Universiteit Leiden

Master Computer Science

Quantification of Parkinson's disease symptoms during activities of daily living using accelerometry-based wearable sensors

Name:	Raffi Mirzoyan
Student ID:	2902869
Date:	18/7/2022
Specialisation:	Artificial Intelligence
1st supervisor:	Mitra Baratchi
2nd supervisor:	Vasileios Exadaktylos

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Contents

1 Introduction **5**

2 Related Work **8**
2.1 Quantification of Parkinson's disease 8
2.2 Machine learning and Parkinson's disease 9
2.3 Limitations and Contributions 10

3 Methods **11**
3.1 Data 11
3.2 Pre-Processing 15
3.3 Features 22
3.4 Model Training and Validation 23

4 Results **25**
4.1 Results - Random Forest 26
4.2 Results - Decision Trees 32
4.3 Results - Review 38

5 Discussion **40**
5.1 Tremor 40
5.2 Dyskinesia 41
5.3 Bradykinesia 42
5.4 Summary 43
5.5 Limitations and Future Work 44

6 Conclusion **45**

Abstract

Quantifying Parkinson's disease (PD) symptoms' presence is challenging as it may manifest differently depending on various factors. With the use of modern wearable sensors and machine learning techniques, it has become easier to monitor motor fluctuations during the daily lives of Parkinson's disease patients. This work identifies which sensor placement and daily life activity combinations better predict Parkinson's disease symptoms. A wide range of results was achieved by implementing Random Forest and Decision Tree classifiers using the data from a wearable smartwatch and a smartphone, and the results are compared. Optimal placements of a single wearable sensor and daily life activity combinations per symptom are proposed for quantifying Parkinson's disease symptoms. In some cases, average accuracy scores of up to 75% were achieved, while in other cases, one sensor proved insufficient for objective outcomes. The results were validated with a grouped nested 10-fold cross-validation approach.

1 Introduction

Movement is one of the most important aspects of our nervous system. Motor neurons are among the largest neurons in the central nervous system and are the final common pathway through which the brain controls all of our skeletal muscle movement [25]. Due to our complex nervous system, humans can achieve far more variation in movement, and different types and speeds of movements than any other animal in the animal kingdom can perform. These nerve cells, or neurons, produce an important brain chemical called dopamine. Dopamine drives us to do the things that we do. It is closely associated with motivation, desire, and craving, and it relates to satisfaction and our feelings of well-being. However, our nervous system is not perfect. Diseases such as depression, schizophrenia, and Parkinson's disease are linked to, among other things, extreme changes in dopamine levels in the brain. On one hand, schizophrenia involves elevated levels of dopamine. On the other hand, Parkinson's disease is caused by a loss of dopaminergic neurons in the part of the brain called the substantia nigra [14].

Parkinson's disease (PD) is a neurological disorder characterized by uncontrollable or unintended movements, as well as complications in the mental health. People initially start to quake, cannot generate smooth movements, have issues with speed, and lose the sense of coordination and balance. It is a progressive disease caused mainly by a lack of dopamine in the brain. People with Parkinson's (PwP) disease also develop behavioral and mental issues such as depression, anxiety, fatigue, sleep problems, and memory difficulties. Parkinson's disease is progressive in nature and gets worse over time, affecting more than 10 million people globally. People with Parkinson's disease may not always get all the symptoms; it is very challenging to identify how bad the symptoms will be and how fast the disease will progress. It is hard to pinpoint precisely how it will progress as the symptoms vary tremendously from person to person, and early signs may be unnoticeable. For example, some people may only have minor shaking in their hands (tremor) which does not bother them in their daily life activities, whereas others may be affected so much that they cannot drink a glass of water on their own. The same issue relates to non-motor symptoms. Some individuals may have severe tremors but no issues with memory or thinking, while others may not have any motor movement complications but suffer from dementia. Overall, there are many gray areas when defining Parkinson's disease, but the usual motor symptoms include tremor, bradykinesia, dyskinesia, gait impediments, posture instability, freezing of gait, and rigidity.

A tremor (shaking) is an involuntary movement or twitching of a limb. Tremor is rhythmic to a certain degree and often begins in a limb such as a hand or fingers, but it can also affect the chin, lips, face, and legs [20]. There are two types of tremor — resting tremor and action tremor. Resting tremor occurs in a body part that is entirely supported against gravity, for example, someone sitting while resting their hands on the arms of the couch. On the other hand, an action tremor arises when an individual is trying to contract a muscle voluntarily and includes postural and kinetic tremors. A postural tremor is any tremor that is produced while a person is trying to maintain a position, such as standing still. Moreover, kinetic tremor occurs during any voluntary movement and may include guided movements, which may turn into a task-specific tremor [4]. Task-specific tremors are essentially kinetic tremors that usually appear during specific activities or movements.

Furthermore, tremor is not the only symptom that limits the movements of people with Parkinson's disease. One of the cardinal symptoms of Parkinson's disease diagnosis is bradykinesia, which is usually described as slowness of movements. Weakness, tremor, and rigidity may contribute to but do not fully explain bradykinesia [3]. In Parkinson's, the slowness or weakness may

happen in various ways, such as fewer facial expressions, difficulty initiating movements, and reduced automatic movements such as swinging the arms when walking. Moreover, bradykinesia is very unpredictable. It is possible to walk normally without any issues and then suddenly freeze or need help to move further. This happens because of overactivity in the lateral premotor areas during task performance, and movements can be speeded by giving sensory cues to people with severe bradykinesia [3]. It was also shown that there is a decrease in performance (slowness of movements or freezing) in patients with Parkinson's who may be executing two tasks simultaneously, especially fine motor movements[3].
Motor/non-motor fluctuations and dyskinesia are common serious late effects of dopaminergic therapy in PD. Most people at mid-and advanced stages of the disease experience these motor complications. Dyskinesias are abnormal, involuntary movements of different body parts that, in many cases, occur during the peak effect when patients take Levodopa or any medication of kind [16]. It is a byproduct of using Levodopa to help alleviate tremor severity. Doctors usually refer to it as ON state when someone takes such medication and OFF state if not taken anything. Almost all people with Parkinson's disease take some derivative of the medication to help them with the symptoms. However, dosing the medication is very tricky and often leads to having too much dopamine in the brain, which leads to dyskinesia. After several years of smooth and stable response to Levodopa, motor fluctuations from on to off states are experienced as Levodopa wears off between doses and the PD symptoms reappear. Levodopa is a prodrug that is converted to dopamine by DOPA decarboxylase and can cross the blood-brain barrier. Levodopa still remains the gold standard for the treatment of motor symptoms of PD in advanced stage [41].

Making an accurate diagnosis of Parkinson's disease can be complicated because of the complexity of the symptoms. No specific test (e.g., a blood test) can give a definite result. Instead, the physical symptoms mentioned above must be present to diagnose the disease and its severity, and doctors heavily rely on manual visual examinations and assessment procedures. In the early 19th century, James Parkinson published his essay on "The Shaking Palsy," in which he described in detail the clinical features of what is now known the Parkinson's disease. He mentions that the disease is of long duration and that long-term follow-ups are needed to accurately detect it and estimate the severity [30]. However, no formal diagnostic guidelines were proposed before the last three decades. Subsequently, United Kingdom Parkinson's Disease Society Brain Bank[1] and other diagnostic criteria were introduced later in the hopes of ameliorating the diagnostic accuracy. Giovanni Rizzo et al. published a meta-analysis in 2016 with the primary objective of evaluating the diagnostic accuracy of Parkinson's disease [37]. They looked into published documents during 1988-2014 and selected 20 studies that report diagnostic parameters regarding clinical diagnosis of PD. The authors concluded that the overall validity of clinical diagnosis accuracy did not significantly improve in the last 25 years.

Conversely, The International Parkinson and Movement Disorder Society (MDS)[2] published clinical diagnostic criteria to measure the disease severity and named it the Unified Parkinson's Disease Rating Scale (UPDRS). The UPDRS includes a questionnaire consisting of 65 items and has four parts: non-motor experiences of daily living, namely sleep problems, mood swings, fatigue along with others; motor experiences of daily living, such as speech, eating, dressing, and so on; motor examination, such as toe-tapping, arising from a chair, gait

[1] Parkinson's UK Brain Bank, *https://www.parkinsons.org.uk/research/parkinsons-uk-brain-bank*

[2] International Parkinson and Movement Disorder Society (MDS), *https://www.movementdisorders.org/*

and more; and motor complications such as time spend with dyskinesias, time spent in the OFF state, and so forth [17]. All tasks and questions are scored in the range of 0-4: 0 for normal (no symptom/problem), 1 for minimal problems, 2 for mild problems, 3 for moderate problems, and 4 being severe, meaning the patient was unable to do some activity (for example wearing clothes or getting out of bed). Ultimately, doctors sum those scores for all questions and give the patients a final severity score. Ronald B Postuma et al. published a validation study showing that the MDS-UPDRS clinical diagnostic criteria demonstrated high sensitivity and specificity compared to the United Kingdom Brain Bank criteria [34].

As mentioned earlier, Parkinson's disease symptoms vary from person to person. Two people may have the same UPDRS score, yet it does not mean they have the same daily life challenges. On the one hand, doctors may be inattentive during manual examinations and miss out on visual cues that would otherwise identify a symptom. On the other hand, Parkinson's disease symptoms fluctuate based on many things such as medication intake, diet, good night's sleep, and overall stress. A patient going through a check-up may feel anxious and perform worse than the symptom during everyday living. Therefore, frequent and objective assessment of the symptoms is critical to manage the health of people with Parkinson's disease effectively. Most research focuses on identifying the symptoms during daily living with various machine learning techniques. Rightly so, since the best way of knowing how badly the disease affects people is to see if they have issues in their daily life. Nevertheless, people are different, and their daily activities may vary a lot, depending on where they live and what kind of lifestyle they have. Thus continuous, accessible, and long-term monitoring of the disease can be very helpful in obtaining information even from the very beginning of the disease. For this reason, growing research involves the assessment of the disease and its intensity through sensor devices, namely, smartwatches, smartphones, inertial measurement units (IMU), and gyroscopes.

The goal of this research is two folds. First, specific daily life activities are thoroughly analyzed, and two machine learning classifiers are implemented to determine which activities are good predictors of Parkinson's disease symptoms. The results are then compared and the best performing model is mentioned based on the accuracy over all activities. This was achieved by looking at activity-level data instead of unknown sensor data from the whole day. Secondly, the positioning of wearable sensors are investigated to learn its impacts on the accuracy of these predictions. To encapsulate this work, the results of both classifiers are compared — optimal placements of a single wearable sensor and daily life activity combinations per symptom are proposed for quantifying Parkinson's disease symptoms using a combination of signal processing and machine learning techniques.

The rest of this paper is structured as follows. Section 2 discusses related work, namely the literature surrounding sensor use in Parkinson's disease estimation, the most common machine learning techniques used and recent advancements, and the contributions of this research in more detail. Section 3 talks about the methods used in this research, namely, data description (3.1), data preprocessing (3.2), overview of the generated features (3.3), finishing with model training and internal validation of the machine learning techniques (3.4). Section 4 includes tables of the final results, and Section 5 discusses the findings and limitations of this research. Section 6 concludes the paper.

2 Related Work

2.1 Quantification of Parkinson's disease

Quantifying Parkinson's disease symptoms or severity has been around for a while. To this day, these assessments mainly happen during periodic in-person visits to the clinic. However, these in-clinic assessments can be subjective. Silvia Del Din and colleagues have done a rigorous review and analysis of the possibilities of free-living monitoring of Parkinson's disease [13]. Their research's main goal was to determine the feasibility of utilizing wearable technology and sensors in estimating and monitoring Parkinson's disease. They explicitly mention that this is important because symptoms are often triggered by task and free-living challenges that cannot be replicated in clinics or controlled environments.

The practicality of utilizing accelerometer data to assess severity and motor complications in patients with Parkinson's disease has been shown by multiple papers such as [43] [7] [28]. Surface EMG is the most common method for diagnosing tremor, but it is generally inappropriate for constant monitoring and lacks certainty and accuracy [42]. On the contrary, needle EMG is highly accurate yet costly and invasive [12]. Inertial sensing based on Micro-Electro-Mechanical Systems (MEMS) is also very common for measuring tremors; however, the accuracy that is achieved using MEMS is much lower than any type of EMG signal analysis [20]. MEMS sensors can achieve high levels of functionality and reliability, yet most MEMS devices contain movable parts, and long-term repeated use will decrease the device reliability [45].

A review published in 2010 compared the pros and cons of quantifying and monitoring tremor using the most commonly used sensors such as accelerometers, gyroscopes, video, EMGs (including surface, needle, fine-wire EMGs), and force transducers [19]. Conventional accelerometers are the most commonly used sensors because of their practical size, price, and ease of use [43]. For that reason, an increasing number of research works indicate that portable systems such as body sensors are helpful for GAIT analysis in medical applications, for example, monitoring patients' recovery after a treatment [28]. Jorge Cancela et al. showed the feasibility of using a Body Network Area (BAN) of wireless accelerometers in real environments [7]. The analysis mostly focused on validating the system performance of continuous gait monitoring of PD patients. They attached a gyroscope to the belt of patients and a set of four tri-axial accelerometers, one on each limb. They found that despite being able to achieve a good understanding of the pathology in the patients, context information is crucial for making accurate classifications. The lack of context makes working with free-living data a very challenging task, and there is a need for more accurate activity recognizers.

While sensor usage can be beneficial, most studies focusing on symptom estimation and severity assessment from sensors are applied in clinical environments, which may not be a true representation of everyday symptoms. However, Clinicians mostly rely on patient recalls for out-of-clinic symptoms [29], which is fault-prone, especially for motor fluctuations arising from medication intake [33]; self-reported data can be inaccurate, biased, and is hard to fill in every few hours. For example, a study focused on measuring patient activity, gait, and tremor inside and outside the clinic found that tremor was only present for 1.6 [0.4–5.9] hours (median [range]) per day in most-affected hands and only 0.5 [0.3–2.3] hours per day in less-affected hands [1]. Regardless, Christopher W Hess and colleagues recruited seven Parkinson's disease patients to look at the possibility of using inertial sensors for feedback on medication effects to examine rest, postural, and action tremors [20]. The authors validated their sensor-based method by comparing it to an EMG motion tracking system and concluded that sensor-based

technologies could achieve reasonable tremor quantification and severity monitoring.

