INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Explainable Artificial Intelligence for Parkinson's Disease Diagnosis Model

Beatriz Rodrigues Godinho

Master in Data Science,

Supervisor:
Doctor Diana Aldea Mendes, Associate Professor,
Iscte – Instituto Universitário de Lisboa

September, 2025

BUSINESS SCHOOL

TECNOLOGIAS E ARQUITETURA

Department of Quantitative Methods for Management and Economics

Department of Information Science and Technology

Explainable Artificial Intelligence for Parkinson's Disease Diagnosis Model

Beatriz Rodrigues Godinho

Master in Data Science,

Supervisor:
Doctor Diana Aldea Mendes, Associate Professor,
Iscte – Instituto Universitário de Lisboa

September, 2025

Acknowledgment

Like everything in life, this work was not accomplished alone. I was fortunate to be surrounded by people who supported me in different ways, some through their knowledge and guidance, others through their encouragement, kind words, and motivation.

Firstly, I want to express my appreciation to my supervisor, Professor Diana Mendes, for accepting to guide me through this project and for sharing insightful feedback and knowledge. It was a pleasure to work and learn with you.

To my parents, João and Ana, who have always believed in my capabilities and encouraged me to follow this academic journey. Especially, for teaching me that with effort and determination, we can achieve anything, we need to believe in ourselves and keep working. I am truly grateful to you both.

To Nuno, thank you for always supporting everything I do and for being there to listen whenever I need it. Part of this would not have been possible without your constant support, patience, and comprehension. Thank you for always managing to put a smile on my face, even during difficult moments. I am deeply grateful to have you in my life.

To my family and friends, for your care, friendship and encouragement that helped me stay grounded and motivated throughout this journey.

Resumo

A doença de Parkinson caracteriza-se pela presença de sintomas motores e não motores, sendo o seu diagnóstico complexo e suscetível a erros. De forma a apoiar este processo, têm sido desenvolvidos modelos de *Machine Learning*, que permitem distinguir doentes com Parkinson de outras patologias com sintomas semelhantes. O modelo em estudo utiliza dados de movimento para a deteção de sintomas motores e dados de um questionário para a deteção de sintomas não motores. Devido à sua complexidade, acrescem problemas de interpretabilidade nas previsões obtidas pelo modelo, sendo este um fator crucial para médicos e clínicos. O presente trabalho incidiu na aplicação de técnicas de inteligência artificial explicável, mais precisamente, S*hapley additive explanations.* Os resultados obtidos indicam que os sintomas não motores mais importantes para o modelo são a perda ou alteração da capacidade de saborear ou cheirar, disfunção urinária, perda de interesse e baba durante o dia. A análise local permitiu identificar padrões nas classificações incorretas, especialmente para doentes com Parkinson, enquanto a heterogeneidade do grupo com doenças diversas dificultou esta análise. Para o modelo com os dados do movimento, não foram obtidos resultados devido à heterogeneidade do pré-processamento de *features* entre *folds e* modelos*.*

PALAVRAS CHAVE: Doença de Parkinson, apoio à decisão clínica, sensores, inteligência artificial explicável, S*hapley additive explanations*

Abstract

Parkinson's disease is characterised by the presence of motor and non-motor symptoms within a complex diagnosis prone to errors. To support this process, Machine Learning models have been developed to distinguish Parkinson's patients from other pathologies with similar symptoms. The model under study uses movement data to detect motor symptoms and questionnaire data to identify non-motor symptoms. These developments enable clinical decision support; however, due to their complexity, interpreting the results can be challenging, which is a crucial factor for doctors and clinicians. This work focused on applying explainable artificial intelligence techniques, specifically, Shapley additive explanations. The results indicate that the most important non-motor symptoms for the model are loss or alteration of the ability to taste or smell, urinary dysfunction, loss of interest, and drooling during the day. Local analysis helped identify patterns in incorrect classifications, especially for patients with Parkinson's disease, while the heterogeneity of the group with different diseases made this analysis challenging. For the model with movement data, no results were obtained due to heterogeneity in feature preprocessing across folds and models.

KEYWORDS: Parkinson’s disease, clinical decision support, wearable sensors, explainable artificial intelligence, Shapley additive explanations

Contents

Acknowledgment i

Resumo iii

Abstract v

Contents vii

List of Figures ix

List of Tables xiii

List of Acronyms xv

Chapter 1. Introduction 1

1.1. Motivation 2

1.2. Objectives and Research Questions 2

1.3. Methodology 3

1.4. Document Structure 4

Chapter 2. Literature Review 5

2.1. Wearable Devices 5

2.2. Related Studies 6

2.3. Case Study 6

2.4. Explainability 7

Chapter 3. Data 9

3.1. Questionnaire Data Analysis 11

3.1.1. Personal Information 11

3.1.2. PDNMS 14

3.2. Smartwatch Data Analysis 16

Chapter 4. Modelling and Explainability 21

4.1. SHAP Explainer 22

4.1.1. Global Level 23

4.1.2. Local Level 24

Chapter 5. Discussion and Results 25

5.1. SHAP for the Questionnaire Model 25

5.1.1. Global Level 25
5.1.2. Local Level 26
5.2. SHAP for the Movement Model 30
Chapter 6. Conclusion 31
References 33
Annex A. Questionnaire Data Analysis 39
Annex B. Movement Data Analysis 45

List of Figures

Figure 3.1. Distribution of participants across groups 10
Figure 3.2. Distribution of diseases in the DD group 11
Figure 3.3. Participants' gender across groups 12
Figure 3.4. Responses to the PDNMS questionnaire across groups, aggregated by symptom category 15
Figure 3.5. V-Cramer correlation between the variables of the PDNMS questionnaire categories and the participant condition 16
Figure 3.6. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (first 10 seconds of the 1a step, see Table 3.5) by a participant from the HC group 18
Figure 3.7. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (first 10 seconds of the 1a step, see Table 3.5) by a participant from the PD group 18
Figure 3.8. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (first 10 seconds of the 1a step, see Table 3.5) by a participant from the DD group 19

Figure 5.1. SHAP top ten features for questionnaire model in training and test sets 26
Figure 5.2. SHAP local level for the ten misclassified PD participants on the questionnaire model 28
Figure 5.3. SHAP local level for the ten misclassified DD participants on the questionnaire model 29

Figure A.1. Boxplot of participants' age across groups 39
Figure A.2. PD and DD participants' age at diagnosis 39
Figure A.3. Boxplot of participants' height across groups 40
Figure A.4. Boxplot of participants' weight across groups 40
Figure A.5. Participants' handedness across groups 41
Figure A.6. Participants' appearance in kindship across groups. 41
Figure A.7. Participants' appearance in the first grade kindship across groups. 42
Figure A.8. Participants' effect of alcohol on tremor across groups 42
Figure A.9. Responses to the PDNMS questionnaire in the DD group, aggregated by symptom category 43

Figure B.1. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (last 10 seconds of the 1a step, see Table 3.5) by a participant from the HC group 45

Figure B.2. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (last 10 seconds of the 1a step, see Table 3.5) by a participant from the PD group 46
Figure B.3. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (last 10 seconds of the 1a step, see Table 3.5) by a participant from the DD group 46
Figure B.4. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (first 10 seconds of the 1b step, see Table 3.5) by a participant from the HC group 47
Figure B.5. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (first 10 seconds of the 1b step, see Table 3.5) by a participant from the PD group 47
Figure B.6. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (first 10 seconds of the 1b step, see Table 3.5) by a participant from the DD group 48
Figure B.7. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (last 10 seconds of the 1b step, see Table 3.5) by a participant from the HC group 48
Figure B.8. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (last 10 seconds of the 1b step, see Table 3.5) by a participant from the PD group 49
Figure B.9. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (last 10 seconds of the 1b step, see Table 3.5) by a participant from the DD group 49
Figure B.10. Acceleration and rotation on the three-dimensional axis during the execution of the "StretchHold" task (step 2, see Table 3.5) by a participant from the HC group 50
Figure B.11. Acceleration and rotation on the three-dimensional axis during the execution of the "StretchHold" task (step 2, see Table 3.5) by a participant from the PD group 50
Figure B.12. Acceleration and rotation on the three-dimensional axis during the execution of the "StretchHold" task (step 2, see Table 3.5) by a participant from the DD group 51
Figure B.13. Acceleration and rotation on the three-dimensional axis during the execution of the "HoldWeight" task (step 4, see Table 3.5) by a participant from the HC group 51
Figure B.14. Acceleration and rotation on the three-dimensional axis during the execution of the "HoldWeight" task (step 4, see Table 3.5) by a participant from the PD group 52
Figure B.15. Acceleration and rotation on the three-dimensional axis during the execution of the "HoldWeight" task (step 4, see Table 3.5) by a participant from the DD group 52
Figure B.16. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (first 10 seconds of the 10 step, see Table 3.5) by a participant from the HC group 53
Figure B.17. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (first 10 seconds of the 10 step, see Table 3.5) by a participant from the PD group 53
Figure B.18. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (first 10 seconds of the 10 step, see Table 3.5) by a participant from the DD group 54

Figure B.19. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (last 10 seconds of the 10 step, see Table 3.5) by a participant from the HC group 54
Figure B.20. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (last 10 seconds of the 10 step, see Table 3.5) by a participant from the PD group 55
Figure B.21. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (last 10 seconds of the 10 step, see Table 3.5) by a participant from the DD group 55
Figure B.22. Acceleration and rotation on the three-dimensional axis during the execution of the "DrinkGlass" task (step 6, see Table 3.5) by a participant from the HC group 56
Figure B.23. Acceleration and rotation on the three-dimensional axis during the execution of the "DrinkGlass" task (step 6, see Table 3.5) by a participant from the PD group 56
Figure B.24. Acceleration and rotation on the three-dimensional axis during the execution of the "DrinkGlass" task (step 6, see Table 3.5) by a participant from the DD group 57
Figure B.25. Acceleration and rotation on the three-dimensional axis during the execution of the "CrossArms" task (step 7, see Table 3.5) by a participant from the HC group 57
Figure B.26. Acceleration and rotation on the three-dimensional axis during the execution of the "CrossArms" task (step 7, see Table 3.5) by a participant from the PD group 58
Figure B.27. Acceleration and rotation on the three-dimensional axis during the execution of the "CrossArms" task (step 7, see Table 3.5) by a participant from the DD group 58
Figure B.28. Acceleration and rotation on the three-dimensional axis during the execution of the "TouchNose" task (step 9, see Table 3.5) by a participant from the HC group 59
Figure B.29. Acceleration and rotation on the three-dimensional axis during the execution of the "TouchNose" task (step 9, see Table 3.5) by a participant from the PD group 59
Figure B.30. Acceleration and rotation on the three-dimensional axis during the execution of the "TouchNose" task (step 9, see Table 3.5) by a participant from the DD group 60

List of Tables

Table 3.1. Pre-processed questionnaire and smartwatch data files characteristics 9
Table 3.2. Field description about the participants' personal information 11
Table 3.3. Participants' age and age at diagnosis across groups, in years 13
Table 3.4. Non-motor symptoms questions based on PDNMS categorised (adapted from Varghese *et al.*, 2024b) 14
Table 3.5. Smartwatch-based assessment steps with the duration in seconds (adapted from Varghese *et al.*, 2024b) 16

Table 4.1. Classification results in each outer test fold after using the best setup (adapted from Varghese *et al.*, 2024a) 21
Table 4.2. Average and standard deviation classification performance using cross-validation and classifier stacking of the questionnaire and smartwatch (adapted from Varghese *et al.*, 2024a) 22

List of Acronyms

BOSS	Bag-of-Symbolic-Fourier-Approximation-Symbols
CatBoost	Categorical Gradient-Boosting
CNN	Convolutional Neural Network
CRISP-DM	Cross-Industry Standard Process for Data Mining
CT	Computed Tomography
DD	Differential Diagnoses
DL	Deep Learning
GradCAM	Gradient-Weighted Class Activation Mapping
HC	Healthy Control
HY	Hoehn and Yahr
LIME	Local Interpretable Model-agnostic Explanations
LSTM	Long Short-Term Memory
MDS-UPDRS	Movement Disorder Society-Unified Parkinson's Disease Rating Scale
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NMS	Non-Motor Symptoms
NN	Neural Network
PADS	Parkinson’s Disease Smartwatch
PD	Parkinson’s Disease
PDNMS	Parkinson’s Disease Non-Motor Symptoms
PET	Positron Emission Tomography
RQ	Research Questions
SHAP	Shapley Additive Explanations
SVM	Support Vector Machine
XAI	eXplainable Artificial Intelligence

CHAPTER 1

Introduction

Parkinson's Disease (PD) is a progressive neurodegenerative disorder characterised by the loss of dopaminergic neurons in the *substantia nigra pars compacta* (Tansey *et al.*, 2022). This region, located in the midbrain, supplies the basal ganglia, which is crucial for motor control, with dopamine (Tansey *et al.*, 2022). When there is a deficit in dopamine production, typically when 50% to 70% of these cells have died, there is insufficient dopamine available, leading to the onset of PD symptoms (Beitz, 2014).

PD ranks as the second most prevalent neurodegenerative condition, with approximately six million individuals impacted in 2016 (Dorsey *et al.*, 2018). The disease primarily affects the elderly population, aged 65 to 70, with a higher prevalence in men than in women (Tysnes and Storstein, 2017). The main risk factors are associated with exposure to toxic substances, dairy products, melanoma, and traumatic brain injury (Ascherio and Schwarzschild, 2016).

Currently, no treatment can slow down the progress of the disease (Bloem *et al.*, 2021). Since there is no known cure, treatments, such as pharmacological, surgical, and therapeutic, can help to manage the symptoms and improve patients' well-being (Lee and Yankee, 2021). Levodopa is the most effective medication used as a dopamine replacement for the disease (Lee and Yankee, 2021). Other treatments can help to manage motor symptoms, such as exercise, physiotherapy, and speech therapy (Armstrong and Okun, 2020).

Although diagnosis is often associated with motor symptoms, such as resting tremor, rigidity, and bradykinesia, the disease also presents non-motor symptoms, including anosmia, constipation, depression, and sleep disorders (Simon *et al.*, 2020). As the disease progresses, additional complications arise, such as autonomic dysfunction, pain, and cognitive decline, significantly impacting the quality of life of patients (Simon *et al.*, 2020).

Due to the inexistence of a specific test for diagnosing PD, clinical diagnosis is a combination of the patient's medical history, clinical signs, and physical examination. Additionally, brain imaging techniques such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) scans, Positron Emission Tomography (PET) scans, and radiographs can aid in diagnosis (Tolosa *et al.*, 2006).

Diagnosis is further complemented by determining the disease stage after detecting PD. The most common scales are the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and the Hoehn and Yahr (HY) Rating Scale (Prashanth and Roy, 2018). The MDS-UPDRS comprises 65 items that evaluate both motor and non-motor features but does not indicate the stage of the disease (Prashanth and Roy, 2018). Otherwise, the HY scale categorises the stages of PD but

does not evaluate pertinent features (Prashanth and Roy, 2018).

1.1. Motivation

Assessment methods rely on subjective evaluations, making the diagnosis particularly challenging. Thus, achieving an accurate clinical diagnosis of PD remains hard, principally due to symptom overlaps with other Parkinsonian syndromes, variability in disease presentation, and the absence of definitive biomarkers (Tolosa *et al.*, 2021). Adler *et al.* (2014), cited in Belić *et al.* (2019), reported that clinical assessments for PD diagnosis exhibit a sensitivity[1] of 88% and a specificity[2] of 68%. Moreover, this error rate becomes even more pronounced in the early stages of the disease as the diagnosis is based on symptoms that appear in later stages (Rizzo *et al.*, 2016).

Machine Learning (ML) has emerged as a powerful tool to complement traditional clinical assessments, enabling clinicians to detect patterns that are often not evident. This allows for an early diagnosis, contributing to mitigating disease progression through personalised treatment (Avutu and Paul, 2022; Calderone *et al.*, 2024) (Gaba and Kaur, 2024).

Recently, wearable devices, such as smartwatches, have led to significant advancements in disease detection through affordable portable devices. These devices employ advanced sensors, such as accelerometers and gyroscopes, to record motor parameters, enabling the monitoring of tremors and other motor symptoms associated with PD (Anikwe *et al.*, 2022; Jafleh *et al.*, 2024). Monitoring the disease brings new possibilities for telemedicine, allowing clinicians to manage the medication effects and track disease progression. Also, it brings advantages for managing more productive follow-up appointments (Del Din *et al.*, 2021)

The study developed by Varghese *et al.* (2024a) explored smartwatch data for the detection of PD. Data was collected at a hospital in Münster, Germany, distributed across three groups: PD, Differential Diagnoses (DD), and Healthy Control (HC). In addition to the smartwatch data, participants completed a questionnaire based on Parkinson's Disease Non-Motor Symptoms (PDNMS). ML models were applied after processing the data, achieving a balanced accuracy of 91.16% for the PD *vs.* HC task and 72.42% for the PD *vs.* DD task.

1.2. Objectives and Research Questions

The interpretation of the predictions generated by ML models can be complex, especially in black-box models like ensemble models. This is critical in the medical field, as it is essential to ensure clinicians' trust in these technologies in disease diagnosis support (Rana *et al.*, 2022). Without this

[1] Performance metric that evaluates the positive labels correctly classified.
[2] Performance metric that evaluates the negative labels correctly classified.

trust, acceptance in the clinical area might be compromised. Consequently, eXplainable Artificial Intelligence (XAI) methods have been explored to enhance this level of interpretability and address model transparency. These methods provide explanations of the contribution of each variable to the model and the interpretability of a generated prediction.

This research aims to investigate the principal XAI methods to enhance model explainability. The case study that will be explored is based on the model proposed by Varghese *et al.* (2024a) for the PD *vs.* DD task. This application will ensure the transparency of the predictions generated by the model, which is essential for clinicians and developers.

To achieve this, the research questions (RQ) guiding the present dissertation are:

- **RQ1:** What are the most suitable XAI methods to interpret disease diagnosis ML models?
- **RQ2:** How do XAI methods enhance the understanding of the model Varghese *et al.* (2024a) developed in the PD *vs.* DD task?

These questions will be analysed and answered in the following chapters.

1.3. Methodology

The methodology employed to develop this research was based on the Cross-Industry Standard Process for Data Mining (CRISP-DM). This approach is commonly used in data science projects, which consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.

The Business Understanding phase involved comprehension of the current state-of-the-art applications for detecting PD. Additionally, we explored the most common XAI techniques and the different types of explainability.

Data Understanding focused on exploring the two distinct data sources: the questionnaire and the smartwatch. This phase aims to comprehend data distribution and identify possible patterns across variables.

The Data Preparation was not necessary since the original data source was used. Considerable effort had already been invested in preprocessing the data, thereby obviating the need for further transformation.

In the Modelling phase, the proposed model for the PD *vs*. DD task was analysed to understand the training process. This was followed by the selection and application of the XAI technique to improve explainability.

Finally, in the Evaluation phase, the results produced by the explainer were analysed and discussed to assess their relevance and interpretability. The last phase, Deployment, is outside the scope of this work.

1.4. Document Structure

The present dissertation is organised into six chapters. Chapter 2 provides a literature review aligned with the objectives. Chapter 3 presents the data analysis of the questionnaire and smartwatch data. Chapter 4 describes the modelling and application of explainers. The obtained results are discussed in Chapter 5. Finally, Chapter 6 concludes the main findings of the proposed RQ and explores possibilities for future research.

CHAPTER 2

Literature Review

ML models have been explored for detecting PD, leveraging different possibilities of data sources that combine motor and non-motor symptoms. The efficacy of these models depends on the quality of the data, which is critical to prevent learning incorrect patterns. This is a significant challenge, as the labelled data relies on clinical diagnoses with a considerable error rate (Adler *et al.*, 2014). To mitigate this issue, the diagnosis should ideally be conducted by movement disorders specialists with slightly lower error rates (Rizzo *et al.*, 2016). Still, guaranteeing accurately labelled data remains a challenge, particularly in the early stages of the disease.

The most frequent data sources explored in this field are based on motor symptoms, including voice recordings, movement data, handwriting patterns, and imaging modalities (Mei *et al.*, 2021). Integrating diverse data sources into ML models led to promising results in recognising a broad spectrum of symptoms for PD (Rana *et al.*, 2022). Pereira *et al.* (2019) highlighted that signal data collected via sensors is the most common approach in computer-assisted diagnosis.

2.1. Wearable Devices

In recent years, advancements in the Internet of Things have enabled the collection of substantial quantities of data from wearable devices. For instance, smartwatches have demonstrated a crucial role in detecting subtle tremors that may sometimes go undetected through clinical assessment (van Alen *et al.*, 2024).

These devices generally incorporate inertial sensors that require calibration to guarantee data quality. Varghese *et al.* (2021) did experimental tests to validate the accuracy of the smartwatch sensors in the Apple Watch Series 3 and 4. The authors performed 43 measurements, each lasting 20 seconds, using a Trillium Compact seismometer, two smartwatches and a high-precision shaker. The results demonstrated a strong concordance between the smartwatch sensors and the reference devices**.**

The optimal number of sensors used during the assessment is not yet consensual (Rovini *et al.*, 2017). While using more sensors provides a larger dataset and ensures data quality by allowing comparisons or serving as a backup in case a sensor fails, it makes the patient's experience less comfortable (Rovini *et al.*, 2017).

To the best of our knowledge, no studies have implemented this technology beyond clinical assessments. Despite the potential of these approaches, real-world applications remain limited,

primarily due to challenges such as the need for large-scale validation (Giannakopoulou *et al.*, 2022).

2.2. Related Studies

According to Mei *et al.* (2021), the most often used models with higher accuracy for detecting PD are Support Vector Machine (SVM), ensemble learning, and Neural Networks (NN). Additionally, Fraiwan *et al.* (2016), as cited in Belić *et al.* (2019), observed that NN is the best model for early diagnosis with an accuracy above 95%. However, it is essential to note that this study utilised a limited dataset of 42 patients. Although Deep Learning (DL) models have demonstrated reliable performance, they are time-consuming and require additional computer resources due to the complex architecture of NN. Furthermore, most data collected have less than 15 samples (Belić *et al.*, 2019), which tends to cause DL models to overfit in such small datasets (Aggarwal *et al.*, 2024).

Most studies focused on distinguishing PD from HC; however, differentiating PD from other Parkinsonism symptoms remains specifically challenging. Khalil *et al.* (2025) developed a model for this task and employed feature importance analysis, identifying the timed up-and-go feature as the most significant feature. With this model, the authors obtained a balanced accuracy of 82.4%.

2.3. Case Study

The case study explored in the present dissertation, initially developed by Varghese *et al.* (2024a), has a larger dataset of patients distributed across three groups: PD, DD, and HC. The data was collected under the supervision of nurses. The investigation focused on evaluating both motor and non-motor symptoms. To analyse non-motor symptoms, the participants answered 30 questions in an electronic questionnaire based on PDNMS from the International Parkinson and Movement Disorder Society. Additionally, a questionnaire with the participant's personal information was provided. For motor symptoms, sensor data were collected from two smartwatches (Apple Watch 4), one twisted during the execution of eleven motor tasks. The smartwatches collected data from acceleration and rotation in the 3-dimensional coordinate axes (x, y, and z). After processing and balancing the data, ML models were applied, achieving a balanced accuracy of 91.16% for the PD *vs*. HC task, and 72.42% for the PD *vs*. DD task.

