To the Graduate Council:

I am submitting herewith a thesis written by Jeremy Watts entitled "Machine Learning and Decision Making to Optimize Treatment Planning in Parkinson's Disease." I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Systems Engineering.

Anahita Khojandi, Major Professor

We have read this thesis
and recommend its acceptance:

John Kobza

Jim Ostrowski

Ritesh Ramdhani

Accepted for the Council:

Dixie Thompson

Vice Provost and Dean of the Graduate School

To the Graduate Council:
I am submitting herewith a thesis written by Jeremy Watts entitled "Machine Learning and Decision Making to Optimize Treatment Planning in Parkinson's Disease." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Systems Engineering.

Anahita Khojandi, Major Professor

We have read this thesis
and recommend its acceptance:

John Kobza

Jim Ostrowski

Ritesh Ramdhani

Accepted for the Council:

Dixie Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Machine Learning and Decision Making to Optimize Treatment Planning in Parkinson's Disease

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Jeremy Watts

May 2023

Acknowledgements

I would like to thank my advisor, Anahita Khojandi, for her support in preparing this dissertation.

The research in Chapter 1 is partially supported by the Science Alliance, the University of Tennessee, the Parkinson's Alliance, and by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory managed by UT-Battelle, LLC, for the U.S. Department of Energy (DOE). Materials from this chapter have previously appeared in the MDPI journal Sensors. For all articles published in MDPI journals, copyright is retained by the authors. Articles are licensed under an open access Creative Commons CC BY 4.0 license. Materials have been reprinted with permission from author Jeremy Watts, (Improving Medication Regimen Recommendation for Parkinson's Disease Using Sensor Technology, 5/2021).

The research in Chapter 2 is partially supported by the Science Alliance, the University of Tennessee, the Parkinson's Alliance, and by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory managed by UT-Battelle, LLC, for the U.S. Department of Energy (DOE).

The research in Chapter 3 is supported by the National Institutes of Health (NIH/NINDS) 1R21NS111301-01. Materials from this chapter have previously appeared in the MDPI journal Brain Sciences. For all articles published in MDPI journals, copyright is retained by the authors. Articles are licensed under an open access Creative Commons CC BY 4.0 license. Materials have been reprinted with

permission from author Jeremy Watts, (Machine Learning's Application in Deep Brain Stimulation for Parkinson's Disease: A Review, 11/2020).

The research in Chapter 4 is supported by the National Institutes of Health (Project number: 1R21NS111301-01) and the Parkinson's Alliance. The authors would like to thank Dr. Martin Niethammer and Toni Fitzpatrick at Northwell Health Hospital, for assisting with data collection.

The research in Chapter 5 is partially supported by NHGRI's Genomic Data Science Analysis, Visualization, and Informatics Lab-space's (AnVIL) Cloud Credits (AC2) pilot program. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP accession number phs000424.v8.p2 on 09/07/2021. Materials from this chapter have previously appeared in the 44^{th} Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) © [2022, 2023] IEEE. Reprinted, with permission, from author Jeremy Watts, (Adapting Random Forests to Predict Obesity-Associated Gene Expression, 7/2022) and from author Jeremy Watts (Towards Faster Gene Expression Prediction via Dimensionality Reduction and Feature Selection, 3/2023). In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of the University of Tennessee Knoxville's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

The research in Chapter 6 used data from the Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and

NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP accession number phs000424.v8.p2 on 09/07/2021. Materials from this chapter have previously appeared in the 45^{th} Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) © [2022, 2023] IEEE. Reprinted, with permission, from author Jeremy Watts, (Adapting Random Forests to Predict Obesity-Associated Gene Expression, 7/2022) and from author Jeremy Watts (Towards Faster Gene Expression Prediction via Dimensionality Reduction and Feature Selection, 3/2023). In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of the University of Tennessee Knoxville's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

Abstract

Parkinson's disease (PD) is a progressive, neurodegenerative disorder resulting from the loss of dopaminergic neurons. The disease is characterized by four cardinal motor symptoms, bradykinesia (slowing of movement), muscle rigidity, tremor, and postural instability/gait disorder. Early in the disease course, dopaminergic medication can effectively manage the cardinal motor symptoms for the majority of patients. The primary medication, which underpins all long-term treatment planning in PD, is levodopa (L-dopa). However, increased L-dopa typically results in increased dyskinesia (a side effect of L-dopa). As the adverse side effects of L-dopa increase in severity, patients may need advanced therapies such as deep brain stimulation (DBS), a neurosurgical treatment. Advanced therapies have been shown effective at alleviating the side effects associated with L-dopa. However, DBS requires the additional tuning of stimulation parameters to optimize treatment plans. Further, the interaction between L-dopa and DBS remains an open question. This dissertation proposes methodological and applied analysis to personalize treatment planning for PD patients throughout their disease course.

First, we optimize early-stage PD medication regimens using machine learning clustering techniques and Markov decision processes incorporating pharmacokinetic considerations for immediate release L-dopa. Further, these models are calibrated using wearable sensor data from a multi-visit cohort of PD patients. We apply reinforcement learning techniques to determine the optimal treatment plans for patient subtypes.

Second, we develop machine learning approaches to improve the efficiency of parameter tuning for DBS as well as predict potential advanced therapy candidates. We begin with a review of the literature. Then we examine a primary cohort of chronic PD patients who have received DBS. We determine the gait parameters impacted by DBS (both high and low-frequency stimulation) both with and without the influence of dopaminergic medication. This study represents the largest study of its type, focusing on low-frequency stimulation in combination with medication.

Finally, we examine the role of machine learning and dimensionality reduction in the prediction of gene expression. Genetics influence a variety of chronic diseases, including Parkinson's disease. This work sought to improve the computational efficiency of gene expression prediction models for use in the progression modeling of several chronic diseases.

Table of Contents

List of Tables

List of Figures

Chapter 1

Improving Medication Regimen Recommendation for Parkinson's Disease Using Sensor Technology

1.1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder resulting from the loss of dopaminergic neurons. It is characterized by four cardinal motor symptoms: Bradykinesia (slowing of movement), muscle rigidity, tremor, and postural instability/gait disorder. Additionally, symptoms, such as rapid eye movement sleep behavioral disorder (RBD), anosmia, and constipation, can present as prodromes, while other nonmotor symptoms— bladder dysfunction, dysphagia, orthostatic hypotension, and cognitive impairment—can manifest later in the disease course (198). The number of individuals diagnosed with PD is estimated to be 6.2 million globally (70), and approximately 60,000 individuals are diagnosed with PD annually in the U.S. alone (61). Studies have reported demographic differences with respect to sex and race in PD diagnosis frequency (123; 124; 146). Individuals living with PD are expected to increase as global life expectancy increases; this will place additional strain on

the medical system. Due to the shortage of neurologists and logistical challenges, including extended travel time, patient disability, and prolonged clinic wait times, PD patients tend to have fewer clinic visits (62; 234). Women, racial minorities, and rural communities have less access to care and lower quality of specialist care (62), causing delays in diagnosis and higher long-term disability (47; 197). It is further expected with the increasing number of PD patients, these inequities to the quality of care will become more prevalent (61). Currently, these clinic visits are critical to improving an individual's treatment planning and represent a potential bottleneck in the quality of patients' care.

Levodopa remains the gold standard therapy for treating the cardinal motor symptoms of PD. As Parkinson's progresses, the duration of levodopa's dose efficacy shortens with the emergence of motor complications, such as: "Wearing-off" episodes (a return of motor symptoms prior to taking the next dose); "delayed ON periods" (slow onset of dose benefit); "On-Off cycling" or "motor fluctuations" (symptomatic benefits are achieved during the ON phase of the dose followed by OFF periods characterized by uncontrolled motor symptoms prior to the next dose); "dyskinesia" (involuntary movements affecting the limb). These complications result from various factors, including disease progression and pulsatile stimulation of dopamine receptors, due to lack of continuous levodopa administration (163). Typically, a patient's medication regimen is optimized by fragmenting and increasing levodopa dosages, while utilizing monoamine oxidase B (MAO-B), dopamine agonists, or catechol-O-methyl transferase (COMT) inhibitors as adjunctive therapies to provide dopamine replacement. The primary goal for PD treatment is to optimize symptom control, while minimizing off periods and medication side effects.

Currently, assessments of the efficacy of patients' treatments are based on a clinician's overall impression of motor disability as determined by clinical assessment tools, such as the MDS-Unified Parkinson's disease Rating Scale (UPDRS) (79) and Hauser paper-based diaries (94). The lack of continuous motor assessment coupled with recall bias and limited integration of nonmotor symptomatology into the

treatment paradigm present real-world limitations in managing such a heterogeneous condition. Sensor-based technology offers a real-time mechanism to objectively measure motor performance in PD (187; 140), moving beyond the "snapshot" clinical assessment of impairment.

Specialty PD motor sensors have demonstrated 70–90% accuracy in measuring fluctuations and dyskinesia in patients' medication response (105; 217; 85; 182). One such inertial sensor is the Personal KinetiGraph™ (PKG) sensor (Global Kinetics Corporation (GKC), Melbourne, Australia). This wrist-worn logger utilizes an accelerometer to collect movement information in two-minute intervals and reminds patients to register when taking their prescribed dopaminergic medication. The raw data are converted into summary dyskinesia and bradykinesia scores (averaged single value assessments over the entire wear period), as well as time-series data, curated into a report (85) using validated algorithms (85; 117; 99; 22). The report shows the continuous changes of dyskinesia and bradykinesia scores, as it relates to levodopa timing as the median, 25th, and 75th percentile, compared to a non-PD control group over six days. A sample PKG report is provided in Figure 1.1.

By examining a spectrum of patients with clinical variability and gauging their responsivity to dopaminergic medication with sensor technology, inherent medication similarities may be present within specific clinical subtypes-offing an opportunity to cluster patients in a treatment-related manner. This approach could serve to predict optimal regimens, potentially reducing the length process of optimizing medication for patients. Additionally, this approach could improve the equity in PD treatment planning by utilizing remote monitoring to reduce the need for difficult clinic visits. Strategic treatment planning using PD patient subtyping has been shown effective (144; 35) and stands to offer a data-driven approach to refining clinical management. Therefore, this study is a proof-of-concept to examine the feasibility of determining clinically relevant patient regimen clusters and identifying these clusters based on symptoms measured by wearable sensors to be utilized in the determination of future patients' treatment plans.

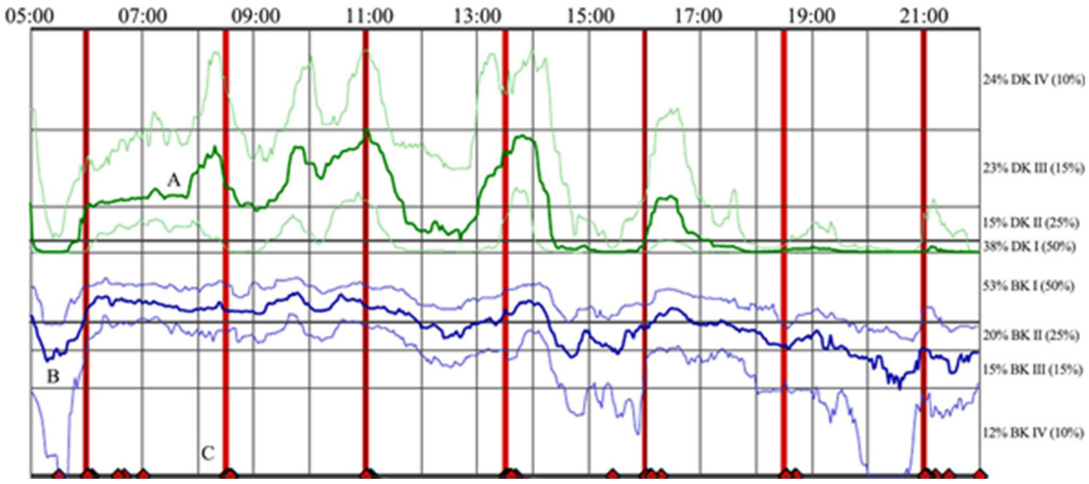

Figure 1.1: Sample PKG sensor time-series data (157), displaying an individual's change in dyskinesia and bradykinesia scores in response to medication with the median, 25th, and 75th percentile, compared to a non-PD control group averaged over six days. (A) The dyskinesia time-series, (B) the bradykinesia time-series, and (C) the patient's self-reported acknowledgment of medication administration.

1.2 Materials and Methods

1.2.1 Study Cohort

Characteristics of the patient cohort and selection process are thoroughly described in the study by Nahab et al. (157), which explored the clinical utility of the PKG in the routine care of Parkinson's patients. All patients were selected from the UCSD Movement Disorder Center from June 2016 to March 2017. The study's inclusion criteria included: An age range of 46–83, being on levodopa, and Hoehn and Yahr stages 1–3 (157). Patients were excluded if they had been previously diagnosed with dementia that could impact their use of the wearable sensor (157). The participants underwent two clinical assessment visits. Before each visit, the PKG sensor was worn by the patient for a six-day period, over which patients' key symptoms, namely, dyskinesia and bradykinesia, were scored every two minutes throughout the patient's full monitoring day (approximately 17 h). During the clinic visits, the physician assessed MDS-UPDRS motor subscales III and IV (79) were conducted. After the first study visit, a management plan, including an updated medication regimen, was developed based on the PKG report and clinical assessments (157). The patient then followed the updated management plan, while monitored by the PKG sensor for another six-day period. The patients were then evaluated in the second visit by the same clinical metrics, including the PKG. The change in patient's symptom control based on the updated management plan was determined. In this study, we retrospectively evaluate patients' symptom control under both management plans in the cohort assessed by Nahab et al. (157).