2.2 Machine learning and Parkinson's disease

The analysis focused on continuous monitoring of patients performing a variety of free body movements observed that a good combination of machine learning algorithms and wearable sensors can be a good estimator of Parkinsonian tremors in their natural environment [22]. A pilot study done in 2009 showed the feasibility of using accelerometer data to estimate motor complications in Parkinson's disease [31]. They also showed how the window length and the selection of different motor tasks affect the estimates of clinical movement scores (UPDRS) derived via analysis of the accelerometer data. The study achieved this by using the data of 8 SHIMMER[3] accelerometers on the upper and lower extremities of patients (2 on each limb) and a support vector machine (SVM) classifier [39]. A recent longitudinal study tried to continuously track fluctuations in resting tremor and dyskinesia for Parkinson's disease patients by using only an Apple Watch[4] and mapping the sensor reading to the MDS-UPDRS ratings of these patients [35]. They employed data of 343 participants with PD, of which 225 were from a study lasting six months. It was found that this approach correlated greatly to clinical evaluations of tremor severity and was roughly accurate to expert ratings of dyskinesia presence during in-clinic tasks. However, the authors mention that the accuracy of severity assessment of the proposed monitoring system has limited use because the assessments are rated by the MDS-UPDRS scale.

Some researchers [18] [27] [38] have applied a linear regression model to estimate dyskinesia severity scores (MD-UPDRS or mAIMS) from daily life using the signals one or two worn sensors (wrist, ankle, wrist, and ankle) using various estimation intervals lasting from 5s to 30 minutes. Rodriguez-Molinero et al. [38] concluded that the magnitude of dyskinesia measured by a waist-worn sensor correlates well with that observed by a physician. Moreover, Mera et al. [27] suggest that only one wrist-worn sensor is enough to predict dyskinesia severity when the arms are resing or in extended positions. Despite all of them using a unified scoring system as their gold standard, they all came up with strong conclusions. Furthermore, a recent study published in 2021 showed promising results with a bidirectional LSTM [10] model estimation of dyskinesia severity scores by using readings of two IMU sensors (wrist, ipsilateral ankle) [23]. The authors showed the medication ON/OFF state effects on the total mAIMS score of various daily life activities. Another study proposed using statistical and spatiotemporal gait features from wearable sensors to automatically detect the medication (ON/OFF) state validated and found out that the Random Forest classifier [6] outperformed other classifiers [2]. The study was done on 20 PD patients with two knee-worn tri-axial accelerometer sensors who were asked to perform a walking task, once in the ON state and once in the OFF state. Furthermore, Hyoseon Jeon et al. ran a rigorous research on estimating tremor from only one wrist-worn smart watch, and found that the Decision Tree classifier showed the highest accuracy and the lowest RMSE compared to other classifiers [24].

[3] `SHIMMER Research Website`, *https://shimmersensing.com*
[4] `Apple Official Website`, *https://www.apple.com*

2.3 Limitations and Contributions

Robust accuracy and validity metrics for various types of signals have been reported throughout all reviewed studies. These findings allow for wearable technologies to be used in specific cases. However, testing and validating the developed algorithms and metrics on a big group of subjects has still not been done and is required to have fully robust applications. One limitation of previous research is the emphasis on a single motor complication or symptom. Moreover, most studies use the MDS-UPDRS or similar unified scoring ratings, yet these severity rating scales also include mental symptom scores, and sensors cannot accurately estimate these symptoms. It is well known that the MDS-UPDRS is not a uniformly reliable tool to measure the severity of motor and non-motor symptoms [36] [15]. All in all, it is worth mentioning that no single best-performing algorithm or the best placement of a sensor for overall good performance is possible — there are many variables in the disease as well as in the daily lives of people with Parkinson's. This work fills the gap of which sensor placement and daily life activity combinations are the better predictors of Parkinson's disease symptoms by using only one wearable sensor and symptom presence scores for each limb per activity rated by clinicians in real time.

3 Methods

3.1 Data

The Michael J. Fox Foundation[5] is an organization committed to finding a cure for Parkinson's disease by funding research and ensuring the development of improved therapies for people with the disease. A few years back, they supported conducting a study named The Levodopa Response Study, which aimed to collect wearable sensor data from individuals with Parkinson's Disease (PD). They have done this to estimate clinically relevant measures of the severity of Parkinson's disease symptoms, including motor symptom fluctuations. After the data collection the data was published for public use. It is freely available on Synapse[6], which is a collaborative, open-source research platform aiming to support scientific collaborations centered around shared biomedical data sets. The website is operated by the nonprofit organization Sage Bionetworks[7], which is partnering with institutions such as the U.S. National Heart, Lung, and Blood Institute[8], U.S. National Institute of Mental Health[9], Children's Tumor Foundation[10], and more.

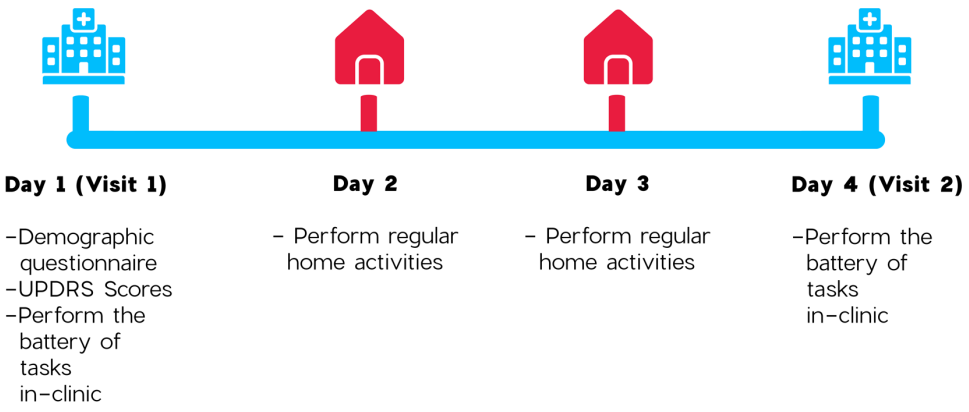


Figure 1: Timeline of the locations and relevant procedures during data collection.

The data was collected from two clinical sites — New York City and Boston. Subjects were monitored for four days, two days in-clinic, and two days at home. Figure 1 shows the timeline and the relevant activities performed during these days. On Day 1 of data collection (Visit 1), the participants went to the laboratory in an ON state for initial data collection, such as demographic and clinical data, then went through the MDS-UPDRS assessment. The study includes 28 participants fitted with either 3 or 8 sensors throughout their body. All subjects were fitted with the same three sensors; meanwhile, only a subset of participants had the five additional sensors. Since the study cohort had a small number of participants, this research looks at the data from the three sensors all participants wore during data collection. The participants were 19 male and 9 female Parkinson's disease patients, with an average age

[5] The Michael J. Fox Foundation for Parkinson's Research, *https://www.michaeljfox.org*
[6] Synapse, *https://www.synapse.org*
[7] Sage Bionetworks, *https://sagebionetworks.org*
[8] U.S. National Heart, Lung, and Blood Institute, *https://www.nhlbi.nih.gov*
[9] U.S. National Institute of Mental Health, *https://www.nimh.nih.gov*
[10] Children's Tumor Foundation, *https://www.ctf.org*

of 69.5 ± 8.96 (mean±std) and an average score of 2.25 ± 0.59 (mean±std) on the Hoehn and Yahr scale [21]. The initial data collection also includes a general symptom occurrence questionnaire, where patients specify which symptoms are usually present in daily life. Figure 2 shows the distribution of each symptom occurrence from the initial questionnaire and the number of patients who reported these symptoms.

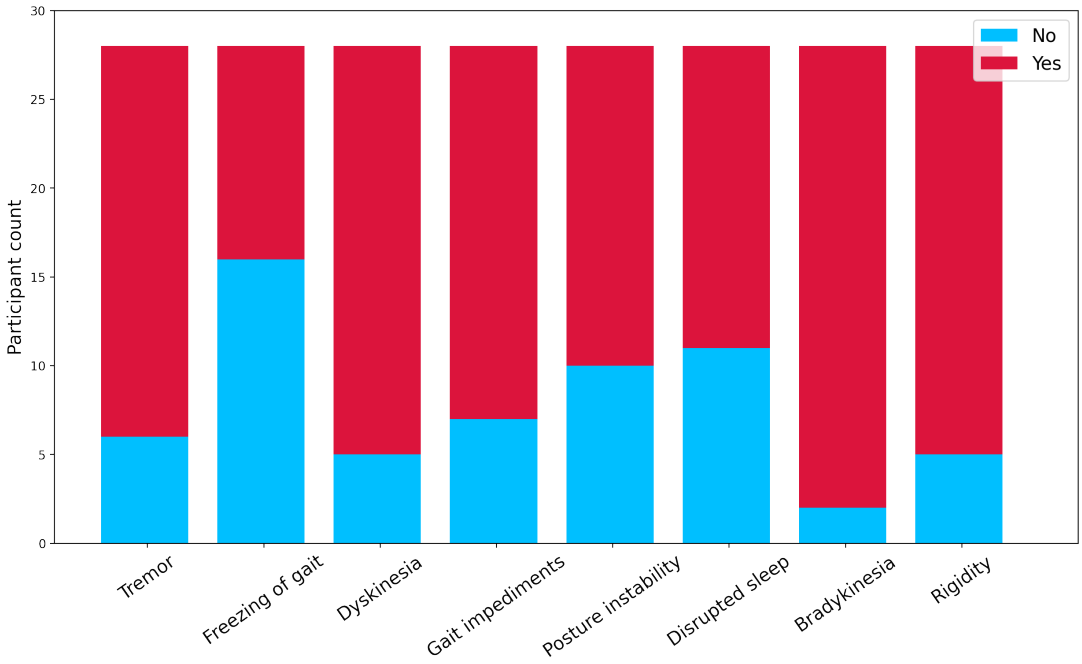


Figure 2: Symptom occurrence distribution among patients.

Afterwards, the participants were fitted with the sensors to start the in-clinic activities. These activities can be divided into 4 categories — functional, clinical, gross motor movements, and fine motor movements. Clinical tasks include standard tests such as a finger-to-nose test with each hand for 15s and alternating hand movements (repeated arm movement) with each hand for 15s. Functional movements include standing, sit to stand, walking in a straight line for 30s, walking in a straight line for 30s while counting backwards, and walking through a narrow corridor. Gross motor movements include pouring water from a bottle and drinking three times, arranging sheets of paper in a folder twice, and folding towels for three times. Fine motor movements include drawing on a piece of paper, typing on a computer keyboard for 30s, and assembling nuts and bolts for 30s. For three movements, sitting, going upstairs, and going downstairs, there were not enough repetitions and clinical labels for the symptoms, therefore the data for these movements were discarded in this study. The battery of tasks were completed one after another, which lasted around 20 minutes, then repeated 6-8 times (depending on the participant) at 30-minute intervals. After completing the first day of the set of tasks in the clinic, participants went home and conducted their usual unscripted daily activities without additional instructions from the doctors. On Day 4 (Visit 2), subjects visited the laboratory in an OFF state and performed the same battery of motor tasks in the same way as on Day 1 (Visit 1). After completing the set of tasks once, the subjects took their medication and continued performing the battery of tasks.

For all participants, limb-specific (right arm, left arm, lower limbs) scores for each task repetition were provided during both laboratory visits. Severity scores from 0-4 were provided

for tremor, whereas for bradykinesia and dyskinesia, only the presence or absence of the phenotype. The scores of this scale correlate with the MDS-UPDRS severity scoring system scale (0-4), with 0 being no symptoms and 4 being severe.

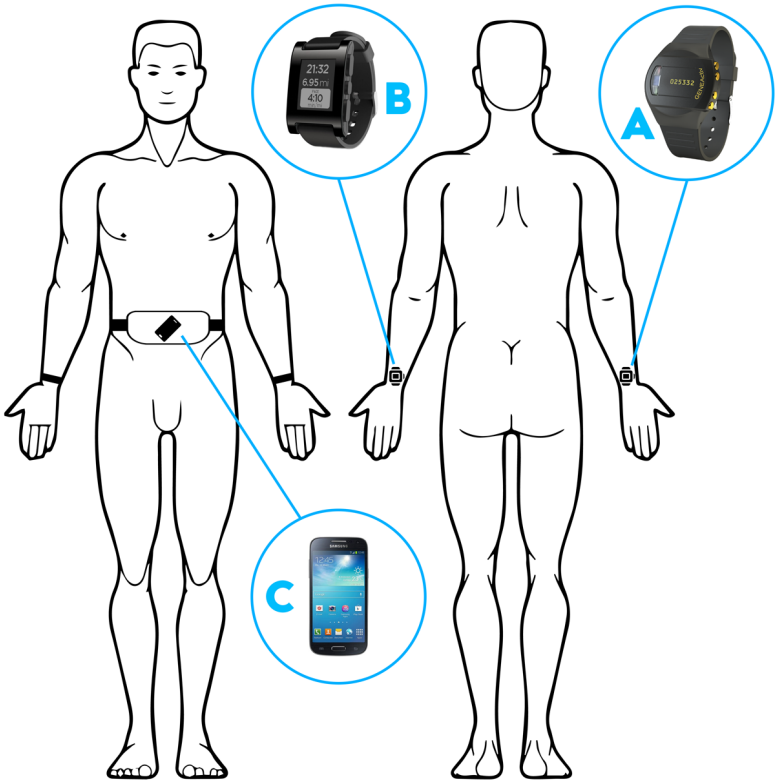


Figure 3: Sensor locations on participants. **A** - GENEActiv raw data accelerometer, **B** - Pebble Smartwatch, **C** - Samsung Galaxy Mini. **NOTE:** The sensors on left and right wrists were interchangeably used depending on the most affected side of the patient.

