tran, "A study on human activity recognition using accelerometer data from smartphones," in *Procedia Computer Science*, vol. 34. Elsevier B.V., 2014, pp. 450–457.

[52] H. Vähä-Ypyä, P. Husu, J. Suni, T. Vasankari, and H. Sievänen, "Reliable recognition of lying, sitting, and standing with a hip-worn accelerometer," *Scandinavian Journal of Medicine & Science in Sports*, vol. 28, no. 3, pp. 1092–1102, mar 2018.

[53] M. Nimwegen, A. D. Speelman, E. J. M. Hofman-van Rossum, S. Overeem, D. J. H. Deeg, G. F. Borm, M. H. L. Horst, B. R. Bloem, and M. Munneke, "Physical inactivity in Parkinson's disease," *Journal of Neurology*, vol. 258, no. 12, pp. 2214–2221, dec 2011.

[54] F. B. Horak, M. Mancini, P. Carlson-Kuhta, J. G. Nutt, and A. Salarian, "Balance and Gait Represent Independent Domains of Mobility in Parkinson Disease." *Physical therapy*, vol. 96, no. 9, pp. 1364–71, 2016.

[55] A. Millecamps, K. A. Lowry, J. S. Brach, S. Perera, M. S. Redfern, and E. Sejdić, "Understanding the effects of preprocessing on extracted signal features from gait accelerometry signals." *Computers in biology and medicine*, vol. 62, pp. 164–74, jul 2015.

[56] R. San-Segundo, R. Torres-Sánchez, J. Hodgins, F. De la Torre, F. D. la Torre, R. San-Segundo, H. Navarro-Hellín, R. Torres-Sánchez, J. Hodgins, and F. De la Torre, "Increasing Robustness in the Detection of Freezing of Gait in Parkinson's Disease," *Electronics*, vol. 8, no. 2, p. 119, jan 2019.

[57] N. A. Capela, E. D. Lemaire, and N. Baddour, "Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients." *PloS one*, vol. 10, no. 4, p. e0124414, 2015.

[58] V. Lugade, E. Fortune, M. Morrow, and K. Kaufman, "Validity of using tri-axial accelerometers to measure human movement—Part I: Posture and movement detection," *Medical Engineering & Physics*, vol. 36, no. 2, pp. 169–176, feb 2014.

[59] L. Atallah, B. Lo, R. King, and G. Z. Yang, "Sensor positioning for activity recognition using wearable accelerometers," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 5, no. 4, aug 2011, pp. 320–329.

[60] S. T. Moore, T. T. Pham, S. J. G. Leong, S. J. G. Lewis, D. N. Nguyen, and E. D. A. J. F. v. d. . A. L. M. . P. H. Dutkiewicz, "Freezing of Gait Detection in Parkinson's Disease: A Subject-Independent Detector Using Anomaly Scores," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 11, pp. 2719–2728, nov 2017.

[61] P. Gupta and T. Dallas, "Feature selection and activity recognition system using a single triaxial accelerometer," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1780–1786, 2014.

[62] U. Maurer, A. Rowe, A. Smailagic, and D. Siewiorek, "Location and activity recognition using eWatch: A wearable sensor platform," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3864 LNAI. Springer Verlag, 2006, pp. 86–102.

[63] A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin, "Direct measurement of human movement by accelerometry," *Medical Engineering & Physics*, vol. 30, no. 10, pp. 1364–1386, dec 2008.

[64] C. C. Hsu, Y. P. Huang, and K. W. Chang, "Extended Naive Bayes classifier for mixed data," *Expert Systems with Applications*, vol. 35, no. 3, pp. 1080–1083, oct 2008.

[65] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato, "Monitoring Motor Fluctuations in Patients With Parkinson's Disease Using Wearable Sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 864–873, nov 2009.

[66] E. Rovini, C. Maremmani, and F. Cavallo, "How Wearable Sensors Can Support Parkinson's Disease Diagnosis and Treatment: A Systematic Review," *Frontiers in Neuroscience*, vol. 11, p. 555, oct 2017.