1.2.2 Study Design

We seek to group patients between clusters based on their optimized clinical medication regimens to determine if clinically relevant clusters exist within the patient cohort. Such patient clusters are thought to exist within cohorts, but may not be

identifiable by demographic information alone. To this end, we also examine the role of MDS-UDPRS-III scores and PKG time-series data to identify these clusters. These clusters would allow for the rapid estimation of near-optimal medication regimens for new patients. By examining within-subject symptom change during the optimization process of patients' medication regimens, we are seeking the patients' "best" performing medication regimens. The cluster allocation of new patients could then be predicted, placing the new patients within clusters that, on average, perform best (i.e., minimizes patients' symptoms based on cohort level estimations). While an individual's PD symptoms are unique, such an average best performing regimen could provide a clinician with an improved starting point for treatment planning, reducing the need for lengthy clinical assessments. The study design is shown in Figure 1.2.

We apply a statistical clustering technique to group patients based on their medication regimens under various conditions. Specifically, we examine their visit two regimens (during the physician-led optimization process) and their best performing medication regimen (being visit 1 or visit 2 when their symptoms were best controlled). These clusters and regimens are then compared to identify significant features which may aid in their consistent prediction.

We perform a comparative analysis between patients' MDS-UPDRS-III scores and PKG's summary dyskinesia and bradykinesia scores during treatment optimization to determine the efficacy of using wearable sensors for symptom management. To accomplish this, we examine the within-subject symptom change under both visits' regimens and determine which visit regimen best controlled a patient's symptoms as assessed by both MDS-UPDRS-III scores and the PKG's summary dyskinesia and bradykinesia scores. We compare these optimized regimens for each patient to examine discrepancies in patients' symptom assessments between the MDS-UPDRS-III and the PKG scores. We then examine patients' demographic information (study age, age at diagnosis, years of PD, and gender) under each clustering condition to identify statistically significant differences between similarly optimized regimens

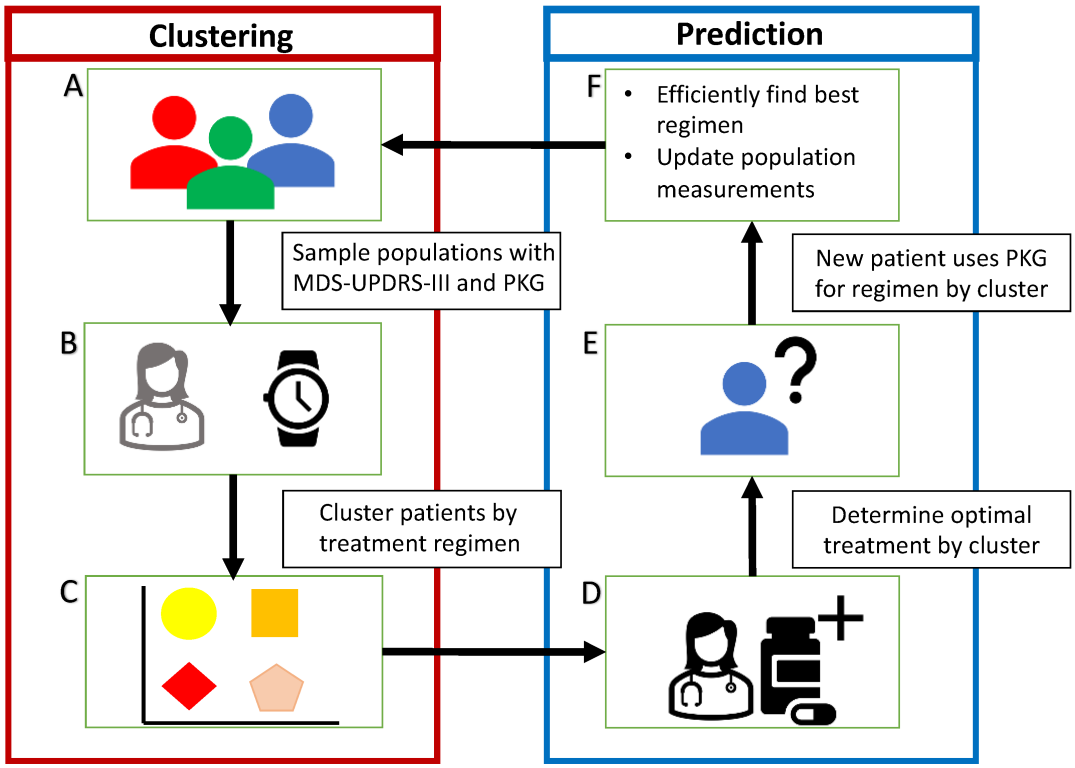

Figure 1.2: The study design is divided into two parts as labeled by the blocks "Clustering" and "Prediction". In the clustering block, (A) we begin with a diverse cohort of PD patients; (B) each patient is assessed by MDS-UPDRS-III scores and PKG summary dyskinesia and bradykinesia scores; the physician icon is greyed out as in the future for some remote contexts this may be accomplished using only the PKG time-series data, but we currently collect both for validation purposes; (C) we identify similar medication regimen clusters through k-means clustering. These clusters are used in the prediction block; (D) we optimize medication regimens and perform statistical analysis on demographic similarities for each group—to create a decision support tool to provide enhanced initial regimen estimates; (E) machine learning methods, specifically random forest, are applied to predict an unknown patient's optimized regimen cluster based on physician assessment and/or wearable sensor measurements depending on the context; (F) the new patient's data will be incorporated to improve the accuracy of the decision support tool.

to determine if demographic information alone may uniquely identify an optimized cluster.

Following the identification of optimized patient regimens based on the best clustering scheme, we apply machine learning techniques to predict the optimal medication regimens of patients through a combination of features. We examine the role of demographic information, MDS-UPDRS-III scores, and PKG time-series data in predicting the cluster allocation of patients. Such a prediction would create a decision support tool that could estimate a patient's optimized regimen aiding physicians. Further, we examine the potential of predictive algorithms without using traditional clinical symptom assessment methods (MDS-UPDRS-III) instead of based solely on wearable sensor measurements. We provide a machine learning algorithm with the patients' visit 1 PKG time-series data and predict their generalized optimal regimen. Being able to predict accurate estimates of a patient's optimal regimen remotely would save clinical time, equalize healthcare opportunities, and place less burden on patients during the process of medication optimization.

1.2.3 K-means Clustering

K-mean clustering is an unsupervised machine learning algorithm that partitions patients into a predetermined number of clusters (k) without a hierarchical structure (222). In this algorithm, clusters are initially formed, and each patient is grouped into their nearest cluster (with respect to Euclidian distance to cluster centroid). The clusters' centroids are then recalculated, seeking to minimize the distance between patients and their assigned centroid. Patients are then reassigned to the nearest clusters. This process is performed iteratively and continues until no patients are reassigned in an update (222).

Daily total levodopa equivalent dose (calculated by converting each PD drug to levodopa equivalent doses (LED) and cumulating them), daily total carbidopa/levodopa IR (immediate release) dose, which is the common dopamine replacement agent

utilized in PD drug regimens, and levodopa administration frequency were used in k-means clustering. These regimen features were used as each is likely to be modified in the physician-led optimization process. The number of clusters (k) was determined per the Within Cluster Sum of Squares (WCSS) measurement, which minimizes the within-cluster variance (e.g., the results of this analysis are provided in Figure A.1). This technique has been effectively used in healthcare applications for clustering data (238). The WCSS resulted to identify four clusters to meaningfully separate the patient cohort. Consequently, we evaluated demographic information (patient's age at visit 1, age at diagnosis, number of years experiencing PD symptoms, and gender) for each cluster under three clustering schemes.

1.2.4 Random Forest

A random forest classifier is a supervised machine learning algorithm that utilizes a large number of decision trees working as an ensemble (23). We opt to use the random forest classifier in this study as it is generally very robust against noisy or high dimensional datasets; it is not susceptible to overfitting (130). Four random forest classifiers were trained to stratify subjects into their designated clusters—identified based on the patients' best medication regimen according to PKG's summary dyskinesia and bradykinesia scores—using combinations of demographic information, visit 1 MDS-UPDRS-III scores, and visit 1 PKG time-series data. The PKG time-series data (representing two-minute increment measurements of the dyskinesia and bradykinesia scores) were extracted from the PKG report. Features were extracted and engineered from the PKG time-series via TSFresh (175), which calculates various time-series characteristics frequently used in classification tasks. Features importance ranking was conducted using the Gini index (214). The topmost important features were identified prior to cross-validation during preliminary experiments. These features are provided to the reader in Table A.1.

Each random forest model's performance was evaluated using leave-one-out cross-validation. In each set, a single patient was left out of the training data on which the random forest learned then that patient's cluster allocation as determined by their best medication regimen was predicted. This process is repeated until all patients have been used for testing, retraining the random forest model each time to prevent contamination between training and testing sets. Due to the unbalanced representation of clusters inherent in the dataset, repeated downsampling was used in each model. Repeated downsampling results in a balanced dataset for use in learning such that during training, the model did not favor the more representative cluster. Specifically, we randomly sampled from the more representative cluster to create a subset equal to the number of the less represented cluster for training each decision tree. This downsampling process was repeated for each decision tree. The total number of decision trees for each of the four random forest models was determined through preliminary experiments. Specifically, 200 decision trees were used in the demographic information, and the demographic information and visit 1 PKG time-series models, whereas 500 decision trees were used in the demographic information and visit 1 MDS-UPDRS-III model, and 100 decision trees were used in the demographic information, MDS-UPDRS-III, and visit 1 PKG time-series model.

Performance metrics included sensitivity, specificity, accuracy, positive predictive value (PPV), F1 score, and the area under the receiver operating characteristic (AUC). Sensitivity is the proportion of positives that are correctly identified. Specificity is the proportion of negatives that are correctly identified. PPV is the measurement of positive and negative results that are true positives. Accuracy is the measurement of correct predictions out of all predictions. The F1 score is the harmonic average of precision and recall. The AUC is the aggregate comparison of the true positive rate and the false positive rate at different classification thresholds and provides an overall performance metric for the model. We determined confidence intervals for each metric by repeating the random forest analysis 100 times under different initial random seeds.

1.3 Results

1.3.1 Cohort Characteristics

A total of 26 subjects (17 male and 9 female) clinical evaluations and PKG reports were included from the study by Nahab et al. (157). The PKG reports consisted of time-series data; specifically, dyskinesia and bradykinesia scores assessed every two minutes averaged over six days, along with medication administration times. The PKG time-series data were extracted from the PKG's reports. The patient cohort utilized in this study is a subset of that presented in (157). Two participants were excluded from the evaluation: One participant did not have corresponding PKG reports; the other had dosage inconsistencies in the recorded medication regimen. During visit 2, the overall mean MDS-UPDRS-III score was significantly reduced (visit 1: 28.9 ± 14.1, visit 2: 24.1 ± 13.5, p-value < 0.028 (157)). Demographic information and clinical characteristics of the participants are provided in Table 1.1. This retrospective study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the University of Tennessee (UTK-IRB-20-06007-XP).

1.3.2 Patient Clustering Using Medication Regimen

K-means clustering was utilized to allocate patients into one of four clusters based on their prescribed daily total levodopa equivalent dose, daily total carbidopa/levodopa IR dose, and levodopa administration frequency. In the first experiment, subjects were clustered by their visit two regimens, in which physicians had adjusted individualized regimens to optimize motor symptoms based on the clinical MDS-UPDRS-III scores and PKG report. In the second experiment, subjects were clustered according to the regimen associated with best motor function improvement (i.e., minimizes patients' symptoms), as defined by the MDS-UPDRS-III scores or PKG's summary dyskinesia and bradykinesia scores, respectively.

Table 1.1: Patient demographics and clinical characteristics.

	Mean ± SD
Gender (female/male)	9/17
Age in years	71.19 ± 9.70
Age at diagnosis (years)	65.77 ± 10.37
Visit 1 MDS-UPDRS-III	28.89 ± 14.07
Visit 2 MDS-UPDRS-III	24.12 ± 13.50
Visit 1 H&Y stage	1.77 ± 0.71
Visit 2 H&Y stage	1.85 ± 0.78
Visit 1 levodopa equivalent dose (mg)	498.94 ± 309.88
Visit 2 levodopa equivalent dose (mg)	637.40 ± 322.37
Time between clinical visits (days)	65.62 ± 26.46

Figure 1.3 presents the clusters based on visit 2's medication regimen. According to Figure 1.3.a, when MDS-UPDRS-III scores are used as the comparison metric: Eighteen subjects show symptom improvements, while seven demonstrated symptom worsening, and one remained unchanged. As shown in Figure 1.3.b, when the PKG's summary dyskinesia and bradykinesia scores are used: Seventeen subjects show symptom improvements, eight demonstrated symptom worsening, and one remained unchanged. The demographic information associated with each cluster is provided in Table 1.2. Cluster D was statistically different from clusters A and B with respect to disease duration and age at diagnosis ($p < 0.05$); no other clusters were statistically different in terms of demographic parameters ($p > 0.05$).