All participants wore the GENEActiv[11] raw data accelerometer on the wrist of the most affected arm, a Pebble[12] smartwatch on the wrist of the least affected arm, and a Samsung Galaxy Mini[13] smartphone in a fanny pack worn in front at the waist. Figure 3 shows the wearable sensors and their respective locations. The dataset includes raw signal files from these wearable sensors, separate for each day. All participants had the sensors attached to their bodies at all times while completing the set of tasks in the laboratory and at home. Sensor data includes the raw $X-Y-Z$ wearable acceleration signals in m/s^2, the magnitude (norm) of the signal, and timestamps in the UNIX time system. UNIX time system is the number of seconds that have passed since the UNIX epoch. The UNIX epoch is 00:00:00 UTC on 1 January 1970[14]. Signals were recorded in $50Hz$, meaning that the data records 50

[11] `GENEActiv Website`, *https://activinsights.com/technology/geneactiv/*
[12] `Pebble Smartwatch Wiki`, *https://en.wikipedia.org/wiki/Pebble_(watch)*
[13] `Samsung Official Galaxy Mini Overview`, *https://www.samsung.com/galaxyace/mini_overview.html*
[14] `UNIX Timestamp`, *https://www.unixtimestamp.com*

readings of the sensor for each second. In total, 12 raw files are accessible per participant — readings of 3 sensors over 4 days. Figure 4 shows an example of raw signals in $X - Y - Z$ axis and the magnitude of that signal from one of the samples of the "walking straight" task.

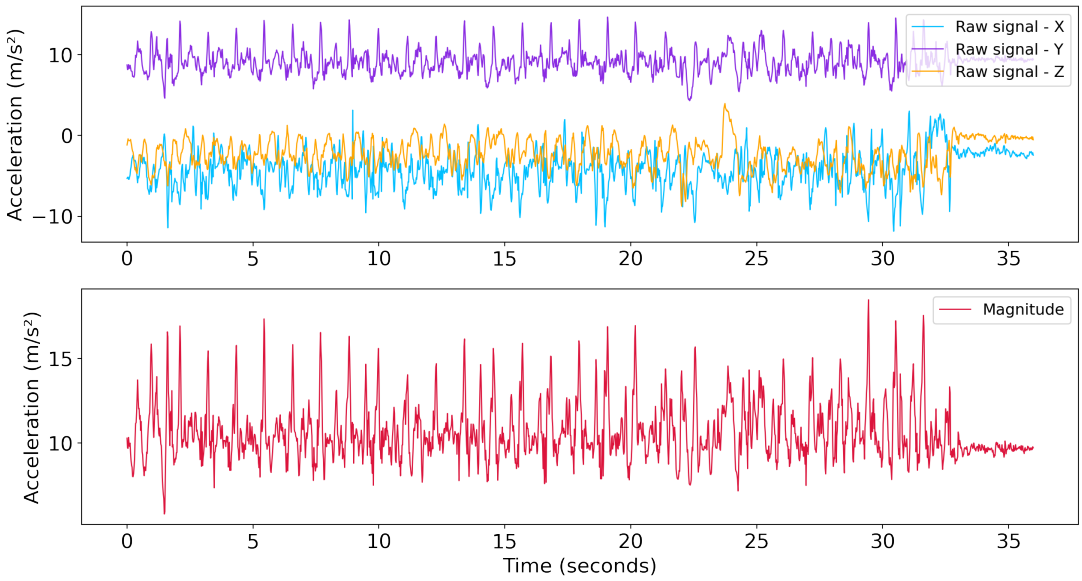


Figure 4: Raw acceleration signals from one instance of an in-clinic task. The top plot shows raw $X - Y - Z$ signals in blue, purple, and yellow, respectively; the bottom plot (red) shows the magnitude of those signals.

In order to label the raw accelerometer data, a task score sheet was provided, which includes subject IDs, visit, session and repetition numbers, task codes, timestamp start and end, and clinical labels for every phenotype (tremor, bradykinesia, dyskinesia) severity/presence for each limb. Table 1 shows a part of the task score sheet. Activity start and end times are given in the UNIX time system. Some of the information mentioned previously is not added in Table 1 to preserve space; however, the table is similar to the real dataset used in this analysis. The data in the task score table is randomized due to data privacy issues. A separate file of task code dictionary 2 is also provided.

Subject ID	Task Code	Timestamp_Start	Timestamp_End	Tremor
49_BOS	stndg	1003000344.96	1003000377.94	1
100_BOS	wlkgs	1040000416.96	1040000449.94	0
63_NYC	wlkgc	1000440499.96	1000440532.94	1
32_BOS	drawg	1000120696.96	1000120721.94	2
55_NYC	ntblt	1234000733.96	1234000766.94	0

Table 1: Sample task score sheet. **NOTE:** This sample table only replicates a part of the actual table.

Task Code	Task Name
ftnr	Finger to nose – right arm
ftnl	Finger to nose – left arm
ramr	Repeated arm movement – right arm
raml	Repeated arm movement – left arm
stndg	Standing
ststd	Sit to stand
wlkgs	Walking in a straight line
wlkgc	Walking in a straight line while counting backwards
wlkgp	Walking through a narrow passage
drnkg	Pouring water from a bottle and drinking
orgpa	Organizing sheets in a folder
fldng	Folding towels
drawg	Drawing and writing on a paper
typng	Typing on a computer keyboard
ntblt	Assembling nuts and bolts

Table 2: Task code dictionary. The rules in between the activities show the different categories of the tasks.

3.2 Pre-Processing

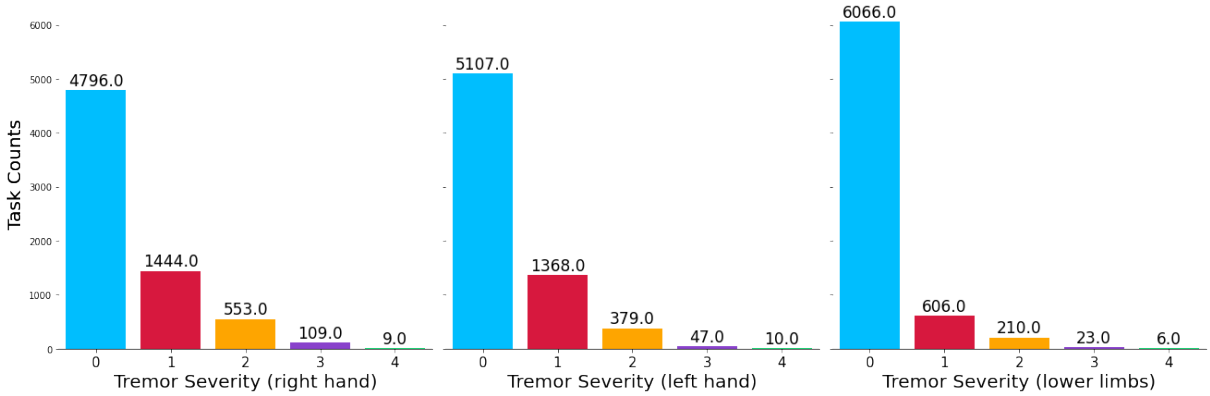
For each subject and repetition of each activity, the raw accelerometer signals were timestamped and extracted per activity by using the timestamp start and end and the clinical labels from the task score sheet. This was achieved by using the activity start and end times from the task score sheet, looping over the raw files, and taking the signal values between the timestamps. As previously mentioned, the battery of tasks completed in the clinic have specific durations, for example, 30 seconds for walking in a straight line. However, upon closer inspection, it was found that the durations do not hold for any of the activities in the actual dataset. Table 3 shows the tasks, and the average time it took to perform that activity.

Task/Activity	Average time sec.(std)
Finger-to-nose – right arm	18.56 (0.80)
Finger-to-nose – left arm	18.58 (0.72)
Repeated arm movement – right arm	18.50 (0.75)
Repeated arm movement – left arm	18.54 (0.64)
Standing	33.63 (0.75)
Sit to stand	18.38 (6.68)
Walking in a straight line	33.64 (0.70)
Walking in a straight line while counting backwards	33.37 (1.24)
Walking through a narrow passage	43.98 (15.45)
Pouring water from a bottle and drinking	36.26 (8.21)
Organizing sheets in a folder	36.96 (9.29)
Folding a towel	36.55 (10.35)
Drawing and writing on a paper	43.00 (19.49)
Typing on a computer keyboard	33.78 (1.29)
Assembling nuts and bolts	34.48 (2.64)

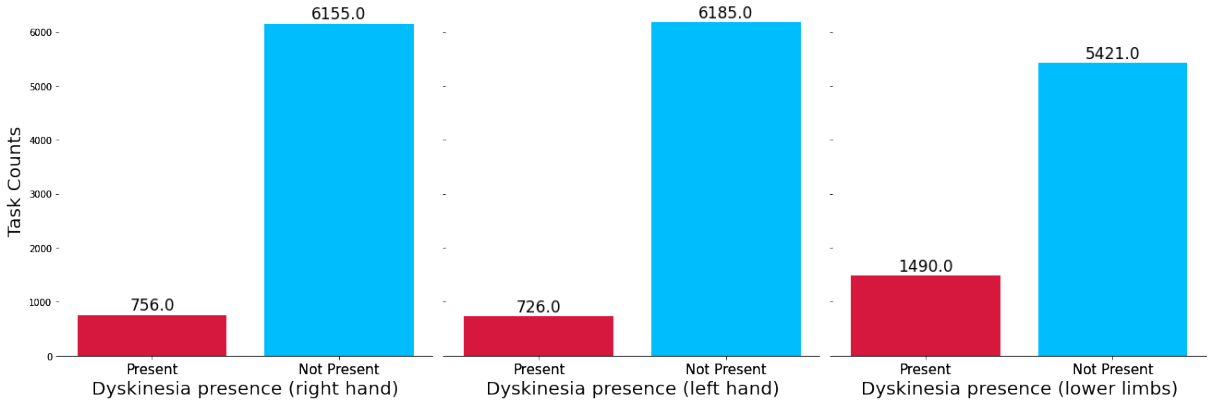
Table 3: Activity duration analysis. The rules in between the activities show the different categories of the tasks.

The times are averaged over all participants and all repetitions. These differences in activity durations could be due to various manifestations of the symptoms and severities — some patients took longer to complete a task than others; hence the duration of task completion is not uniform.

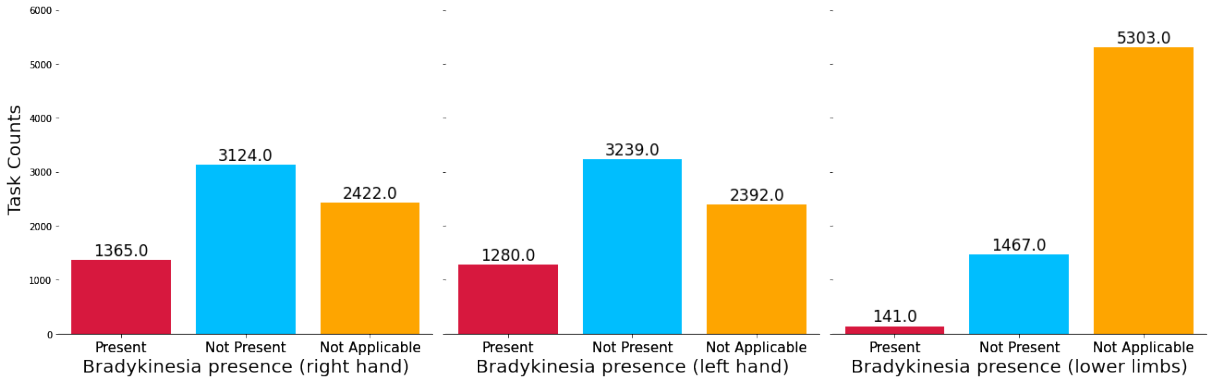
Moreover, the counts of symptoms were analyzed from the data of the task score table, where clinicians scored the symptoms of participants per each limb per each repetition of an activity. It was found that the reported symptom occurrence mentioned in the initial questionnaire is not very accurate (Figure 2). A simple analysis showed that symptoms reported during a one-time clinical questionnaire is not an accurate assessment of the disease. The plots in Figure 5 show the actual distribution of symptoms labeled by clinicians during in-clinic task repetitions.



(a) Task counts for tremor severity per body part.



(b) Task counts for dyskinesia presence per body part.



(c) Task counts for bradykinesia presence per body part.

Figure 5: Symptom task counts per body part. Figures 5a 5b and 5c show the amount of tasks where tremor, dyskinesia, and bradykinesia were present accordingly.

It is clear from Figure 5 that the distribution of symptoms is very imbalanced. Most of the clinical labels given to participants during the activities mention that there was no symptom, either "Not Present" for bradykinesia and dyskinesia, or 0 for tremor severity. In total, 20733 clinical labels were given per body part, out of which 15969 was no tremor, 17947 was no/not applicable bradykinesia, and no dyskinesia present in 17761 cases.

Moreover, symptom severity/presence are plotted per activity repetition to identify how much each activity contributed to the symptom presence estimation. For tremor severity, it is

noticeable from Figure 5a that tremor was not present during the completion of most activities, and severe tremor was only present in very few cases (labels 3 or 4). This indicates very few true positive labels, especially for severe tremor. The plot in Figure 6 shows the disproportional distribution of severity scores per task.