[67] T. T. Pham, D. N. Nguyen, E. Dutkiewicz, A. L. McEwan, and P. H. Leong, "Wearable healthcare systems: A single channel accelerometer based anomaly detector for studies of gait freezing in Parkinson's disease," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, may 2017, pp. 1–5.

[68] M. Bächlin, D. Roggen, M. Plotnik, J. M. Hausdorff, N. Giladi, and G. Tröster, "Online Detection of Freezing of Gait in Parkinson's Disease Patients: A Performance Characterization," in *Proceedings of the 4th International ICST Conference on Body Area Networks*. ICST, 2009, p. 11.

[69] L. Gao, A. Bourke, and J. Nelson, "Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems," *Medical Engineering & Physics*, vol. 36, no. 6, pp. 779–785, jun 2014.

Appendix A
Predictor Characteristics and Prevalence in Combinations Obtaining the Best Classification Results

Table A.1: Characteristics of the fourteen predictors with highest feature selection scores for all nodes. Y-axis predictors are numerous at node 3 (classification of ambulatory tasks) and at node 4 (classification of static tasks). Z-axis predictors were among the highest scoring predictors at node 1 (separating static and dynamic tasks) and node 2 (seperating cycling and ambulatory tasks). Frequency domain predictors resulted in high feature selection scores, especially at node 1.

Predictor type	Node 1	Node 2	Node 3	Node 4	Total
3D	1	1	1	2	5
x-axis	6	6	3	5	20
y-axis	2	1	10	6	19
z-axis	5	6	0	1	12
Time domain	7	10	9	12	38
Frequency domain	7	4	5	2	18