Using DL, Yu *et al.* (2024) improved the accuracy compared to the model proposed by Varghese *et al.* (2024a) by 4.25% and 8.03% for PD *vs*. HC and PD *vs*. DD tasks, respectively. Also, the authors developed a triple classification task (PD *vs*. HC *vs*. DD) with an accuracy of 78.23%. Data augmentation was applied to balance the dataset. The developed model uses a Convolutional Neural Network (CNN) for extracting special features and a Long Short-Term Memory (LSTM) for capturing temporal features.

2.4. Explainability

Non-linear models tend to have higher accuracy and tend to lack transparency and explainability. Also, dimension reduction techniques can compromise the interpretability of the model by transforming it into a lower-dimensional representation, potentially obscuring critical relationships between variables (Burkart and Huber, 2021). This weakness can make it difficult for clinicians to accept and trust the predictions generated by these models (Adadi and Berrada, 2020). Also, the European Commission (2019) mentioned that a trustworthy model must respect all applicable laws and the seven proposed rules, one of which underlines the importance of transparency. To address this issue, XAI methods have been explored as they have the potential to give the necessary explainability of the model and the generated predictions (Markus *et al.*, 2021). In the context of healthcare, XAI is particularly important in clinical decision support systems, as it provides a rationale that allows users to understand why a model has produced a prediction, which is essential to trust and adoption among clinicians (Antoniadi *et al.*, 2021).

XAI applications have been widely explored in the health domain (Saarela and Podgorelec, 2024). Explanations can be implemented through model-specific or model-agnostic techniques (Rai, 2020). Model-specific techniques integrate interpretability directly into the model structure; conversely, model-agnostic techniques analyse the model's inputs and outputs without modifying its internal structure.

XAI techniques are divided into three levels of explanations: global, cohort-specific, and local (Ahmad *et al.*, 2018). Global explanations provide an overview of how the entire model makes predictions. Cohort-specific explanations target a specific subgroup; for example, in a model predicting PD, the cohort-specific group could focus on how the model predicts whether a woman has PD. Local explanations aim to clarify the explainability of a single prediction.

It is fundamental to follow the necessary levels of explainability for each application. According to Ahmad *et al.* (2018), disease progression may require explanations at global, cohort, and local levels.

Loh *et al.* (2022) conducted a systematic review of 99 articles using XAI in ML and DL models. Most of the explainers used were Shapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and Gradient-Weighted Class Activation Mapping (GradCAM).

SHAP has gained particular attention among these methods due to its strong mathematical foundations (Saarela and Podgorelec, 2024). It leverages concepts from cooperative game theory, specifically the Shapley value, to fairly give the contribution of each feature to a particular prediction in global or local explanation (Winter, 2002). More specifically, SHAP assigns an importance value to each feature by calculating its marginal contribution across all possible combinations of features. Despite its advantages, the exact computation of SHAP values can be computationally intensive. This

makes it not viable for complex models with many features. To address this, Lundberg and Lee (2017) proposed several approximation methods, including Shapley sampling values and Kernel SHAP, which are model-agnostic, as well as model-specific approaches such as deep SHAP and tree SHAP. These methods make SHAP feasible for practical use while preserving its interpretability.

Unlike SHAP, LIME provides only local explanations, making it helpful in understanding individual predictions (Ribeiro *et al.*, 2016). This explainer can be used in different types of models (Dieber and Kirrane, 2020).

GradCAM provides explanations for CNN models. It is applied to the final convolutional layer and generates a heatmap that highlights the areas of the image that are most important for the model's predictions (Loh *et al.*, 2022).

The field of explainability is limited by the lack of a standardised metric for quantifying and comparing the various techniques (Burkart and Huber, 2021). This absence of a universal evaluation framework makes it challenging to assess the effectiveness of different explainability methods.

CHAPTER 3

Data

Data[3] were collected at the University Hospital Münster in Germany, one of the country's largest centers for PD assessments (Varghese *et al.*, 2024a; Varghese *et al.*, 2024b; Goldberger *et al.*, 2000). This phase lasted three years, from 2018 to 2021, and involved two main data sources. The first was an electronic questionnaire, which included (1.1) participants' personal information and (1.2) 30 yes-or-no questions based on the PDNMS. The second was a movement-based assessment, in which participants performed 11 motor tasks while wearing smartwatches on both wrists.

Following the authors' pre-processing phase, the dataset comprises data from 469 participants. Table 3.1 summarises the structure of the pre-processed questionnaire and smartwatch data. Personal information is stored in a .csv file, with each row corresponding to a single participant. PDNMS responses are saved in .bin files with one file per participant, containing answers to the 30 questions. The smartwatch data follow the same format, with one .bin file per participant. Each file contains 132 rows, reflecting the combination of motor tasks performed with both wrists, recorded across three axes using two sensor types (acceleration and rotation). The 976 columns correspond to signal values recorded throughout the task duration.

Table 3.1. Pre-processed questionnaire and smartwatch data files characteristics

Data	**Number of files**	**Type**	**Shape per file[4]**	**Memory per file in kilobyte**
(1) Questionnaire data				
(1.1) Personal information	1	.csv	(469, 15)	46
(1.2) PDNMS	469	.bin	(30, 1)	1
(2) Smartwatch data	469	.bin	(132, 976)	504

Participants were clinically diagnosed by a board of neurologists in collaboration with a movement disorder senior specialist and grouped as follows:

- PD, encompassed a wide age range and varied times of diagnosis onset;

[3] https://physionet.org/content/parkinsons-disease-smartwatch/1.0.0/
[4] The number of rows and columns.

- DD, which includes neurological diseases such as essential tremor, multiple sclerosis, and atypical Parkinsonism, like progressive supranuclear palsy or multiple system atrophy, secondary causes of Parkinsonism;
- HC, were age-matched with the PD to ensure the validity of intergroup comparisons.

The distribution of the 469 participants across groups is shown in Figure 3.1. The PD group accounts for the majority of the sample (58.85%), followed by the DD group (24.31%), and the HC group (16.84%).

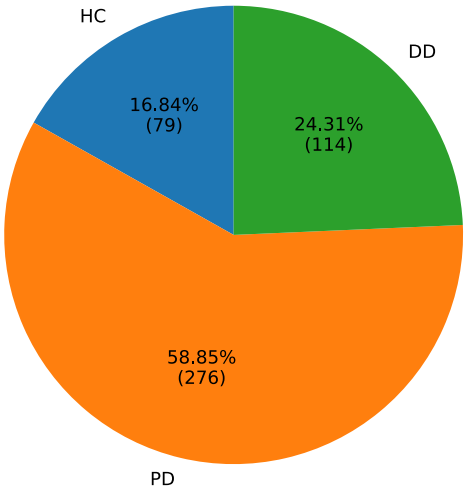


Figure 3.1. Distribution of participants across groups

DD group participants include several neurological disorders, with the most common being illustrated in Figure 3.2. Essential tremors are the most prevalent disease, affecting 24.56% of the group, followed by atypical Parkinsonism, 13.16%, and multiple sclerosis, 9.56%. Despite this, most DD participants are affected by other types of movement disorders, which are less prevalent in the dataset.

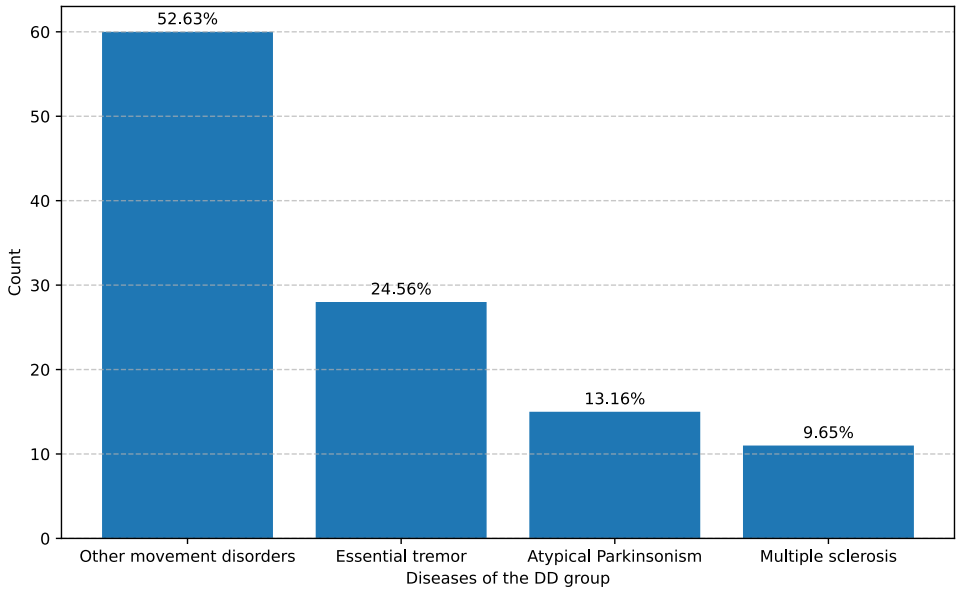


Figure 3.2. Distribution of diseases in the DD group

Each dataset will be analysed across the different groups in the following sections.

3.1. Questionnaire Data Analysis

Participants completed an electronic questionnaire divided into two steps: personal information and PDNMS.

3.1.1. Personal Information

The first part of the electronic questionnaire corresponds to questions related to personal information like age, height, weight, gender, handedness, kinship with PD, and the effect of alcohol on tremors. The columns of this dataset are described in Table 3.2.

Table 3.2. Field description about the participants' personal information

Field	Description
resource_type	The resource type. In this case, it is always "patient".
id	The participant's identification.
study_id	The identification of the study. In this case, it is "PADS[5]".
condition	The participant's condition.
disease_comment	Additional comments or notes on the disease, if any.
age_at_diagnosis	The participant's age at diagnosis with the condition.
age	The participant's current age.

[5] Parkinson's Disease Smartwatch (PADS).

Field	Description
height	The participant's height, in centimetres.
weight	The participant's weight, in kilograms.
gender	The participant's gender.
handedness	The participant's dominant hand (e.g., left or right).
appearance_in_kinship	Whether symptoms appear in any family member.
appearance_in_first_grade_kinship	Whether symptoms appear in first-degree relatives.
effect_of_alcohol_on_tremor	Indicates whether the participant reports any effect of alcohol consumption on tremor symptoms.
label	The diagnostic label assigned to the participant.

The gender distribution in each group is represented in Figure 3.3. Participants are predominantly male, corresponding to 59.91% of the total sample. The PD group has the most noticeable difference, with 41.58% males and 17.27% females of the total sample. This imbalance reflects the reality of PD, which tends to affect men more frequently than women. In contrast, the HC group has a higher proportion of females than males, whereas the DD group is equally balanced by gender.

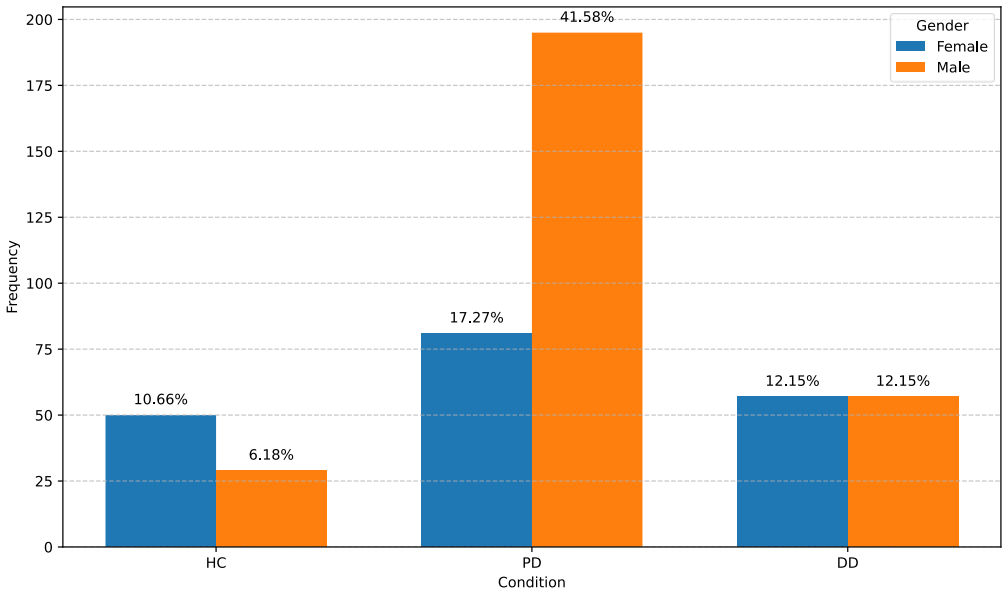


Figure 3.3. Participants' gender across groups

Table 3.3 describes the basic statistics (mean, standard deviation, minimum, and maximum values) of the numerical variables age and age at diagnosis. The participants' age ranges from 40 to 90 years old, and the diagnosis ages range from 0 to 82 years old. A minimum diagnosis age of 0 years

indicates the patient was diagnosed at birth, which is doubtful for these neurodegenerative conditions since they manifest in older ages. This likely reflects a data entry error, such as missing data.

HC has a mean age of 62.95 years (± 12.45) and no age at diagnosis, since they are not affected by a neurological condition.

PD has a slightly older age, with a mean of 65.36 years (± 9.64), and a mean age at diagnosis of 56.61 years (± 13.94), indicating that most were diagnosed approximately 9 years before the assessment.

DD has a similar mean age to the HC group, with a lower standard deviation (62.43 ± 1.55). The diagnosis age mean is 10 years less than that of the PD group, at 46.73 years (± 22.63). In the DD group, multiple sclerosis participants have a younger age and age at diagnosis with a mean of 54.09 (± 7.92) and 33.27 (± 19.78), respectively. On the other hand, atypical Parkinsonism has the oldest group ages (66.00 ± 7.96) and age at diagnosis (62.60 ± 8.21).

In conclusion, the mean ages are approximately consistent across the groups, suggesting that participants in each group are nearly the same age. Furthermore, the median ages across groups are close, which provides a comparable age distribution essential for interpreting the age-related effects (see Figure A.1 in Annex A). The mean age at diagnosis is lower in the DD group; however, it is essential to note that this variable exhibits an anomaly at age zero, which is present in both groups (see Figure A.2 in Annex A).

Table 3.3. Participants' age and age at diagnosis across groups, in years

Group	**Age** **mean ± std[6] (min; max)**	**Age at diagnosis** **mean ± std (min; max)**
HC	62.95 ± 12.45 (40; 90)	-
PD	65.36 ± 9.64 (42; 85)	56.61 ± 13.94 (0; 80)
DD	62.43 ± 1.55 (40; 85)	46.73 ± 22.63 (0; 82)
Atypical Parkinsonism	66.00 ± 7.96 (54; 79)	62.60 ± 8.21 (50; 75)
Essential tremor	66.18 ± 10.84 (45; 85)	46.96 ± 22.89 (3; 82)
Multiple sclerosis	54.09 ± 7.92 (40; 67)	33.27 ± 19.78 (0; 62)
Other Movement Disorders	61.32 ± 12.32 (41; 85)	45.12 ± 23.63 (0; 80)

These variables, along with the personal information (age, height, weight, gender, handedness, kinship with PD, and the effect of alcohol on tremor), were not included in the ML model since they yielded no significant results. For further analysis of these variables, see Annex A.

[6] Standard deviation.

3.1.2. PDNMS

The second step of the electronic questionnaire consists of 30 yes-or-no questions, where a value of 1 indicates that the participant has manifested the corresponding symptom, and 0 indicates the absence of that symptom. Based on symptom types, these questions were categorised into 10 categories as shown in Table 3.4.

Table 3.4. Non-motor symptoms questions based on PDNMS categorised (adapted from Varghese *et al.*, 2024b)

Category	**Question Numbers**	**Symptoms[7]**
Gastrointestinal tract	1, 3, 4, 5, 6, 7	Dribbling, swallowing, vomiting, constipation, bowel inconsistence, bowel emptying incomplete
Urinal tract	8, 9	Urgency, nocturia
Pain	10	Pains
Miscellaneous	11, 28, 29	Weight, sweating, diplopia
Apathy/attention/memory	12, 13, 15	Remembering, loss of interest, concentrating
Distortion of perception	2, 14,30	Taste/smelling, hallucinations, delusions
Depression/anxiety	16, 17	Sad blues, anxiety
Sexual function	18, 19	Sex drive, sex difficulty
Cardiovascular	20, 21, 27	Dizzy, falling, swelling
Sleep/fatigue	22, 23, 24, 25,26	Daytime sleepiness, insomnia, intense vivid, acting out during dreams, restless legs

The answers to the PDNMS questionnaire were analysed based on the symptom category (Table 3.4). A category was considered endorsed if a participant reported at least one symptom. The answers according to the distinct groups are presented in Figure 3.4. Most PD participants confirmed that they manifest symptoms in several categories, such as gastrointestinal, urinal tract, miscellaneous, apathy/attention/memory, distortion of perception, cardiovascular issues, and sleep/fatigue. Some of these symptoms overlap with those of the DD group, including gastrointestinal, urinal tract, apathy/attention/memory, cardiovascular, and sleep/fatigue.

Within the DD group (see Figure A.9 in Annex A), participants with atypical Parkinsonism also presented miscellaneous symptoms. Multiple sclerosis participants did not report gastrointestinal symptoms, but reported symptoms related to miscellaneous and sexual function categories.

[7] The symptoms are organised with the corresponding question number.

Additionally, participants with essential tremor did not report any cardiovascular symptoms.

In contrast, most of the HC participants did not report symptoms in any of these categories (Figure 3.4).

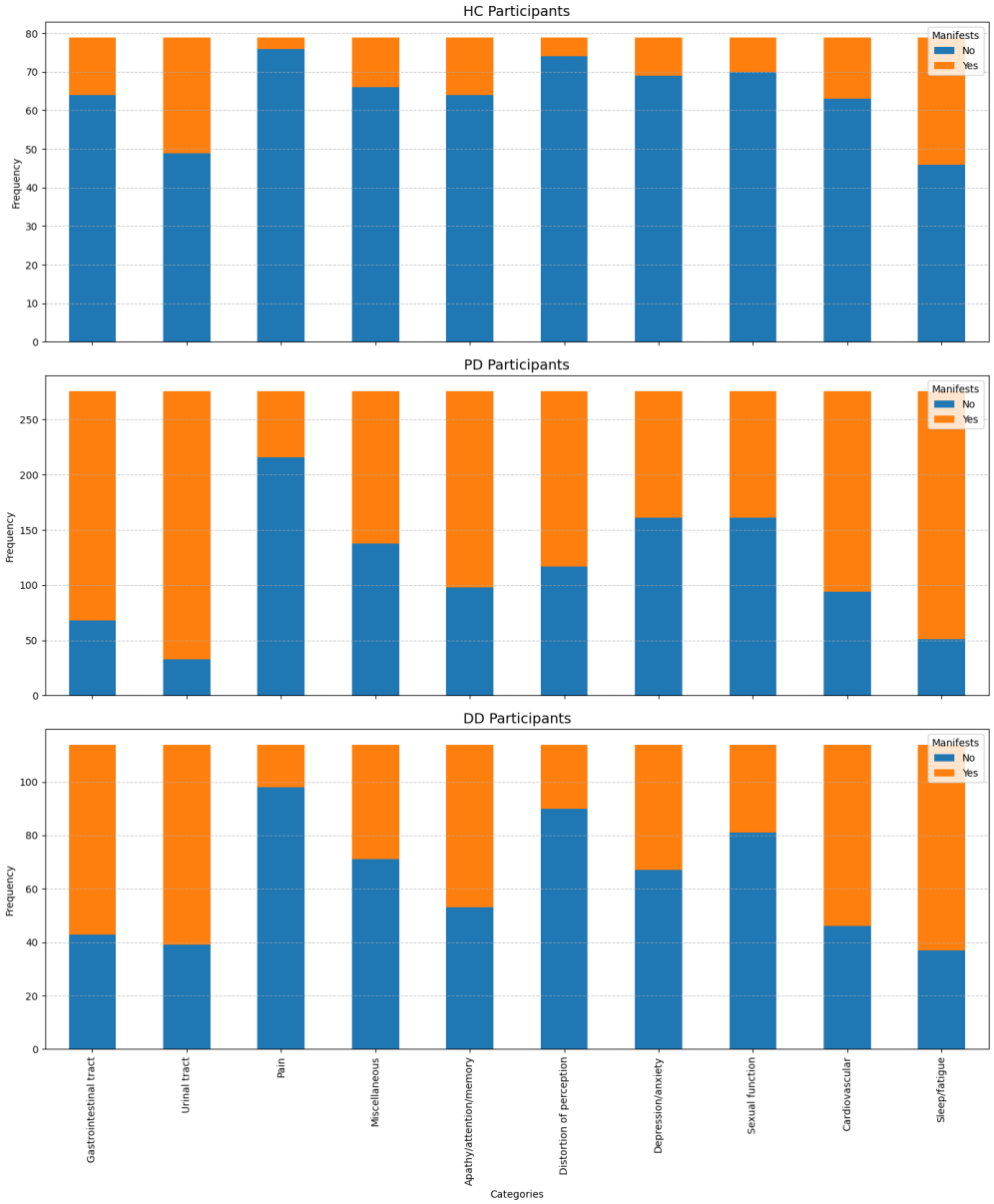


Figure 3.4. Responses to the PDNMS questionnaire across groups, aggregated by symptom category

To understand the association between the variables, the V-Cramer correlation was applied. The correlation values range from 0 to 1, where 0 indicates no association and 1 indicates a perfect association. Figure 3.5 shows the V-Cramer correlation matrix. The condition was a significant association (above 0.4) with the following categories: gastrointestinal, urinal tract and distortion of perception. Additionally, depression/anxiety showed a relationship with apathy/attention/loss of memory. The other variables did not show any significant relationships.

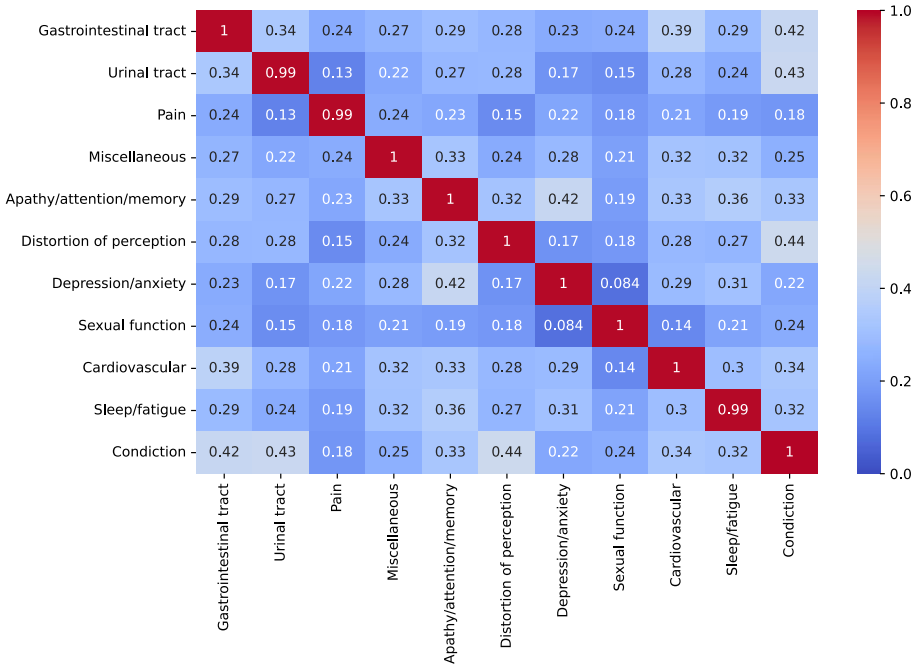


Figure 3.5. V-Cramer correlation between the variables of the PDNMS questionnaire categories and the participant condition

3.2. Smartwatch Data Analysis

Smartwatch data were collected using two Apple Watch Series 4 devices worn on each wrist while executing eleven motor steps. Table 3.5 describes the assessment steps, durations, and the respective categories (resting, postural, and kinetic). During each step, acceleration and rotation were recorded along the 3-dimensional axes, with durations of either 10 or 20 seconds, measured in milliseconds. To ensure consistent input windows for model training, the 20-second steps were divided into two 10-second segments. Steps three (lifting the arms), five (pointing the index finger), and eight (bringing the index fingers together) were removed based on the analysis by Varghese *et al.* (2024a), as they did not add additional value to the classification task. Consequently, considering the two wrists, with three axes and two measures of acceleration and rotation across eleven 10-second segments, we have 132 features for model training. Each of the features has 976 values for the signal-data.