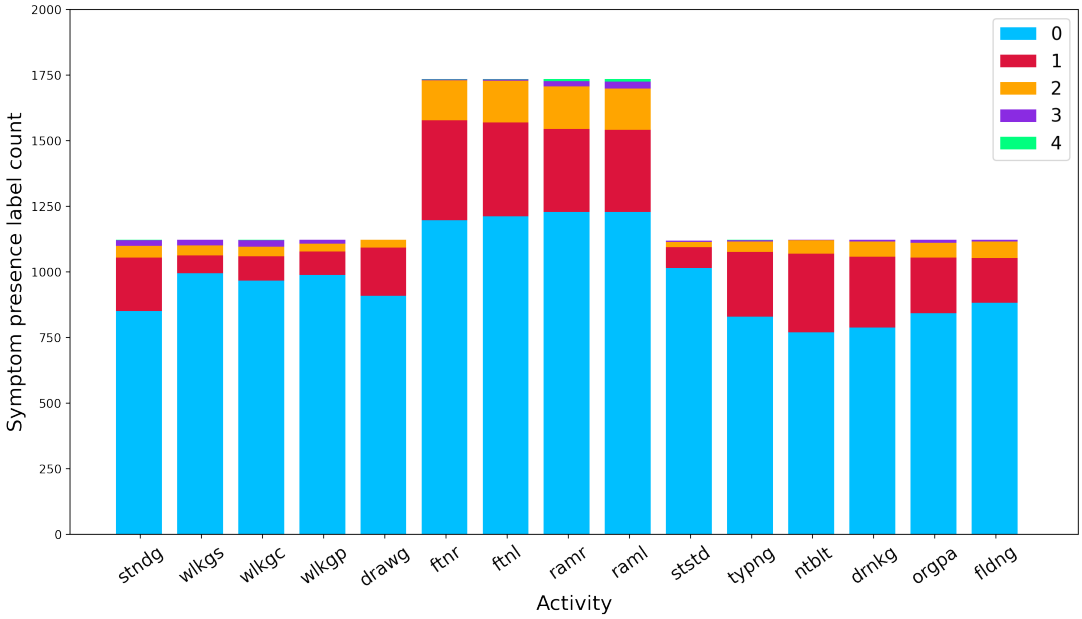


Figure 6: Tremor labels per task repetition count before severity aggregation.

For this reason, clinical labels for tremor are aggregated and transformed into "Present" / "Not Present" $(1/0)$ labels, just like other symptoms. This was done by keeping all the labels for 0 severity ("Not Present") and changing other severity scores into 1 ("Present"), given by equation 1:

$$tremor = \begin{cases} 0, & \text{if } severity = 0 \\ 1, & otherwise. \end{cases} \tag{1}$$

This aggregation step resulted in a new distribution of labels, changing severity scores to symptom presence scores, as visualized in Figure 7.

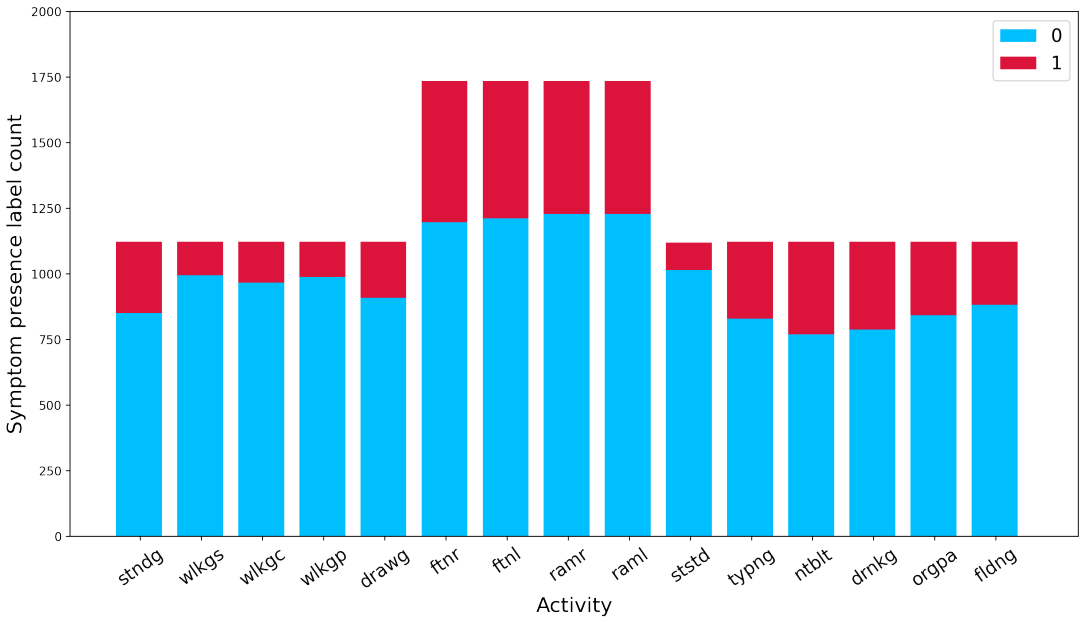


Figure 7: Tremor labels per task repetition count after severity aggregation. The severity scores were aggregated into one score, which indicates tremor presence.

Moreover, some movements do not include the motion of particular body parts; hence the label "Not Applicable" was given for bradykinesia. For example, drawing does not include any movement from the lower body since the participants were sitting during that activity. Therefore, there could not be any slowness of movements in the lower limbs, and these were labeled as "Not Applicable." As visualized in Figure 8, there are many task repetitions where bradykinesia is not applicable. For this reason, the data for the "Not Applicable" label for bradykinesia was discarded since these labels do not hold any value. Figure 9 shows the data distribution after removing the not applicable data.

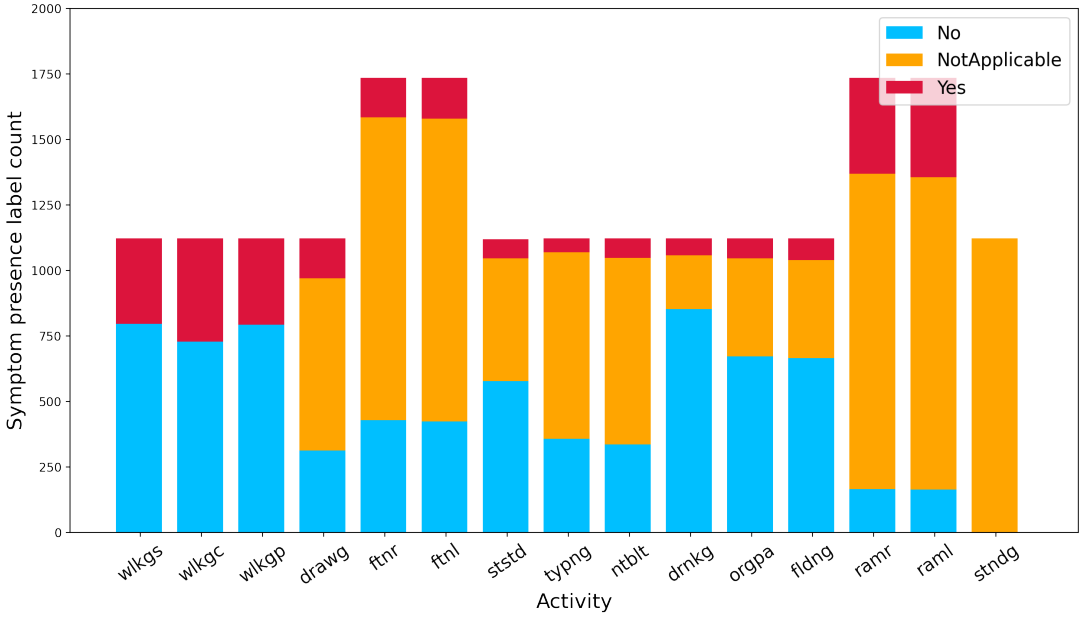


Figure 8: Bradykinesia labels per task repetition count before the removal of not applicable data.

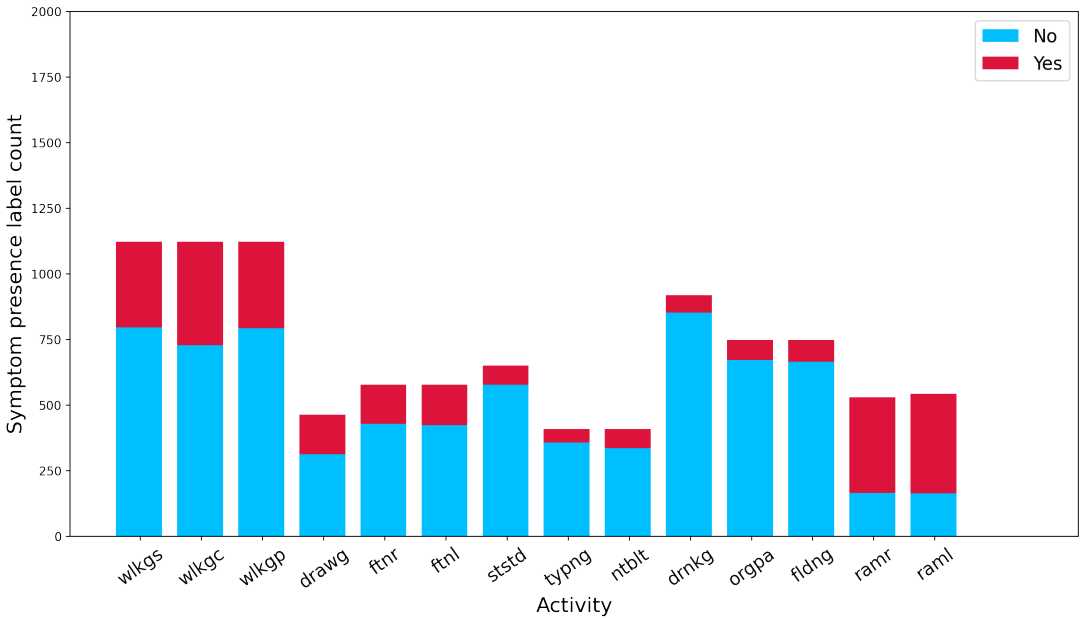


Figure 9: Bradykinesia labels per task repetition count after the removal of not applicable data.

No data manipulations are needed for dyskinesia since there is only data for symptom presence and absence. Figure 10 shows the distribution of symptom presence/absence per activity repetition count for dyskinesia.

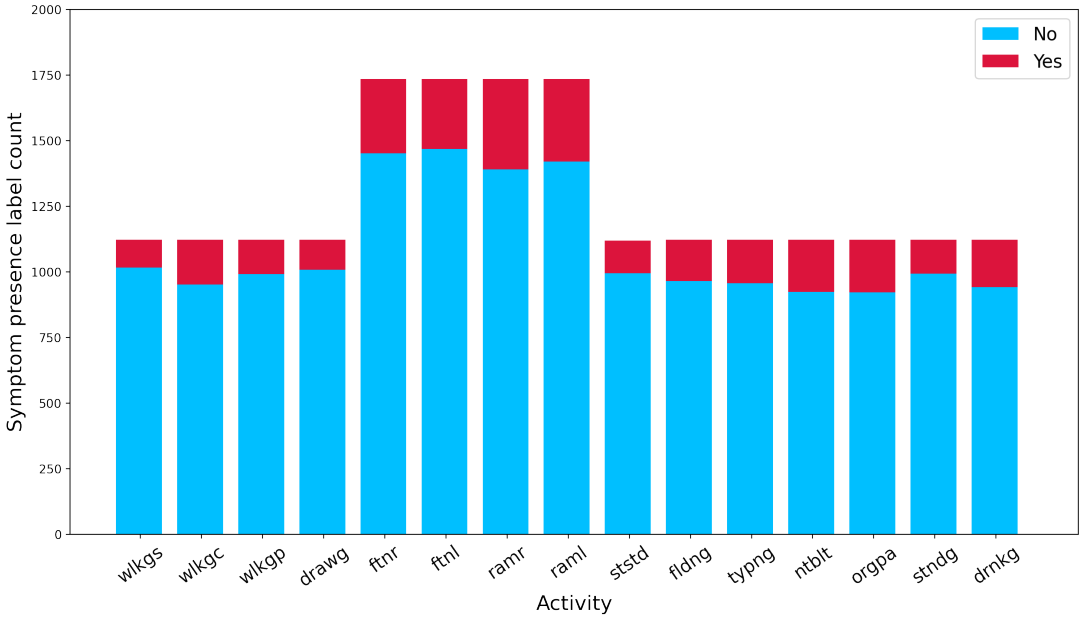


Figure 10: Dyskinesia labels per task repetition count.

This research looks into a combination of one activity and one sensor; therefore, only the data from the sensor of the dominant hand (right hand) and the phone in the fanny pack are considered. The choice of using the right-hand sensor lies in the assumption that the dominant hand of most participants (except one) is the right hand, meaning that most participants will give priority to doing daily life activities with their right hand. Although only one sensor is used

at a time, the goal is to detect symptoms in body parts to which a sensor is not attached. This was achieved using a logical disjunction operator ($\vee$) to aggregate the symptom labels from different body parts. In other words, the symptom label for an activity is "Present" if at least one body part has symptom presence. The symptom labels were combined for each activity and are given by the logical formulas 2 and 3,

$$score(L - R) = \mu \vee \nu, \tag{2}$$

$$score(L - R - LO) = (\mu \vee \nu) \vee \eta, \tag{3}$$

where μ and ν are the clinical labels for the right and left hands accordingly, and η is the clinical label for the lower limbs during an activity. The logical operation returns the truth value "Present" unless all of its arguments are false ("Not Present"). First, only clinical scores from the right hand were used to match the data from the right-hand and phone sensors, resulting in 15 datasets for each task/sensor combination (30 total). Next, the clinical labels for the right and left hands were aggregated into one score 2 to label the sensor data, creating 30 new datasets. In addition, the clinical score of the lower limbs and left-right hand combination were aggregated into one label 3, creating 30 additional datasets with the combination of three labels. These labels are then matched with both timestamped right hand and phone sensor readings, thus totaling 90 datasets for the classification task.

During signal timestamping, it was noticed that some signals were much shorter than the average duration analysis mentioned in Table 3. After visual inspection, it was found that the dataset contained raw sensor files full of NaN values in between the sensor readings, and some were even completely filled with NaNs. This resulted in issues when pre-processing the signals.