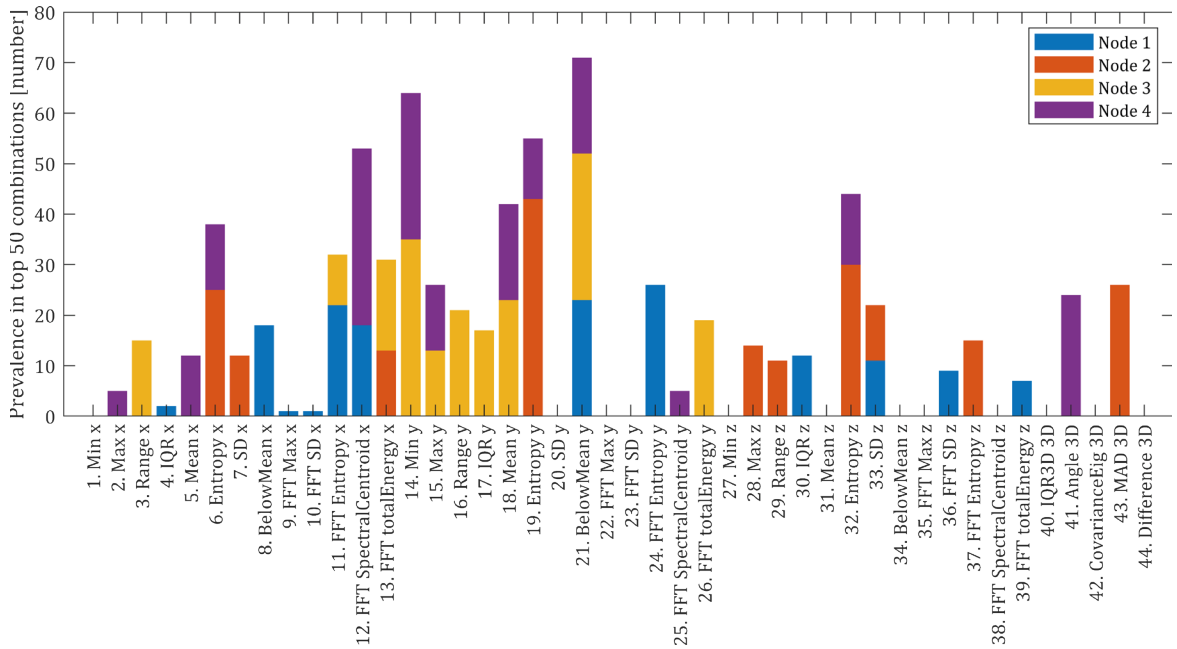


Figure A.1: Prevalence of predictors in the 50 predictor combinations which obtained the highest sensitivity per node. At node 3 (classification of specific ambulatory tasks), Y-axis predictors were the most prevalent. Z-axis predictors were prevalent at node 1 (classification of static and dynamic tasks) and node 2 (classification of cycling versus ambulatory tasks). Furthermore, this figure illustrates that certain predictors show high prevalence at a certain node, while not being prevalent at other nodes. This is where a decision tree model is advantageous: specific predictors can be used for specific classification steps.

Appendix B
Predictor Combinations used in the Decision Tree Model for both Healthy Subjects and PD Patients

Table B.1: Predictor vectors resulting in the highest classification sensitivity per node. These predictor combinations were used at their respective nodes to train and test the complete decision tree model.

Node 1	Node 2
FFT SpectralCentroid x	MAD 3D
IQR z	Entropy y
FFT Entropy y	FFT Entropy z
	Entropy z
Node 3	**Node 4**
Range y	Min y
FFT totalEnergy x	BelowMean y
BelowMean y	Max y
IQR y	FFT SpectralCentroid x

Table B.2: Five predictor vectors resulting in the highest classification sensitivity per node in healthy subjects. The combinations were subsequently used to validate the model on PD patients.

Node 1	Node 2	Node 3	Node 4
FFT SpectralCentroid x	MAD 3D	Range y	Min y
IQR z	Entropy y	FFT totalEnergy x	BelowMean y
FFT Entropy y	FFT Entropy z	BelowMean y	Max y
	Entropy z	IQR y	FFT SpectralCentroid x
SD z	MAD 3D	Min y	Mean y
FFT SpectralCentroid x	Entropy y	FFT totalEnergy x	BelowMean y
FFT Entropy y	Entropy x	BelowMean y	Max y
	Range z	IQR y	FFT SpectralCentroid x
BelowMean y	MAD 3D	Min y	Min y
IQR z	Entropy y	Mean y	BelowMean y
FFT Entropy y	FFT Entropy z	FFT totalEnergy x	Entropy z
	Entropy x	IQR y	FFT SpectralCentroid x
SD z	MAD 3D	Min y	Min y
BelowMean y	Entropy y	Mean y	BelowMean y
FFT Entropy y	Entropy z	FFT totalEnergy x	Entropy x
	Range z	FFT totalEnergy y	FFT SpectralCentroid x
FFT SpectralCentroid x	MAD 3D	Min y	Mean y
FFT Entropy x	Entropy y	FFT totalEnergy x	Min y
IQR z	Entropy z	FFT totalEnergy y	BelowMean y
	FFT totalEnergy x	BelowMean y	FFT SpectralCentroid x

Table B.3: Predictor vectors resulting in the highest classification sensitivity per node when the proposed decision tree model was transferred from healthy subjects to PD patients. As can be seen when comparing to Table B.1, most predictors are equal to the ones used for healthy subjects.

Node 1	Node 2
FFT SpectralCentroid x	MAD 3D
FFT Entropy x	Entropy y
IQR z	Entropy z
	Range z
Node 3	**Node 4**
Min y	Mean y
FFT totalEnergy x	BelowMean y
FFT totalEnergy y	Max y
BelowMean y	FFT SpectralCentroid x