Figure 1.3.c,d presents the medication regimens' clusters associated with the best motor function between the two study visits. The centroid positions of the clusters are in different locations, since medication regimens related to improved clinical function differ according to MDS-UPDRS-III scores and PKG's summary dyskinesia and bradykinesia scores. The patients' demographic information for each cluster is provided in Table 1.2, while the breakdown of PD medication and dosing is provided in Table 1.3 . It should be noted that for MDS-UPDRS-III scores and PKG scores that were unchanged between visit 1 and visit 2, the regimen associated with visit 2 was considered "best" and used in this clustering. No two clusters are statistically different ($p > 0.05$) in terms of gender or age at diagnosis. Cluster D was statistically different from cluster A with respect to patients' age ($p < 0.05$) for the best MDS-UPDRS-III scores. Likewise, cluster D was statistically different from clusters A and B with respect to disease duration ($p < 0.05$) for the best PKG's summary dyskinesia and bradykinesia scores.

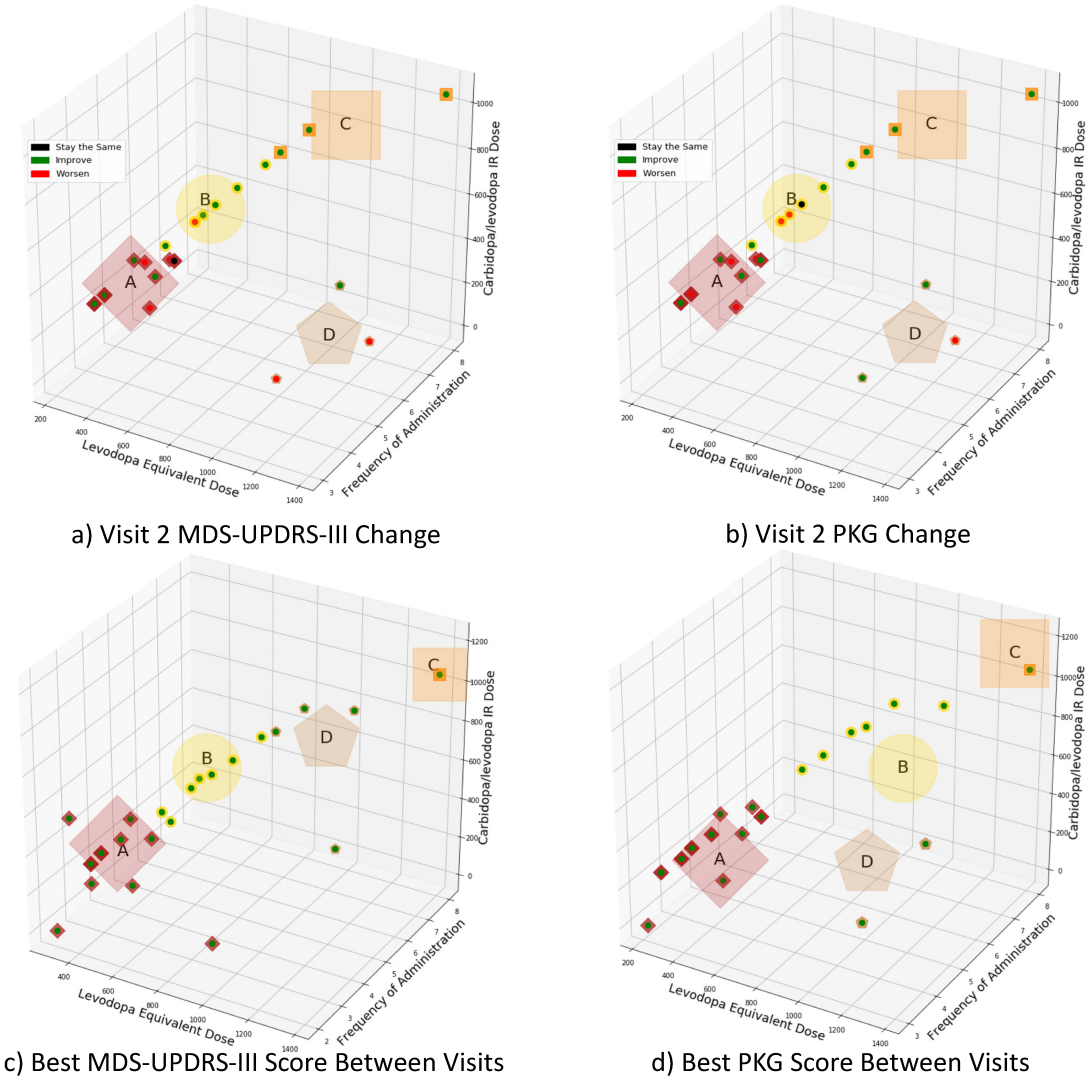

Figure 1.3: demonstrates the clusters for the motor function changes between visit 1 to visit 2 based on the MDS-UPDRS-III scores and the PKG's summary dyskinesia and bradykinesia scores, respectively. (c,d) highlight the subjects' best symptom control recorded using MDS-UPDRS-III scores and PKG's summary dyskinesia and bradykinesia scores, respectively. The large shapes denote each cluster's centroid, and the exterior marker of each point corresponds to the cluster centroid shape. The capital letters (A–D) are used to refer to each cluster. Each point's interior maker in (a,b) represents the state of MDS-UPDRS-III and PKG change for each patient, where a black interior-point denotes the patient stayed the same between visits, a green point denotes that the patient's visit 2 scores were better than their visit 1 scores. A red point denotes that the patient's visit 2 scores were worse than their visit 1 scores. Note that due to the three-dimensional projection of the plot, the distance between points may appear skewed.

Table 1.2: Demographic and clinical information for clusters depicted in Figure 1.3.a–d. † Pairwise p-value < 0.05 when compared with both Cluster A and B. †† Pairwise p-value < 0.05 when compared with only Cluster A. Unless denoted by † or ††, each pairwise cluster comparison yields no statistical difference (p-value > 0.05).

	Cluster A	Cluster B	Cluster C	Cluster D
Visit 2 MDS-UPDRS-III & PKG Scores				
Study age (years)	74.31 ± 9.99	67.96 ± 9.70	68.15 ± 3.96	68.29 ± 5.68
Age at diagnosis (years)	69.69 ± 9.22	65.41 ± 9.08	60.48 ± 10.32	54.95 ± 4.15 †
Years of PD	4.62 ± 3.54	2.55 ± 1.78	7.67 ± 6.65	13.33 ± 2.05 †
Gender (female/male)	5/8	2/5	1/2	1/2
Number of participants	13	7	3	3
Best MDS-UPDRS-III				
Study age (years)	75.62 ± 7.84	65.59 ± 10.86	63.54 ± 0.00	67.40 ± 3.74 ††
Age at diagnosis (years)	70.26 ± 7.67	62.90 ± 10.35	46.54 ± 0.00	59.90 ± 8.17
Years of PD	5.36 ± 4.55	2.69 ± 1.77	17.00 ± 0.00	7.50 ± 4.61
Gender (female/male)	5/9	2/5	0/1	2/2
Number of participants	14	7	1	4
Best PKG Scores				
Study age (years)	72.92 ± 10.21	67.66 ± 7.54	63.54 ± 0.00	70.93 ± 5.24
Age at diagnosis (years)	68.69 ± 9.47	63.49 ± 8.78	46.54 ± 0.00	57.43 ± 2.74
Years of PD	4.23 ± 3.31	4.17 ± 4.14	17.00 ± 0.00	13.50 ± 2.50 †
Gender (female/male)	6/11	3/3	0/1	0/2
Number of participants	17	6	1	2

Table 1.3: Dosage and medication types for clusters depicted in Figure 1.3.c,d. Medications are aggregate results for each cluster and do not represent a single subject's regimen.

	Cluster A	Cluster B	Cluster C	Cluster D
Best MDS-UPDRS-III				
LEDD	387 ± 151	643 ± 127	1380 ± 0	1157 ± 183
Carbidopa/levodopa IR	279 ± 159	629 ± 150	1050 ± 0	925 ± 299
Carbidopa/levodopa CR	21 ± 80	–	200 ± 0	300 ± 476
Ropinirole	–	1 ± 4	4 ± 0	–
Selegiline	–	–	10 ± 0	1 ± 1
Rasagiline				
Rytary	194 ± 547	–	–	–
Best PKG Scores				
LEDD	381 ± 105	942 ± 233	1380 ± 0	1131 ± 206
Carbidopa/levodopa IR	335 ± 147	917 ± 183	1050 ± 0	250 ± 354
Carbidopa/levodopa CR	18 ± 73	33 ± 82	200 ± 0	500 ± 707
Ropinirole				
Selegiline	0.6 ± 2.4	–	10 ± 0	1 ± 2
Rasagiline	0.1 ± 0.3	–	–	0.3 ± 0.4
Rytary	45 ± 184	–	–	1170 ± 1655

1.3.3 Random Forest Classification Using PKG Readouts

Four random forest classifiers were trained to examine the efficacy of using combinations of demographic information (patient's age at visit 1, age at diagnosis, number of years experiencing PD symptoms, and gender), visit 1 MDS-UPDRS-III scores, and visit 1 PKG time-series data, to stratify the subjects in clusters A and B, as identified through k-means clustering using the best PKG score (see Figure 1.3.d). As shown in Table 1.2, "Best PKG Score," clusters A and B contain 17 and 6 participants, respectively. As noted in Section 1.3.2 ("Patient Clustering Using Medication Regimen"), clusters A and B were the most statistically similar with respect to the demographic information and contained the majority of participants. The performance of each of the classifiers is presented in Table 1.4.

The random forest classifier using solely demographic information achieved a sensitivity of $61.3 \pm 1.0\%$, a specificity of $62.3 \pm 1.5\%$, an accuracy of $61.6 \pm 0.8\%$, a PPV of $82.2 \pm 0.6\%$, an F1 score of $70.1 \pm 0.8\%$, and an AUC of 0.618 ± 0.008. Whereas the random forest classifier using both demographic information and visit 1 MDS-UPDRS-III scores achieved a sensitivity of $65.2 \pm 0.8\%$, a specificity of $66.0 \pm 0.7\%$, an accuracy of $65.4 \pm 0.6\%$, a PPV of $84.4 \pm 0.3\%$, an F1 score of $73.5 \pm 0.6\%$, and an AUC of 0.656 ± 0.005.

The random forest classifier using demographic information and visit 1 PKG time-series data had superior performance to the subjective MDS-UPDRS-III-based classifier. To train this random forest classifier, over 1000 features were extracted from PKG sensors' dyskinesia and bradykinesia time-series for each patient, of which the top ten most important features were included in the analysis. These features are provided in Table A.1. This random forest classifier achieved a sensitivity of $84.5 \pm 0.7\%$, a specificity of $81.7 \pm 2.2\%$, an accuracy of $83.8 \pm 0.7\%$, a PPV of $93.1 \pm 0.8\%$, an F1 score of $88.5 \pm 0.5\%$, and an AUC of 0.831 ± 0.011.

The random forest classifier using demographic information, visit 1 MDS-UPDRS-III and visit 1 PKG time-series data had the best overall performance with a

Table 1.4: Random forest classifier performance identifying subjects in clusters A and B, as stratified through k-means clustering under the best PKG score. Each classifier was trained with a combination of features: Demographic information, demographic information and MDS-UPDRS-III scores (Visit 1), demographic information and PKG time-series data (Visit 1), and demographic information, visit MDS-UPDRS-III scores (Visit 1), and PKG time-series data (Visit 1).

	Demographics Alone	Demographics and MDS-UPDRS-III	Demographics and PKG	Demographics, MDS-UPDRS-III and PKG
Sensitivity	61.3 ± 1.0%	65.2 ± 0.8%	84.5 ± 0.7%	86.5 ± 0.5%
Specificity	62.3 ± 1.5%	66.0 ± 0.7%	81.7 ± 2.2%	87.7 ± 1.6%
Accuracy	61.6 ± 0.8%	65.4 ± 0.6%	83.8 ± 0.7%	86.9 ± 0.6%
PPV	82.2 ± 0.6%	84.4 ± 0.3%	93.1 ± 0.8%	95.3 ± 0.6%
F1 Score	70.1 ± 0.8%	73.5 ± 0.6%	88.5 ± 0.5%	90.7 ± 0.4%
AUC	0.618 ± 0.008	0.656 ± 0.005	0.831 ± 0.011	0.871 ± 0.008

sensitivity of $86.5 \pm 0.5\%$, a specificity of $87.7 \pm 1.6\%$, an accuracy of $86.9 \pm 0.6\%$, a PPV of $95.3 \pm 0.6\%$, an F1 score of $90.7 \pm 0.4\%$, and an AUC of 0.871 ± 0.008. The PKG time-series features included in this random forest model are identical to those listed in Table A.1.