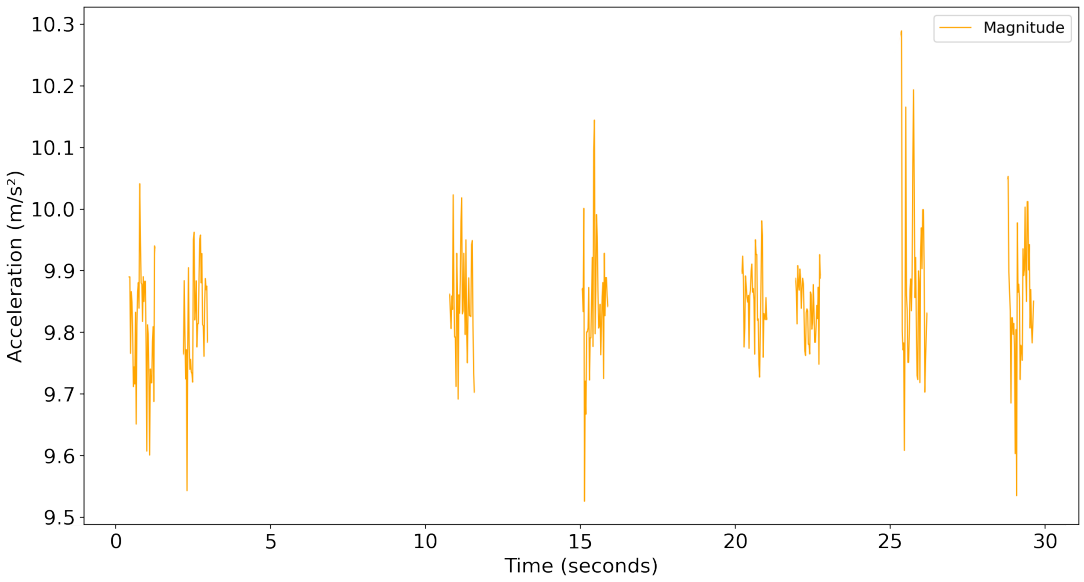


Figure 11: Example of a signal with missing data.

Figure 11 shows an example of a signal with missing data of an activity from the right-hand sensor. As it can be seen, the signal is 30 seconds long, but the readings make "jumps" because of missing values in between. Such faulty signals were 3% of the overall dataset

and were carefully inspected and removed for further analysis. Some signals also had missing values in the duration of the activity but were not as extreme. Any signal with more than 10% missing values was removed from the dataset because these are unsuitable for this research. Furthermore, raw data from wearable devices usually have sensor noise and must be pre-processed. For this reason, each accelerometer signal $X - Y - Z$ was smoothed separately with a second-order Butterworth band-pass filter given by Equation 4.

$$\alpha, \beta, \gamma = f_{Butter}(X, Y, Z, \psi, \phi) \tag{4}$$

Band-pass filters are designed to pass frequencies within a specific range and attenuate frequencies outside that range. The lower bound was set to $\psi = 0.1Hz$ to eliminate sensor drift, and the upper bound was set to $\phi = 20Hz$ to smooth noise and gross body movements. In addition, vector magnitude $\| \cdot \|$ of smoothed signals α, β, γ is then calculated by Equation 5 to acquire a scalar signal, which will not be dependent on the orientation and placement of the sensor.

$$M = \left| \sqrt{\alpha^2 + \beta^2 + \gamma^2} \right| \tag{5}$$

All subsequent feature extraction is performed on the magnitude (M) of the signals.

3.3 Features

Two sets of features were generated in the feature engineering step — statistical features and Fast Fourier Transform (FFT) features. The features were extracted using the $tsfresh$ library [9]. Statistical features are the most common features used in the literature, whereas FFT features are additionally added for this research. Fast Fourier Transform is the discrete Fourier transform of a sequence. The amplitude of the FFT can be used to calculate the power spectrum of a time series, which converts the signal from the time domain into the frequency domain by describing the distribution of power into frequency components composing that signal. The power spectral density is defined as the discrete-time Fourier transform of the covariance sequence [40]. The symptoms of Parkinson's disease have specific intervals (bands) in the frequency domain. Tremor typically occurs with a detected peak within the range of $3.5 - 12Hz$ frequency band [26]. Meanwhile, it was shown that bradykinesia and dyskinesia are mostly present in the $0.5 - 1.17Hz$ and $1 - 4Hz$ frequency bands, respectively [11] [38]. Therefore, the FFT features are expected to have some predictive power.

Due to varying signal and activity lengths, no windowing techniques were used. As a result, all features are generated from the magnitude (M) of the full-length signal data of each unique task repetition, and ground truth was labeled based on scores given by the clinicians ($+$ combination scores). Table 4 lists all relevant features and their descriptions.

Feature	Description
Standard Deviation	*Standard deviation of the time series*
Root Mean Square	*Root mean square (RMS) of the time series*
Skewness	*Sample skewness of the time series*
Kurtosis	*Kurtosis of the time series*
Maximum	*Highest value of the time series*
Mean Absolute Change	*Average over first differences*
Mean 10 Absolute Max	*Arithmetic mean of the 10 absolute maximum values*
Mean Second Derivative Central	*Mean value of a central approx. of the* 2^{nd} *derivative*
First Location of Maximum	*The first location of the maximum value*
Last Location of Maximum	*The last location of the maximum value*
FFT Centroid (mean)	*Spectral centroid of the Fourier transform spectrum*
FFT Variance	*Variance of the Fourier transform spectrum*
FFT Skew	*Skew of the Fourier transform spectrum*
FFT Kurtosis	*Kurtosis of the Fourier transform spectrum*

Table 4: Feature table with descriptions.

3.4 Model Training and Validation

As visualized in Figures 7, 9, 10, the dataset is extremely non-uniformly distributed. Therefore, before the model training, the majority class (no symptom) was randomly undersampled to have equal parts of symptom presence and absence. This was done to have the same percentage of samples for each class and keep the dataset balanced. Several techniques were used for the process of model training. For the classification models, a Random Forest classifier [6] and a Decision Trees classifier [44] were implemented for the estimation of symptom presence using the $scikit-learn$ library [32]. These algorithms were chosen for a few reasons. First, Jeon et al. [24] showed that decision trees outperformed other methods when using only one wearable sensor on the wrist, including SVMs with various kernels. Discriminant Analysis model was very close in performance followed by a Random Forest classifier. Second, a recent paper [2] showed that Random Forest outperformed other classifiers such as SVMs, K-nearest neighbour, and Naïve Bayes for estimating dyskinesia in 20 subjects. Many researchers compare their results with the SVMs for the estimation of the symptoms, yet their sample size is always limited to only a few subjects or features. In this case, even though the study cohort was not very small, the dataset did not consist of many severe cases; thus a mild-state and a no-symptom state may not be linearly separable. For non-linear SVM classifiers, the complexity of the algorithm is exponential. Kernelized SVMs compute the distance function between each point in the dataset, and storing all of these distances will generally require between $O(n^2)$ to $O(n^3)$ comparisons [5]. However, the time complexity for a Decision Tree classifier is in the range of $O(n)$-$O(n^2)$ and it is $O(nlogn)$ for a Random Forest classifier. Some comparisons for the classifiers implemented in the studies mentioned in Related Work 2 showed that tree based models outperformed other methods in performance and speed, and for that reason the classifiers which need less computational power were chosen. Furthermore, the classifiers were

used in combination with 10-fold nested cross-validation, which has an inner and outer loop. Nested cross-validation was chosen to train the model on 10-folds of train/validation/test set splits, and to optimize the hyperparameters simultaneously with a grid search technique using the $scikit-learn$ library [32]. The outer loop is the commonly known standard evaluation loop of a cross-validation, which trains each split with optimal hyperparameters and averages test errors for each split to properly evaluate the generalization performance. In the meantime, the inner loop is responsible for maximizing the score by fitting the model to each training set, and selecting the hyperparameters during the training on the validation set. In other words, during each fold, each set of hyperparameters are evaluated by using the 10-fold cross-validation that splits the train dataset into 10 folds and the accuracy for the best performing model is chosen. In the end, the mean accuracy from all folds is calculated. Tables 5 and 6 lists the hyperparameters for the classifiers and their values used for optimization. It is worth mentioning that not all hyperparameters were taken into account for the minimization of model training time and that may have resulted in slightly worse model performance.

Hyperparameter	**Values**
Number of trees in the forest	*5, 10, 20, 30, 40, 45, 50, 100, 200, 400, 500*
Number of features to consider when looking for the best split	*'sqrt', 'log2'*

Table 5: The range of hyperparameter values for the Random Forest classifier.

Hyperparameter	**Values**
Minimum number of samples to split an internal node	*2, 3, 5, 10, 20, 50, 100*
Number of features to consider when looking for the best split	*'sqrt', 'log2'*
Criterion (quality of the split)	*'gini', 'entropy', 'log_loss'*

Table 6: The range of hyperparameter values for the Decision Trees classifier.

Since there were many repetitions of the same activity by the same participant, there was a possibility of data leakage. Using the same participant's data in any of the training/test/validation sets will result in overfitting even though those are different repetitions of the same movement. Therefore, the algorithm could potentially learn the movement pattern of the participant, and for this reason, grouping was introduced inside the nested cross-validation. The grouping procedure would return non-overlapping ID's of the participants to the cross-validation step, such that the same group (participant) does not appear in two different train/validation/test set splits. The algorithm performance was evaluated by considering the average accuracy and standard deviation over the 10 folds retrieved from the nested cross-validation step.

4 Results

The Random Forest and Decision Trees algorithms both provided interesting results in determining a good combination of sensor and activity. The results are divided into two subsections, one for each classifier, and the result tables are divided into six parts. For each sensor and phenotype pair, a separate table is made stating the sensor, clinical label for the phenotype (right, right+left, right+left+lower), tasks, and their respective mean accuracy scores and standard deviations over ten folds. In some cases, "No Data" is reported in the tables for dyskinesia. This is because there were not enough positive clinical labels for these activities, making the 10-fold cross-validation meaningless. Any activity with less than 25 samples per class has been discarded and thus marked as "No Data." Moreover, the "Not Applicable" label was used instead of an accuracy score in the case of bradykinesia. In the case of "finger to nose left" and "repeated arm movement left," there potentially could not be any slowness of movements in the right hand since the right hand was not being used during those tasks. On the other hand, "standing" is a stationary task; therefore, no slowness of movements is possible. It is assumed that the participants were stationary during the completion of most tasks, either sitting or standing. For this reason, lower limbs do not play any role in non-functional movements; thus, the reported accuracy scores for the combination labels are the same in some cases. Ultimately, two final tables are created for each classifier to emphasize the results. No statistical tests were performed during this analysis and accuracy scores in bold show the best performing scores ($> 70\%$) for clarity.

4.1 Results - Random Forest

Task	**Accuracy % (std) per body part**		
	R	*R-L*	*R-L-LO*
Finger to nose Left	52.17 (15.05)	40.18 (6.74)	46.45 (10.51)
Finger to nose Right	55.40 (11.57)	58.18 (13.43)	59.14 (13.26)
Repeated arm movement Left	51.54 (12.61)	52.13 (12.88)	48.49 (10.27)
Repeated arm movement Right	47.19 (15.90)	52.88 (15.88)	51.18 (20.53)
Standing	**71.76 (12.60)**	67.55 (8.56)	65.09 (12.35)
Sit to Stand	68.23 (16.46)	61.02 (13.79)	62.18 (12.81)
Walking straight	**75.72 (12.38)**	**73.95 (19.67)**	**74.11 (17.83)**
Walking counting	**75.68 (22.20)**	61.25 (22.38)	66.38 (18.37)
Walking narrow passage	63.24 (18.61)	61.56 (19.98)	63.67 (11.56)
Drinking	64.48 (16.12)	57.49 (14.39)	52.58 (5.49)
Organizing papers	55.35 (12.77)	64.21 (15.82)	59.49 (8.99)
Folding a towel	63.23 (18.29)	**71.31 (13.47)**	**73.74 (21.03)**
Drawing	**77.12 (9.75)**	67.16 (24.40)	62.28 (14.65)
Typing	55.90 (12.04)	51.64 (8.68)	50.06 (14.86)
Nuts and bolts	41.60 (11.23)	49.95 (13.01)	55.84 (10.14)

Table 7: **Random Forest:** results for detecting **tremor** per body part per task using the **right hand sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

Task	**Accuracy % (std) per body part**		
	R	*R-L*	*R-L-LO*
Finger to nose Left	40.26 (18.03)	42.75 (9.36)	43.85 (13.17)
Finger to nose Right	43.95 (12.33)	47.28 (19.88)	56.13 (18.26)
Repeated arm movement Left	53.77 (19.54)	48.40 (14.86)	51.85 (15.26)
Repeated arm movement Right	55.77 (15.40)	54.85 (17.63)	62.08 (18.27)
Standing	53.35 (20.85)	59.82 (9.89)	60.66 (9.36)
Sit to Stand	43.51 (34.12)	32.23 (24.31)	43.79 (23.22)
Walking straight	41.97 (17.09)	43.01 (31.19)	49.06 (29.52)
Walking counting	59.41 (29.58)	46.51 (24.20)	44.12 (25.00)
Walking narrow passage	48.26 (23.47)	44.99 (23.26)	46.99 (24.61)
Drinking	**72.76 (17.80)**	**73.27 (12.75)**	**73.48 (19.29)**
Organizing papers	59.73 (11.86)	57.03 (13.88)	**71.25 (12.17)**
Folding a towel	57.27 (11.10)	59.97 (23.28)	53.03 (11.25)
Drawing	60.67 (21.56)	63.37 (19.41)	61.38 (14.33)
Typing	59.42 (17.08)	63.05 (17.76)	59.43 (12.54)
Nuts and bolts	50.05 (19.03)	63.12 (14.87)	60.17 (16.93)

Table 8: **Random Forest:** results for detecting **tremor** per body part per task using the **lower body sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