Table 3.5. Smartwatch-based assessment steps with the duration in seconds (adapted from Varghese *et al.*, 2024b)

Step	Duration	Description	Category
1a	20	Resting with closed eyes while sitting, standardised to Zhang *et al.* (2018).	Resting
1b	20	Resting while patient is calculating serial sevens.	Resting

Step	Duration	Description	Category
2	10	Lift and extend arms according to Zhang *et al.* (2018).	Postural
3	10	Remain arms lifted.	Postural
4	10	Hold one-kilogram weight in each hand for 5 s. Start with the right hand. Then, have the arm rested again as in 1a.	Postural
5	10	Point index finger to the examiners lifted hand. Start with right index, then left, repeat the movement.	Kinetic
6	10	Drink from glass. Grasp an empty glass as if drinking from it. Start with the right hand. Then repeat with the left hand.	Kinetic
7	10	Cross and extend both arms.	Kinetic
8	10	Bring both index fingers together.	Kinetic
9	10	Tap own nose with index finger. Start with the right, then with left index. Then extend the arms.	Kinetic
10	20	Entrainment. The examiner stomps on the ground, setting the pace. Start stomping with the right foot according to the pace. Leave the arms extended during the movement. Repeat this with the left foot.	Postural

To analyse the smartwatch data, one participant was selected from each group, HC, PD, and DD (essential tremor), identified by IDs 295, 19, and 66, respectively. All three participants are male and 64 years old, ensuring consistency in age and gender across the groups for this comparison. The results for the *Relaxed1* task, corresponding to the first 10 seconds of step 1a (Table 3.5), are shown in Figures Figure 3.6, Figure 3.7 and Figure 3.8 for the HC, PD, and DD participants, respectively. The HC displayed low-amplitude oscillations in both acceleration and rotation, while the PD and DD participants showed broader variations during the task execution. Additionally, the PD participant exhibited more pronounced tremor in the left wrist, while the DD participant displayed tremor in both wrists. Further analysis of other tasks is provided in Annex B.

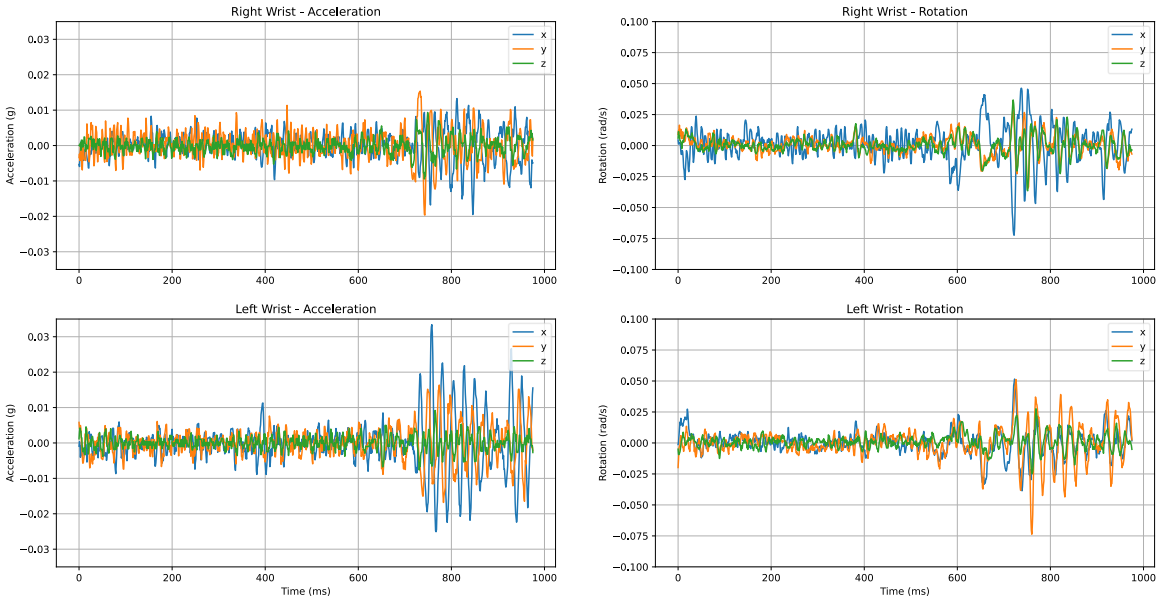


Figure 3.6. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (first 10 seconds of the 1a step, see Table 3.5) by a participant from the HC group

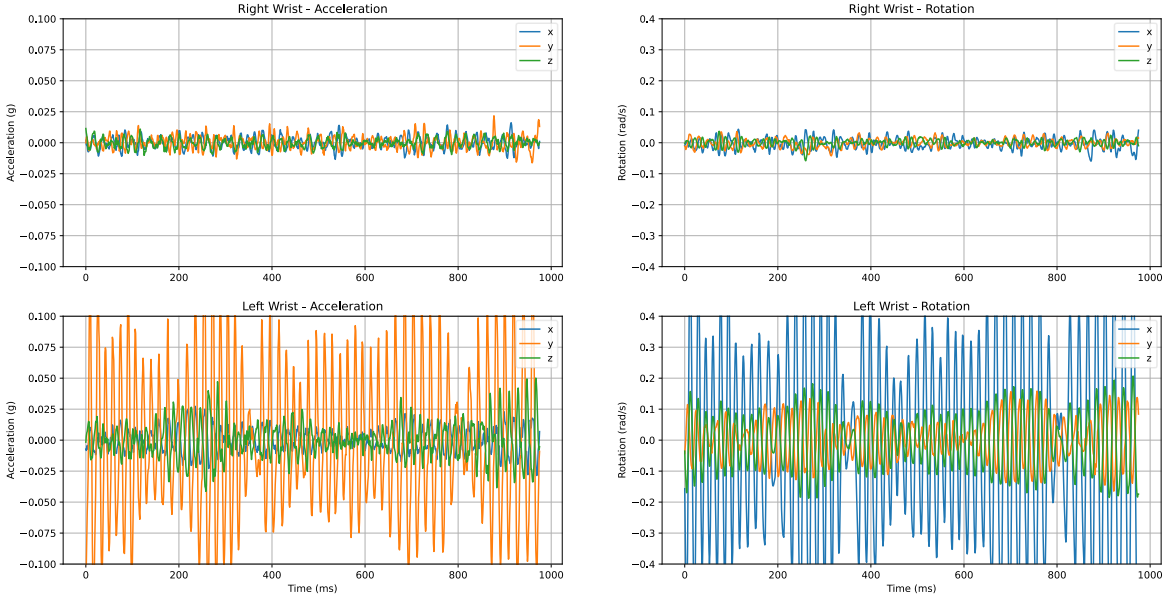


Figure 3.7. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (first 10 seconds of the 1a step, see Table 3.5) by a participant from the PD group

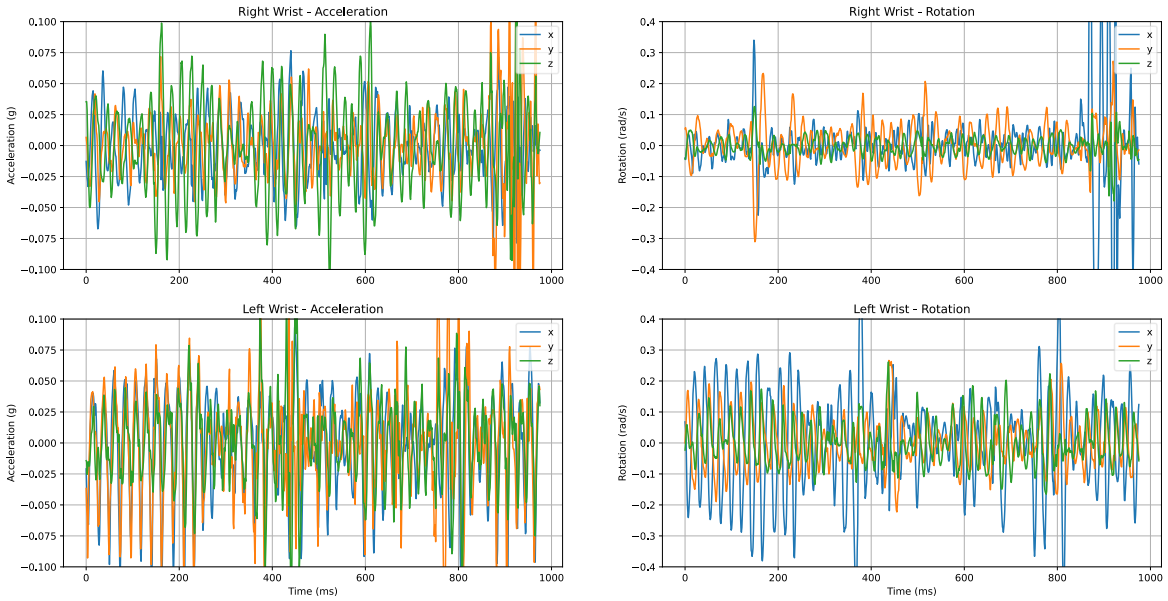


Figure 3.8. Acceleration and rotation on the three-dimensional axis during the execution of the “Relaxed” task (first 10 seconds of the 1a step, see Table 3.5) by a participant from the DD group

CHAPTER 4

Modelling and Explainability

Varghese *et al.* (2024a) performed a ML model for two distinct tasks: PD *vs*. HC and PD *vs*. DD. In each task, they trained two different models, one for the questionnaire data and the other for the smartwatch data. A stacking classifier technique was used to combine the outputs from the two models (questionnaire and movement) into a final model output.

In this dissertation, we will focus solely on analysing the task of PD *vs*. DD for improving model interpretability. Class labels were defined such that PD corresponds to class 0 and DD to class 1. The dataset includes 390 participants, comprising 276 with PD and 114 with DD.

To identify the best-performing model and classification pipeline, the authors used a nested five-fold validation strategy. Balanced accuracy served as the primary metric for model selection in each fold. Table 4.1 describes the best classification pipeline for each fold with the corresponding balance accuracy, F1-score, precision, and recall.

The questionnaire model employed Categorical Gradient-Boosting (CatBoost) which is well suited to handling categorical features, in this case, the 30 responses from the PDNMS questionnaire.

The movement model used smartwatch acceleration and rotation signal data collected during an 11-segment motor task. These signals yield 66 input features (11 segments × 2 wrists × 3 axes). In fold 2, where both acceleration and rotation were combined, the input dimensionality increased to 132. Across folds, the most effective movement models were SVMs and NN, each relying on distinct preprocessing data tecniques.

Table 4.1. Classification results in each outer test fold after using the best setup (adapted from Varghese *et al.*, 2024a)

Test fold	Input	Classification pipeline	Balanced accuracy	F1	Precision	Recall
Questionnaire data only						
#1	PDNMS	(D) CatBoost	74.68%	63.64%	63.64%	74.68%
#2	PDNMS	(D) CatBoost	60.47%	42.86%	47.37%	60.47%
#3	PDNMS	(D) CatBoost	68.62%	56.00%	51.85%	68.62%
#4	PDNMS	(D) CatBoost	68.06%	55.56%	48.39%	68.06%
#5	PDNMS	(D) CatBoost	67.00%	53.33%	54.55%	67.00%
Smartwatch data only						
#1	Rotation	(B) SVM	77.35%	68.29%	73.68%	63.64%
#2	Acceleration + Rotation	(B) SVM	74.98%	65.12%	70.00%	60.87%
#3	Acceleration	(A) NN	58.46%	43.14%	39.29%	47.83%

Test fold	Input	Classification pipeline	Balanced accuracy	F1	Precision	Recall
#4	Rotation	(B) SVM	68.46%	55.00%	64.71%	47.83%
#5	Rotation	(B) SVM	66.64%	52.38%	57.89%	47.83%
Options A, B, and C were applied to smartwatch-data only, option D was applied to questionnaire data only. A: Manually defined features, shallow machine learning, B: Automatic feature extraction via classical signal processing, C: Automatic feature extraction via deep learning (not listed due to underperformance), D: Decision-tree based classifier.						

The overall performance of the model is the average of the results obtained across the five folds. The final model results in a logistic regression stacking classifier between the questionnaire and movement models. Results of the questionnaire, movement and the final model are described in Table 4.2.

On average, the questionnaire model achieved a balanced accuracy of 67.77%, while the movement model slightly outperformed it, reaching 69.18%. When combined through stacking, the final model improved performance to a balanced accuracy of 72.42%, with notable gains in recall (67.71%), suggesting enhanced detection of DD cases.

Table 4.2. Average and standard deviation classification performance using cross-validation and classifier stacking of the questionnaire and smartwatch (adapted from Varghese *et al.*, 2024a)

Input	Balanced accuracy	F1	Precision	Recall
Questionnaire data	67.77% (5.06%)	54.28% (7.48%)	53.16% (6.51%)	56.21% (10.80%)
Smartwatch data	69.18% (7.45%)	56.79% (10.13%)	61.11% (13.57%)	53.6% (7.96%)
Questionnaire + Smartwatch	72.42% (12.33%)	60.45% (16.7%)	54.92% (15.05%)	67.71% (19.45%)

The models and experiments by the authors are available in a GitLab repository[8]. The code, implemented in Python, was used to replicate the PD *vs*. DD task, enabling further analysis and interpretation of the model results.

4.1. SHAP Explainer

To enhance model explainability, we utilised SHAP, which is a recognised method that operates at both global and local levels. At the global level, SHAP reveals how the model generates predictions across the entire dataset, highlighting the overall importance of features. At the local level, it allows a detailed examination of the features impacting an individual model prediction.

[8] https://imigitlab.uni-muenster.de/published/pads-project (last accessed on 15/10/2025)

In this study, SHAP values were calculated for each model, the questionnaire and movement, in both training and test datasets. For the questionnaire model, we calculated the SHAP[9] values. However, due to the high dimensionality of the movement model, we applied a variant called KernelSHAP, which is model-agnostic and significantly reduces computational costs. This variation can be applied to any model and reduces the computer costs. All calculations were performed using the *shap*[10] Python library (version 0.48.0).

SHAP values have positive and negative values that impact the final model prediction. A negative SHAP value indicates that it tends to be a PD (class with the lower value); on the opposite, a positive value indicates that it tends to be a DD (class with the higher value).

4.1.1. Global Level

To obtain global level explanations, we followed the approach proposed by Scheda and Diciotti (2022), which involved averaging the SHAP values across the training and test sets within each fold. Considering *N* to be 390 as the number of samples, *M* as the number of features, and *K* to be 5 as the number of folds. Algorithm 4.1 describes the pseudocode (Scheda and Diciotti, 2022). Aggregated SHAP values are stored in *train_folds_shap_values* and *test_folds_shap_values* for the training and test datasets, respectively.

Algorithm 4.1. Pseudocode for nested cross-validation SHAP values aggregation (adapted from Scheda and Diciotti, 2022)

```
X ← Dataset    # Data table with N samples and M features (N rows × M columns)
y ← Target
model ← Classifier
Explainer ← SHAP.Explainer()  # Shap function which computes SHAP values

train_folds_shap_values ← 0  # Initialised matrix (N rows × M columns)
test_folds_shap_values ← 0    # Initialised matrix (N rows × M columns)
fold_splits ← split(K)
innerCV (fold_splits, model)
outerCV (innerCV, K). fit(X,y)

for k in fold_splits do
   X_train, y_train ← fold_splits.train(k)
```

[9] We also know that there are some approximations of SHAP, like the Tree SHAP method. However, in this case, it was not necessary because the model used was simple enough and the exact SHAP values could be computed efficiently.

[10] https://pypi.org/project/shap/ (last accessed on 15/10/2025)

X_test, y_test ← fold_splits.test(k)

best_k_model ← outer_CV[k].best_model

best_k_model.fit(X_train, y_train)

train_shap_values ← Explainer(best_k_model(X_train))

test_shap_values ← Explainer(best_k_model(X_test))

train_folds_shap_values ← train_ folds_shap_values + train_shap_values/(K-1)

test_folds_shap_values ← test_ folds_shap_values + test_shap_values

end for

4.1.2. Local Level

Local explainers are used to explain individual predictions based on the contribution of the model's input variables.

In this research, we utilised a local explainer to gain a deeper understanding of the model's misclassification. Therefore, two types of errors were considered: misclassification of PD cases (where the model predicted DD, but the true label was PD) and misclassification of DD cases (where the model predicted PD, but the true label was DD). For each error type, a random sample of two misclassified participants per fold was selected, totalling ten participants. This approach enabled a detailed investigation of the possible factors driving incorrect predictions.

CHAPTER 5

Discussion and Results

This chapter is divided into two sections, presenting the results of the SHAP application for the questionnaire and movement models. Explainability is examined at both global and local levels. The global level analysis helps to identify which features most significantly influence the model's predictions. The local level explanations focus on understanding individual predictions, which were used to investigate the model's misclassifications. This analysis is relevant for both clinicians and developers, as it improves transparency regarding the model's predictions.

The SHAP analyses were executed locally on a computer with 64 GB of RAM. Computing the SHAP values for the questionnaire model required 9 hours and 14 minutes. For the movement model, the computation time increased substantially to 68 hours and 32 minutes, a difference that can be attributed to the model's complexity and the high dimensionality of the input features.

5.1. SHAP for the Questionnaire Model

5.1.1. Global Level

Firstly, we started by calculating the SHAP values for the questionnaire model within both training and test sets. Figure 5.1 illustrates the SHAP values for the top ten features at a global level for the training set (represented on the left) and the test set (represented on the right). The features are listed in order of overall importance and are labelled according to the Non-Motor Symptoms (NMS) questionnaire, followed by the corresponding question number (e.g., NMS_01). The top row is a beeswarm summary plot, where the x-axis represents the SHAP values and the y-axis lists the top ten most important features. Each point of the plot determines the SHAP for a feature row. The colour reflects the original feature value. The bottom row is a summary bar plot representing global feature importance, where the x-axis represents the mean absolute SHAP value for each feature across all given samples, and the y-axis represents the features.

As shown in Figure 5.1, there is no significant difference between the train and test sets in terms of SHAP values. The top-most important features are the same except for the last feature, which is NMS_22 in the training set and NMS_20 in the test set.

In the beeswarm summary plots, lower values on the features NMS_13 (loss of interest), NMS_03 (swallowing) and NMS_20 (dizzy), the model tends to predict the lower-class value; in other words, if the participant does not manifest that symptom, the model tends to classify as PD. On the other hand,

the features with higher values, such as NMS_02 (taste/smelling), NMS_08 (urgency), NMS_09 (nocturia), NMS_01 (dribbling), NMS_25 (acting out during dreams), NMS_05 (constipation), NMS_28 (sweating) and NMS_22 (daytime sleepiness), the model tends to predict that they have PD.

According to the summary bar plot, the most important features to the model are the related to the following categories: distortion of perception (feature NMS_02), urinal tract (features NMS_08 and NMS_09), apathy/attention/memory (feature NMS_13), gastrointestinal tract (features NMS_01, NMS_03 and NMS_05), sleep/fatigue (features NMS_25 and NMS_22 for the training set only), miscellaneous (NMS_28), and cardiovascular for the test set only (NMS_20).

In summary, the most important features that distinguish whether a participant has PD or DD are related to loss or change in the ability to taste or smell, urinary dysfunction, loss of interest, and saliva dribbling.

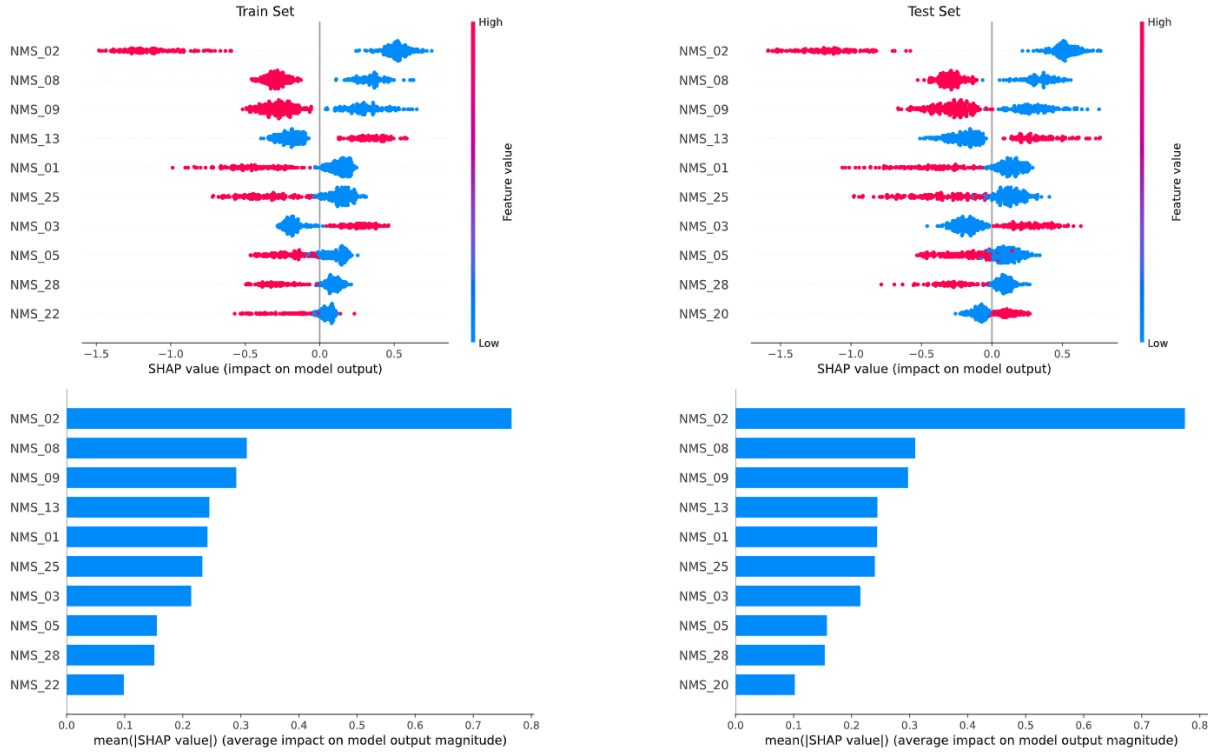


Figure 5.1. SHAP top ten features for questionnaire model in training and test sets

5.1.2. Local Level

For local level explainability, a waterfall SHAP plot was used. This plot displays how individual features impact the model's prediction from the baseline to the final output. Each row represents the feature contribution, with red bars indicating a positive contribution and blue bars indicating a negative contribution. Features are ordered by SHAP value magnitude, with the least important ones grouped at the bottom.