1.4 Discussion

Utilizing a Parkinson's patient cohort dataset consisting of within-subject medication regimen titrations—clinically assessed by the MDS-UPDRS-III scores and PKG's summary dyskinesia and bradykinesia scores—k-means clustering was used to group patients in terms of daily total levodopa equivalent dose, daily total carbidopa/levodopa IR dose, and levodopa administration frequency. We demonstrate that subjects can be meaningfully clustered based on longitudinal dopaminergic treatment regimens. The sensor-based assessments of the PKG can estimate patient symptoms corresponding to similar MDS-UPDRS-III scores. Further, the PKG sensor can be thought of as enhancing the granularity of this clustering method compared with the MDS-UPDRS-III scores: When referencing cluster D, the PKG clustering has statistical significance between clusters A and B, whereas the MDS-UPDRS-III clustering only has statistical significance between cluster A.

Figure 1.3.a,b show the difference between MDS-UPDRS-III scores and PKG's summary dyskinesia and bradykinesia scores when determining subject improvement. This difference is quite minor between the two assessment instruments with respect to the regimens yielding the best motor function. This result supports the growing body of literature that the MDS-UPDRS-III score can be adequately determined by wearable sensor estimates (85; 157; 168). This comparison shows that determining the optimization of a patient's medication regimen may be effectively estimated using sensors. However, since the cohort was treated, considering both PKG changes and traditional clinical assessments, further conclusions regarding the robustness of the treatment approaches cannot be drawn. Therefore, our results suggest that efficiently

establishing a patient's best performing regimen could be improved by objective measurements.

Additionally, the cohort's demographics and clinical characteristics are generally statistically indistinguishable from the MDS-UPDRS-III and PKG clustering methods. Only subjects with the longest disease duration were grouped into a separate cluster. This group required greater dosages and more frequent administration of dopaminergics for symptom control—a treatment strategy aligned with current clinical practice. Furthermore, using demographic information along with PKG time-series data yielded a classification model that enabled the random forest classifier to predict the cluster allocation of patients with high accuracy. A classification algorithm, such as one utilizing PKG measurements, could be used to streamline medication regimen optimization by providing a clinician with an estimate of a patient's optimal regimen prior to clinical assessments. This would allow for a more complete view of the patient's symptoms and medication response as the patient is continuously monitored throughout their daily lives. It is worth noting that using demographic information, MDS-UPDRS-III scores, and PKG time-series data resulted in the best performance and an incremental improvement (3%) over the model that only used demographic information and PKG time-series data. However, models that do not include MDS-UPDRS-III scores can be used in remote settings where physicians may not have direct access to patients. Hence, the restrictions and considerations around access can select the right predictive model as part of a flexible decision support tool.

While the prognostication of disease progression is evident in clinical subtyping (222; 155; 65), the implications on treatment have yet to be established. Similar to the phenotypic variability of PD, the treatment approaches are also heterogeneous. Therefore, a continuous assessment of treatment response not only offers the possibility of more robust medication titrations, but the ability to cluster these sensor-based responses may help potentiate the impact of the emerging clinical phenotypes. This proof-of-concept study establishes that rich information can be extracted from

time-series data collected from wearable sensors, such as the PKG—that measures both motor function and medication responses—and incorporated into ML algorithms to build predictive models capable of expanding the clinical treatment platform.

1.4.1 Limitations

The small patient sample is a limitation of this work. Due to the size of the examined patient cohort, a fully representative cluster of medication regimens was not likely achieved. This may be further biased, due to the underrepresentation of rare subtypes. By incorporating additional medication regimens in the future would provide a more comprehensive clustering scheme and would improve the optimized regimens estimates.

Additionally, several subjects' motor control symptoms within the cohort were never successfully controlled as measured by the PKG's summary dyskinesia and bradykinesia scores (168). Such subjects, when further optimized clinically, may be placed within a different cluster—altering the demographic and clinical information associated with that cluster. A more representative patient cohort followed longitudinally stands to enhance the clustering and possibly reveal other inherent treatment clusters.

It should also be noted that the patients' treatment regimens were optimized by clinicians considering both MDS-UPDRS-III scores and PKG's summary dyskinesia and bradykinesia scores. This could further introduce bias into the classification algorithm as PKG measurements may perform better at predicting patients, in which physicians heavily utilized the PKG (namely those with motor fluctuations and dyskinesia) to determine their optimal regimens. Leave-one-out cross-validation was used in the analysis as each patient predicted by the random classifier is effectively a completely new patient—thus minimizing this potential source of bias.

Finally, nonmotor symptoms were not directly considered in the determination of the best regimen. Therefore, regimen estimates that utilize sensors, such as the PKG,

which primarily measures changes in motor symptoms, will need to be holistically considered by a physician.

1.5 Conclusions

Clustering patients in a treatment-related manner has the potential to determine inherent regimen similarities within clinically relevant clusters. These inherent regimens are not necessarily specific to demographic groups. The identification of such regimen similarities could guide strategic treatment planning for a data-driven approach to refine clinical management, namely, by providing patients with wearable sensors to monitor their symptoms and medication responses. This approach can provide clinicians with estimated regimens based on the patients' demographic information, clinical assessments, and wearable sensor measurements. The most accurate regimen estimations are achieved using a combination of clinical assessments and wearable sensors. This potentially expands the current approach for optimizing drug regimens in the clinic and remotely.

Chapter 2

Partially Observable Markov Decision Process for Optimizing Medication Regimens via Wearable Sensors in Parkinson's Disease

2.1 Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disease that manifests with the deterioration of motor control (52; 12). As PD is progressive, symptom management poses a significant healthcare challenge as it requires frequent reevaluation. In the United States, 60,000 individuals are diagnosed with PD annually (61). The global number of individuals living with PD is estimated to be 6.2 million (70). Additionally, the global rate of PD is expected to increase as life expectancy increases, putting a strain on medical systems globally (61). PD patients tend not to receive consistent follow-up care due to logistical challenges, including a shortage of neurology specialists, extended travel time, patient disability, and prolonged clinical wait times (62; 234). Recurring clinical visits and symptom evaluations are critical to

maintaining patients' quality of life as PD medication regimening requires frequent adjustments.

PD manifests largely with motor symptoms, including bradykinesia (a stiffness of movement), rigidity, shuffling gait, and hand tremor (52; 12). In the majority of cases, these cardinal symptoms of PD can be effectively managed through dopaminergic medications (78; 161). Levodopa (L-dopa) is one such dopaminergic medication and drives the majority of PD patient medication regimens (78; 161). However, the regimening of L-dopa and other PD medications is dynamic as the patients' symptomologies change during the course of the disease. In more advanced stages of the disease, patients may experience "wearing-off" periods, characterized by the re-emergence of PD symptoms between medication dosages (78; 161; 45). These periods prompt physicians to increase the titration and frequency of dopaminergic medications (45). However, as the dosage of medication increases, side effects such as medication-induced dyskinesia (involuntary movements) become more prominent (45). These competing interests of symptom control (bradykinesia) and reduction in side effects (dyskinesia) necessitates complex dynamic and patient-specific medication regimen optimization.

Currently, in clinical practice, patients' regimen efficacy is based on physicians' overall impression of a patient's motor disability as assessed by the Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) (79). Non-motor symptoms are largely captured by subjective, patient self-reported Hauser paper diaries and questionnaires (94). The lack of objective, continuous monitoring of symptoms have limited the efficacy of treatment planning. However, developments in sensor technology has enabled objective day-to-day monitoring in PD (187; 140; 99).

In this study, we present a partially observable Markov decision process (POMDP) to prescribe medication regimens in PD based on wearable sensor measurements. We consider the role of patient subtyping and medication pharmacokinetics in the calibration of these models. We compare the POMDP models' performance in a cohort of PD patients whose physician-assessed medication regimens were optimized

following a two-week assessment of patient symptoms via wearable sensors. The proposed POMDP models could have a widespread impact on the treatment of several progressive chronic diseases and could provide a tool for aiding physicians in medication regimening.

The paper is organized as follows. We first review the existing literature on PD management and the role of POMDPs and reinforcement learning models for the optimization of treatment planning in Section 2.2. Then we introduce the POMDP model and the patient cohort used for calibration and evaluation in Section 2.3. Experimental results and conclusions are provided in Sections 2.4 and 2.5, respectively.

2.2 Background and Relevant literature

In this section, we review the recent developments in the use of data-driven optimization processes in PD treatment planning. We then review the use of wearable sensors in PD. Finally, we review the modeling literature on L-dopa to enhance the accuracy of our POMDP models.

2.2.1 MDPs and Reinforcement Learning in Parkinson's Disease

There is a large body of research using Markov decision processes (MDPs) and reinforcement learning (RL) in optimizing treatment planning. Such studies typically seek to optimize group or patient-specific treatment plans given some combination of patients' medical history, clinical evaluations, wearable sensor measurements, demographics, or genetic biomarkers. As such, PD treatment planning has been modeled in a variety of contexts.

MDPs and POMDPs have been previously leveraged to optimize PD treatment planning. Both Dams et al. and Eggert et al. examined the cost-effectiveness of

different treatment plans for maintaining PD symptom control (48; 64). Additionally, the stochastic nature of PD progression, along with the cost-effectiveness of treatment planning, was explored via POMDPs (228). These articles, however, did not seek to dynamically optimize patients' medication regimens, nor did they model medication response, but instead, medication regimens in the aggregate. Additionally, work has also been done to model real-time medication response optimized treatment plans (233); however, this work did not evaluate real-world data or generate patient-specific models. Moreover, the progression of PD has been modeled using multi-state Markov and hidden Markov models (142; 44; 202). Such progression studies reveal interesting characteristics of the disease, which can be leveraged to better understand patients' medication responses and subtypes.

RL has also been leveraged to optimize PD treatment plans. Kim et al. applied RL to optimize patient treatment plans over a multi-visit process (114). However, this work used subjective clinical evaluations over a short-term optimization processes. To move beyond the 'snap-shot' subjective clinical evaluations, Baucum et al. used RL in combination with wearable sensors to monitor patients' symptoms (9). The work of Baucum et al. was a breakthrough evaluation process for the effective modeling of medication response in PD; however, this work was concerned with the optimization of patients using limited data. Our work aims to extend this existing literature by improving the medication modeling which underpins these RL models, as well as provide a Markov policy that can be evaluated in real-time to dynamically optimize a patient's medication regimen.

2.2.2 Wearable Sensors in Parkinson's Disease

The use of electronic medical devices for disease monitoring and treatment outside hospitals is growing (25; 216; 147). Wearable sensors are one such device used for non-intrusive patient monitoring. Wearable devices have been used for a wide array of health monitoring applications, including medication adherence (37; 185).

Additionally, wearable sensors have been leveraged to aid in personalizing and informing patients' health trajectories (126; 173; 51). Further, a limited body of literature has investigated the use of wearable sensors in treatment planning, e.g., in chronic disease management. For instance, wearable motion sensors have shown promise in monitoring motor symptoms in PD patients, allowing physicians to utilize summaries of patients' wearable sensor data when prescribing medication regimens to improve patient outcomes (157).

In relation to PD, wearable sensors have been studied in clinical environments to assess motor fluctuations in patients (172). Such work allows for the assessment of motor fluctuations during "normal" fluctuation cycles in treatment response (172). There have also been clinical assessments of gait monitoring using wearable sensors (143; 157; 187). The personal kinetigraph (PKG) is one such wearable device that utilizes accelerometers to identify patient symptoms (196). This system proved effective in the diagnostics of clinical symptoms and allowed for informed decision-making for physician treatment planning (196; 157).

2.2.3 Levodopa Response

While PD has no known cure, its cardinal symptoms can be effectively controlled through the use of dopaminergic medication. Since its discovery, L-dopa has been the gold standard in PD treatment planning. However, orally administered L-dopa can have challenges with symptom control, especially in advanced-stage patients (129). These patients frequently have the emergence of "off" states characterized by the re-emergence of symptoms (129). In advanced PD patients, off periods can occupy more than a third of a patients waking hours, which can have a dramatic impact on patients' quality of life (129). Off periods occur in more than 40% of patients after 4 to 6 years of PD (161). These off periods are due to the pharmacokinetics of L-dopa. Traditionally, the pharmacokinetics of interest include the time to peak dose, bioavailability/absorption of L-dopa, and plasma half-life. These pharmacokinetic

features are patient-specific, and it has been shown that patient demographics, including sex, body mass index (BMI) or weight, and disease duration have a direct impact on L-dopa response (129; 43; 10). L-dopa off periods can be reduced by increasing the magnitude or frequency of L-dopa administration. However, a known side effect is L-dopa-induced dyskinesia (161). Medication-induced dyskinesia has been shown to become more intense with increased dosages (161). These trade-offs are a substantial challenge for physicians in the L-dopa optimization process.