Appendix C
Confusion matrices of complete classification

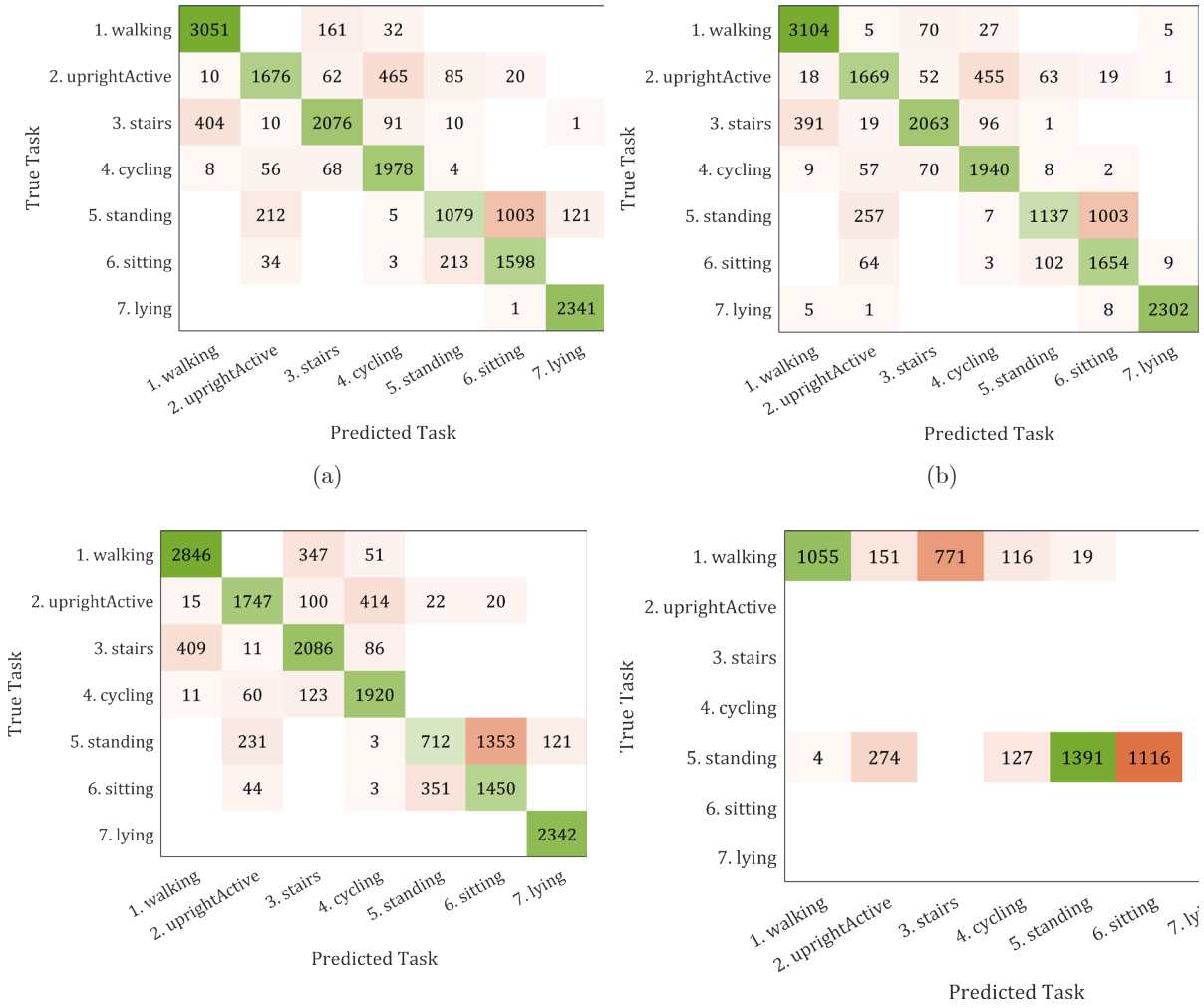


Figure B.1: Confusion matrices showing classification results. The rows show the true class of each instance (window) that was classified. (a) Results of the proposed decision tree method using Bayesian decision node classifiers and leave-one-out cross-validation. (b) Results of the proposed decision tree method using Bayesian decision node classifiers and Monte-Carlo cross-validation. (c) Results of using the Naive Bayes classifier at the decision nodes. (d) Classification results of the proposed decision tree model using dataset 1 as training data and dataset 2 as test data.