Task	Accuracy % (std) per body part		
	R	*R-L*	*R-L-LO*
Finger to nose Left	**75.52 (11.67)**	**71.24 (14.17)**	**70.23 (8.92)**
Finger to nose Right	52.81 (33.64)	49.04 (19.78)	53.67 (8.34)
Repeated arm movement Left	58.40 (19.33)	64.07 (15.14)	55.54 (13.60)
Repeated arm movement Right	55.33 (20.09)	52.67 (14.94)	49.33 (11.07)
Standing	62.53 (18.99)	31.34 (15.31)	51.87 (15.43)
Sit to Stand	59.81 (22.01)	54.00 (23.66)	57.88 (20.26)
Walking straight	55.51 (14.96)	49.01 (20.17)	49.01 (20.17)
Walking counting	53.75 (13.24)	56.98 (9.68)	50.99 (14.56)
Walking narrow passage	55.99 (19.36)	51.87 (14.00)	46.78 (15.99)
Drinking	**70.24 (38.67)**	**70.52 (19.66)**	61.25 (9.78)
Organizing papers	**77.50 (26.89)**	63.00 (38.43)	48.09 (14.36)
Folding a towel	62.56 (34.43)	55.36 (25.20)	59.17 (12.72)
Drawing	No Data	63.17 (31.19)	62.12 (14.77)
Typing	No Data	No Data	45.03 (14.47)
Nuts and bolts	52.08 (38.34)	52.08 (38.34)	51.33 (5.10)

Table 9: **Random Forest:** results for detecting **dyskinesia** per body part per task using the **right hand sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

Task	Accuracy % (std) per body part		
	R	*R-L*	*R-L-LO*
Finger to nose Left	62.76 (20.30)	47.93 (25.91)	50.38 (8.67)
Finger to nose Right	**77.45 (29.18)**	60.54 (18.51)	58.20 (6.24)
Repeated arm movement Left	57.49 (17.83)	60.87 (12.73)	59.31 (20.00)
Repeated arm movement Right	61.96 (27.48)	55.18 (18.03)	59.14 (14.60)
Standing	57.33 (41.97)	56.25 (22.92)	50.44 (22.15)
Sit to Stand	52.79 (42.21)	57.19 (29.10)	58.58 (30.02)
Walking straight	57.69 (23.94)	58.63 (25.93)	58.63 (25.93)
Walking counting	37.97 (20.74)	42.35 (16.51)	42.35 (16.51)
Walking narrow passage	51.85 (22.27)	49.47 (23.07)	58.65 (19.95)
Drinking	51.67 (42.46)	56.75 (36.95)	57.97 (13.39)
Organizing papers	No Data	No Data	43.30 (9.97)
Folding a towel	No Data	No Data	60.30 (13.35)
Drawing	No Data	61.67 (43.49)	57.15 (19.63)
Typing	No Data	No Data	50.79 (15.49)
Nuts and bolts	**74.17 (26.99)**	74.17 (26.99)	50.50 (15.70)

Table 10: **Random Forest:** results for detecting **dyskinesia** per body part per task using the **lower body sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

Task	**Accuracy % (std) per body part**		
	R	*R-L*	*R-L-LO*
Finger to nose Left	Not Applicable	53.11 (7.31)	53.11 (7.31)
Finger to nose Right	**77.10 (11.45)**	77.10 (11.45)	77.10 (11.45)
Repeated arm movement Left	Not Applicable	57.02 (13.28)	57.02 (13.28)
Repeated arm movement Right	46.07 (18.00)	46.07 (18.00)	46.07 (18.00)
Standing	Not Applicable	Not Applicable	Not Applicable
Sit to Stand	54.52 (12.46)	55.46 (15.02)	55.91 (12.59)
Walking straight	**70.25 (13.14)**	58.04 (14.08)	61.23 (15.79)
Walking counting	67.81 (14.66)	60.04 (13.21)	60.93 (17.04)
Walking narrow passage	62.85 (14.19)	60.59 (15.50)	63.16 (7.47)
Drinking	64.83 (19.85)	**70.04 (13.86)**	70.04 (13.86)
Organizing papers	**72.12 (22.98)**	72.12 (22.98)	72.12 (22.98)
Folding a towel	59.13 (21.21)	62.44 (9.61)	62.44 (9.61)
Drawing	**75.38 (15.41)**	61.84 (18.96)	61.84 (18.96)
Typing	**77.71 (30.42)**	**73.24 (27.34)**	73.24 (27.34)
Nuts and bolts	**88.11 (14.44)**	**92.21 (9.97)**	92.21 (9.97)

Table 11: **Random Forest:** results for detecting **bradykinesia** per body part per task using the **right hand sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

Task	**Accuracy % (std) per body part**		
	R	*R-L*	*R-L-LO*
Finger to nose Left	Not Applicable	56.78 (27.08)	56.78 (27.08)
Finger to nose Right	**79.96 (16.24)**	79.96 (16.24)	79.96 (16.24)
Repeated arm movement Left	Not Applicable	49.71 (25.17)	49.71 (25.17)
Repeated arm movement Right	58.17 (18.74)	58.17 (18.74)	58.17 (18.74)
Standing	Not Applicable	Not Applicable	Not Applicable
Sit to Stand	35.20 (21.24)	35.20 (21.24)	46.35 (27.88)
Walking straight	56.39 (20.53)	57.57 (9.79)	63.72 (16.64)
Walking counting	67.88 (19.43)	59.48 (18.72)	60.04 (23.28)
Walking narrow passage	68.49 (22.69)	62.50 (16.67)	60.39 (12.70)
Drinking	57.88 (36.64)	57.88 (36.64)	57.88 (36.64)
Organizing papers	**72.89 (16.17)**	72.89 (16.17)	72.89 (16.17)
Folding a towel	58.77 (30.98)	67.27 (23.83)	67.27 (23.83)
Drawing	**78.18 (21.99)**	52.34 (17.38)	52.34 (17.38)
Typing	50.69 (41.50)	50.69 (41.50)	50.69 (41.50)
Nuts and bolts	36.38 (32.14)	36.38 (32.14)	36.38 (32.14)

Table 12: **Random Forest:** results for detecting **bradykinesia** per body part per task using the **lower body sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

4.2 Results - Decision Trees

Task	**Accuracy % (std) per body part**		
	R	*R-L*	*R-L-LO*
Finger to nose Left	54.89 (12.83)	44.62 (9.53)	46.15 (11.85)
Finger to nose Right	52.49 (9.80)	50.57 (12.83)	54.60 (14.84)
Repeated arm movement Left	49.28 (15.53)	52.24 (11.88)	50.11 (9.67)
Repeated arm movement Right	50.73 (8.31)	47.52 (12.57)	45.21 (16.16)
Standing	63.10 (10.15)	61.50 (10.74)	58.92 (10.81)
Sit to Stand	**73.79 (17.91)**	65.88 (20.04)	55.79 (17.77)
Walking straight	66.76 (13.99)	**76.74 (24.93)**	76.74 (24.93)
Walking counting	**71.59 (21.40)**	61.97 (16.33)	61.97 (16.33)
Walking narrow passage	65.01 (15.59)	62.21 (17.02)	60.73 (16.95)
Drinking	61.73 (8.58)	61.17 (13.62)	50.61 (10.39)
Organizing papers	57.08 (14.31)	59.92 (17.09)	52.34 (13.75)
Folding a towel	61.49 (19.62)	**70.76 (12.83)**	**71.17 (17.77)**
Drawing	65.01 (16.39)	56.94 (18.29)	63.93 (8.32)
Typing	46.65 (12.99)	49.46 (8.81)	47.47 (9.83)
Nuts and bolts	49.77 (15.46)	54.68 (12.78)	53.42 (9.29)

Table 13: **Decision Trees:** results for detecting **tremor** per body part per task using the **right hand sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

Task	Accuracy % (std) per body part		
	R	*R-L*	*R-L-LO*
Finger to nose Left	51.31 (11.30)	49.86 (11.78)	53.79 (13.31)
Finger to nose Right	43.42 (11.64)	49.75 (16.00)	56.45 (19.33)
Repeated arm movement Left	54.21 (18.40)	47.88 (10.16)	53.72 (10.08)
Repeated arm movement Right	56.05 (10.46)	51.35 (12.67)	59.21 (14.88)
Standing	46.32 (16.81)	59.77 (12.12)	54.22 (7.13)
Sit to Stand	56.01 (34.28)	37.49 (27.82)	45.20 (28.31)
Walking straight	51.09 (30.42)	47.41 (20.66)	47.41 (20.66)
Walking counting	49.32 (19.80)	43.33 (18.05)	43.33 (18.05)
Walking narrow passage	47.23 (27.24)	41.67 (20.34)	41.67 (20.34)
Drinking	64.48 (14.87)	66.16 (15.11)	65.02 (16.31)
Organizing papers	64.48 (16.21)	42.36 (9.93)	64.13 (8.42)
Folding a towel	60.35 (11.73)	64.14 (14.49)	56.77 (12.48)
Drawing	63.08 (20.33)	60.68 (18.46)	68.41 (16.12)
Typing	49.70 (12.20)	57.20 (18.89)	59.64 (15.80)
Nuts and bolts	48.82 (19.41)	58.75 (19.43)	69.55 (18.57)

Table 14: **Decision Trees:** results for detecting **tremor** per body part per task using the **lower body sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

Task	Accuracy % (std) per body part		
	R	*R-L*	*R-L-LO*
Finger to nose Left	**70.64 (11.67)**	66.44 (17.48)	63.88 (10.68)
Finger to nose Right	34.10 (30.66)	52.86 (11.09)	48.03 (12.79)
Repeated arm movement Left	53.43 (13.96)	59.51 (15.39)	52.32 (17.73)
Repeated arm movement Right	59.61 (18.44)	54.19 (13.53)	51.09 (8.60)
Standing	51.41 (25.06)	44.90 (18.97)	50.40 (18.37)
Sit to Stand	57.92 (26.45)	52.95 (21.95)	51.73 (16.46)
Walking straight	61.52 (18.00)	36.30 (16.92)	36.30 (16.92)
Walking counting	53.50 (9.64)	53.53 (11.50)	49.74 (13.17)
Walking narrow passage	52.69 (20.43)	54.74 (9.32)	54.08 (17.76)
Drinking	58.99 (37.45)	67.40 (18.56)	56.64 (13.87)
Organizing papers	62.67 (30.37)	47.00 (36.28)	48.82 (8.52)
Folding a towel	55.06 (33.62)	47.62 (32.27)	46.69 (10.45)
Drawing	31.83 (38.99)	57.72 (34.29)	53.84 (12.98)
Typing	No Data	No Data	42.49 (11.72)
Nuts and bolts	47.92 (36.38)	47.92 (36.38)	49.34 (9.11)

Table 15: **Decision Trees:** results for detecting **dyskinesia** per body part per task using the **right hand sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

Task	Accuracy % (std) per body part		
	R	*R-L*	*R-L-LO*
Finger to nose Left	61.18 (23.83)	45.81 (24.12)	46.96 (13.00)
Finger to nose Right	68.12 (30.67)	62.65 (17.64)	62.82 (11.08)
Repeated arm movement Left	53.49 (20.01)	59.13 (9.02)	67.54 (17.77)
Repeated arm movement Right	65.46 (28.46)	48.96 (19.20)	55.74 (13.11)
Standing	61.02 (32.01)	58.33 (30.33)	**70.07 (26.34)**
Sit to Stand	**71.27 (34.13)**	62.24 (26.61)	59.36 (29.14)
Walking straight	57.44 (28.96)	51.87 (26.67)	51.87 (26.67)
Walking counting	48.94 (17.13)	37.38 (19.92)	37.38 (19.92)
Walking narrow passage	51.13 (26.44)	61.44 (22.34)	57.13 (29.24)
Drinking	55.00 (42.20)	68.75 (32.13)	57.44 (9.39)
Organizing papers	No Data	No Data	47.20 (16.53)
Folding a towel	No Data	No Data	57.89 (12.58)
Drawing	No Data	53.33 (39.30)	48.23 (26.32)
Typing	No Data	No Data	56.40 (10.61)
Nuts and bolts	**71.67 (24.78)**	71.67 (24.78)	46.96 (13.00)

Table 16: **Decision Trees:** results for detecting **dyskinesia** per body part per task using the **lower body sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

Task	Accuracy % (std) per body part		
	R	*R-L*	*R-L-LO*
Finger to nose Left	Not Applicable	51.09 (14.52)	51.09 (14.52)
Finger to nose Right	66.51 (11.75)	66.51 (11.75)	66.51 (11.75)
Repeated arm movement Left	Not Applicable	54.10 (16.35)	54.10 (16.35)
Repeated arm movement Right	52.26 (13.70)	52.26 (13.70)	52.26 (13.70)
Standing	Not Applicable	Not Applicable	Not Applicable
Sit to Stand	64.34 (26.56)	64.34 (26.56)	64.77 (11.67)
Walking straight	67.18 (16.07)	60.30 (10.51)	58.36 (15.54)
Walking counting	61.38 (15.11)	55.24 (11.31)	56.67 (18.73)
Walking narrow passage	57.85 (19.89)	58.30 (10.78)	51.89 (15.42)
Drinking	69.50 (20.62)	69.50 (20.62)	69.50 (20.62)
Organizing papers	**70.64 (15.29)**	70.64 (15.29)	70.64 (15.29)
Folding a towel	**70.31 (19.90)**	**74.68 (13.45)**	74.68 (13.45)
Drawing	62.92 (18.60)	56.26 (10.93)	56.26 (10.93)
Typing	**74.29 (30.05)**	46.06 (30.14)	46.06 (30.14)
Nuts and bolts	**72.40 (30.77)**	**76.36 (29.64)**	76.36 (29.64)

Table 17: **Decision Trees:** results for detecting **bradykinesia** per body part per task using the **right hand sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

Task	Accuracy % (std) per body part		
	R	*R-L*	*R-L-LO*
Finger to nose Left	Not Applicable	64.93 (11.05)	64.93 (11.05)
Finger to nose Right	**78.18 (12.92)**	78.18 (12.92)	78.18 (12.92)
Repeated arm movement Left	Not Applicable	43.36 (20.24)	43.36 (20.24)
Repeated arm movement Right	47.65 (22.64)	47.65 (22.64)	47.65 (22.64)
Standing	Not Applicable	Not Applicable	Not Applicable
Sit to Stand	37.73 (24.92)	37.73 (24.92)	38.64 (26.74)
Walking straight	46.16 (16.95)	48.62 (7.90)	52.19 (20.39)
Walking counting	55.36 (22.61)	64.72 (16.82)	68.24 (17.93)
Walking narrow passage	56.39 (22.46)	56.23 (21.03)	50.93 (17.23)
Drinking	**72.28 (26.41)**	72.28 (26.41)	72.28 (26.41)
Organizing papers	**75.88 (18.38)**	75.88 (18.38)	75.88 (18.38)
Folding a towel	52.41 (30.67)	56.35 (27.85)	56.35 (27.85)
Drawing	67.43 (21.50)	56.49 (25.91)	56.49 (25.91)
Typing	58.28 (42.18)	58.28 (42.18)	58.28 (42.18)
Nuts and bolts	38.53 (32.27)	38.53 (32.27)	38.53 (32.27)

Table 18: **Decision Trees:** results for detecting **bradykinesia** per body part per task using the **lower body sensor**. R — Right hand ground truth; L-R — Right/Left hands combined ground truth; R-L-LO — Right/Left/Lower body combined ground truth. Accuracy scores higher than 70% are in bold typeface.