The ten misclassified PD participants are illustrated in Figure 5.2. Each row of the figure represents two examples of misclassified PD participants in a fold. The features NMS_02, NMS_08, NMS_09, and NMS_03 are the most essential features that contribute incorrectly to the model prediction. Looking at Figure 5.1 these features are also the most important for a global-level explanation.

Particularly, NMS_02 is the most influential feature present in nine out of ten of the sample. This feature, related to the loss or change in the ability to taste or smell, tends to manifest in PD participants (Figure 5.1). However, Figure 5.2 shows that in most cases, the value for this feature is zero, contributing to the model's prediction being a DD instead of a PD.

The misclassified DD participants are shown in Figure 5.3. Within this group, one participant has essential tremor (ID 91), two have atypical Parkinsonism (IDs 15 and 114), two have multiple sclerosis (IDs 421 and 432), and five have other movement disorders (IDs 11, 16, 273, 298, and 314).

In contrast to the misclassified PD cases, there is no clear pattern or shared feature among the misclassified DD participants. This may be due to the heterogeneity within the DD group, as it includes a broad spectrum of neurological conditions with diverse symptoms, making this type of analysis more challenging. Additionally, since the number of participants in the sample is small (at most two), it becomes hard to draw any general conclusions.

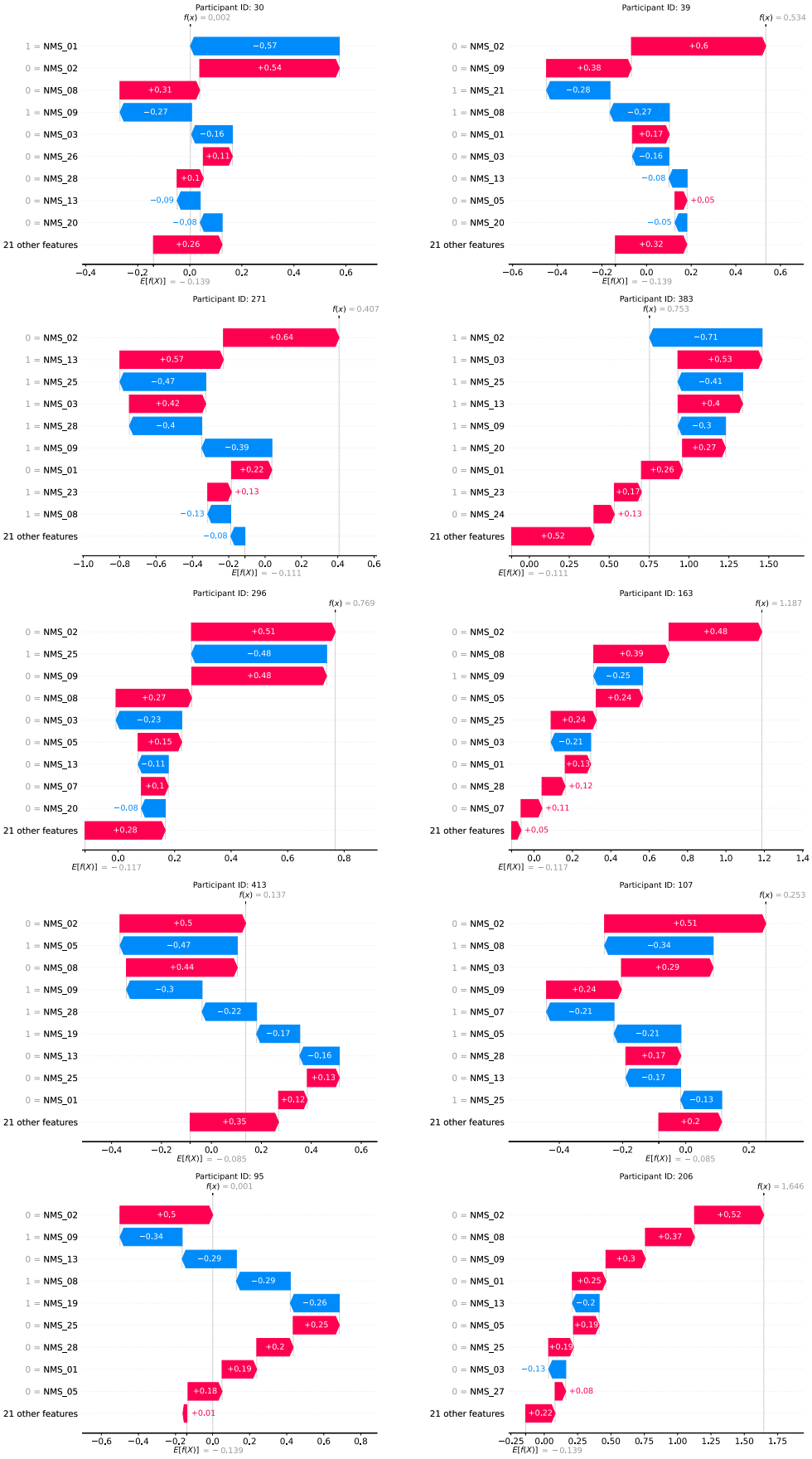


Figure 5.2. SHAP local level for the ten misclassified PD participants on the questionnaire model

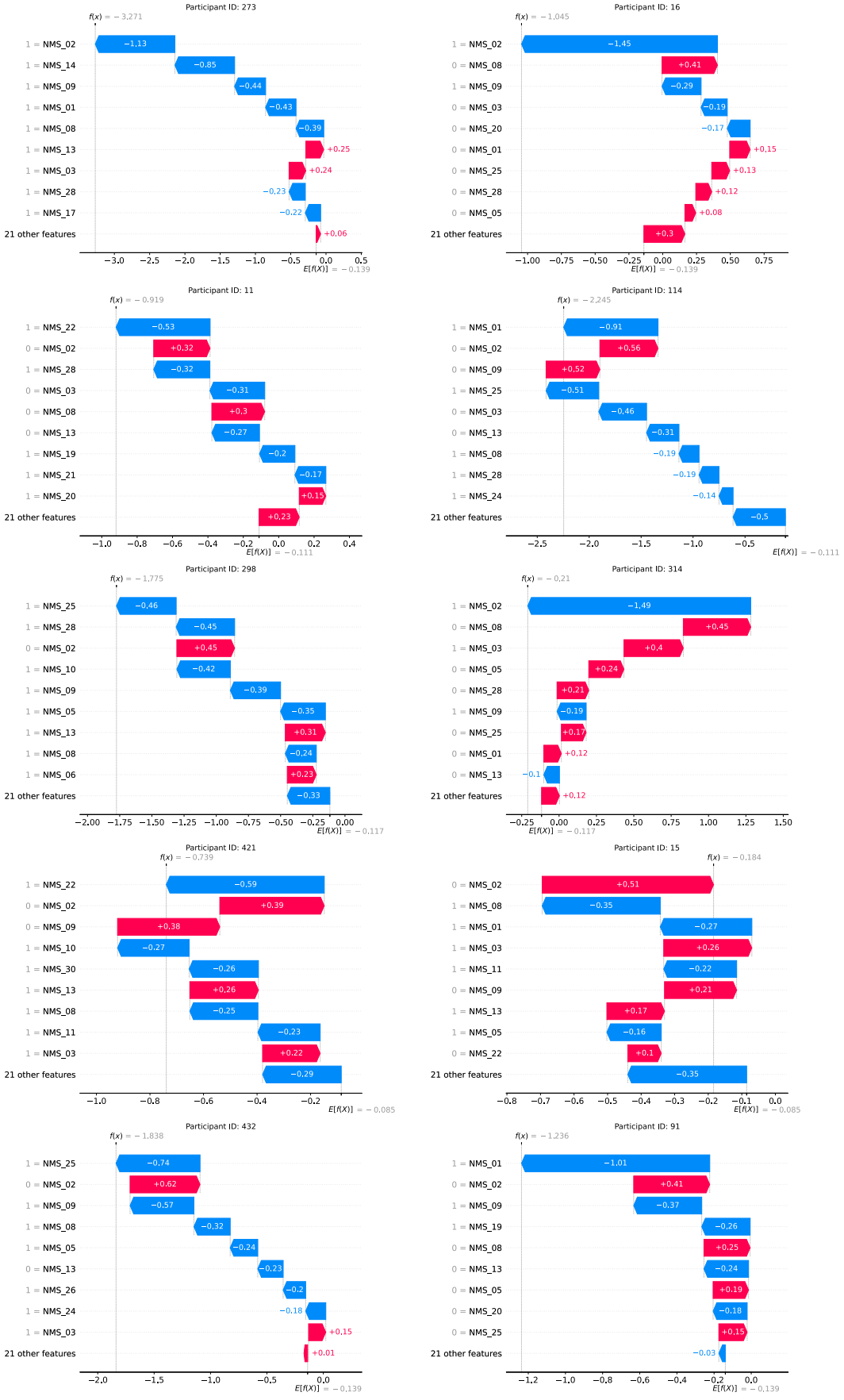


Figure 5.3. SHAP local level for the ten misclassified DD participants on the questionnaire model

5.2. SHAP for the Movement Model

The movement model combined two different ML models with different preprocessing pipelines and three different combinations of input data (Table 4.1).

In the SVM model, the preprocessing incorporated a Bag-of-Symbolic-Fourier-Approximation-Symbols (BOSS) algorithm, which significantly increased the number of features. Specifically, three window sizes, three bin sizes, and two word sizes were considered. In the case of the rotation, the data was transformed from 66 features to 1782 features.

In contrast, the NN model used a different feature extraction strategy. Each original feature went through a series of transformations based on statistical and signal processing descriptors. These included band power (19 features), windowed standard deviation (4 features), windowed absolute energy (4 features), and windowed absolute maximum (4 features). Totalling 2046 input features for model training instead of the original 66 features for acceleration data.

This heterogeneity in model features causes significant challenges for interpreting SHAP values across folds. The transformed features differ not only in structure but also in meaning, which makes the approach used in the questionnaire model unreliable. Although attempts were made to interpret these results, this was ultimately infeasible due to both computational cost and time constraints.

Future work should therefore focus on systematically analysing and normalising the transformed features to improve the explainability and overall interpretability of the movement model.

CHAPTER 6

Conclusion

In this work, we analysed the application of XAI techniques to enhance interpretability in the model's task, PD *vs*. DD. Firstly, we started by analysing the data used for model training. A total of 276 participants with PD and 114 participants with DD were identified. Additionally, the PD group has more men than women. The age across groups seems to be well-balanced.

Secondly, we analysed the proposed model, which uses a stacking classifier that combines the results of two distinct models: a questionnaire-based model and a movement-based model using smartwatch data. Additionally, a nested cross-validation was used during model training and evaluation to select the optimal classification pipeline in each fold.

Lastly, we applied the SHAP explainer to enhance model transparency, which allows for both global and local levels of explainability. While the global level helped to understand the overall features' importance, the local level was used to understand the model's misclassification errors.

The questionnaire model used 30 binary questions related to non-motor symptoms and the Catboost model. The global level results showed that the questions that most contribute to the model prediction are related to loss or change in the ability to taste or smell, urinary dysfunction, loss of interest, and dribbling of saliva during the daytime. For participants who report these symptoms, except for loss of interest, the model tends to predict a diagnosis of PD. On the other hand, when a participant reports loss of interest, the model is more likely to predict a DD diagnosis. These questions are related to the categories of distortion of perception, urinal tract, apathy/attention/memory, and gastrointestinal. All these categories, except apathy/attention/memory, also showed the strongest correlation with the condition variable, according to the V-Cramer correlation matrix.

For the local interpretability of the questionnaire model, we examined cases where the model misclassified PD or DD. This analysis resulted in a ten-sample set for each error type.

In the misclassified PD, the most influential feature that contributes wrongly to the model prediction is the loss of taste and smell. This finding aligns with the global SHAP analysis, which also identified this feature as the most important. Other features also contributed to misclassification, though less frequently and with lower impact. These included urinary urgency, nocturia, and difficulties with swallowing.

In the DD misclassification, no single feature consistently appeared as the main contributor across the sample. This inconsistency may be explained by the heterogeneity within the DD group, which has

different diseases that can manifest different non-motor symptoms. As a result, it becomes challenging to identify which specific symptoms are leading to the model's prediction errors in these cases.

No interpretability results are reported to the movement model due to the high variability and complexity of the extracted features during the model training. Each pipeline used different pre-processing techniques for modelling training, resulting in heterogeneous features between folds.

Therefore, the proposed research questions were almost addressed. The RQ1 was fully addressed with the application of the SHAP explainer. However, RQ2, which focused on enhancing model explainability, was only discussed for the questionnaire model.

Following the present research, it is recommended that future work effectively incorporate explainability into the movement model. A key step in this process is a comprehensive analysis of the preprocessing strategies used during model training. This would enable a more meaningful aggregation of SHAP values, allowing for both global and local interpretability across different folds.

Furthermore, the final model, which combines the questionnaire and movement data using a stacking classifier approach, could be analysed for explainability. This could enhance transparency by clarifying which models and features contribute most to the final prediction.

References

Adadi, A., and Berrada, M. (2020). Explainable AI for healthcare: from black box to interpretable models. In *Embedded systems and artificial intelligence: proceedings of ESAI 2019, Fez, Morocco* (pp. 327-337). Springer Singapore. https://doi.org/10.1007/978-981-15-0947-6_31

Adler, C. H., Beach, T. G., Hentz, J. G., Shill, H. A., Caviness, J. N., Driver-Dunckley, E., … and Dugger, B. N. (2014). Low clinical diagnostic accuracy of early vs advanced Parkinson disease: clinicopathologic study. *Neurology*, 83(5), 406-412. https://doi.org/10.1212/WNL.0000000000000641

Aggarwal, N., Saini, B. S., and Gupta, S. (2024). Role of artificial intelligence techniques and neuroimaging modalities in detection of Parkinson's disease: a systematic review. *Cognitive Computation*, 16(4), 2078-2115. https://doi.org/10.1007/s12559-023-10175-y

Ahmad, M. A., Eckert, C., and Teredesai, A. (2018). Interpretable machine learning in healthcare. In *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics* (pp. 559-560). https://doi.org/10.1145/3233547.3233667

Anikwe, C. V., Nweke, H. F., Ikegwu, A. C., Egwuonwu, C. A., Onu, F. U., Alo, U. R., and Teh, Y. W. (2022). Mobile and wearable sensors for data-driven health monitoring system: State-of-the-art and future prospect. *Expert Systems with Applications,* 202, 117362. https://doi.org/10.1016/j.eswa.2022.117362

Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., and Mooney, C. (2021). Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. *Applied Sciences*, 11(11), 5088. https://doi.org/10.3390/app11115088

Armstrong, M. J., and Okun, M. S. (2020). Diagnosis and treatment of Parkinson disease: a review. *Jama*, 323(6), 548-560. https://doi.org/10.1001/jama.2019.22360

Ascherio, A., and Schwarzschild, M. A. (2016). The epidemiology of Parkinson's disease: risk factors and prevention. *The Lancet Neurology*, *15*(12), 1257-1272. https://doi.org/10.1016/S1474-4422(16)30230-7

Avutu, S.R. and Paul, S. (2022). Artificial Intelligence Algorithms for Healthcare and Neurorehabilitation Engineering. In: Husain, M.S., Adnan, M.H.B.M., Khan, M.Z., Shukla, S., Khan, F.U. (eds) Pervasive Healthcare. EAI/Springer Innovations in Communication and Computing. *Springer, Cham*., 103-118. https://doi.org/10.1007/978-3-030-77746-3_8

Belić, M., Bobić, V., Badža, M., Šolaja, N., Đurić-Jovičić, M., and Kostić, V. S. (2019). Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease—A review. *Clinical neurology and neurosurgery*, 184, 105442. https://doi.org/10.1016/j.clineuro.2019.105442

Beitz, J. M. (2014). Parkinson's disease: a review. *Front Biosci (Schol Ed)*, 6(1), 65-74. https://doi.org/10.2741/S415

Bloem, B. R., Okun, M. S., and Klein, C. (2021). Parkinson's disease. *The Lancet,* 397(10291), 2284-2303. https://doi.org/10.1016/S0140-6736(21)00218-X

Burkart, N., and Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70, 245-317. https://doi.org/10.1613/jair.1.12228

Calderone, A., Latella, D., Bonanno, M., Quartarone, A., Mojdehdehbaher, S., Celesti, A., and Calabrò, R. S. (2024). Towards Transforming Neurorehabilitation: The Impact of Artificial Intelligence on Diagnosis and Treatment of Neurological Disorders. *Biomedicines*, *12*(10), 2415. https://doi.org/10.3390/biomedicines12102415

Del Din, S., Kirk, C., Yarnall, A. J., Rochester, L., and Hausdorff, J. M. (2021). Body-worn sensors for remote monitoring of Parkinson's disease motor symptoms: vision, state of the art, and challenges ahead. *Journal of Parkinson's disease*, 11(s1), S35-S47. https://doi.org/10.3233/JPD-202471

Dieber, J., and Kirrane, S. (2020). Why model why? Assessing the strengths and limitations of LIME. *arXiv preprint arXiv:2012.00093*. https://doi.org/10.48550/arXiv.2012.00093

Dorsey, E. R., Elbaz, A., Nichols, E., Abbasi, N., Abd-Allah, F., Abdelalim, A., ... and Murray, C. J. (2018). Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, 17(11), 939-953. https://doi.org/10.1016/S1474-4422(18)30295-3

European Commission. (2019). Press release Artificial intelligence: Commission takes forward its work on ethics guidelines. https://ec.europa.eu/commission/presscorner/detail/en/ip_19_1893

Fraiwan, L., Khnouf, R., and Mashagbeh, A. R. (2016). Parkinson's disease hand tremor detection system for mobile application. *Journal of medical engineering and technology*, 40(3), 127-134. https://doi.org/10.3109/03091902.2016.1148792

Gaba, S., and Kaur, H. (2024). Machine Learning Techniques for Parkinson's Disease Prediction and Progression: A Comprehensive Review. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)* (pp. 430-436). IEEE. https://doi.org/10.1109/IC3SE62002.2024.10593626

Giannakopoulou, K. M., Roussaki, I., and Demestichas, K. (2022). Internet of things technologies and machine learning methods for Parkinson's disease diagnosis, monitoring and management: a systematic review. *Sensors*, 22(5), 1799. https://doi.org/10.3390/s22051799

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. C*irculation*, 101(23), e215-e220. https://doi.org/10.1161/01.CIR.101.23.e215

Khalil, R. M., Shulman, L. M., Gruber-Baldini, A. L., Reich, S. G., Savitt, J. M., Hausdorff, J. M., ... and Cummings, M. P. (2025). Applying Wearable Sensors and Machine Learning to the Diagnostic Challenge of Distinguishing Parkinson's Disease from Other Forms of Parkinsonism. *Biomedicines*, 13(3), 572. https://doi.org/10.3390/biomedicines13030572

Lee, T. K., and Yankee, E. L. (2021). A review on Parkinson's disease treatment. *Neuroimmunol and Neuroinflamm*, 8, 222-244. http://dx.doi.org/10.20517/2347-8659.2020.58

Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., and Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer methods and programs in biomedicine*, 226, 107161. https://doi.org/10.1016/j.cmpb.2022.107161

Lundberg, S. M., and Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874

Markus, A. F., Kors, J. A., and Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of biomedical informatics*, 113, 103655. https://doi.org/10.1016/j.jbi.2020.103655

Mei, J., Desrosiers, C., and Frasnelli, J. (2021). Machine learning for the diagnosis of Parkinson's disease: a review of literature. *Frontiers in aging neuroscience*, *13*, 633752. https://doi.org/10.3389/fnagi.2021.633752

Jafleh, E. A., Alnaqbi, F. A., Almaeeni, H. A., Faqeeh, S., Alzaabi, M. A., and Al Zaman, K. (2024). The Role of Wearable Devices in Chronic Disease Monitoring and Patient Care: A Comprehensive Review. *Cureus*, 16(9), e68921. https://doi.org/10.7759/cureus.68921

Pereira, C. R., Pereira, D. R., Weber, S. A., Hook, C., De Albuquerque, V. H. C., and Papa, J. P. (2019). A survey on computer-assisted Parkinson's disease diagnosis. *Artificial intelligence in medicine*, 95, 48-63. https://doi.org/10.1016/j.artmed.2018.08.007

Prashanth, R., and Roy, S. D. (2018). Novel and improved stage estimation in Parkinson's disease using clinical scales and machine learning. *Neurocomputing*, 305, 78-103. https://doi.org/10.1016/j.neucom.2018.04.049

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the academy of marketing science*, 48, 137-141. https://doi.org/10.1007/s11747-019-00710-5

Rana, A., Dumka, A., Singh, R., Panda, M. K., Priyadarshi, N., and Twala, B. (2022). Imperative role of machine learning algorithm for detection of Parkinson's disease: review, challenges and recommendations. *Diagnostics*, 12(8), 2003. https://doi.org/10.3390/diagnostics12082003

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778

Rizzo, G., Copetti, M., Arcuti, S., Martino, D., Fontana, A., and Logroscino, G. (2016). Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis. *Neurology*, 86(6), 566-576. https://doi.org/10.1212/WNL.0000000000002350

Rovini, E., Maremmani, C., and Cavallo, F. (2017). How wearable sensors can support Parkinson's disease diagnosis and treatment: a systematic review. *Frontiers in neuroscience*, 11, 555. https://doi.org/10.3389/fnins.2017.00555

Saarela, M., and Podgorelec, V. (2024). Recent Applications of Explainable AI (XAI): A Systematic Literature Review. *Applied Sciences*, 14(19), 8884. https://doi.org/10.3390/app14198884

Scheda, R., and Diciotti, S. (2022). Explanations of machine learning models in repeated nested cross-validation: an application in age prediction using brain complexity features. *Applied Sciences*, 12(13), 6681. https://doi.org/10.3390/app12136681

Simon, D. K., Tanner, C. M., and Brundin, P. (2020). Parkinson disease epidemiology, pathology, genetics, and pathophysiology. *Clinics in geriatric medicine*, 36(1), 1-12. https://doi.org/10.1016/j.cger.2019.08.002

Winter, E. (2002). The shapley value. *Handbook of game theory with economic applications*, 3, 2025-2054. https://doi.org/10.1016/S1574-0005(02)03016-3

Tansey, M. G., Wallings, R. L., Houser, M. C., Herrick, M. K., Keating, C. E., and Joers, V. (2022). Inflammation and immune dysfunction in Parkinson disease. *Nature Reviews Immunology*, *22*(11), 657-673. https://doi.org/10.1038/s41577-022-00684-6

Tolosa, E., Garrido, A., Scholz, S. W., and Poewe, W. (2021). Challenges in the diagnosis of Parkinson's disease. *The Lancet Neurology*, 20(5), 385-397. https://doi.org/10.1016/S1474-4422(21)00030-2

Tolosa, E., Wenning, G., and Poewe, W. (2006). The diagnosis of Parkinson's disease. *The Lancet Neurology*, 5(1), 75-86. https://doi.org/10.1016/S1474-4422(05)70285-4

Tysnes, O. B., and Storstein, A. (2017). Epidemiology of Parkinson's disease. *Journal of neural transmission*, 124, 901-905. https://doi.org/10.1007/s00702-017-1686-y

van Alen, C. M., Brenner, A., Warnecke, T., and Varghese, J. (2024). Smartwatch Versus Routine Tremor Documentation: Descriptive Comparison. *JMIR Formative Research*, 8(1), e51249. https://doi.org/10.2196/51249