There is a growing body of literature to estimate the pharmacokinetic parameters of L-dopa (1; 6; 8; 166; 149; 218). Many of these studies have collect primary data sets of patient plasma and symptom response to the administration of L-dopa. Existing pharmacokinetic response models are then fit to such plasma levels to estimate L-dopa's characteristic pharmacokinetic parameters. However, a variety of models have been utilized for this task with different number of components and absorption assumptions (102; 243). This is further complicated as each formulation of L-dopa will have different specific characteristics (189).

2.3 Model formulation

In this section, we will first describe our patient cohort used in the calibration and testing of the model. We will then outline the POMDP model used in the optimization of treatment regimens. Next we will describe the process of identifying patient subtypes, which was used to define group-level POMDPs policies to move beyond cohort level treatment planning. Finally, we will describe our modeling of L-dopa's pharmacokinetics process.

2.3.1 Patient Cohort

Data used in the study were originally reported by Nahab et al. in (157). The patient cohort consisted of 26 PD patients, whose symptoms (bradykinesia and dyskinesia)

were monitored by wrist-worn wearable sensors, the Personal Kinetigraph (PKG; Global Kinetics Corporation). The PKG is a FDA-approved PD symptom monitor, which uses built-in accelerometers to collect gait, tremor, and movement pattern data; these data are summarized using previously-validated algorithms (165; 99; 85; 168) to obtain clinically-relevant bradykinesia and dyskinesia scores. The PKG additionally, serves as a patient aid providing medication vibration alerts on a clinician prescribed fixed schedule.

In the study by Nahab et al., data was collected over a two visit period. Patients were initially assessed under a prescribed L-dopa based medication regimen, these patients then wore the PKG sensor for a six-day period. The patients were then re-evaluated using clinical metrics, as well as using measurements from the PKG sensor. The patients' medication regimens were updated following the clinical visit, and this second regimen was similarly monitored by the PKG sensor over an additional six-day period. We refer to the initial six-day PKG period as 'Visit 1' and the second period as 'Visit 2'.

A sample PKG report is provided in Figure 2.1. In this summary plot, data was collected between 5:00am and 10:00pm. This summary plot consists of two scores, the top score represents patients' dyskinesia and the bottom represents patients' bradykinesia. Both are rated on scales from I to IV representing symptom severity with I being least severe to IV being most severe. Note that following the administration of medication this patient's bradykinesia score improves (from BK II to BK I), whereas their dyskinesia worsens (from DK I to DK III). However, these trends do not perfectly follow the prescribed times for medication administration, this is likely due to patients taking their medication before or after the prescribed times. Finally, note that patients' symptom response to L-dopa is varied across the assessment week, suggesting the need for a stochastic modeling approach.

Lastly note that patients were taking one of three L-dopa formulations during this study L-dopa IR ('immediate release'), L-dopa CR ('controlled release'), and

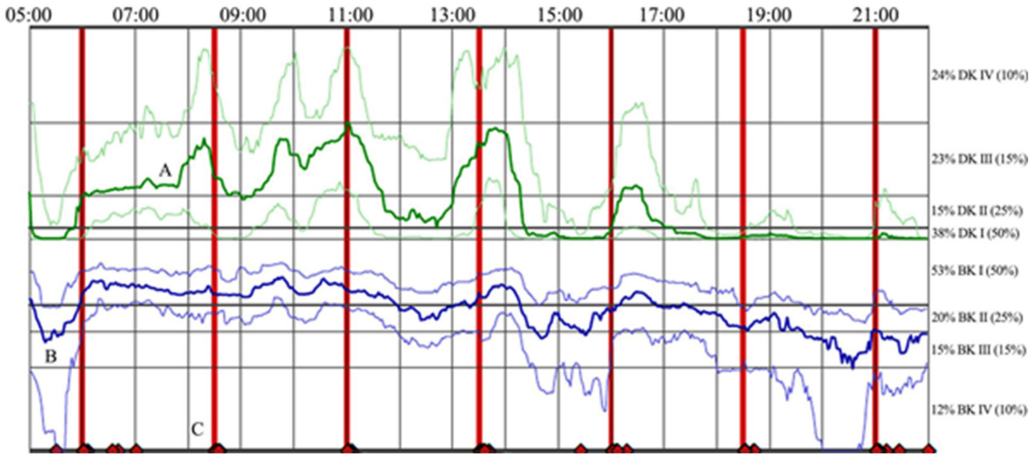

Figure 2.1: Example of a PKG summary plot (157), presenting both dyskinesia (top) and bradykinesia (bottom) scores over a six-day period. In both plots the 'I' region represents the lowest symptom severity and the 'IV' region represents the greatest symptom severity. The bold lines represent the median scores, whereas the lower and upper faded lines represent the 25th and 75th percentiles, respectively. The vertical red lines represent clinician prescribed medication administration times, whereas the red diamonds at the bottom of the image present patients' 'self-reported' medication administration times.

Rytary (which is an extended-release form of L-dopa). In addition, patients may have received dopamine agonists to alter the absorption time of L-dopa. In this study, we restrict our patient cohort to 15 patients of the original cohort, who received only L-dopa IR over both evaluations periods. Our patient cohort is made up of 10 males/5 females, with an average age of 71 years (min: 53 years, max 87 years), with anb average time since PD diagnosis of 4 years (min: 1 year, max: 11 years), and a Visit 1 average L-dopa IR dosage of 380mg (min: 200mg, max: 800mg) and an average Visit 2 dosage of 537mg (min: 300mg, max: 1000mg). Note that none of our patients have an 'early PD' diagnosis with the minimum age of PD diagnosis of 53 years. A summary of the patient cohort is provided in Table 2.1.

2.3.2 Model Definition

In this section we formulate a partially observable Markov decision process (POMDP) model to optimize patients' medication regimens (i.e., to maximize the patient's utility in any given day). We describe a POMDP model using a seven-tuple (S, A, O, p, K, Ψ, r), with a set of states S, actions A, observations O, state transition probabilities p, patient subtypes C, an information matrix Ψ, and a reward function r. Consistent with PKG time series, we discretize time into 2-min time segments, which we refer to as time epochs.

The proposed model aims to describe the pharmacokinetic process of medication from which patient medication regimens can be optimized over the course of many chronic diseases. Without invasive clinical measurements at time specific plasma concentration of medication cannot be obtained. However, as medication impacts the symptomology of patients, by using wearable sensors it is possible to estimate patients' current concentration of medication. This leads to a partially observable model in which wearable sensor measurements give insights into patients' overall state. In the case of PD, we model patients' symptomatic response to L-dopa. Specifically, this

Table 2.1: Summary of the patient cohort

	Average	Min	Max
Gender	10M/5F	-	-
Study Age (years)	71 ± 11	53	87
Age at Diagnosis (years)	68 ± 9	52	80
Years of PD	4 ± 3	1	11
H&Y Stage	1.6 ± 0.7	1	3
V1 L-dopa IR Dose (mg)	380 ± 150	200	800
V2 L-dopa IR Dose (mg)	540 ± 220	300	1000

model allows for the administration of various dosage of medication and simulates the varied probabilistic responses to each medication dosage and their interactions.

- **Decision epochs:** To preserve the Markovian assumption, we assume that the pharmacological response is history independent. Let t denote the amount of time elapsed from the beginning of the monitoring day and T denote the length of the monitoring horizon (i.e., a day) in units of 2-minutes. Here, $T = 510$ representing the patients' monitored day from 5:00am to 10:00pm.

- **Actions:** Let action $a \in A = \{0, 100, 200, 250, \ldots, 1000\}$ denote the amount of medication administered (i.e., L-dopa dose), with $a = 0$ representing a 'wait' action (no medication). We assume the patient has perfect adherence to the prescribed regimen (all actions taken on time and without fail). We consider dosages in 100mg and 250mg steps as those are the commonly available orally administered dosages of L-dopa IR, our considered dosages stop at 1000mg as that is the largest single dosage within our cohort.

- **States:** Let $s_t = (c_p, \tau_1, d_1, \tau_2, d_2) \in S$ denote the medication state of the patient at time t. Here $c_p \in C$ denotes the patients' plasma concentration of medication. Additionally, $d_i \in A$ represent the dosage of medication which has a therapeutic effect on symptom control, (i.e., $a \neq 0$. In the context of L-dopa IR, the medication has an uptake period during which the medication is stochastically processed by the body to provide therapeutic effect. We allow for the model to simultaneously track two dosages of medication. Here τ_i represents the last non-wait action (i.e., $\tau_i = a_t$ where $a \neq 0$. If $\min(t - \tau_i) < 30$ (i.e., 60-minutes) then additional action can not be taken. This 'blackout period' prevents the over-administration of medication by the model, which may result in medication adherence fatigue or pose dangerous health outcome from overdosing. Similarly, as the medication is used by the body, the bioavailability of the medication reduces in a stochastic manner. Further, the value τ_i is

replaced for the lower of the two τ values,

$$\tau_1 = t_{a\neq 0}, \quad \text{if} \quad \tau_1 < \tau_2 \quad \text{or}, \tag{2.1}$$

$$\tau_2 = t_{a\neq 0}, \quad \text{if} \quad \tau_2 < \tau_1. \tag{2.2}$$

- **Patient subtypes:** Patients are given a subtype of L-dopa response $k \in K = \{1, 2, 3, 4\}$ this subtype is based on patients demographic information and previous L-dopa response curves. These subtypes enable subtype-specific polices to better optimize patients' medication regimens.

- **Transition probabilities:** At each time epoch the patient will probabilistically transition to a medication state defined as $p_t(j|s, a)$ which denotes the probability that the embedded Markov process transitions to state $j \in S$ from state $s_t \in S$, under action a. Note that this process is independent of subtype as we assume that patient subtypes are characteristic of symptom response (i.e., observations).

- **Observations:** Let $o_t \in O$ denote the observation made at time t which provides a hint as to the patient's underlying medication state. Let the emission matrix $\Psi_{t,k}(o|s)$ denote the probability of making observation o, given the true core state s at time t, for a subtype k.

- **Beliefs:** Let π_t denote the belief of the underlying state at time t. Formally, let the belief be a vector where $\pi_t = [\pi(1), \pi(2), \ldots, \pi(31)]$, where $\pi(\cdot)$ corresponds to beliefs in states s and π_s be any element of π_t. The belief π_t is updated at the beginning of the next decision epoch using Bayes' theorem (74).

$$\pi_k(j) = \frac{\Psi_k(o|s) \sum_{j\in S} \pi_t \cdot p_t(j|s, a)}{\sum_{j\in S} \Psi_k(o|j)\pi_t \cdot p_t(j|s, a)}$$

- **Total expected reward:** Let $r_t(\pi)$ denote the immediate reward corresponding to symptom control in belief state π at time t. The goal is to maximize the

total expected reward to go, $V_{t,k}(\pi)$ by selecting the optimal action, however this is made based on decisions its belief. Hence, for $t \leq T$, we have

$$V_{t,k}(\pi) = \max_{a \in A} \left\{ \sum_{j \in S} \sum_{s \in S} \sum_{o \in O} \pi_t(s) \cdot p_t(j|s,a) \cdot \Psi_{t,k}(o|j) \cdot \left(\sum_{i=t}^{t} r_i(\pi) + V_{t,k}(\pi') \right) \right.$$
$$\left. + \left(1 - \sum_{j \in S} \sum_{s \in S} \pi_t(s) \cdot p_t(j|s,a) \right) \cdot \left(\sum_{k=t}^{T} r_k(\pi) \right) \right\},$$

where π' is the updated belief at time t. Starting from state s_t, the next decision epoch. Alternatively, for the event where $t = T$, the process will extend past the monitoring horizon, the process terminates at time T and a lump sum reward is received. This lump sum reward is heavily discounted for the remainder of the medication response, as it is scaled per the lower symptom control importance during non-waking hours/sleep.

2.3.3 Patient Subtyping

To generate group level treatment planning policies, patients were subtyped based on their response to L-dopa. It has been shown that sex, BMI, and years with PD all impacted patient's response to L-dopa (129; 43; 10). Using patients' demographic information (sex, and years with PD), peak single dose symptom severity per milligrams of L-dopa, and time to peak dose per milligrams of L-dopa, patients were clustered into responder subtypes. Specifically, we used patients' demographic information collected at Visit 1. To obtain patients' peak dosage information, we evaluated their first daily response to L-dopa from Visit 1. Specifically for this subtyping analysis, patients' PKG scores were only considered from their first dosage to their second. This was done to mitigate the impact of additional bioavailability of L-dopa in the body following patients' first daily dose. Peak symptom severity was the maximum bradykinesia and dyskinesia scores obtained during this time period. Time to peak dose was determined as the time in which peak bradykinesia control

was achieved. Note that the time period evaluated will differ between patients. Additionally, we assume patients took their medications on time and without fail as prescribed.