4.3 Results - Review

Task	**Right Hand Sensor**		
	R	*R-L*	*R-L-LO*
Finger to nose Left	✓	✓	✓
Finger to nose Right	✓		
Standing	✓		
Walking straight	✓ ✓	✓	✓
Walking counting	✓		
Drinking	✓	✓ ✓	
Organizing papers	✓ ✓		
Folding a towel		✓	✓
Drawing	✓ ✓		
Typing	✓	✓	
Nuts and bolts	✓	✓	

Table 19: Final results for the likelihood ($> 70\%$ accuracy) of symptom detection with the **Random Forest** classifier using the right hand sensor. ✓ — Tremor; ✓ — Dyskinesia; ✓ — Bradykinesia. The notation of ground truth columns is the same as in all other result tables.

Task	**Lower Body Sensor**		
	R	*R-L*	*R-L-LO*
Finger to nose Right	✓ ✓		
Drinking	✓	✓	✓
Organizing papers	✓		
Drawing	✓		
Nuts and bolts	✓		

Table 20: Final results for the likelihood ($> 70\%$ accuracy) of symptom detection with the **Random Forest** classifier using the lower body sensor. ✓ — Tremor; ✓ — Dyskinesia; ✓ — Bradykinesia. The notation of ground truth columns is the same as in all other result tables.

Task	**Right Hand Sensor**		
	R	*R-L*	*R-L-LO*
Finger to nose Left	✓		
Sit to Stand	✓		
Walking straight		✓	
Walking counting	✓		
Organizing papers	✓		
Folding a towel	✓	✓✓	✓
Typing	✓		
Nuts and bolts	✓	✓	

Table 21: Final results for the likelihood ($> 70\%$ accuracy) of symptom detection with the **Decision Trees** classifier using the right hand sensor. ✓ — Tremor; ✓ — Dyskinesia; ✓ — Bradykinesia. The notation of ground truth columns is the same as in all other result tables.

Task	**Lower Body Sensor**		
	R	*R-L*	*R-L-LO*
Finger to nose Right	✓		
Standing			✓
Sit to Stand	✓		
Drinking	✓		
Organizing papers	✓		
Nuts and bolts	✓		

Table 22: Final results for the likelihood ($> 70\%$ accuracy) of symptom detection with the **Decision Trees** classifier using the lower body sensor. ✓ — Tremor; ✓ — Dyskinesia; ✓ — Bradykinesia. The notation of ground truth columns is the same as in all other result tables.

5 Discussion

Various Parkinson's disease symptom detection techniques have been investigated in many works included in the Related Work section of this research. Some studies have employed models to estimate the disease with the help of two or more sensors for one symptom only [23] [22]. However, it is very impractical for people to have multiple sensors on their bodies all day. This research aimed to discover the possibility of detecting Parkinson's disease symptoms using only one sensor and acquiring information about which daily activities are good predictors of the presence of each symptom. Daily task observations, such as those mentioned in this study, allow us to look at the symptoms from the perspective of known body movements instead of unknown daily data. The models used to achieve these results were the same Random Forest and Decision Tree classifier models for all tasks and symptoms, with minor adjustments in the hyperparameters that were optimized during the nested cross-validation. Two aggregated tables are created to emphasize the results for each algorithm — Table 19 and 21 for the right hand sensor and Table 20 22 for the lower body sensor results. All tables indicate which motor tasks are good predictors of the presence of Parkinson's disease symptoms per activity per body part. Each symptom is marked with a colored check mark to highlight the model's effectiveness for the mentioned task, body part, and symptom. Only the activities having at least 70% accuracy are mentioned with colored ticks. The *orange* tick represents tremor; *red* tick represents dyskinesia; *blue* tick represents bradykinesia. Comparing the results from these summarized tables, it is obvious that the sensor on the dominant hand is superior in detecting Parkinson's disease symptoms in the upper extremities of the body compared to results from the lower body sensor. Notably, the right-hand sensor could detect dyskinesia and tremor in the lower body in some cases, whereas the lower-body sensor only in one case for both classifiers. In the case of the Decision Tree classifier, the detection of dyskinesia suffered while some of the fine and gross motor movements showed similar results to the Random Forest algorithm. Unfortunately, the Decision Tree algorithm was not able to detect tremor in any of the cases using lower body sensor. However, these results are not significant to conclude that using the right-hand sensor will always be accurate enough to detect symptoms in the body's lower extremities.

5.1 Tremor

Tables 7 and 8 show the results of the right hand and lower body sensors for detecting tremor using the Random Forest classifier. Considering the fact that before the pre-processing step most common tremor severity label was 1 (slight tremor), the results are not surprising. Accuracy scores above 70% are reached in some activities for detecting tremor in the right hand using the right-hand sensor. These activities include "standing", "walking counting", "walking straight", "drawing" with accuracy scores of 71.76%, 75.68%, 75.72%, 77.12% respectively. On the other hand, the right-hand sensor data did not perform as well in detecting tremor on the other parts of the body, except when the participants performed the "walking straight" and "folding a towel" activities with mean accuracy scores ranging from 71% up to 74%. In the "walking straight" task, the high accuracy score could be because the activity is usually done in a controlled manner. However, as seen in all walking tasks, the accuracy scores improved and had a lower standard deviation when the symptom presence score of the lower limbs was aggregated to the ground truth. This shows that the right-hand sensor can detect tremor in the lower extremities with better accuracy than on the left hand. In the case of the lower

body sensor, some interesting results have emerged. While the lower body sensor showed poor results for most activities, the "drinking" task showed 73% accuracy for detecting tremor anywhere in the body. This is because pouring water and drinking it from glass requires immense concentration and balance, so the water does not get spilled. We can speculate that tremor manifests in the whole body, and the body trembles when bringing the arm close to the body, for example, when drinking water. Other performed tasks are more or less free body movements. For example, dropping a towel or a piece of paper is not as crucial as dropping a glass of water; thus, such free movement activities have more room for error. For this reason, tremor did not manifest in the signals of the lower body sensor in most activities. Moreover, the Decision Tree classifier showed similar results for "sit to stand" and "walking counting" activities, with accuracy scores of 73.79% and 71.59% respectively to detect tremor in the right-hand using the right hand sensor. These results are represented in the Tables 13 and 14. In the case of detecting tremor on the other parts of the body with the right hand sensor, "walking straight" and "folding a towel" activities performed relatively well in detecting tremor in any of the upper extremities of the body with accuracy scores of 76.74% and 70.76% accordingly. Meanwhile, the Decision Tree classifier was able to detect tremor in the lower body only during the "folding a towel" activity with a mean accuracy score of 71.17% when using the right-hand sensor. Unfortunately, the Decision Tree classifier showed no consistent results in accurately detecting tremor using the lower body sensor; however, some promising results in detecting tremor when performing gross body movements such as "drinking", "organizing papers" and "folding a towel" activities. Generally, the Random Forest classifier showed to be effective in detecting tremor during more activities than the Decision Tree classifier when using the right-hand sensor. The Decision Tree classifier showed no promising results in detecting tremor when using the lower body sensor.

5.2 Dyskinesia

For dyskinesia presence detection with the Random Forest classifier, right-hand and lower-body sensor results are interpreted from Tables 9 and 10 accordingly. For the right-hand sensor, only three activities showed promising results — "finger to nose left," "drinking," and "organizing papers." It was explicitly mentioned in one of the studies that only one wrist-worn sensor is enough to predict dyskinesia when the arms are in extended positions [27]. The "finger to nose left" task involves the extension of the left arm, and the acquired results for this task validate the findings of previous researchers. Accuracy of 75% was achieved for detecting dyskinesia on the right hand, 71% if we consider both hands, and 70% if we consider dyskinesia presence in the lower body too. However, no such results are achieved for the same task with the right hand ("finger to nose right"). If we follow the logic, the extension of the right hand triggers dyskinesia in the left hand, but since the right-hand sensor is far from the left hand, the sensor did not record those symptom manifestations. Furthermore, the model performed relatively well when assessing the gross motor skills with an accuracy score of 77.50% for the "organizing papers" task. For the "drinking" activity, the right-hand sensor showed similar results of 70% accuracy in detecting dyskinesia in the right hand and the left hand. However, the standard deviation changed from 38.67 for only the right hand to 19.66 for both hands, showing that the right-hand sensor is a good sensor placement for detecting dyskinesia in the upper limbs of the body when drinking water. In addition, the lower-body sensor shows interesting results in detecting dyskinesia in the upper body. In this case, the "finger to nose" movements again showed high accuracies, with 62.76% and 77.45% accuracy for the left and right hands. This

result shows that the lower-body sensor can detect dyskinesia in the upper extremities when the limbs are fully extended. However, the accuracy is much higher during the extension of the dominant hand ("finger to nose right"). Moreover, the lower body sensor showed an accuracy score of 74.14% in detecting dyskinesia in the upper limbs during the "assembling nuts and bolts" task, which requires fine motor skills. Once more, we may speculate that dyskinesia can manifest anywhere in the body during fine motor movements. Unfortunately, the model showed no potential results in dyskinesia presence detection in the lower extremities of the body. Furthermore, Tables 15 and 16 show the mean accuracy scores for the Decision Tree classifier results. Most notably, the "finger to nose left" task accuracy for the right hand was very close to the Random Forest classifier with an accuracy score of 70.64%. However, detecting dyskinesia in the other parts of the body with the right-hand sensor for the same activity, although similar, was not as successful with this algorithm. Accuracy score of 66.44% was achieved if we consider both hands, and 63.88% if we consider dyskinesia presence in the lower body. In general, the Decision Tree classifier performed poorly during other tasks compared to the Random Forest classifier. Moreover, in the case of the lower-body sensor, mean accuracy scores of 71.27% and 71.67% were achieved for detecting dyskinesia in the right hand during the "sit to stand" and "nuts and bolts" tasks accordingly. No such results were achieved in other movements, yet the combination of the lower-body sensor and the Decision Tree classifier was able to detect dyskinesia in the lower body of the participants when they were performing the "standing" activity with an accuracy score of 70.07%. What's more, accuracy scores ranging from 62.65% to 68.12% were achieved for various body parts when performing the "finger to nose right" movement, and an accuracy score of 68.75% was achieved for detecting dyskinesia in both hands during the "drinking" task. In general, the Random Forest classifier outperformed the Decision Tree classifier when using the right hand sensor. However, although the accuracy scores for some tasks were better when using the Random Forest classifier, the Decision Tree classifier showed more consistent results in detecting dyskinesia using the lower-body sensor.

5.3 Bradykinesia

Bradykinesia results for the Randoom Forest classifier are represented in the tables 11 and 12. The tables indicate that the model performed best in detecting bradykinesia using the right-hand sensor. The best results are achieved during the execution of gross and fine motor skills tasks such as "drinking," "organizing papers," "drawing," "typing," and "assembling nuts and bolts." Accuracy of 77.10% was achieved for the "finger to nose right" task using the right-hand sensor and the right hand score. These results are expected since this clinical task is a very controlled movement, and the sensor on the limb will register any distorted movement patterns. For the fine motor skills, accuracy results of 75.38% for "drawing," 88.11% for "assembling nuts and bolts," and 77.71% for "typing" were achieved in detecting bradykinesia in the right hand. For gross motor skills, accuracy results of 64.83% for "drinking" and 72.12% for "organizing papers" were achieved. Additionally, taking into account the symptom presence in the left hand increased the accuracy of bradykinesia detection to 92.21% while doing the "assembling nuts and bolts" activity, and to 70.04% for the "drinking" task. Fair accuracy was also achieved during the walking movements with 70.25% for "walking straight," 67.81% for "walking while counting," and 62.85% for "walking through a narrow passage" tasks. The accuracy decreased in all other cases when left- and right-hand scores were combined; however, there was a slight increase when the score for the lower body was also taken into

account. In the case of the lower body sensor, although with a slightly increased standard deviation, similar results were achieved for the activities "finger to nose right," "organizing papers," and "drawing," with accuracy scores of 79.96%, 72.89%, 78.18% accordingly. These results show that slowness of movements in the upper limbs may be possible to detect using a sensor located in the lower body of the patient. It is worth mentioning that the lower body was inactive during non-functional movements. Therefore, the scores reported in the third column (R-L-LO) are not considered practical results. The same principle is applied to the second column of results tables (R-L) because, in some cases, these tasks were performed only with the right hand. Overall, these results show that slowness of movements (bradykinesia) is most detectable with the right-hand sensor when performing fine motor movements, which need attention to detail. For the Decision Tree classifier, Tables 17 and 18 show the results for detecting bradykinesia using the right-hand sensor and the lower-body sensor accordingly. Although some of the results are comparable to the Random Forest classifier, the Decision Tree classifier performed worse in most cases. In detail, the performance for detecting bradykinesia in the right hand for "finger to nose right", "organizing papers", "typing" and "nuts and bolts" tasks got worse by a couple percent with accuracy scores of 66.51%, 70.64%, 74.29%, 72.40% accordingly. The "walking straight" tasks performed similarly but worse, resulting in a sub-70% accuracy for this task. The "drinking" task showed more consistent results with the use of the Decision Tree classifier. However, accuracy scores of 70.31% and 74.68% were achieved for the "folding a towel" task when considering bradykinesia in the right hand and both hands accordingly, using the right-hand sensor. In the case of the lower-body sensor, the results are mixed. The Decision Tree classifier showed similar results for the "finger to nose right" and "organizing papers" tasks with accuracy scores of 78.18% and 75.88% for detecting bradykinesia in the right hand respectively. However, the Decision Tree classifier outperformed Random Forest on the "drinking" task for the right hand with an accuracy score of 72.28%. As a whole, the results of the two algorithms for bradykinesia, both resulted in comparable accuracy scores for most tasks. Although the Decision Tree classifier performed similarly to the Random Forest classifier for the lower-body sensor, overall, the Random Forest classifier could detect bradykinesia during more activities when the right-hand sensor was used.