Appendix D
Classification Results

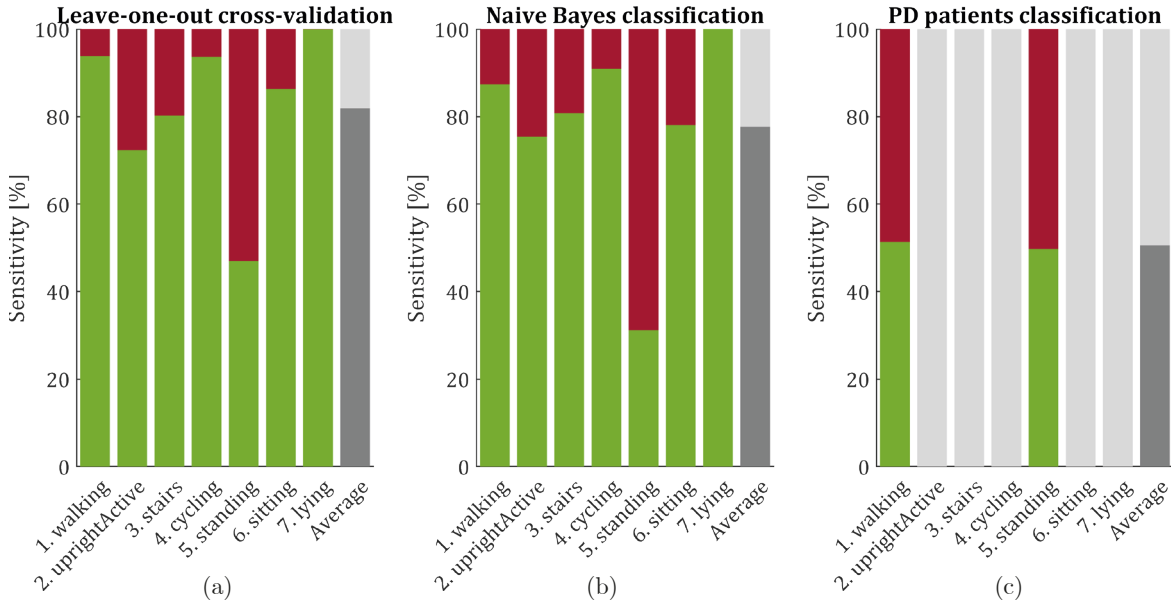


Figure D.1: A visual representation of the classification resuls. (a) Results of the proposed decision tree method using Bayesian decision node classifiers and leave-one-out cross-validation. (b) Results of using the Naive Bayes classifier at the decision nodes. (c) Classification results of the proposed decision tree model using dataset 1 as training data and dataset 2 as test data.

Closing Remarks

The result of this thesis is a solid foundation for a task classification model for PD patients. Since this model has not been trained on PD patients, classification results were not satisfactory: the transferability of the model from healthy individuals to PD patients was low. However, upon training this model with data obtained from PD patients, I believe that the model will perform adequately. To increase sensitivity, the model can even be trained on the patients for whom it will be used. I am certain this will provide Cue2Walk with an effective solution to the misclassifications in FoG detection. Knowing that in the near future, I contribute to helping PD patients to walk more confidently, and to live a more active life, I truly feel like the Biomedical Engineer that I (almost) am.

Sources - Report

[1] RIVM: Bilthoven, "Volksgezondheidenzorg.info." [Online]. Available: www.volksgezondheidenzorg.info/onderwerp/ziekte-van-parkinson/cijfers-context/huidige-situatie#node-prevalentie-ziekte-van-parkinson-huisartsenzorg. [Accessed: 28-Apr-2020].

[2] C. Punin et al., "A Non-Invasive Medical Device for Parkinson's Patients with Episodes of Freezing of Gait," Sensors, vol. 19, no. 3, p. 737, Feb. 2019.

[3] A. E. Lang and A. M. Lozano, "Parkinson's Disease," N. Engl. J. Med., vol. 339, no. 15, pp. 1044–1053, Oct. 1998.

[4] D. M. Tan, J. L. McGinley, M. E. Danoudis, R. Iansek, and M. E. Morris, "Freezing of Gait and Activity Limitations in People With Parkinson's Disease," Arch. Phys. Med. Rehabil., vol. 92, no. 7, pp. 1159–1165, Jul. 2011.