Varghese, J., Alen, C. M. V., Fujarski, M., Schlake, G. S., Sucker, J., Warnecke, T., and Thomas, C. (2021). Sensor validation and diagnostic potential of smartwatches in movement disorders. *Sensors*, 21(9), 3139. https://doi.org/10.3390/s21093139

Varghese, J., Brenner, A., Fujarski, M., van Alen, C. M., Plagwitz, L., and Warnecke, T. (2024a). Machine Learning in the Parkinson's disease smartwatch (PADS) dataset. *npj Parkinson's Disease*, 10(1), 9. https://doi.org/10.1038/s41531-023-00625-7

Varghese, J., Brenner, A., Plagwitz, L., van Alen, C., Fujarski, M., and Warnecke, T. (2024b). PADS - Parkinsons Disease Smartwatch dataset (version 1.0.0). PhysioNet. https://doi.org/10.13026/m0w9-zx22

Yu, J., Meng, K., Liang, T., Liu, H., and Wang, X. (2024). Improved Deep Learning for Parkinson's Diagnosis Based on Wearable Sensors. *Electronics*, 13(23), 4638. https://doi.org/10.3390/electronics13234638

Zhang, B., Huang, F., Liu, J., and Zhang, D. (2018). A novel posture for better differentiation between Parkinson's tremor and essential tremor. *Frontiers in neuroscience*, 12, 317. https://doi.org/10.3389/fnins.2018.00317

ANNEX A

Questionnaire Data Analysis

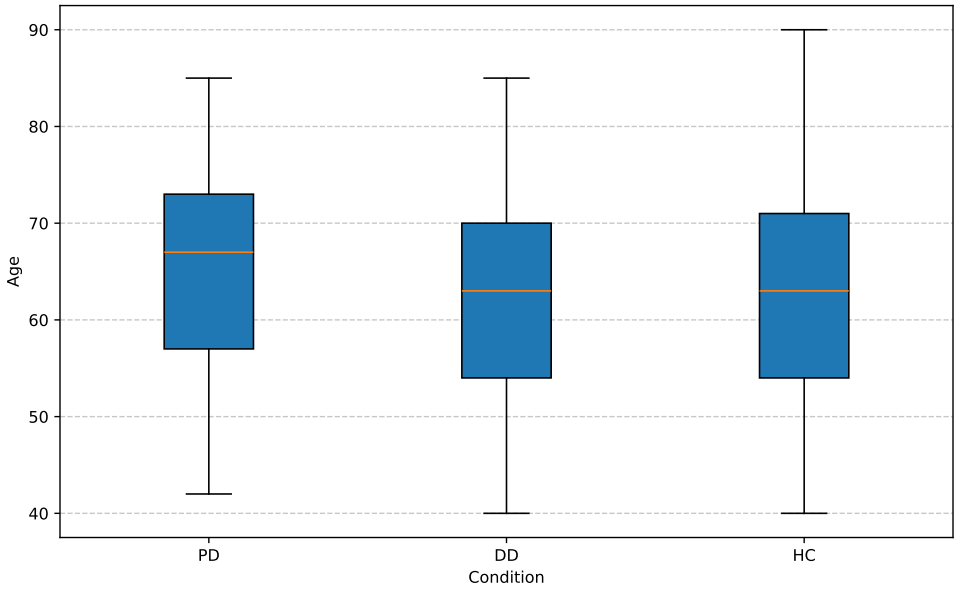


Figure A.1. Boxplot of participants' age across groups

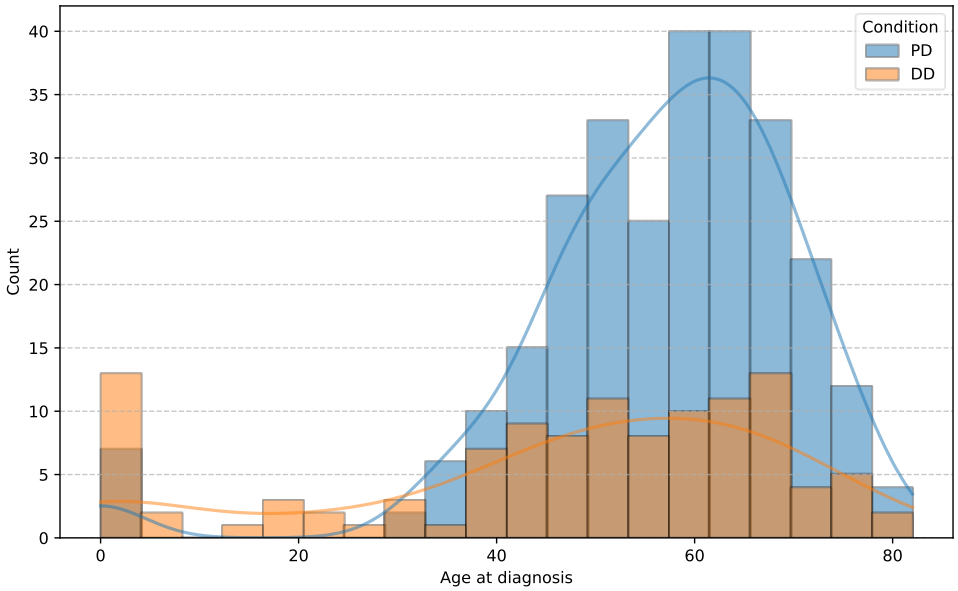


Figure A.2. PD and DD participants' age at diagnosis

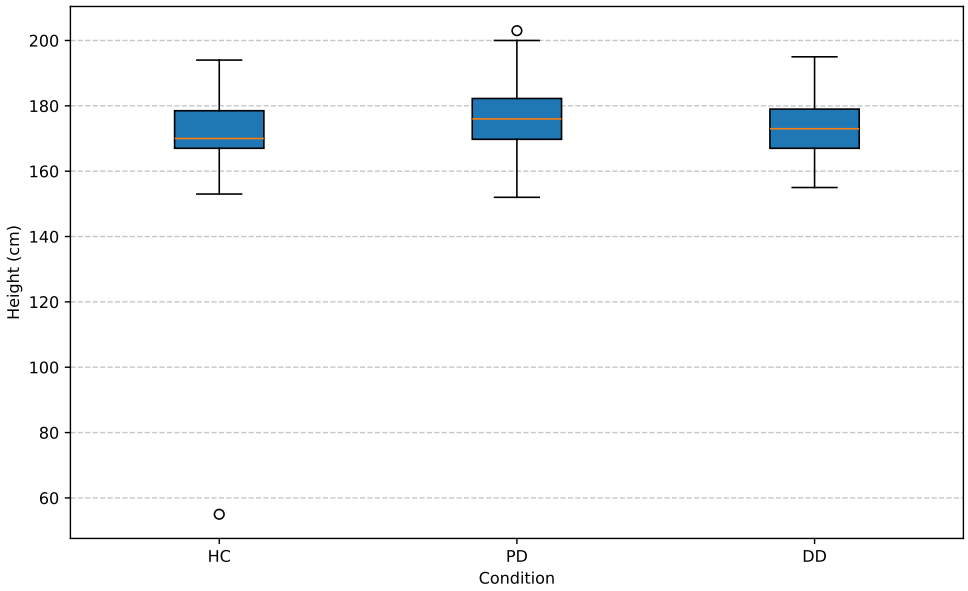


Figure A.3. Boxplot of participants' height across groups

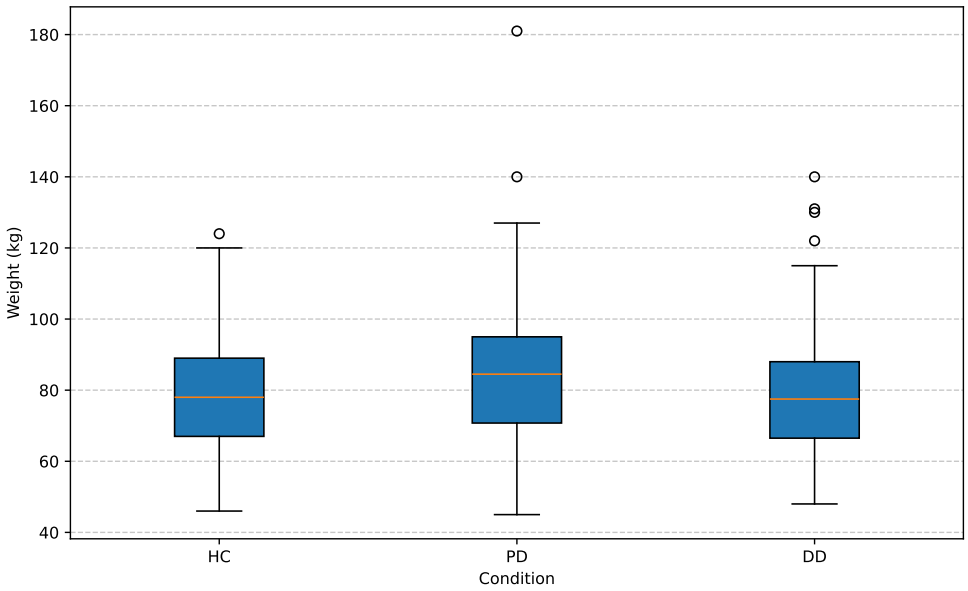


Figure A.4. Boxplot of participants' weight across groups

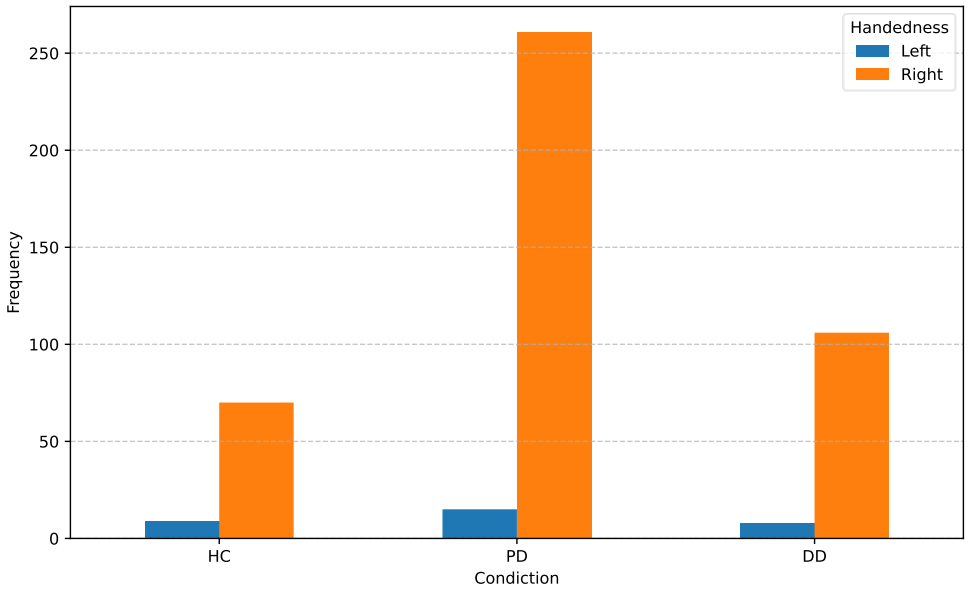


Figure A.5. Participants' handedness across groups

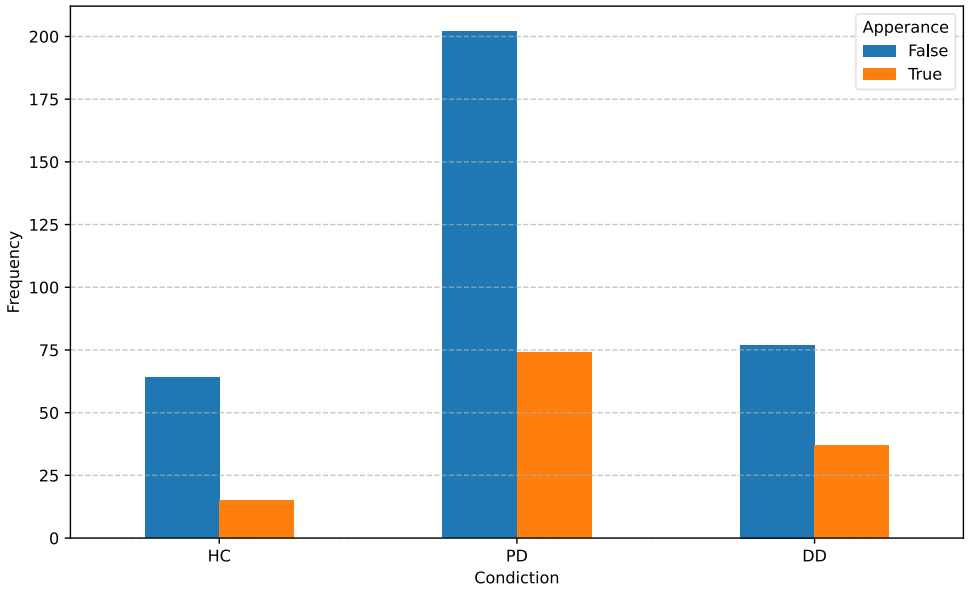


Figure A.6. Participants' appearance in kindship across groups.

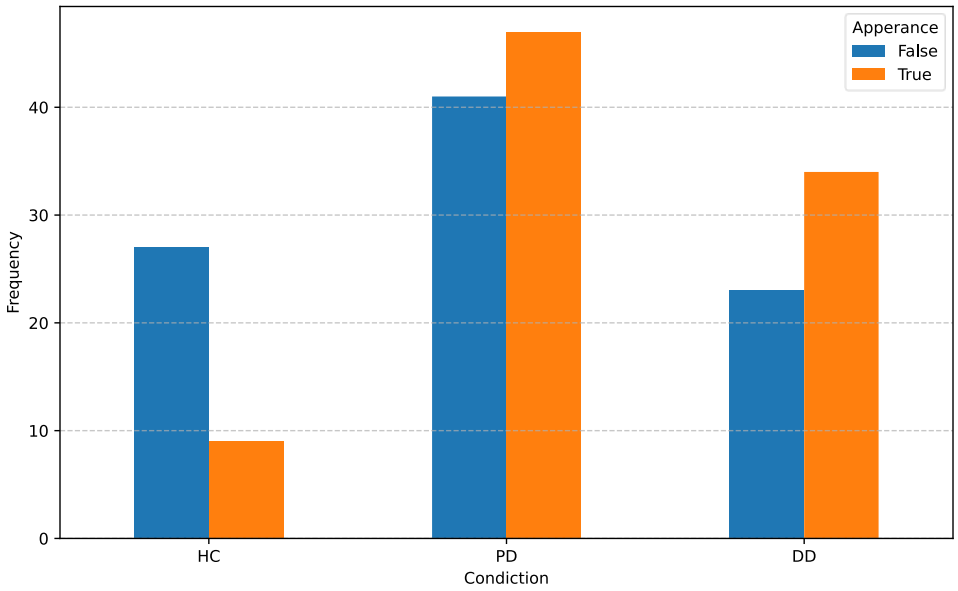


Figure A.7. Participants' appearance in the first grade kindship across groups.

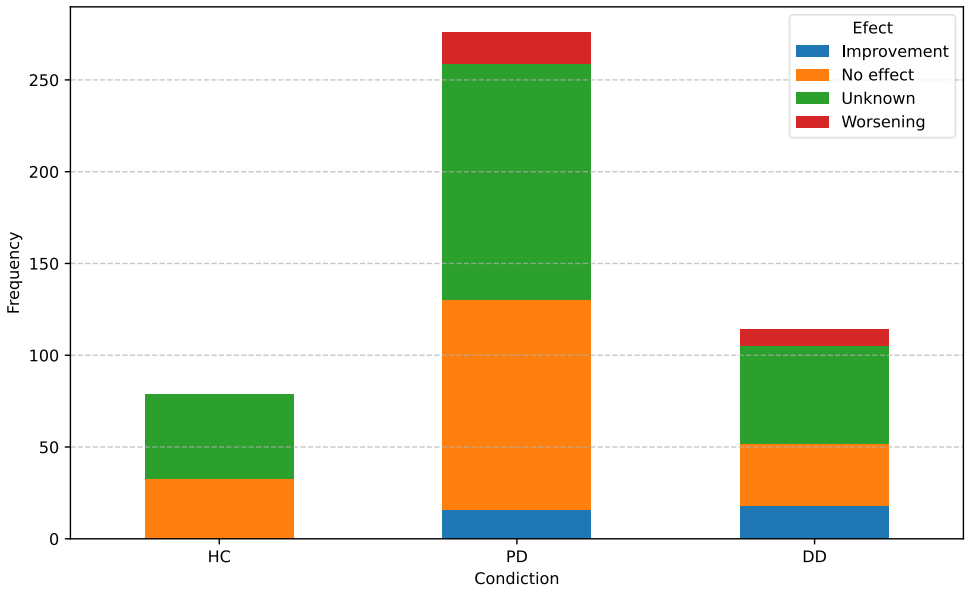


Figure A.8. Participants' effect of alcohol on tremor across groups

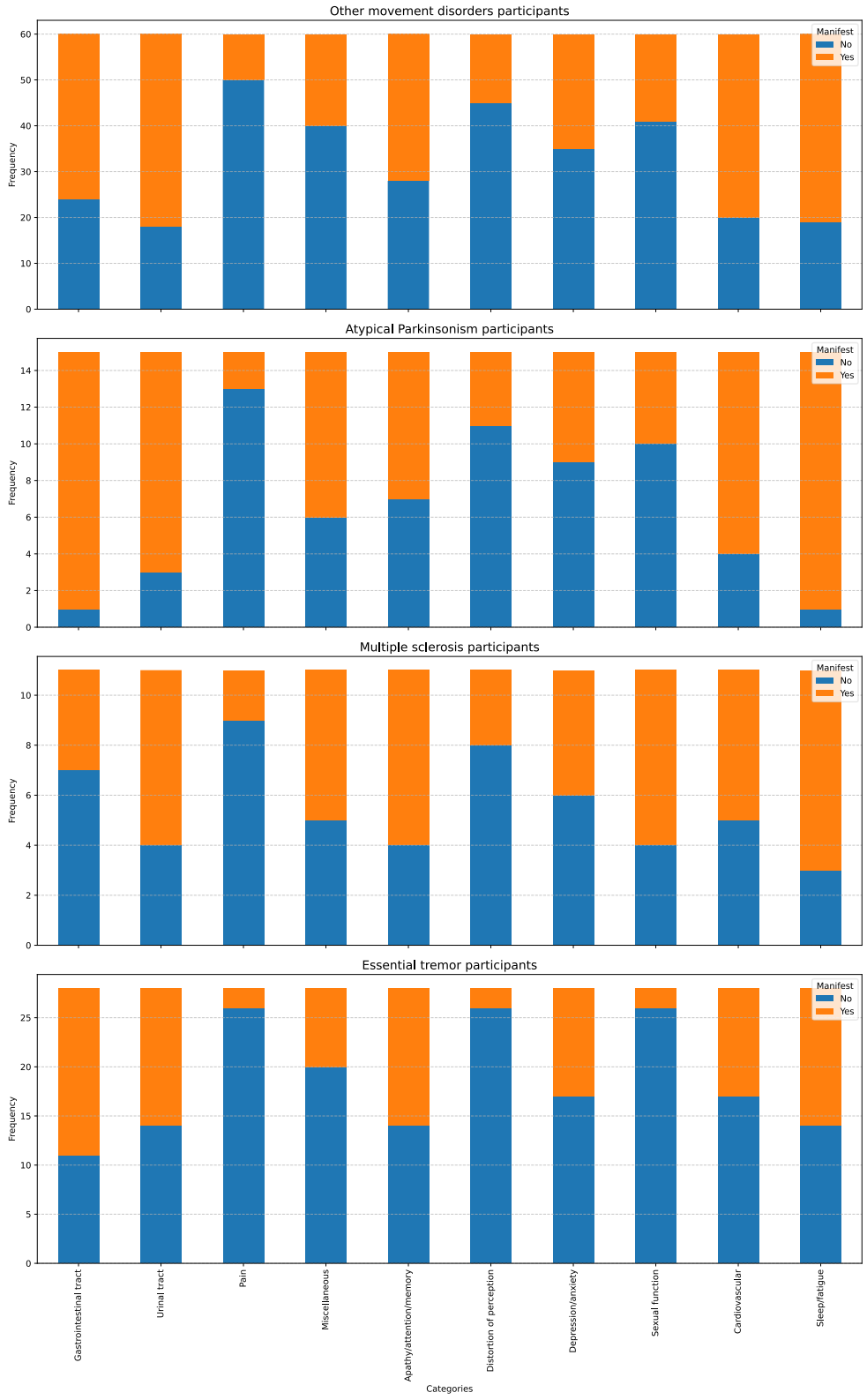


Figure A.9. Responses to the PDNMS questionnaire in the DD group, aggregated by symptom category

ANNEX B

Movement Data Analysis

The participants with IDs 293, 17, and 64 were considered for the HC, PD, and DD groups, respectively.