We apply k-means clustering to separate the patients into subtypes based on L-dopa response. The performance of the k-means clustering model was evaluated over different numbers of clusters. The calibration is shown in Figure 2.2. In this 'elbow' plot, the k-means clustering performance can be roughly described in three parts between 1-4, 4-6, and 6-15. While we desire to maximize explainability due to the limited number of patient data, we determine to use 4 subtypes. The characteristics of these subtypes are presented in Table 2.2. Comparing the subtypes it is interesting to note that Subtype B contains all female members of the cohort, suggesting that gender is a major factor in L-dopa response. Additionally, Subtype D contains only one member who is the oldest member of the cohort, this patient also has the most severe dyskinesia while receiving a relatively low dose of L-dopa. Interestingly, both Subtype A and C are more similar in demographic make up; however, while both groups have similar bradykinesia scores, members of Subtype C have less dyskinesia. These differences in how patients respond has significant implications in how their medication treatment plans should be optimized.

2.3.4 Pharmacokinetic Modeling of Levodopa

Pharmacokinetic modeling is the process of quantifying the dose-concentration time course relationship of medication (102; 243). Orally administrated medication (such as commonly prescribed L-dopa formulations), are typically modeled as first-order or zero-order processes (243). Zero-order processes assume that medication is absorbed at a constant rate, whereas first-order processes describe drug absorption through differential equations accounting for bioavailability (243). Beyond accounting for bioavailability, the method of compartmental modeling is also leveraged to more accurately model drug absorption (243; 104). One-compartment models assume that

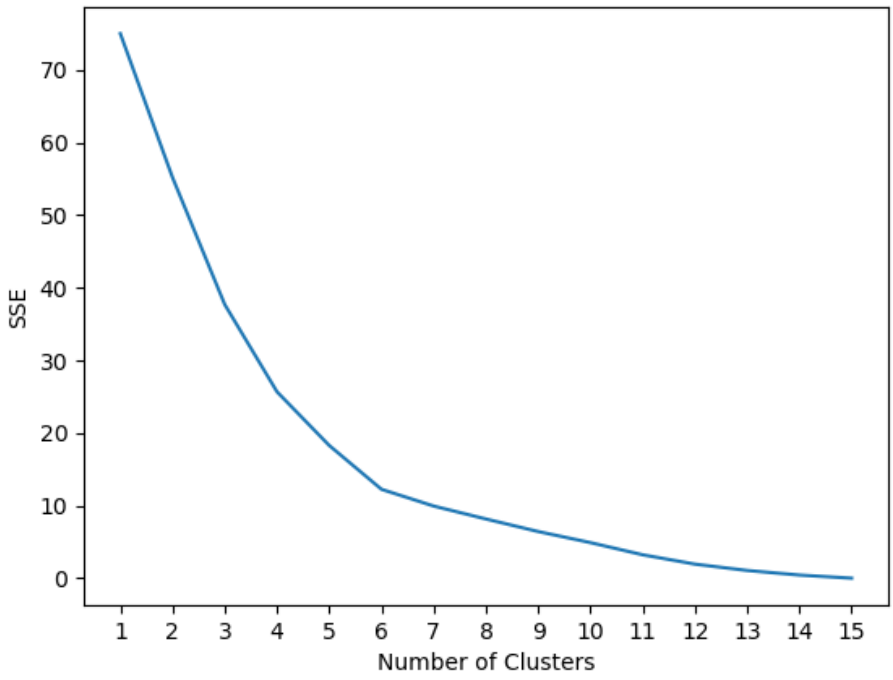

Figure 2.2: WCSS as a function of the number of clusters when clustering patients based on their best symptom control medication regimens per L-dopa dosage and demographic information (sex, and years with PD) based on their first daily L-dopa response from visit 1.The number of clusters is set to four, as a number towards the bottom of the "elbow" is considered best, but further increasing the number of clusters has been shown to reduce the model's robustness.

Table 2.2: Comparison of the L-dopa responder subtypes as identified by k-means clustering. Note the values represent the mean and standard deviation across all members of the subtype.

	Subtype A	Subtype B	Subtype C	Subtype D
Gender (M/F)	(4/0)	(0/4)	(4/0)	(1/0)
Study Age (years)	67 ± 14	69 ± 5	74 ± 11	87 ± 0
Years of PD (years)	3 ± 2	4 ± 4	3 ± 2	10 ± 0
V1 L-dopa (mg)	425 ± 125	440 ± 207	300 ± 100	300 ± 0
V2 L-dopa (mg)	500 ± 183	620 ± 286	550 ± 179	300 ± 0
Peak Bradykinesia	-0.40 ± 0.06	-0.43 ± 0.10	-0.51 ± 0.15	-0.141 ± 0.0
Peak Dyskinesia	0.14 ± 0.14	0.09 ± 0.12	0.03 ± 0.02	0.79 ± 0.0
Time to Peak (minutes)	110 ± 74	242 ± 107	273 ± 142	129 ± 0

the whole body acts a single compartment, this is a common approach for drug's with a monoexponential pharmacokinetic decline (243). However, two or three-compartment models typically assume that blood and organs belong to a compartment whereas, tissues which slower absorption rates/elimination rates belong to a second compartment, finally other peripheral compartments can be model according the specific drug (243).

L-dopa's pharmacokinetic process has been modeled with a variety of underlying assumptions. Typically, L-dopa pharmacokinetic studies measure intravenous L-dopa plasma levels in a cohort of patients, while performing a symptom response task. Specifically, L-dopa's pharmacokinetics have been previously modeled as a one-compartment, first-order process (1; 6), a two-compartment, first-order process (8; 166), and as a multi-compartment (three or more-compartment), first-order process (149; 218). Each of these studies sought to effectively model their own collected data; however, there is no single pharmacokinetic model which used for L-dopa modeling. Due to the availability of pharmacokinetic data in the literature, we will assume that L-dopa follows an one-compartment, first-order absorption with first-order elimination model. An overview of the pharmacokinetic process is provided in Figure 2.3.

Following the oral administration of L-dopa the pharmacokinetic process can be modeled by an one-compartment, first-order absorption and elimination model. Mathematically this is described as,

$$\frac{dC_1}{dt} = -k_l \cdot C_1 \tag{2.3}$$

$$\frac{dC_2}{dt} = -k_l \cdot C_1 - (\frac{CL}{M}) \cdot C_2 \tag{2.4}$$

$$C_p = (\frac{C_2}{M}) \tag{2.5}$$

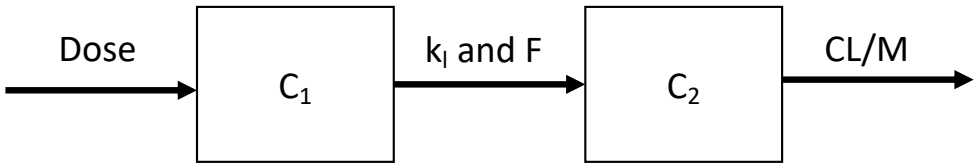

Figure 2.3: A diagram overviewing the pharmacokinetic process for L-dopa response. This process represents the first-order absorption followed by elimination, with the corresponding pharmacokinetic parameters denoted.

here, C_1 denotes the amount of the administered drug, k_l is the absorption rate, C_2 denotes the amount of the drug in the body, CL represents the clearance, M represents the volume of the distribution, and C_p denotes the plasma drug concentration. Following these definitions, the initial conditions for C_1 and C_2 are.

$$C_{1_0} = \text{Dose} \cdot F \quad \text{and} \quad C_{2_0} = 0 \tag{2.6}$$

here Dose denotes the dosage of the administered drug and F represents the bioavailability. Thereby, the the drug concentration, C_p, is the solution to the above differential equations,

$$C_p = \frac{F \cdot k_l \cdot \text{Dose}}{M(k_l - \frac{CL}{M})} \cdot (e^{-t\frac{CL}{M}} - e^{-k_l \cdot t}) \tag{2.7}$$

To calibrate this model, we utilize values from the existing literature. Specifically, we incorporate the pharmacokinetic parameters obtained by (166), while this model represents a two-compartment model, its parameters are consistent with other first-order, one-compartment models (1; 6) The benefit of the work by Othman et al. is their inclusion of a median and 95% confidence interval with their parameter estimations. The inclusion of this confidence interval, allows for the modeling of a distribution to our pharmacokinetic response. Patients have a varied response to L-dopa based on several confounders; therefore, we account for this varied response via the pharmacokinetic distributions. Consistent with the literature, we assume that the bioavailability of the system, F, is consistent throughout the system and pharmacological process (i.e., $F = 1$) (166). The pharmacokinetic parameters used in this study are presented in Table 2.3. Note that these parameters are estimates to model the complex nature of L-dopa response, for a particular pharmacokinetic modeling framework.

To allow for a larger distribution of potential responses, we utilize the 95% confidence interval pharmacokinetic values as a reference for two standard deviation from the median parameter to create a set of normally distributed responses. This

Table 2.3: Pharmacokinetic parameters for the modeling of L-dopa, the parameters where obtained from the literature (166).

Parameter	Median	95% Confidence Interval
k_l (hr^{-1})	2.2	0.85 to 5.2
CL (l/hr)	24.4	20.4 to 26.8
M (l)	56.0	36.6 to 71.0
F	1	–

enables a variety of responses. As the our data set is collected over a six-day period, we assume for all simulations that pharmacokinetic parameters are constant through each day, regardless of time or dosage. The L-dopa pharmacokinetic response is shown in Figure 2.4. Each plasma concentration represents a separate state in the POMDP model. To calibrate transition probabilities, 500 days are simulated for pair of actions and τ_i.

2.3.5 Wearable Sensor Calibration

As the drug concentration and bioavailability of L-dopa is not measurable in real-time, this dictates the formulation of the model to a POMDP. This partial observability can take the form of wearable sensor measurements. As the uptake and use of L-dopa in the body is characteristic of symptom control/response, we are thereby able to estimate the current state drug concentration.

Here, we use patients' Visit 1 regimens to estimate these observation probabilities. These observation probabilities are subtype-specific, as the subtypes characterize patients' response to L-dopa. To determine the observation probabilities, we first use patients' Visit 1 regimens using the known state values, as obtained from the L-dopa model. To calibration the observation probability we assume that patients' level of symptom control is based on concentration of L-dopa. Therefore, the median values of the pharmacokinetic parameters are used in the calibration, all L-dopa plasma concentrations that are not reached within the patient cohort are estimated by linear interpolated. The observation probability is calculated with a mean value of the bradykinesia and dyskinesia values along with the 25^{th} and 75^{th} percentile of symptom control. Each of the four bradykinesia and dyskinesia regions can be separately measured by the sensor. We consider combinations of bradykinesia and dyskinesia scores as the set of observations creating sixteen separate observations.

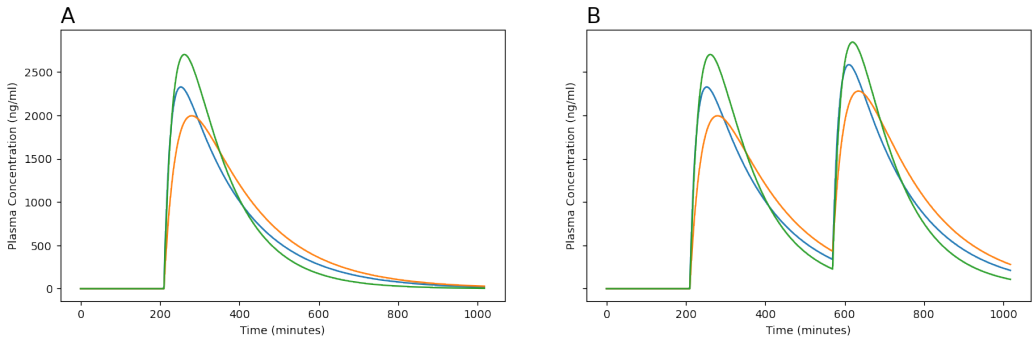

Figure 2.4: The pharmacokinetic absorption and elimination of L-dopa. Plot A, shows the a single administration of 200mg of L-dopa at time 210-minutes. Whereas plot B, shows the administration of 200mg at both time 210-minutes and 570-minutes. These times are representative a specific patient's regimen in the cohort. Each plot represents the varied response to L-dopa, shown over three days, both in magnitude of response and elimination of the medication.

2.4 Numerical analysis

In this section, we evaluate the model's performance. Recall our data set consists of two clinical evaluations and regimens. In our analyses, the Visit 1 analysis was used to calibrate the model, whereas the Visit 2 analysis was used to evaluate the model's overall performance. Note that patients' Visit 2 regimen was adjusted by a physician to optimize patients' symptom control based on clinical evaluation as well as the wearable sensors report. In this way, the Visit 2 regimen is a physician's attempt to perform the same task as the POMDP framework. We evaluate the model's performance at simulating patients' Visit 2 prescribed regimens and associated symptom control: this portion is to validate the model's performance.

2.4.1 Model Validation

To validate the models performance we examine its ability to produce similar L-dopa responses to those observed in the data set. Specifically, we use the L-dopa response model along with the wearable sensor observation process to produce patients' Visit 2 regimens. In this way, we use the physician prescribed medication regimens for each patient. We average patients' symptomatic response to these regimens over a six day period. We compare the time spent in each bradykinesia and dyskinesia region as well as the average daily reward to that of the patients' wearable sensor measured symptoms.