[5] J. G. Nutt, B. R. Bloem, N. Giladi, M. Hallett, F. B. Horak, and A. Nieuwboer, "Freezing of gait: moving forward on a mysterious clinical phenomenon," Lancet Neurol., vol. 10, no. 8, pp. 734–744, Aug. 2011.

[6] D. Sweeney, L. Quinlan, P. Browne, M. Richardson, P. Meskell, and G. ÓLaighin, "A Technological Review of Wearable Cueing Devices Addressing Freezing of Gait in Parkinson's Disease," Sensors, vol. 19, no. 6, p. 1277, Mar. 2019.

[7] J.-P. Azulay, S. Mesure, and O. Blin, "Influence of visual cues on gait in Parkinson's disease: Contribution to attention or sensory dependence?," J. Neurol. Sci., vol. 248, no. 1–2, pp. 192–195, Oct. 2006.

[8] M. H. Thaut, G. C. McIntosh, R. R. Rice, R. A. Miller, J. Rathbun, and J. M. Brault, "Rhythmic auditory stimulation in gait training for Parkinson's disease patients," Mov. Disord., vol. 11, no. 2, pp. 193–200, Mar. 1996.

[9] P. J. McCandless, B. J. Evans, J. Janssen, J. Selfe, A. Churchill, and J. Richards, "Effect of three cueing devices for people with Parkinson's disease with gait initiation difficulties," Gait Posture, vol. 44, pp. 7–11, Feb. 2016.

[10] J. M. Hausdorff, Y. Balash, and N. Giladi, "Time series analysis of leg movements during freezing of gait in Parkinson's disease: akinesia, rhyme or reason?," Phys. A Stat. Mech. its Appl., vol. 321, no. 3–4, pp. 565–570, Apr. 2003.

[11] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in Parkinson's disease," J. Neurosci. Methods, vol. 167, no. 2, pp. 340–348, Jan. 2008.

[12] M. Bächlin, D. Roggen, M. Plotnik, J. M. Hausdorff, N. Giladi, and G. Tröster, "Online Detection of Freezing of Gait in Parkinson's Disease Patients: A Performance Characterization," in Proceedings of the 4th International ICST Conference on Body Area Networks, 2009, p. 11.

[13] E. Jovanov, E. Wang, L. Verhagen, M. Fredrickson, and R. Fratangelo, "deFOG — A real time system for detection and unfreezing of gait of Parkinson's patients," in 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 5151–5154.

[14] C. A. Coste, B. Sijobert, R. Pissard-Gibollet, M. Pasquier, B. Espiau, and C. Geny, "Detection of freezing of gait in Parkinson disease: preliminary results.," Sensors (Basel)., vol. 14, no. 4, pp. 6819–27, Apr. 2014.

[15] "Parkinson Buddy ready for sale - 2M ENGINEERING." [Online]. Available: https://www.2mel.nl/parkinson-buddy-ready-for-sale/. [Accessed: 06-Jun-2020].

[16] "Parkinson Smartwatch BV - Parkinson Smartwatch." [Online]. Available: https://parkinsonsmartwatch.com/nl/. [Accessed: 06-Jun-2020].

[17] "Objective Parkinson's Assessment: Wearables, Ambulatory Monitoring, & Telemedicine - Kinesia." [Online]. Available: http://glneurotech.com/kinesia/. [Accessed: 06-Jun-2020].

[18] "Improve walking for people with Parkinson's disease." [Online]. Available: http://medigait.com/pd/. [Accessed: 06-Jun-2020].

[19] T. Ellis and L. Rochester, "Mobilizing Parkinson's disease: The future of exercise," Journal of Parkinson's Disease, vol. 8, no. s1. IOS Press, pp. S95–S100, 2018.

[20] T. Paillard, Y. Rolland, and P. S. de Barreto, "Protective effects of physical exercise in Alzheimer's disease and Parkinson's disease: A narrative review," Journal of Clinical Neurology (Korea), vol. 11, no. 3. Korean Neurological Association, pp. 212–219, 01-Jul-2015.

[21] "Documentation – MbientLab." [Online]. Available: https://mbientlab.com/documentation/.