Resting Tasks

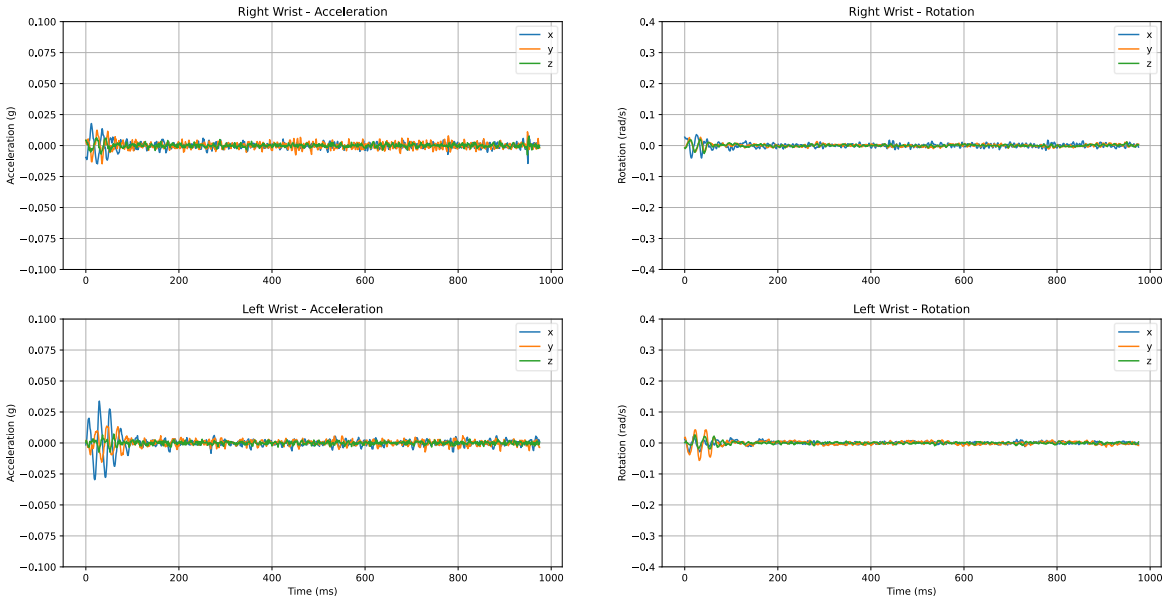


Figure B.1. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (last 10 seconds of the 1a step, see Table 3.5) by a participant from the HC group

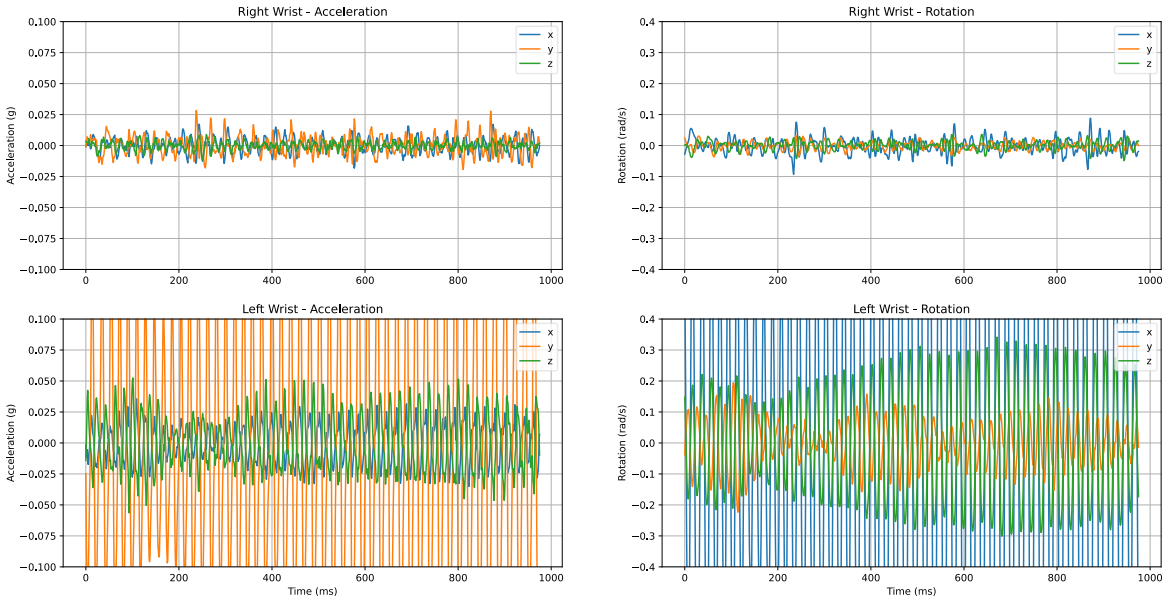


Figure B.2. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (last 10 seconds of the 1a step, see Table 3.5) by a participant from the PD group

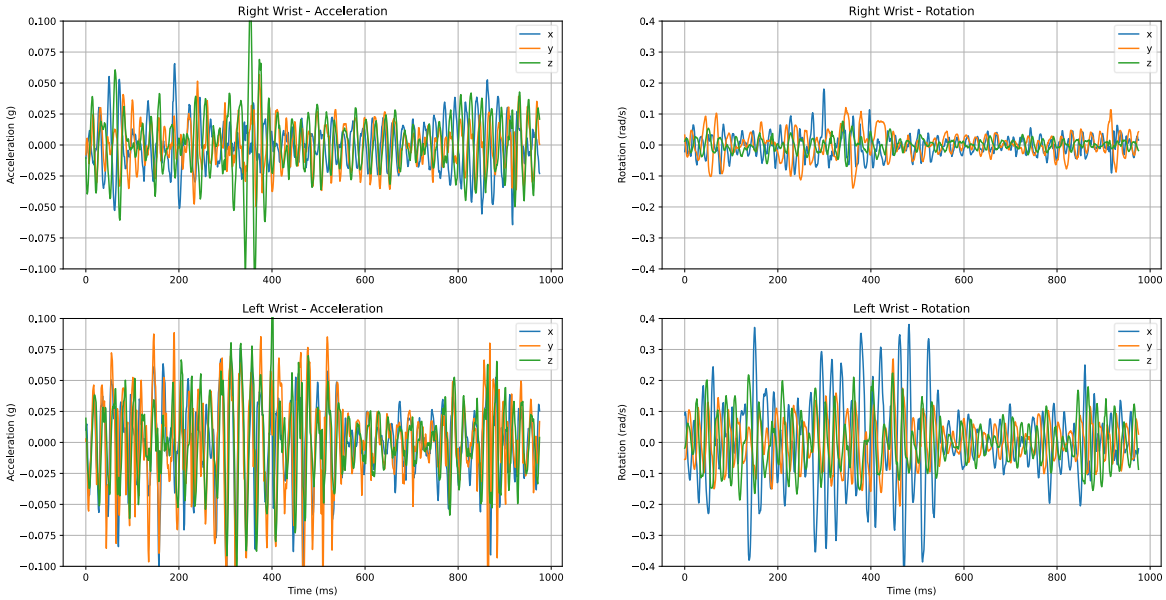


Figure B.3. Acceleration and rotation on the three-dimensional axis during the execution of the "Relaxed" task (last 10 seconds of the 1a step, see Table 3.5) by a participant from the DD group

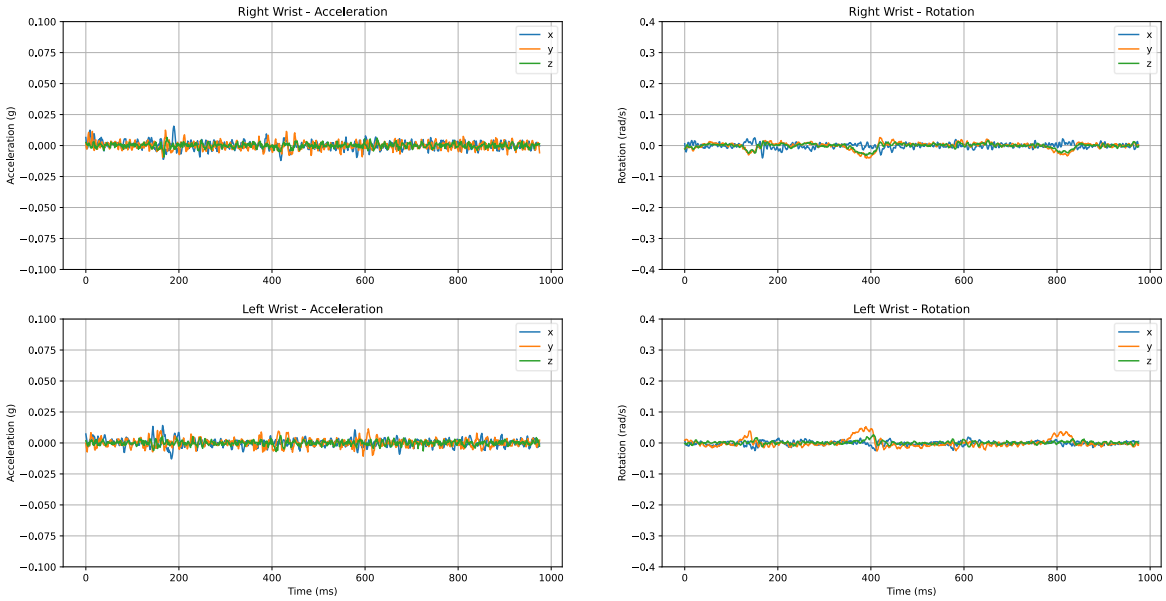


Figure B.4. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (first 10 seconds of the 1b step, see Table 3.5) by a participant from the HC group

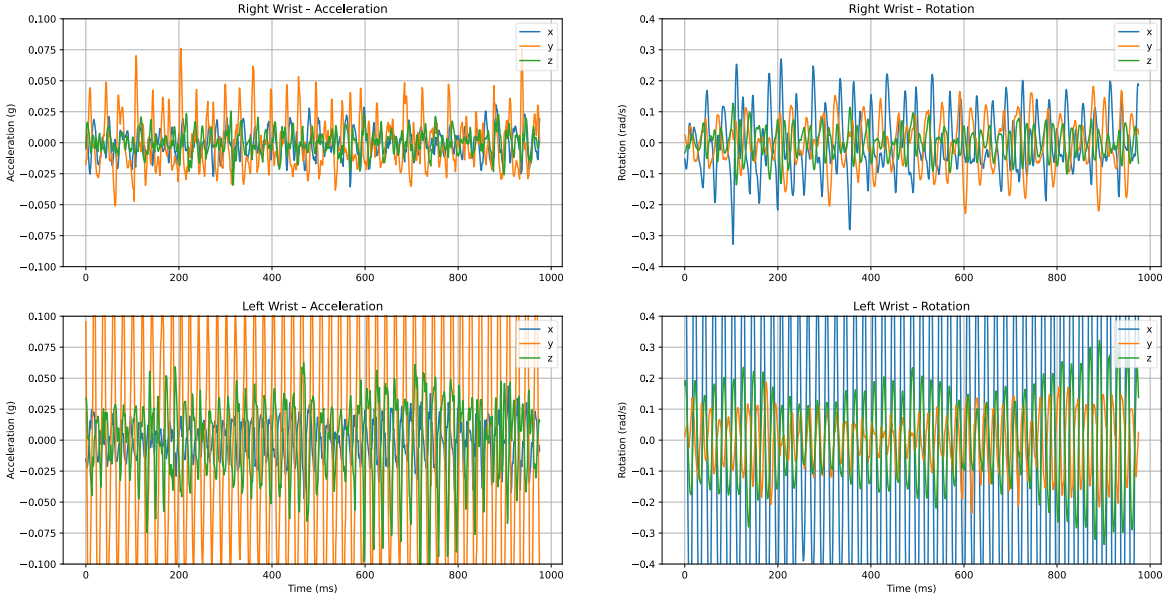


Figure B.5. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (first 10 seconds of the 1b step, see Table 3.5) by a participant from the PD group

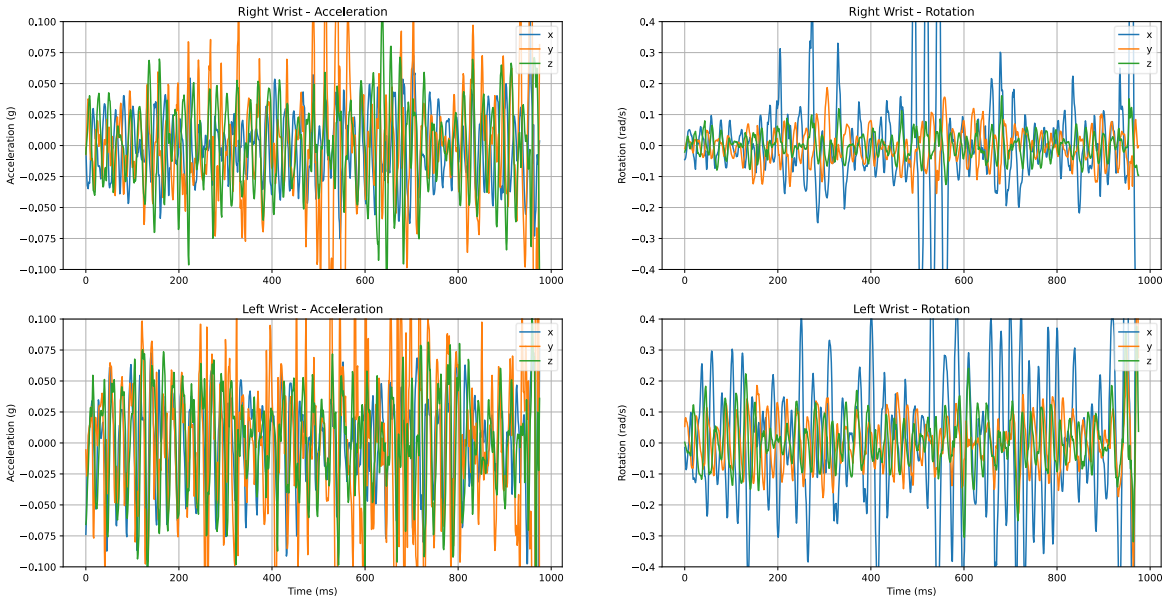


Figure B.6. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (first 10 seconds of the 1b step, see Table 3.5) by a participant from the DD group

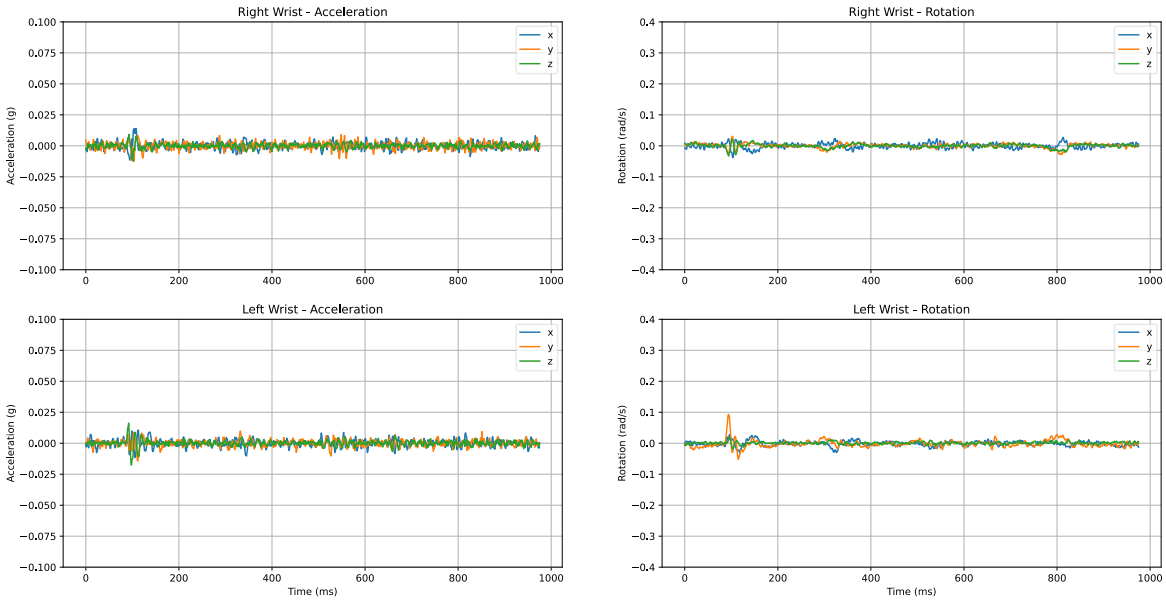


Figure B.7. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (last 10 seconds of the 1b step, see Table 3.5) by a participant from the HC group

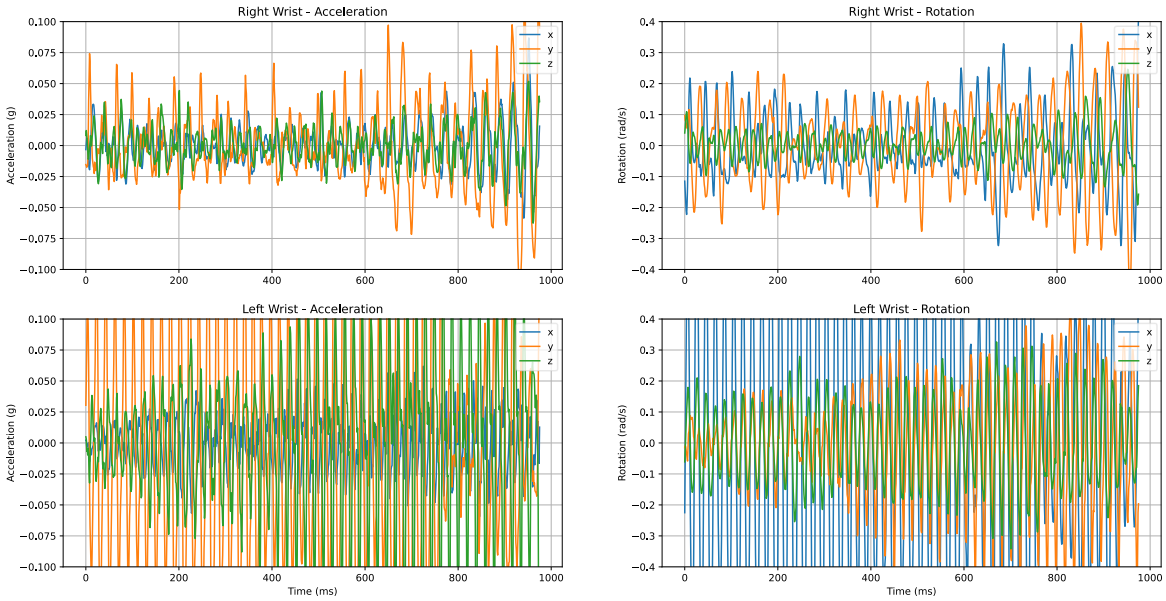


Figure B.8. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (last 10 seconds of the 1b step, see Table 3.5) by a participant from the PD group

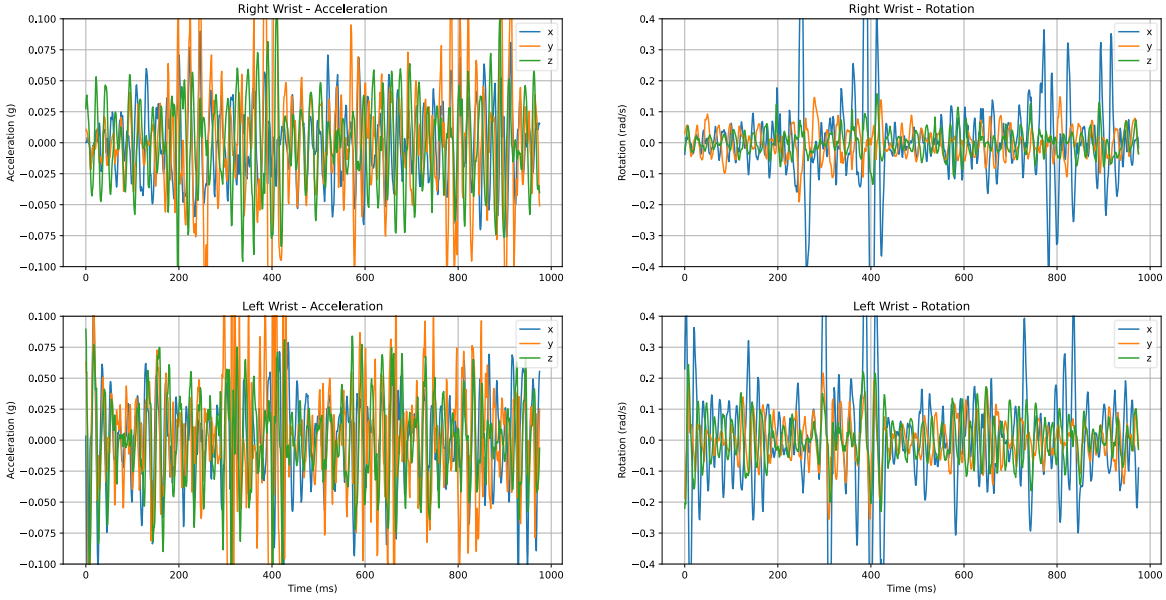


Figure B.9. Acceleration and rotation on the three-dimensional axis during the execution of the "RelaxedTask" (last 10 seconds of the 1b step, see Table 3.5) by a participant from the DD group

Postural Tasks

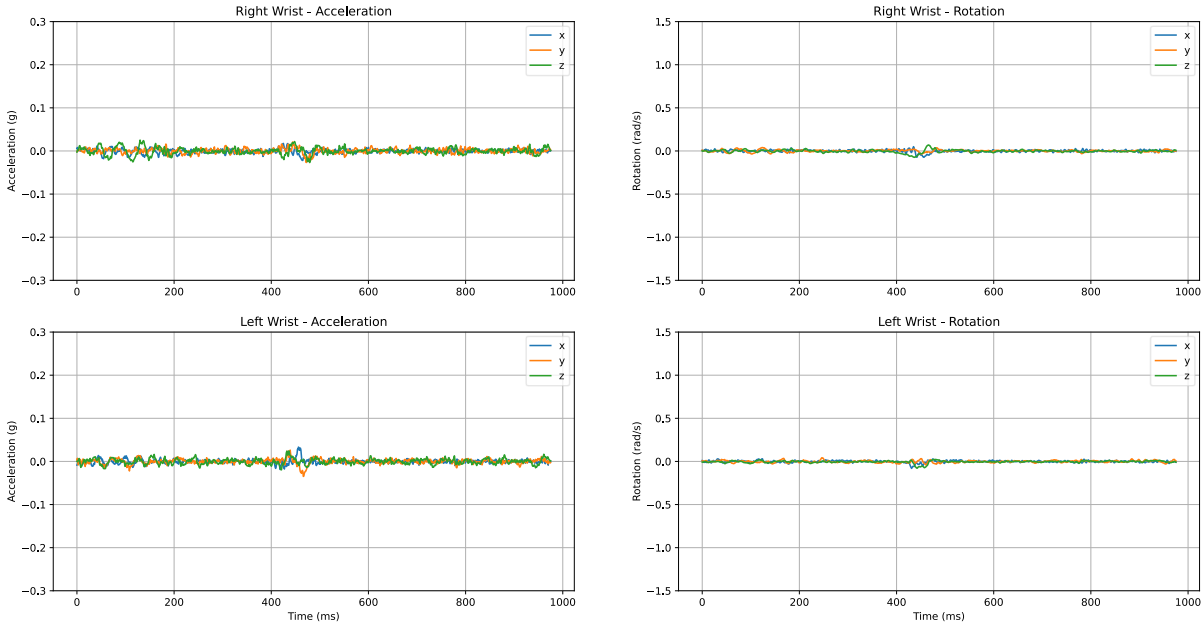


Figure B.10. Acceleration and rotation on the three-dimensional axis during the execution of the "StretchHold" task (step 2, see Table 3.5) by a participant from the HC group

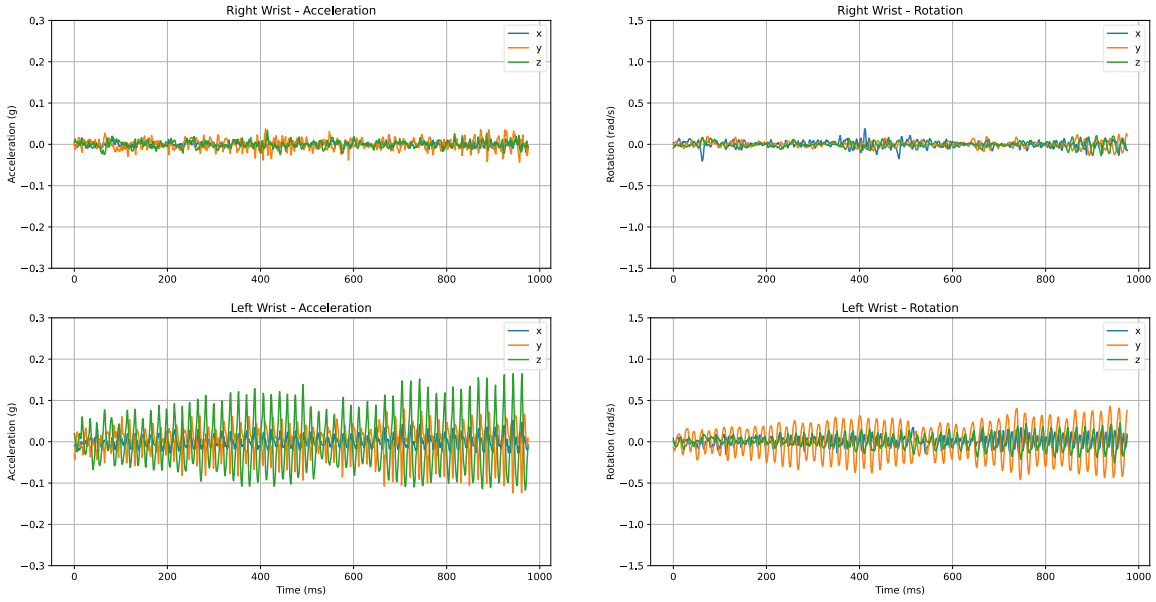


Figure B.11. Acceleration and rotation on the three-dimensional axis during the execution of the "StretchHold" task (step 2, see Table 3.5) by a participant from the PD group