The reward structure for this analysis aims to represent the goal of L-dopa administration. Namely, similar to L-dopa reward structures that appear in the literature (10), we prioritize the control of bradykinesia over that of dyskinesia. This assumption follows that symptom control for bradykinesia is the primary use case of L-dopa and dyskinesia is a side effect. Specifically, observations in bradykinesia regions (I-IV) have a corresponding reward of 1, 0.75, 0.50, 0.25, respectively. Whereas the dyskinesia regions (I-IV) have the corresponding rewards of 0.75, 0.5, 0.25, 0.0, respectively.

To determine the effectiveness of the POMDP framework, each patient's Visit 2 regimen was evaluated both using the POMDP framework in comparison to the observations collected directly by the wearable sensors. The average performance of the L-dopa model is shown in Table 2.4. Overall, the model performs well for Subtypes A, B and D, however each has a larger standard deviation than the wearable sensor values. This increased standard deviation may have multiple sources, namely that the PODMP uses state transition probabilities from a range of pharmacokinetic parameters, which can lead to a large fluctuation in response. Additionally, as the observation probabilities are collected across patients in each subtype. It appears that the POMDP's performance suffered for patients' whose medication dosage was increased between visits, this is especially true for Subtype C. Additional patient data may help improve these results as response trajectories become more realistic with additional patient data.

2.5 Conclusions and Future Work

In this paper, we present a POMDP framework to model medication regimens in a cohort of PD patients. The model was calibrated using a combination of pharmacological data (obtain from the literature) and wearable sensor data. Instead of comparing static medication regimens as has been previously shown, this framework allows for dynamic treatment planning. Additionally, the proposed framework, while developed for PD treatments, could be applied to a variety of progressive, chronic diseases. We demonstrate that the POMDP framework was able to accurately model medication regimens for our study's cohort, per medication response subtypes. However, these models are restricted by the availability of different type of L-dopa responders. Therefore, the framework can be further refined through the inclusion of additional patient data.

There are several interesting future directions for followup research. First, the current framework could be extended to account for limited sensing in which

Table 2.4: Validation of the PODMP framework. Results are presented averaged across all patients in each subtype over a six-day period with the associated standard deviation.

	Subtype A	Subtype B	Subtype C	Subtype D
Wearable Sensor Assessed	1226 ± 54	1231 ± 140	1076 ± 68	1337 ± 0
POMDP Model	959 ± 512	848 ± 504	279 ± 86	1115 ± 112

observations have an associated cost. This process could enable increased battery life as wearable sensors require frequent charging, which can be a burden, especially, for an elderly population. Second, the framework could incorporate physician-in-the-loop decision-making, under which the physician-prescribed medication regimen would be used as a baseline by the framework. The framework's action space would be restricted to a limited deviation from this established regimen. Such a process could provide recommendations to the physician, which can be more easily introduced clinically. Finally, a full-scale longitudinal study in which the POMDP framework determines and prescribes patients' optimal regimens with the oversight of a physician could serve as a precursor to establishing the use of such a framework with L-dopa pumps (3) that would provide immediate, dynamic medication response at any dosage level.

Chapter 3

Machine Learning's Application in Deep Brain Stimulation for Parkinson's Disease: A Review

3.1 Introduction

Deep brain stimulation (DBS) has evolved into an effective therapeutic agent for Parkinson's disease (PD) patients suffering from medication refractory tremor, motor fluctuations, and/or troublesome dyskinesia. The stimulation is delivered through electrodes that are stereotactically implanted into either the subthalamic nucleus (STN) or the globus pallidus internus (GPi). Targeting is achieved with intraoperative image guidance (magnetic resonance imaging (MRI) or computed tomography (CT)) (139; 171; 27). Neurophysiological mapping via microelectrode recording (MER) is often utilized to improve targeting precision. DBS therapy is engineered by customizing the size, shape, and intensity of the stimulation field with newer electrodes (tripartite-segmented electrodes from Abbott©, Medical Infinity™, Austin TX, USA, and Boston Scientific©, Vercise™, Marlborough, MA, USA), producing

larger therapeutic windows and increasing the combinations of programming options (55; 210).

Currently, high frequency continuous stimulation drives therapeutic benefit and requires time-consuming programming sessions, often relying on a trained clinician culling from experience and established programming approaches (227; 154) to achieve the best clinical outcomes. However, stimulation-related side effects and suboptimal clinical responses can be limiting factors. Adaptive DBS (aDBS), which is a closed loop stimulation system (133) that uses a physiological biomarker of the parkinsonian state to automatically adjust the stimulation level and location, can provide an alternative to traditional DBS. This offers the possibility of more efficiently treating patients while reducing the challenges inherent in the current programming approaches. Furthermore, the recent advances in DBStechnology, including improvements in spatial control from segmented electrodes and image guided programming, along with brain sensing (a likely precursor for closed loop DBS) can benefit from the application of data driven modeling to enhance clinical decision making and outcome prediction. In this review, we examine the applications of machine learning approaches for DBS in PD.

Machine learning (ML) is a subfield of artificial intelligence (AI) that identifies patterns within data sets. The application of ML in healthcare is becoming more widespread to augment the work of clinicians (186). For instance, ML models have been created to identify anatomical structures through medical imaging (207), and to optimize pervasive sensing in healthcare informatics (190). In the context of PD, much work has been done to successfully use ML/AI. This includes detecting PD symptoms for diagnosis purposes (e.g., (2)), predicting disease progression and patient response to medication using motor symptoms (e.g., (13)), optimizing clinical assessments through leveraging wearable sensors (e.g., (187)), and providing medication regimen planning (e.g., (233)).

Traditionally, computer algorithms execute explicit, specifically written code to interpret data and make decisions. ML is a computer analysis subfield that

distinguishes itself from traditional methods by seeking to "learn" the relationships that underscore data without exact instructions (204; 156). Simply put, ML is an automated method that uncovers patterns in the data and utilizes those patterns to make predictions and/or facilitate decisions. ML has a wide domain of uses but excels in tasks that have large and complex data sets (204). Generally, ML algorithms are categorized into two subtypes, namely, supervised and unsupervised learning.

In supervised learning, the algorithm's general goal is to learn a mapping of inputs to outputs. To accomplish this task, the algorithm is provided a training set of input–output pairs. These inputs are generally multi-dimensional vectors representing features of the data. The response variable, also known as a dependent variable, is the output of interest that is associated with these features. As the true values of the response variable, namely the ground truth, have already been identified by a human, supervised learning algorithms frequently are attempting to approximate human performance (56). The response variable may be discrete (often called a label) or continuous, resulting in a classification or a regression task, respectively (156). The performance of a supervised learning algorithm is objectively evaluated on a testing set, a subset of the data that is never examined by the algorithm when training. This allows for quantifying the generalizability of the learned model to unseen data. Dependent on the type of task (classification vs. regression), the model performance is described by diagnostical statistic measurements, such as accuracy, F1 score, or the area under the receiver operating characteristic (AUC) for classification, or metrics such as mean squared error or mean absolute error for regression. Examples of supervised learning algorithms include random forests (RF) (130), neural networks (NN) (91), support vector machines (SVM) (212), naïve Bayes (NB) (194), and k-nearest neighbor (KNN) (63).

In unsupervised learning, the algorithm's general goal is to find trends in data given only inputs. Unsupervised learning tends to be a more difficult process, as there are no known ground truths to predict. Examples of unsupervised learning algorithms include clustering (86), density estimation (18), and some neural networks (91).

Each technique requires different methods to determine the model's performance. For instance, in clustering, data is separated into distinct groups that ideally share characteristics. A clustering algorithm's performance may then be analyzed by how close its group members are to one another (intra-cluster distance) or how close the members are to another cluster (inter-cluster distance). While benchmarking unsupervised learning algorithms can be challenging, unsupervised learning allows for the analysis of data sets that are completely incompatible with supervised learning algorithms. Unsupervised learning is well suited for situations where ground truths cannot be identified or in which human experts cannot determine trends. Under these situations, unsupervised learning can propose trends in the data that can lead to meaningful insights when interpreted by experts.

The efficacy of DBS in the treatment of PD is well established (57; 226; 14; 122). Many of the potential benefits of ML in strategic PD treatment planning have been considered (2; 13; 187); however, these reviews do not generally explore the role of ML in DBS. Specific surgical or treatment topics relating to ML in DBS have been explored (34; 95; 231; 178). However, they do not explore all facets of ML in DBS, and frequently only provide brief considerations of its application. As such, identification of the best performing models, as well as gaps in the existing body of literature, could help guide future exploration in using the powerful tool of ML. Here, we review the applications of ML approaches for DBS in PD. We compartmentalize these approaches into four groups: (1) DBS candidate selection; (2) programming approaches; (3) surgical targeting; and (4) insights into DBS mechanisms.

3.2 Materials and Methods

We performed a comprehensive search for studies that applied ML techniques in conjunction with DBS treatment of PD. This search encompassed all relevant journal, review, and conference publications cataloged on four common computerized database search engines: (1) PubMed, (2) the Institute for Scientific Information's Web of

Science, (3) the Cochrane Database of Systematic Reviews, and (4) IEEE's Xplore Digital Library. We utilized the computerized databases' keyword search functions using a combination of three key phrases: (1) "Parkinson's Disease," (2) "Deep Brain Stimulation," and (3) "Machine Learning." We limited our results to publications that were published between January 2010 and August 2020. We further limited our results to English language manuscripts. This initial search identified 76 potential publications. These were evaluated with the predetermined exclusion criteria: (1) PD was not the primary application focus (seven papers were excluded), and (2) papers that contained only published abstracts (one paper was excluded). This process returned 68 publications that were included in this review. The number of publications per year and the number of papers per database are provided in Figures 3.1 and 3.2, respectively.

3.3 DBS Candidate Selection

A key challenge for clinicians is to properly identify DBS candidates who will receive measurable clinical benefits with minimal surgical and stimulation related risk. A large-scale ML assisted analysis was performed to identify patients who may experience complications from DBS surgery. Specifically, Farrokhi et al. (68) examined a retrospective cohort of 501 DBS patients and determined presurgical features that indicated a greater chance of DBS complications. Logistic regression was applied to evaluate risk factors. The model was able to correctly identify subjects that would have any complications (AUC 0.86) and complications within 12 months (AUC 0.91) following DBS surgery. They found that age body mass index, procedure side, gender, and diagnosis of PD were all prominent features.

The screening of potential DBS patients can be limited through visits conducted virtually. There is a growing body of research that utilizes video recordings and ML to determine symptom scores for use in telemedicine. Das et al. (50) compared DBS patients on and off stimulation and classified severe versus mild symptoms

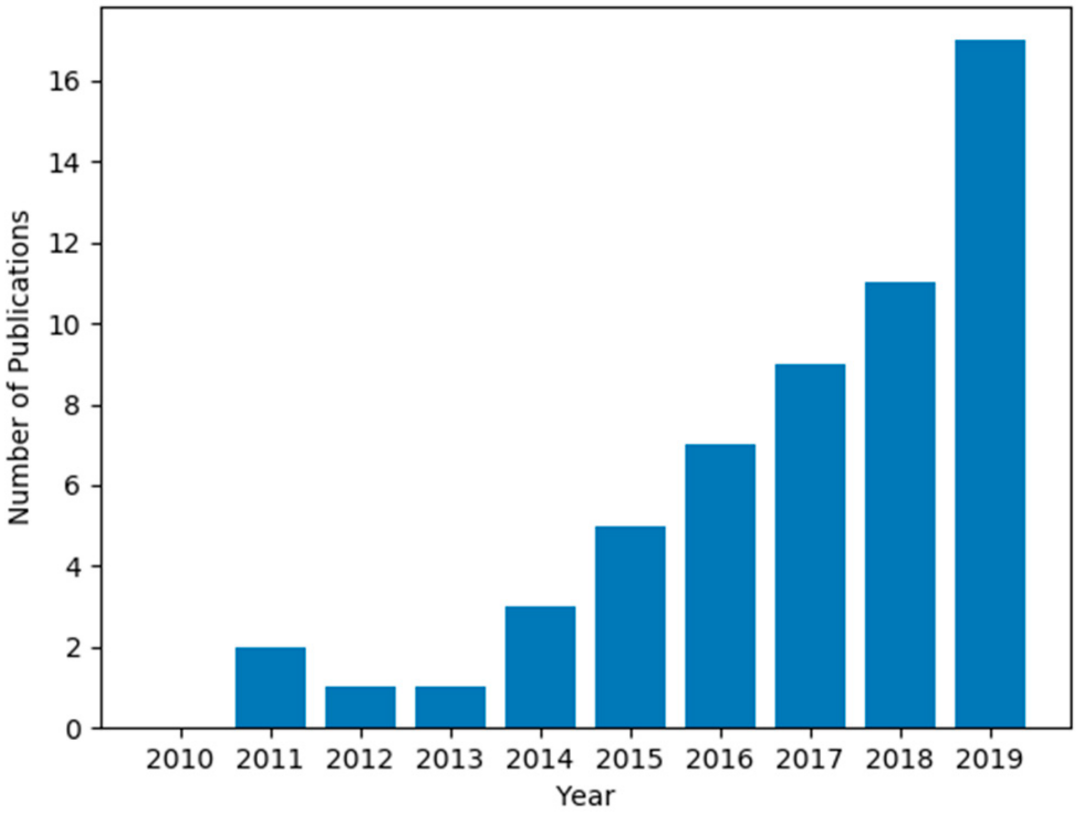

Figure 3.1: The number of publications examined within the review, segmented by the original publication date up to the last full year. There were 12 publications reviewed between January 2020 and August 2020.