5.4 Summary

To summarize the results of both classifiers, the Random Forest classifier showed robust results in detecting Parkinson's disease symptoms during more activities compared to the Decision Tree classifier. The average accuracy scores for both classifiers for all tasks per ground truth were fairly similar yet the Random Forest classifier had higher average accuracy. Both classifiers showed consistent results in detecting bradykinesia when the patients were performing fine or gross motor movements. What's more, both classifiers had comparable results for detecting tremor when the patients were performing functional tasks such as walking or standing. However, the Random Forest classifier could detect all symptoms in more diverse cases (tasks). Furthermore, the Decision Tree classifier was unable to detect tremor when using the lower-body sensor. All in all, the Random Forest classifier showed to be superior in detecting Parkinson's disease symptoms with overall higher accuracy scores.

5.5 Limitations and Future Work

Several limitations of this study should be noted. Not all hyperparameters were optimized for both algorithms during model training, thus resulting in not consistent results in some cases where for some activities there was a big gap in accuracy scores when comparing both classifiers. Accuracy is the main evaluation metric of the algorithm performance used in this work because the majority class was undersampled, resulting in a balanced dataset. Having the same amount of classes makes accuracy a valid metric to use. However, no statistical significance testing was done. Accuracy may not be useful in clinical conditions, and sensitivity and specificity may be better metrics for a general approach. In most cases, the specificity trade-off was too high because most symptom manifestations were very slight. These metrics could potentially be improved by recruiting a cohort with more severe cases, depending on the clinical use of this method. Furthermore, this study looks at only one sensor/task pair, and the observations of a task were limited to a single wearable sensor. The left-hand sensor was not used in this research on account of the assumption that all participants (except one) are right-handed, and they are more likely to perform activities with their right hand. As a result, poor accuracy scores were achieved in particular scenarios to detect the symptoms, especially those not present on the limb where the sensor is located. Wearing more than one sensor on the body may be inconvenient for some people with Parkinson's disease. However, many people these days have both a smartphone in their pocket and a smartwatch worn on their wrist. Future work may combine smartphone and smartwatch observations for a more robust symptom presence classification method. Additionally, there are many similarities in the activities mentioned in this work, such as the walking tasks and gross/fine motor movement activities. It is possible to use the similarities of these activities by learning the tasks in parallel, and this could be achieved by using a multitask learning approach [8].

6 Conclusion

To conclude, this research attempted to discover the possibility of detecting Parkinson's disease symptoms in different body parts using only one sensor, either on the dominant hand (right hand) or a phone. In addition, daily life activities were used to identify during which activities the symptoms are most probably to manifest. The data for this research included clinical scores for each limb, labeled by clinicians while those activities were performed. The accuracy of both implemented classifiers were compared and it was observed that the Random Forest classifier outperformed the Decision Tree classifier. All things considered, the acquired results state that not all daily life activities are good predictors of tremor, dyskinesia, or bradykinesia presence. For example, good results were achieved for bradykinesia and tremor detection in the upper limbs during fine and gross motor movement tasks using the right-hand sensor. However, in most cases, estimating the presence of dyskinesia showed to be inaccurate with this method and additional steps may be needed. All in all, no one daily activity and sensor combination proved to be superior in detecting all symptoms in every part of the body, and an individual approach may be needed depending on clinical and patient needs.

References

[1] Jamie L Adams et al. "A real-world study of wearable sensors in Parkinson's disease". en. In: *NPJ Parkinsons Dis* 7.1 (Nov. 2021), p. 106.

[2] Satyabrata Aich et al. "A Supervised Machine Learning Approach to Detect the On/Off State in Parkinson's Disease Using Wearable Based Gait Signals". en. In: *Diagnostics (Basel)* 10.6 (June 2020).

[3] A. Berardelli et al. "Pathophysiology of bradykinesia in Parkinson's disease". In: *Brain* 124.11 (Nov. 2001), pp. 2131–2146. ISSN: 0006-8950. DOI: 10.1093/brain/124.11.2131. eprint: https://academic.oup.com/brain/article-pdf/124/11/2131/858511/1242131.pdf. URL: https://doi.org/10.1093/brain/124.11.2131.

[4] R Bhidayasiri. "Differential diagnosis of common tremor syndromes". In: *Postgraduate Medical Journal* 81.962 (2005), pp. 756–762. ISSN: 0032-5473. DOI: 10.1136/pgmj.2005.032979. eprint: https://pmj.bmj.com/content/81/962/756.full.pdf. URL: https://pmj.bmj.com/content/81/962/756.

[5] Léon Bottou and Chih-Jen Lin. "Support Vector Machine Solvers". In: Jan. 2007, pp. 301–320.

[6] Leo Breiman. "Random Forests". In: *Machine Learning* 45.1 (Oct. 2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324.

[7] Jorge Cancela et al. "Feasibility Study of a Wearable System Based on a Wireless Body Area Network for Gait Assessment in Parkinson's Disease Patients". In: *Sensors* 14.3 (2014), pp. 4618–4633. ISSN: 1424-8220. DOI: 10.3390/s140304618. URL: https://www.mdpi.com/1424-8220/14/3/4618.

[8] Rich Caruana. "Multitask Learning". In: *Machine Learning* 28.1 (July 1997), pp. 41–75. ISSN: 1573-0565. DOI: 10.1023/A:1007379606734. URL: https://doi.org/10.1023/A:1007379606734.

[9] Maximilian Christ et al. "Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)". In: *Neurocomputing* 307 (2018), pp. 72–77. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2018.03.067. URL: https://www.sciencedirect.com/science/article/pii/S0925231218304843.

[10] Zhiyong Cui et al. *Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction*. 2018. DOI: 10.48550/ARXIV.1801.02143. URL: https://arxiv.org/abs/1801.02143.

[11] Houde Dai, Haijun Lin, and Tim C Lueth. "Quantitative assessment of parkinsonian bradykinesia based on an inertial measurement unit". en. In: *Biomed Eng Online* 14 (July 2015), p. 68.

[12] Jasper R Daube and Devon I Rubin. "Needle electromyography". In: *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine* 39.2 (2009), pp. 244–270.

[13] Silvia Del Din et al. "Free-living monitoring of Parkinson's disease: Lessons from the field". en. In: *Mov Disord* 31.9 (July 2016), pp. 1293–1313.

[14] George DeMaagd and Ashok Philip. "Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis". en. In: *P T* 40.8 (Aug. 2015), pp. 504–532.

[15] Luc J W Evers et al. "Measuring Parkinson's disease over time: The real-world within-subject reliability of the MDS-UPDRS". en. In: *Mov Disord* 34.10 (July 2019), pp. 1480–1487.

[16] Giovanni Fabbrini et al. "Levodopa-induced dyskinesias". en. In: *Mov. Disord.* 22.10 (July 2007), pp. 1379–1389.

[17] Christopher G Goetz et al. "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results". en. In: *Mov Disord* 23.15 (Nov. 2008), pp. 2129–2170.

[18] Robert I Griffiths et al. "Automated assessment of bradykinesia and dyskinesia in Parkinson's disease". en. In: *J Parkinsons Dis* 2.1 (2012), pp. 47–55.

[19] Giuliana Grimaldi and Mario Manto. "Neurological tremor: sensors, signal processing and emerging applications". en. In: *Sensors (Basel)* 10.2 (Feb. 2010), pp. 1399–1422.

[20] Christopher W Hess and Seth L Pullman. "Tremor: clinical phenomenology and assessment techniques". en. In: *Tremor Other Hyperkinet Mov (N Y)* 2 (June 2012).

[21] Margaret M. Hoehn and Melvin D. Yahr. "Parkinsonism". In: *Neurology* 17.5 (1967), pp. 427–427. ISSN: 0028-3878. DOI: 10.1212/WNL.17.5.427. eprint: https://n.neurology.org/content/17/5/427.full.pdf. URL: https://n.neurology.org/content/17/5/427.

[22] Murtadha D Hssayeni et al. "Wearable Sensors for Estimation of Parkinsonian Tremor Severity during Free Body Movements". en. In: *Sensors (Basel)* 19.19 (Sept. 2019).

[23] Murtadha D. Hssayeni et al. "Dyskinesia estimation during activities of daily living using wearable motion sensors and deep recurrent networks". In: *Scientific Reports* 11.1 (Apr. 2021), p. 7865. ISSN: 2045-2322. DOI: 10.1038/s41598-021-86705-1. URL: https://doi.org/10.1038/s41598-021-86705-1.

[24] Hyoseon Jeon et al. "Automatic Classification of Tremor Severity in Parkinson's Disease Using a Wearable Device". en. In: *Sensors (Basel)* 17.9 (Sept. 2017).

[25] Ltd John Wiley Sons. *Motor Neurons and Spinal Control of Movement*. 1st ed. Major Reference Works. John Wiley Sons, Ltd. ISBN: 0-470-01617-5.

[26] Frauke Luft et al. "A Power Spectral Density-Based Method to Detect Tremor and Tremor Intermittency in Movement Disorders". en. In: *Sensors (Basel)* 19.19 (Oct. 2019).

[27] Thomas O Mera, Michelle A Burack, and Joseph P Giuffrida. "Objective motion sensor assessment highly correlated with scores of global levodopa-induced dyskinesia in Parkinson's disease". In: *Journal of Parkinson's disease* 3.3 (Jan. 2013), pp. 399–407. ISSN: 1877-7171. DOI: 10.3233/jpd-120166. URL: https://doi.org/10.3233/JPD-120166.

[28] Alvaro Muro-de-la-Herran, Begonya Garcia-Zapirain, and Amaia Mendez-Zorrilla. "Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications". en. In: *Sensors (Basel)* 14.2 (Feb. 2014), pp. 3362–3394.

[29] Spyridon Spyros Papapetropoulos. "Patient diaries as a clinical endpoint in Parkinson's disease clinical trials". en. In: *CNS Neurosci Ther* 18.5 (June 2011), pp. 380–387.

[30] James Parkinson. "An essay on the shaking palsy". In: *The Journal of neuropsychiatry and clinical neurosciences* 14.2 (2002), pp. 223–236.

[31] Shyamal Patel et al. "Monitoring Motor Fluctuations in Patients With Parkinson's Disease Using Wearable Sensors". In: *IEEE Transactions on Information Technology in Biomedicine* 13.6 (2009), pp. 864–873. DOI: 10.1109/TITB.2009.2033471.

[32] Fabian Pedregosa et al. "Scikit-learn: Machine learning in Python". In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[33] Sara Pietracupa et al. "Poor self-awareness of levodopa-induced dyskinesias in Parkinson's disease: clinical features and mechanisms". en. In: *Parkinsonism Relat Disord* 19.11 (July 2013), pp. 1004–1008.

[34] Ronald B Postuma et al. "Validation of the MDS clinical diagnostic criteria for Parkinson's disease". en. In: *Mov Disord* 33.10 (Aug. 2018), pp. 1601–1608.

[35] Rob Powers et al. "Smartwatch inertial sensors continuously monitor real-world motor fluctuations in Parkinson's disease". en. In: *Sci Transl Med* 13.579 (Feb. 2021).

[36] Antoine Regnault et al. "Does the MDS-UPDRS provide the precision to assess progression in early Parkinson's disease? Learnings from the Parkinson's progression marker initiative cohort". en. In: *J Neurol* 266.8 (May 2019), pp. 1927–1936.

[37] Giovanni Rizzo et al. “Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis”. en. In: *Neurology* 86.6 (Jan. 2016), pp. 566–576.

[38] Alejandro Rodrıguez-Molinero et al. “Estimating dyskinesia severity in Parkinson’s disease by using a waist-worn sensor: concurrent validity study”. en. In: *Sci Rep* 9.1 (Sept. 2019), p. 13434.

[39] Durgesh Srivastava and Lekha Bhambhu. “Data classification using support vector machine”. In: *Journal of Theoretical and Applied Information Technology* 12 (Feb. 2010), pp. 1–7.

[40] P. Stoica and R.L. Moses. *Spectral Analysis of Signals*. Pearson Prentice Hall, 2005. ISBN: 9780131139565. URL: https://books.google.nl/books?id=h78ZAQAAIAAJ.

[41] Kazushi Takahashi. “The therapy of the motor symptoms in the advanced stage of Parkinson’s disease”. In: *Nihon rinsho. Japanese journal of clinical medicine* 75 (Jan. 2017), pp. 77–82.

[42] Basilio Vescio et al. “Wearable Devices for Assessment of Tremor”. en. In: *Front Neurol* 12 (June 2021), p. 680011.

[43] Wai Yin Wong, Man Sang Wong, and Kam Ho Lo. “Clinical applications of sensors for human posture and movement analysis: a review”. en. In: *Prosthet Orthot Int* 31.1 (Mar. 2007), pp. 62–75.

[44] Xindong Wu et al. “Top 10 algorithms in data mining”. In: *Knowledge and information systems* 14.1 (2008), pp. 1–37.

[45] Xingguo Xiong, Yu-Liang Wu, and Wen-Ben Jone. “Material Fatigue and Reliability of MEMS Accelerometers”. In: *2008 IEEE International Symposium on Defect and Fault Tolerance of VLSI Systems*. 2008, pp. 314–322. DOI: 10.1109/DFT.2008.37.