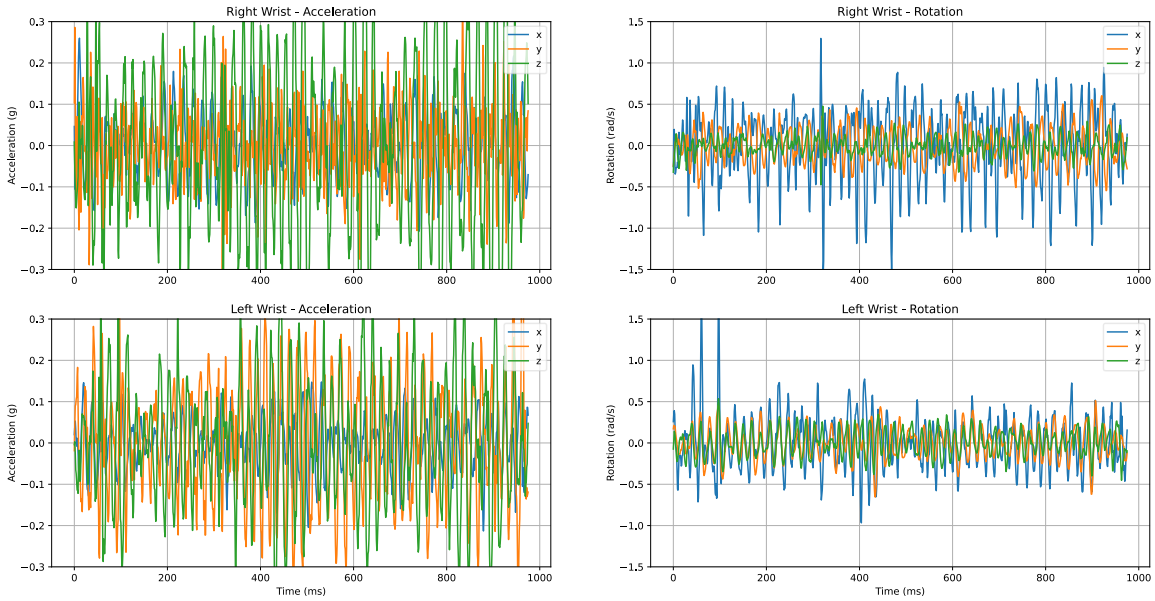


Figure B.12. Acceleration and rotation on the three-dimensional axis during the execution of the "StretchHold" task (step 2, see Table 3.5) by a participant from the DD group

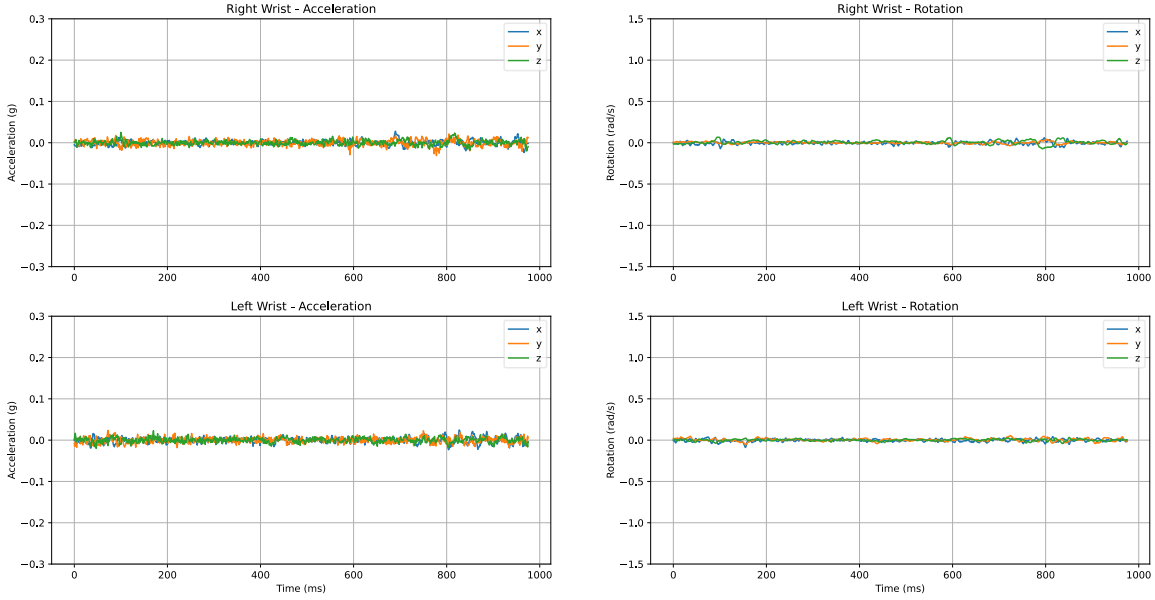


Figure B.13. Acceleration and rotation on the three-dimensional axis during the execution of the "HoldWeight" task (step 4, see Table 3.5) by a participant from the HC group

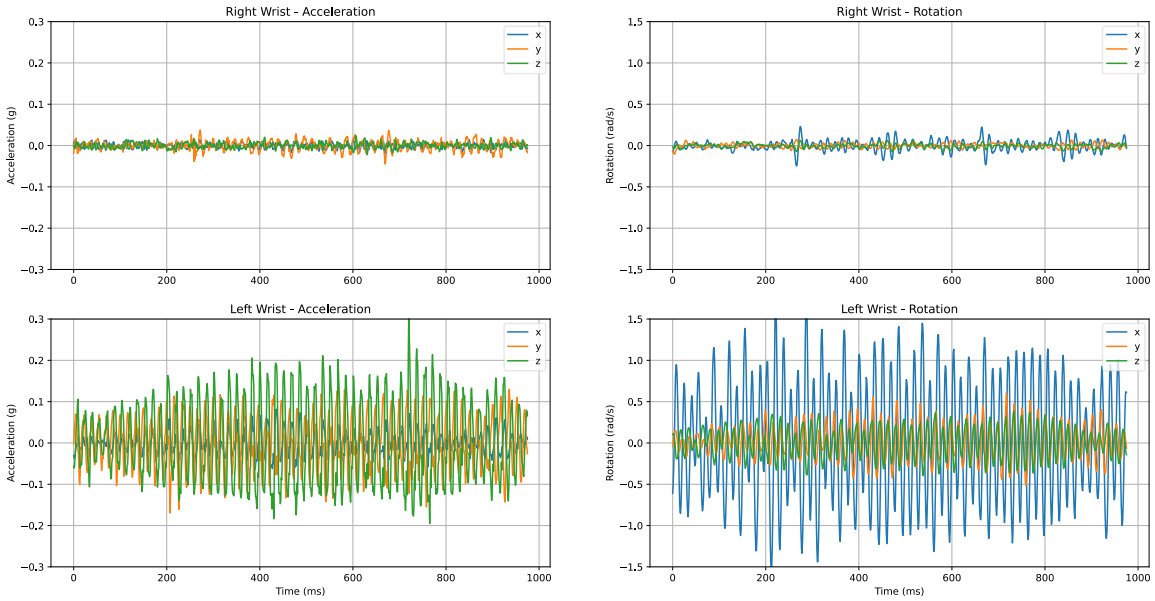


Figure B.14. Acceleration and rotation on the three-dimensional axis during the execution of the "HoldWeight" task (step 4, see Table 3.5) by a participant from the PD group

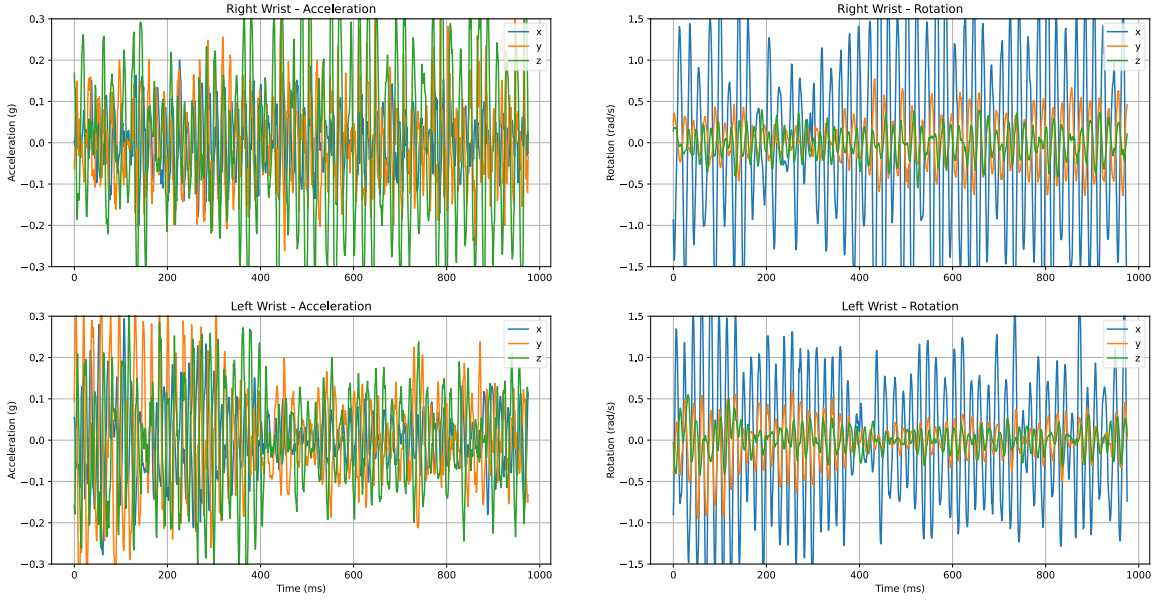


Figure B.15. Acceleration and rotation on the three-dimensional axis during the execution of the "HoldWeight" task (step 4, see Table 3.5) by a participant from the DD group

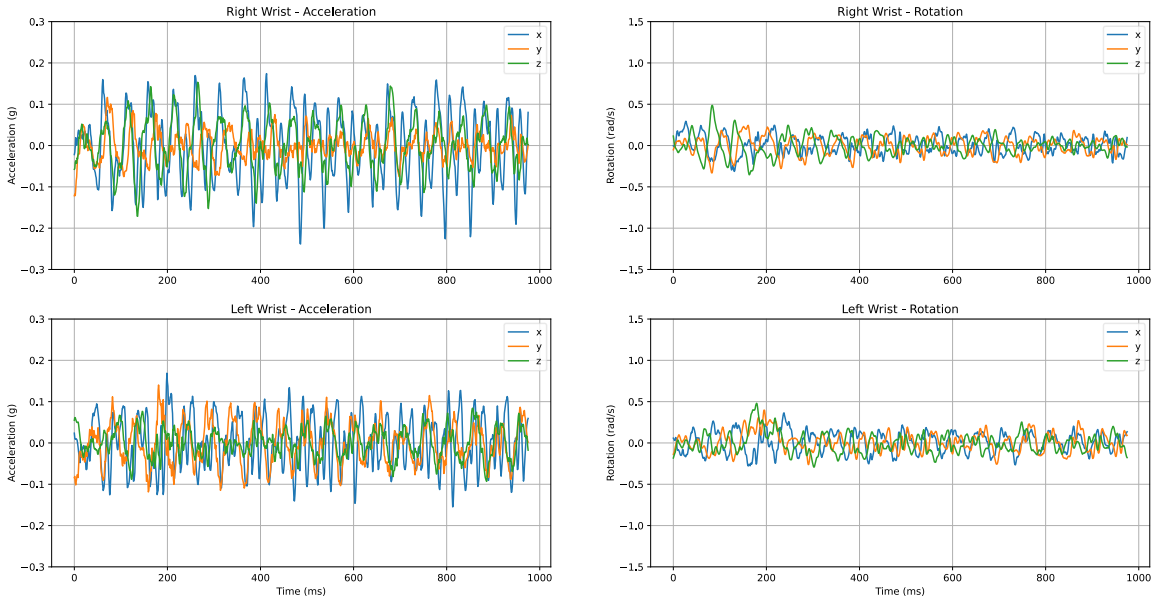


Figure B.16. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (first 10 seconds of the 10 step, see Table 3.5) by a participant from the HC group

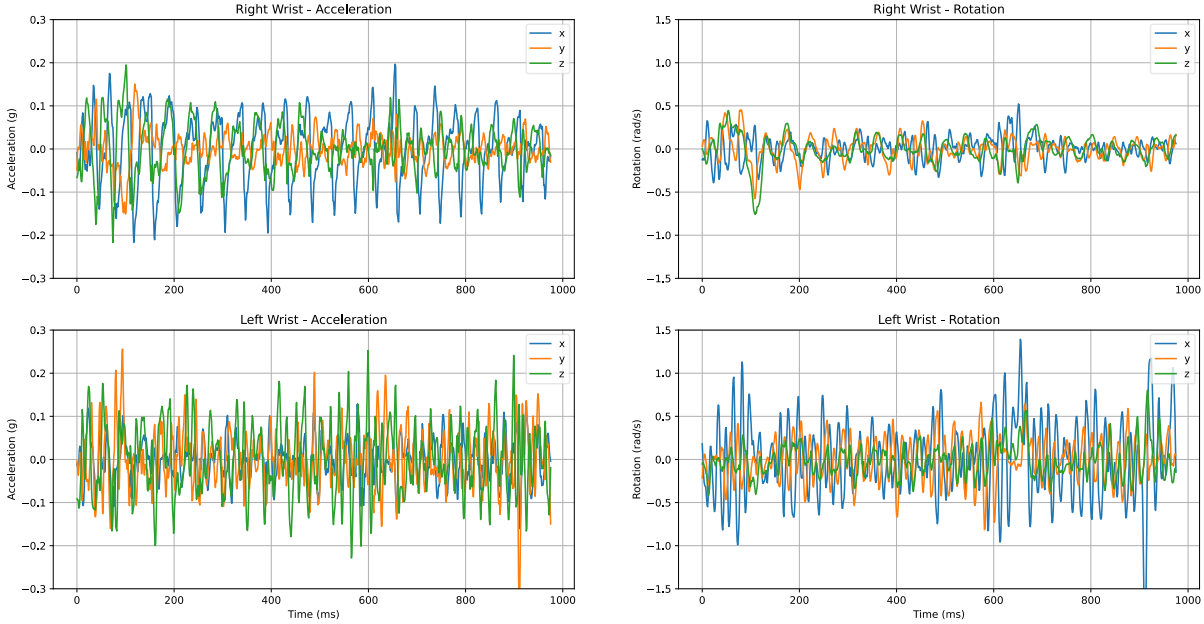


Figure B.17. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (first 10 seconds of the 10 step, see Table 3.5) by a participant from the PD group

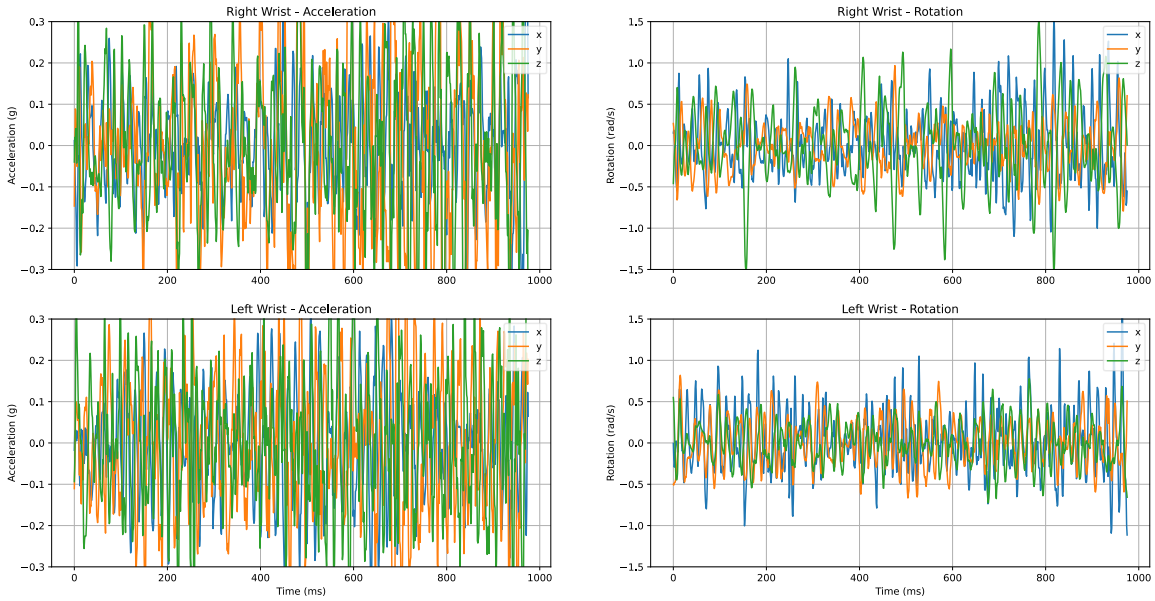


Figure B.18. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (first 10 seconds of the 10 step, see Table 3.5) by a participant from the DD group

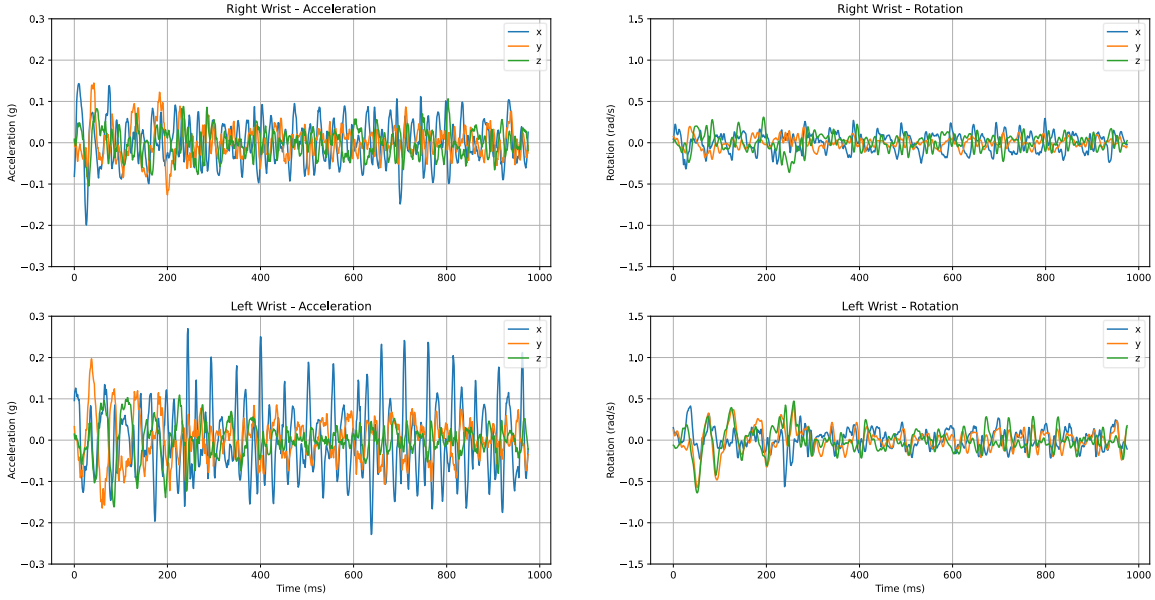


Figure B.19. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (last 10 seconds of the 10 step, see Table 3.5) by a participant from the HC group

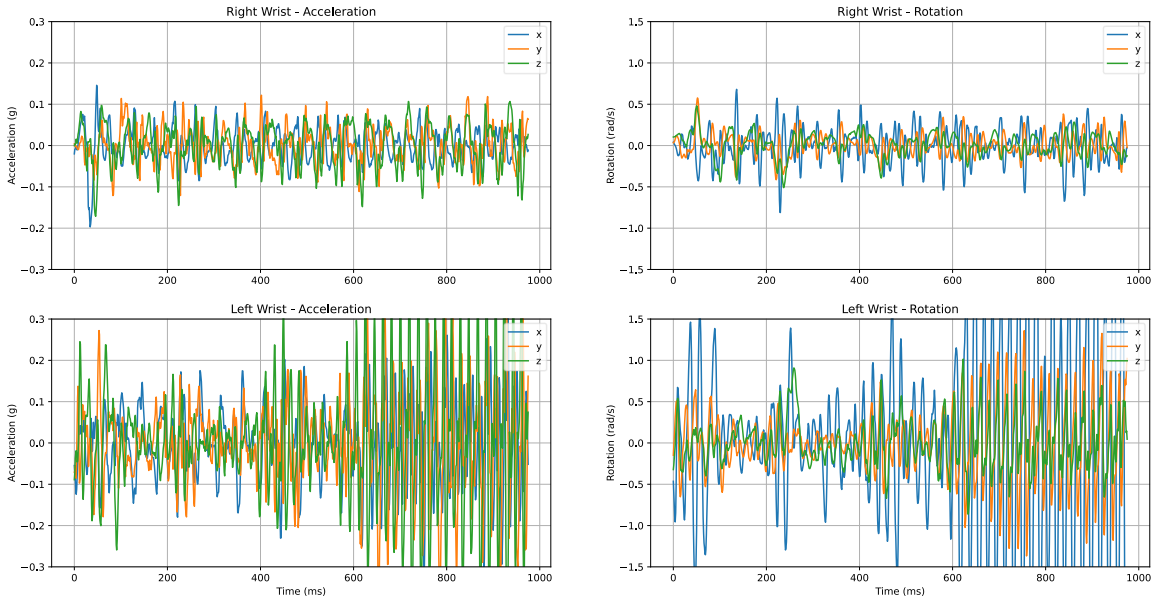


Figure B.20. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (last 10 seconds of the 10 step, see Table 3.5) by a participant from the PD group

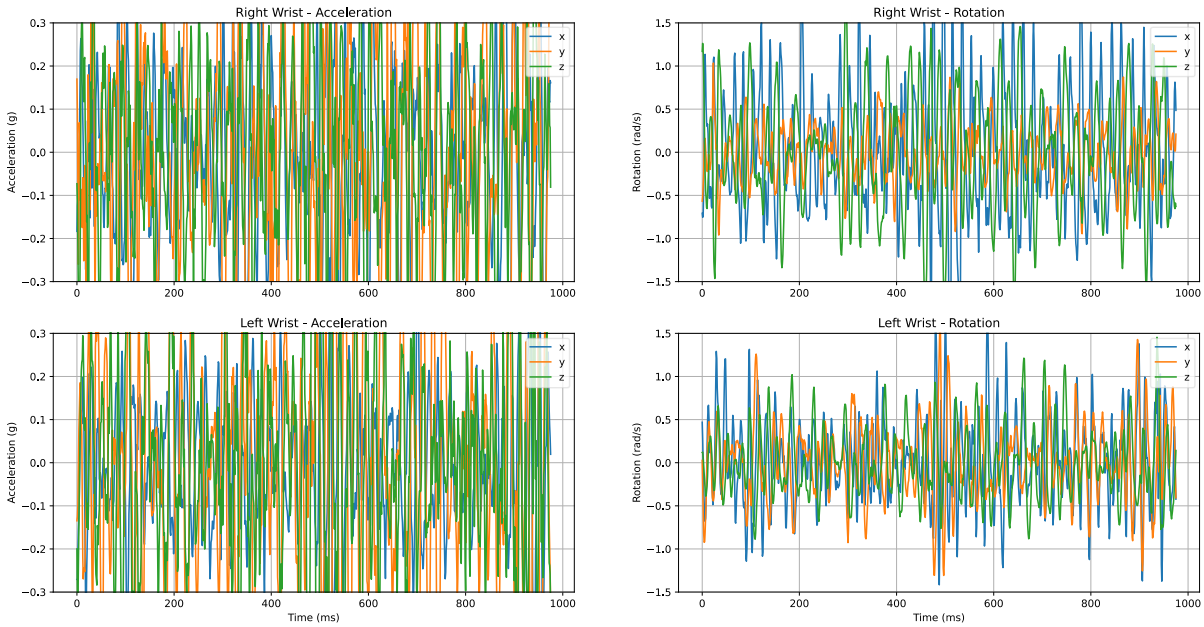


Figure B.21. Acceleration and rotation on the three-dimensional axis during the execution of the "Entrainment" (last 10 seconds of the 10 step, see Table 3.5) by a participant from the DD group

Kinetic Tasks

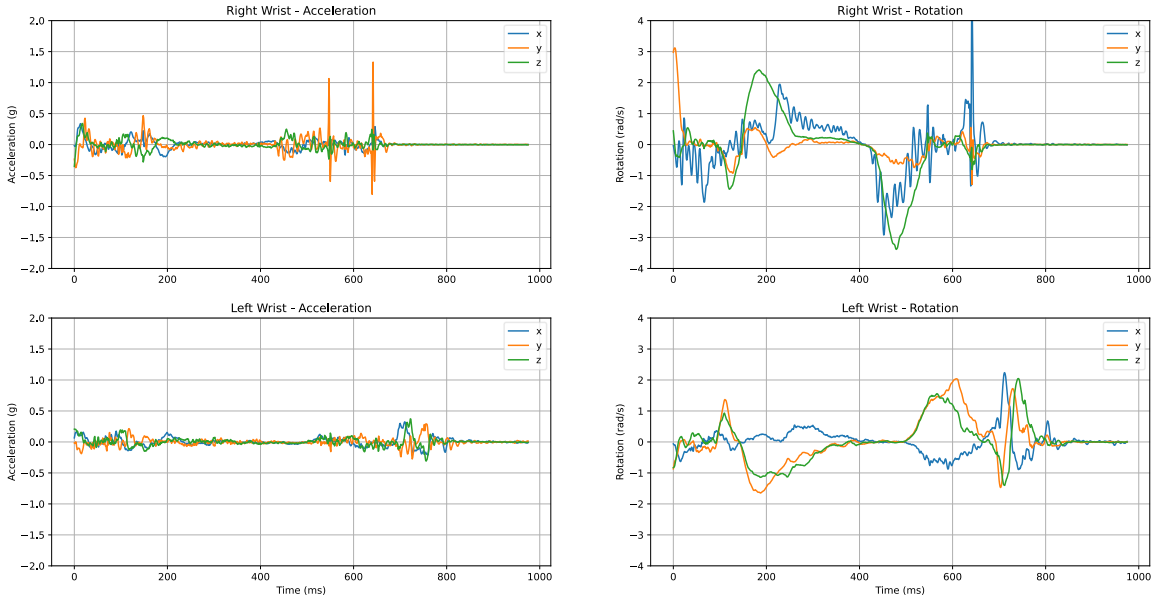


Figure B.22. Acceleration and rotation on the three-dimensional axis during the execution of the "DrinkGlass" task (step 6, see Table 3.5) by a participant from the HC group

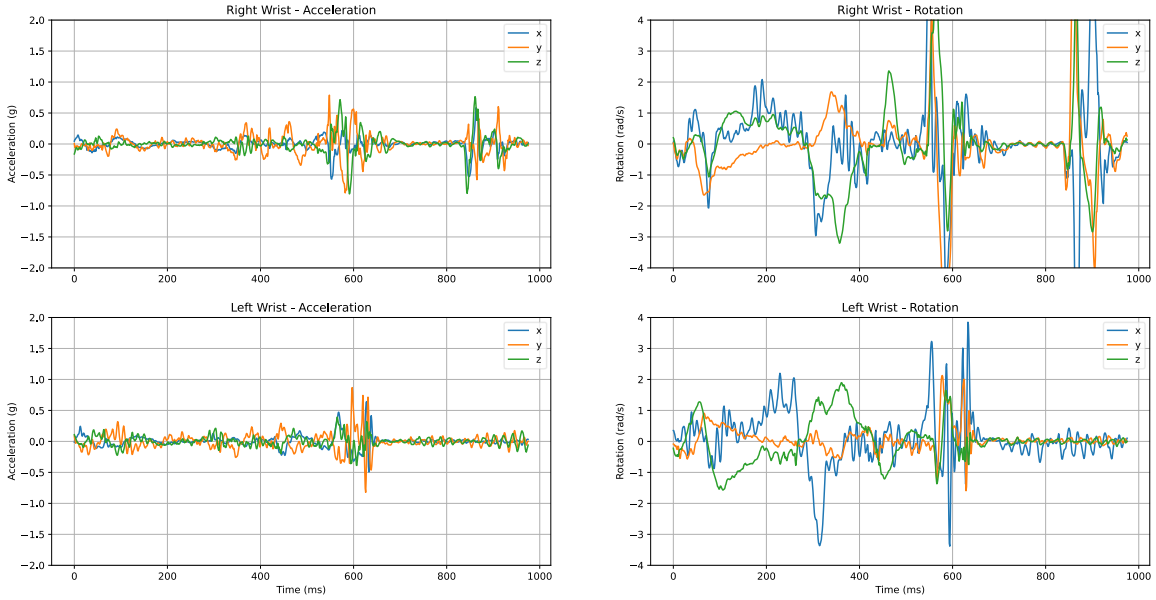


Figure B.23. Acceleration and rotation on the three-dimensional axis during the execution of the "DrinkGlass" task (step 6, see Table 3.5) by a participant from the PD group

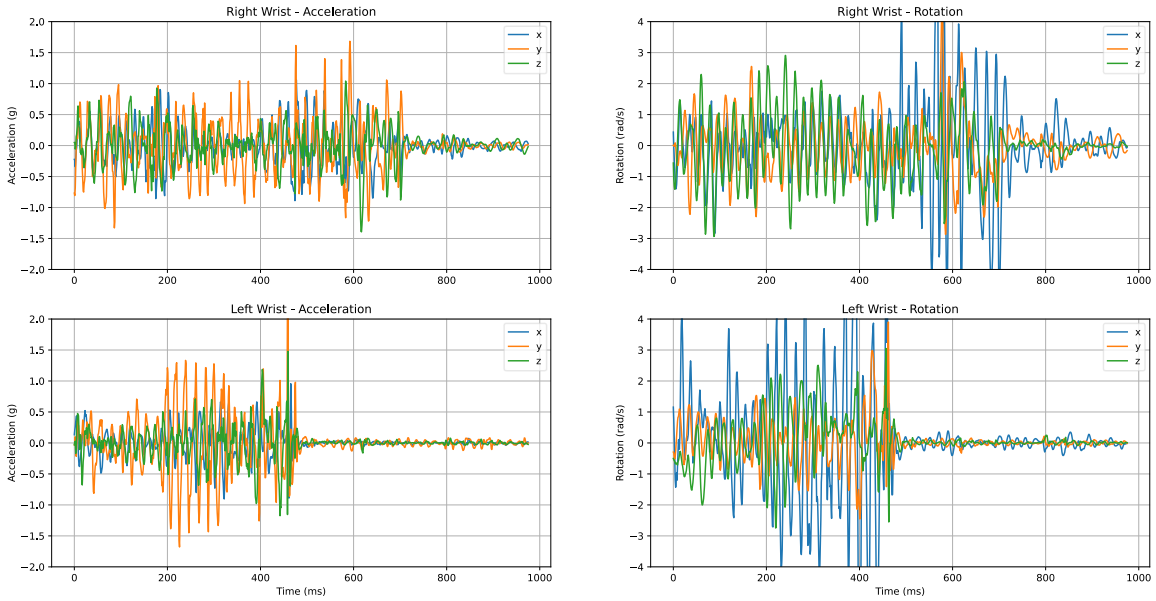


Figure B.24. Acceleration and rotation on the three-dimensional axis during the execution of the "DrinkGlass" task (step 6, see Table 3.5) by a participant from the DD group

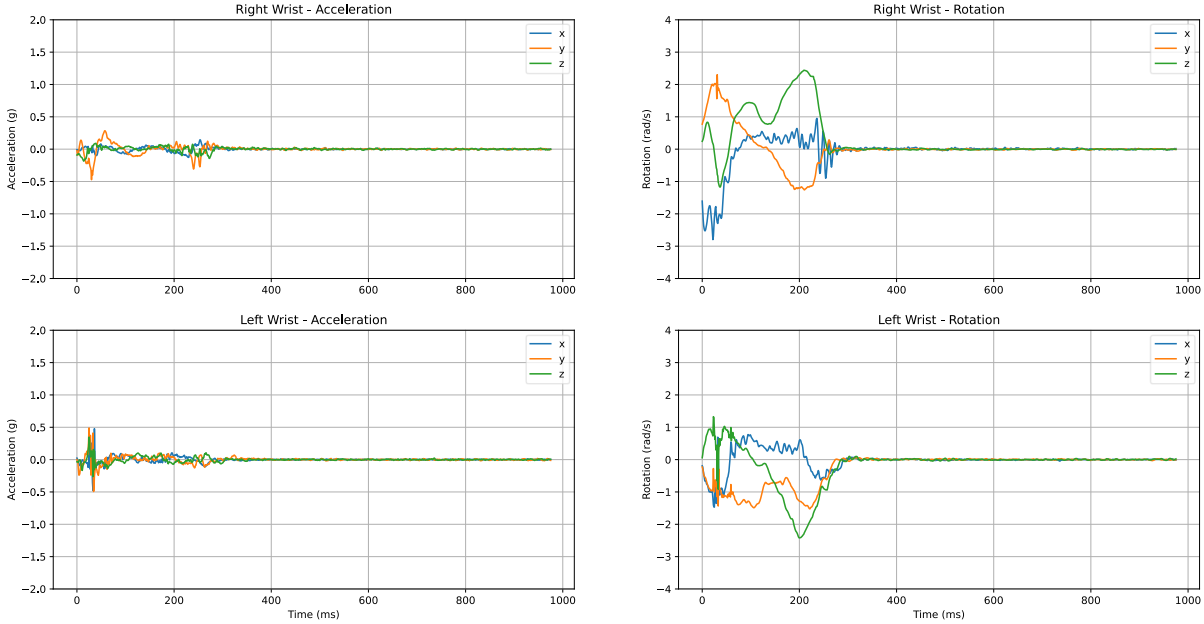


Figure B.25. Acceleration and rotation on the three-dimensional axis during the execution of the "CrossArms" task (step 7, see Table 3.5) by a participant from the HC group

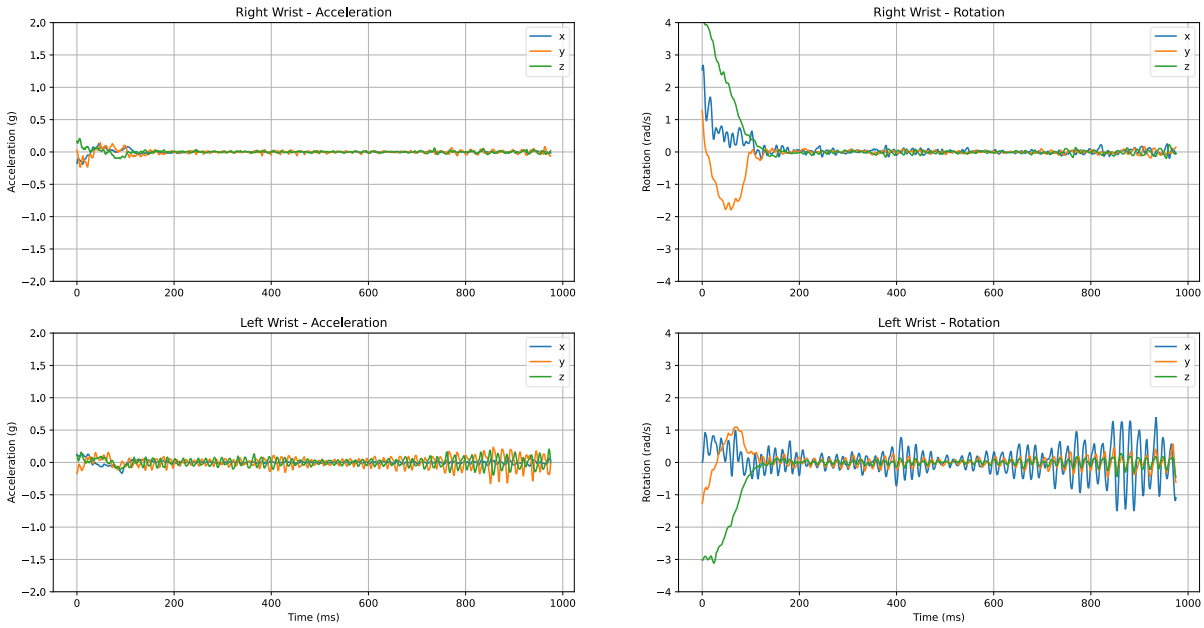


Figure B.26. Acceleration and rotation on the three-dimensional axis during the execution of the "CrossArms" task (step 7, see Table 3.5) by a participant from the PD group

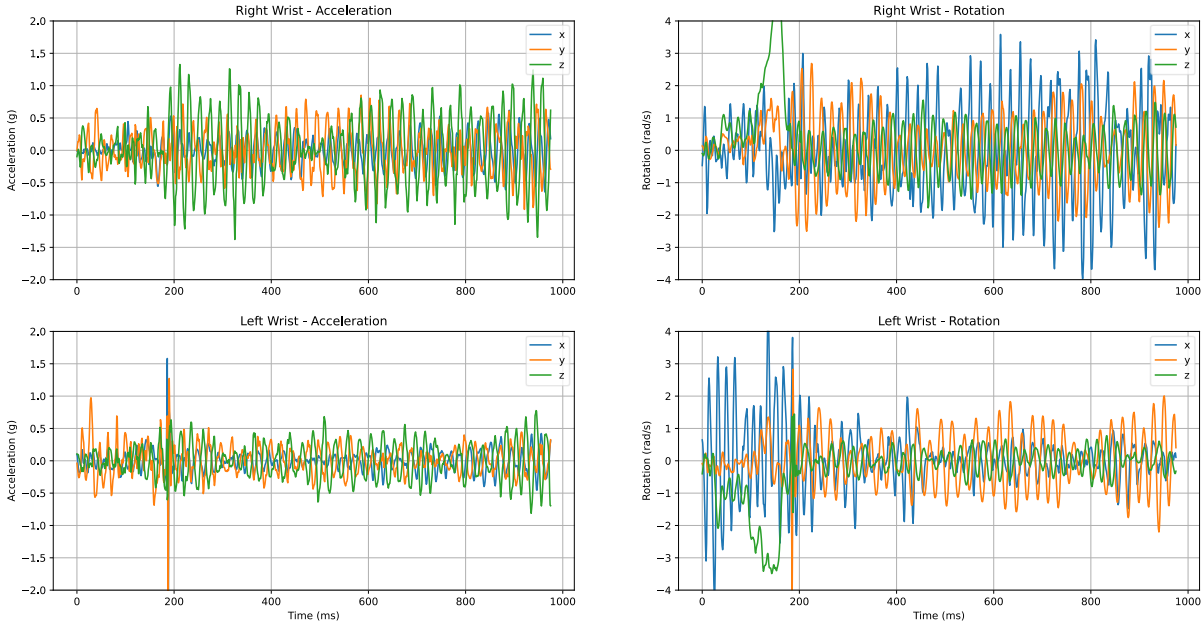


Figure B.27. Acceleration and rotation on the three-dimensional axis during the execution of the "CrossArms" task (step 7, see Table 3.5) by a participant from the DD group

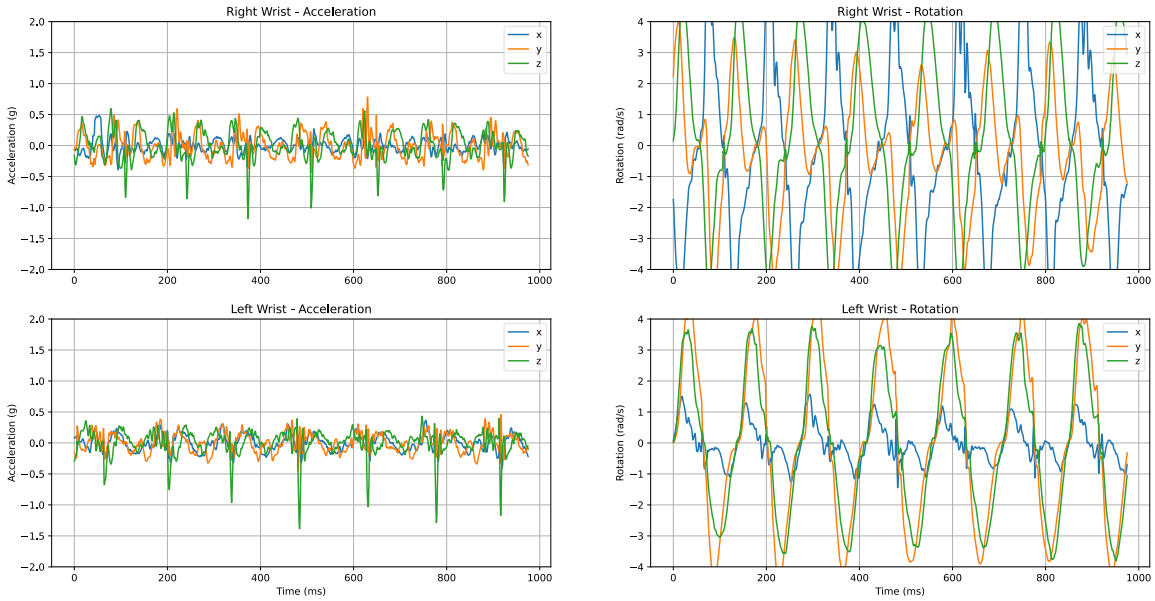


Figure B.28. Acceleration and rotation on the three-dimensional axis during the execution of the "TouchNose" task (step 9, see Table 3.5) by a participant from the HC group

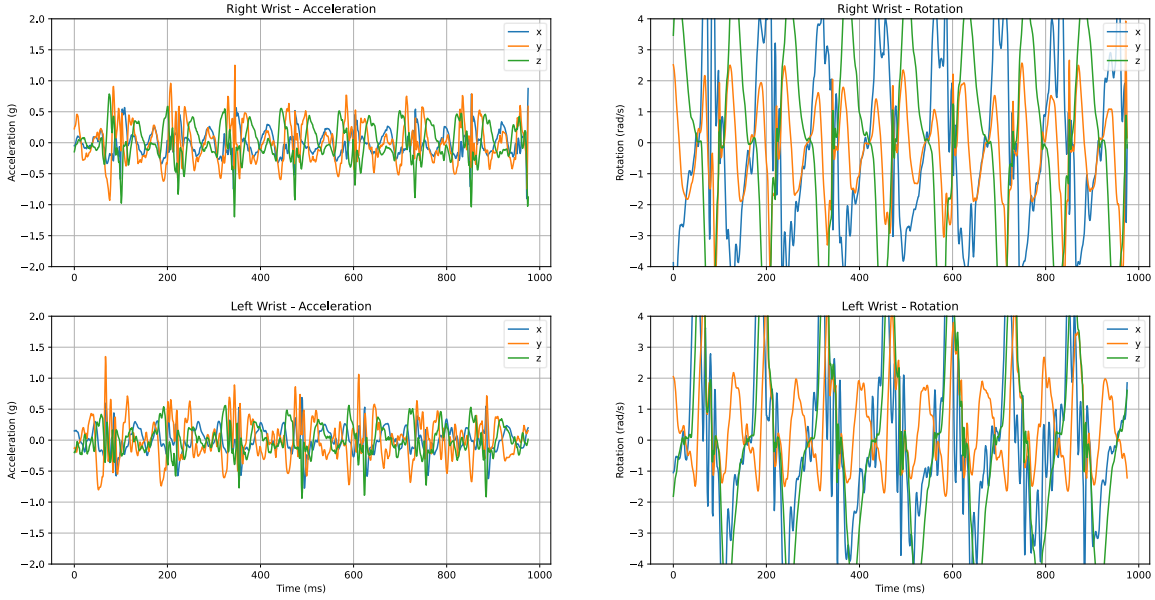


Figure B.29. Acceleration and rotation on the three-dimensional axis during the execution of the "TouchNose" task (step 9, see Table 3.5) by a participant from the PD group

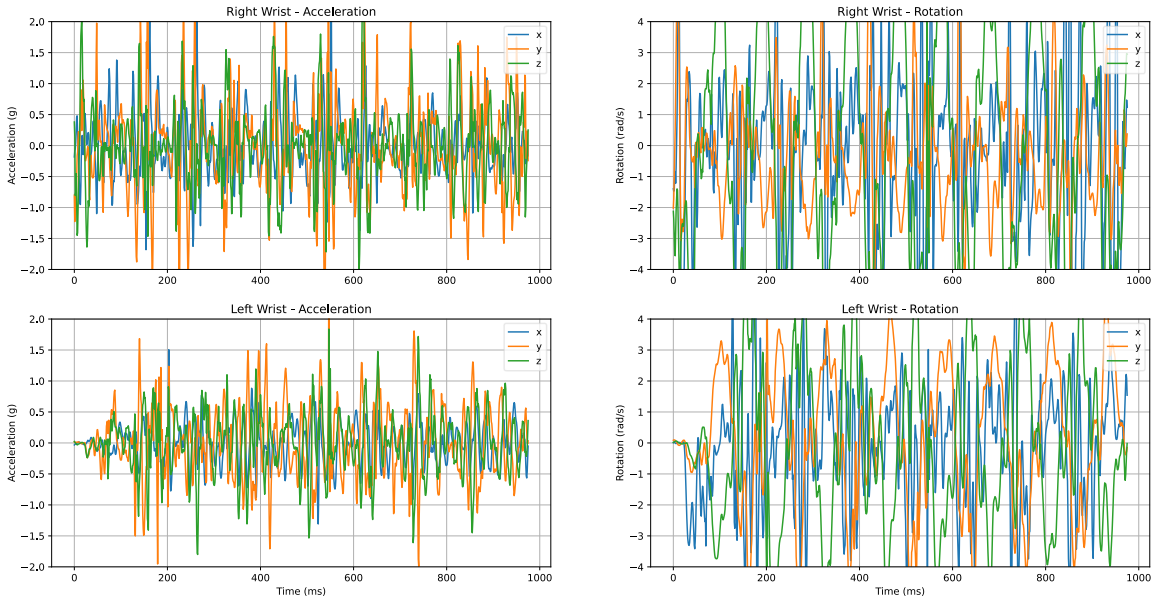


Figure B.30. Acceleration and rotation on the three-dimensional axis during the execution of the "TouchNose" task (step 9, see Table 3.5) by a participant from the DD group