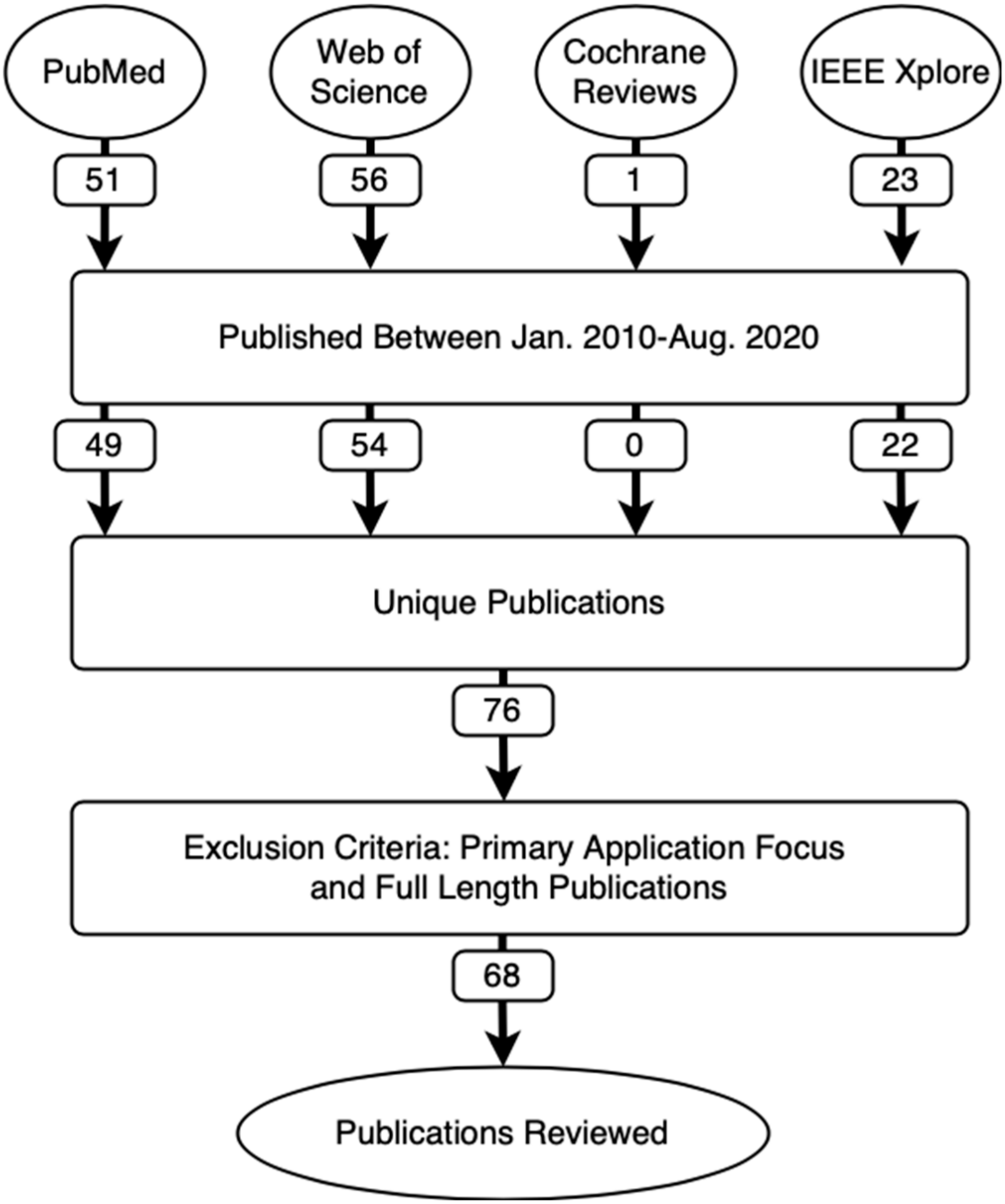

Figure 3.2: The description of the database keyword search results per database. Each number located within a square indicates the number of returned publications. The exlusion criteria are descrbied in Section 3.2.

using full-body capture video. This analysis was performed for six DBS subjects and classified through SVM. Yohanandan and colleagues (239) used wearable sensors with RF models and predicted tremor severity following the Bain–Findley tremor rating scale (BTRS) with 81% accuracy for use in telemedicine. These studies not only demonstrate the feasibility of using ML in remote monitoring platforms, but also provide early insights into developing a way to classify symptom severity or patients' phenotypes virtually.

3.3.1 Predictive Motor Biomarkers

Several studies have deciphered motor classifiers by comparing certain motor response differences between healthy and Parkinson's subjects (120; 121; 93; 164). Kuhner et al. (120; 121) used RF algorithms to stratify 14 PD patients with (DBS-ON) or without (DBS-OFF) stimulation from 26 healthy subjects using a variety of motor control tasks. They found a 94.6% accuracy in distinguishing DBS-OFF patients from healthy subjects when using a standing up test. Another study (121) performed 3D motion capture and gait analysis on 26 PD patients (seven with STN-DBS) and 25 healthy subjects. They identified classifiers (a combination of low-order time derivatives) that performed best at distinguishing DBS-OFF patients from healthy subjects. Such classifying features could then be applied to a large-scale study of gait analysis. Similarly, Huo et al. (100) developed a unique multi-sensor motor response test and utilized KNN to classify 33 subjects as either having PD or being healthy with a 96.6% accuracy.

3.3.2 Non-Motor Considerations in Candidate Selection

Non-motor symptoms (NMS) of PD include neurobehavioral dysfunction, dysautonomia, and sleep dysfunction. NMS treatment response has varied with DBS (201; 36; 69), with several of these symptoms (i.e., sleep, pain, impulse control disorders, anxiety, and nonmotor fluctuations) showing some degree of improvement.

However, with the limited number of studies investigating NMS changes with DBS, the mechanisms underlying the responses are postulated to be related to direct stimulation effects on the anatomic substrate (i.e., STN), postoperative dopaminergic medication reductions, or indirect relation to motoric benefit. Since NMS impacts the quality of life in PD patients, factoring it into the surgical decision-making process is an important clinical consideration.

Koch et al. (115) studied 40 DBS patients' preoperative electroencephalograms (EEGs) to identify those who would have postoperative cognitive deterioration using an RF algorithm. The data set was balanced in cognitive performance outcomes. The best performing model used both EEG and clinical features and achieved an accuracy of 91%. Chen et al. (38) examined the sleep cycle characteristics of 12 DBS patients. These patients were classified by SVM and decision trees based on local field potential (LFP) signals from the implanted target (STN). They found that alpha, beta, and gamma brain waves were the important predictors in sleep monitoring. Paliwal et al. (169) explored the effect that DBS had on impulsivity in decision making. In this study, 38 PD patients played a virtual slot machine before and after beginning DBS treatment. They applied a hierarchical Gaussian filter model that determined gambling behaviors. They found that DBS patients had increased volatility in relationship to impulsivity. Further, Chrabaszcz et al. (41) examined the LFP signals of 11 patients with STN-DBS while they read three-phoneme words. They were able to identify tongue or lip articulation, specific to DBS patients, that showed that DBS affected some patients' speech patterns.

3.4 Programming Optimization

Once potential DBS patients have been identified, the optimization of stimulation oftentimes requires frequent and time-consuming programming sessions conducted over 3–6 months from implantation. The programming strategies, coupled with simultaneous medication adjustments, rely a great deal on clinical trial and error.

The efficient optimization of strategic DBS treatment planning using ML has been investigated. For example, Khojandi et al. (109) performed a retrospective analysis on 20 STN-DBS patients' Unified Parkinson's Disease Rating Scale (UPDRS) III scores. UPDRS (203) is a standard, subjective clinical measurement of PD symptom severity. Khojandi et al. developed an RF algorithm based on preoperative patient and disease characteristics to distinguish patients' optimized stimulation frequency (high and 60 Hz) with 95% accuracy. Such a model could enable more efficient optimization of DBS parameters. Additionally, Shamir et al. (205) predicted the expected stimulation and medication dosage based on three ML methods, specifically SVM, NB, and RF, for ten patients using 89 post-DBS surgical visits. These visits were used to create a clinical decision support system. This system enabled the correct prediction of motor improvement scores one year after surgery at 86% accuracy. The best performing model that classified individual symptoms varied. For example, SVM predicted tremor and speech outcomes with up to 100% and 93% accuracy, respectively, whereas RF outperformed SVM in terms of axial akinetic symptom prediction, with 86% accuracy.

Khodakarami et al. (108) examined the role of wearable sensors, specifically the personal kinetigraph, in an on/off stimulation classification problem. Their linear regression models determined classification with an AUC of 0.92. Baumgarten et al. (11) examined ten patients with GPi implanted DBS for pyramidal tract side effects (PTSE). An NN classified PTSE in patients whose parameters had been designed to trigger PTSE by clinicians. Each PTSE event was identified by professionals. A positive predictive value (PPV) of over 0.8 was achieved in this study.

Due to the clinical significance of the UPDRS, several publications explored the prediction of the UPDRS score as a response variable. Przybyszewski et al. (180; 181) used data mining and decision tree approaches to predict UPDRS III scores based on reflexive saccades with 76% accuracy for 41 PD patients (18 with STN-DBS), and consequently using UPRDRS predictions, they estimated the disease progression in a cohort of PD-DBS patients. Candamil et al. (31) successfully categorized four

subjects' UPDRS scores through linear machine learning based on a copy/draw test that could be applied to calibrate stimulation parameters. A study (5) that applied KNN on seven DBS subjects was able to predict UPDRS symptom severity with over 90% accuracy using wearable sensors on the most affected arm. The authors determined the potential benefit for home-based objective sensor measurements for improving DBS parameter optimization.

3.4.1 Adaptive DBS

Adaptive DBS (aDBS) refers to automated stimulation based on the real-time feedback of electrophysiology or wearable sensor signals (133; 90). Validated input signals, including basal ganglia local field potentials and especially beta band activity (133; 145; 132; 4), are deemed to be associated with rigidity and bradykinesia (further review of electrophysiology-based aDBS is provided in (159)). Alternative input signals rely on wearable sensors (31; 5; 107; 29; 213)that can measure tremor, bradykinesia, dyskinesia, and gait (107). The fidelity of tremor detection in calibrating a closed loop adaptive system has been demonstrated in several studies (97; 84). The use of ML in tremor classification for DBS patients was also explored and deemed as a useful tool for adaptive, patient-specific optimization (203; 107; 29; 127; 235; 236; 134; 150). Khobragade et al. (107) examined the surface electromyography of two patients (one PD-DBS) over two sessions and analyzed the signals using an NN. They achieved over 80% accuracy when detecting tremor using an NN trained on the same session. However, there was a nearly 50% decrease in performance for repeat patient measurement taken a week apart, which indicated the high variability of tremor symptoms and the benefits from aDBS, which is capable of real-time performance. To that point, LeMoyne et al. (127) surveyed the use of the BioStamp nPoint, a data acquisition device for tremor detection, specifically with the use of an NN, to identify stimulation parameters based on tremor response for future use in aDBS. Camara et al. (29) examined four STN DBS patients and classified

tremor based on LFP signals. The authors suggested that a simple threshold would be insufficient to classify tremors. Therefore, they instead used -recurrence networks to classify tremors, and consequently applied an SVM to adaptively start or stop stimulation. This process averaged over 90% accuracy at stopping stimulation and over 78% accuracy at starting the stimulation. Shah et al. (203) used a logistic regression algorithm to classify tremor and non-tremor periods in seven DBS subjects using wearable sensor (accelerometer) data. The data set was labeled based on the frequency and duration of accelerometer events. The model achieved an AUC of 0.78 ± 0.11 for the identification of tremor periods. Additionally, Yao et al. (235; 236) showcased the benefits of Kalman filtering when applied to tremor classification using multiple ML methods (SVM, RF, KNN, etc.), with as high as 11% classification improvement.

Additionally, ML has been applied to freezing of gait (FoG) (213), boundary decisions (151), and PD stages (152) for use in aDBS. However, the clinical feasibility of detecting these non-tremor motor symptoms have yet to be clinically validated in such a system.

Classifying human behavior from brain signals has also been explored in developing closed loop aDBS (80; 81; 83; 82). The real-time classification of human behaviors would allow for adaptively optimizing stimulation parameters per task. Golshan et al. (80) used variants of SVM classifiers to predict a variety of human behaviors, including speech, mouth movements, arm movements, finger movements, and a button press, with 70% accuracy, and demonstrated the effects of medication on STN-LFP (83). Mamun et al. (141) decoded movement behaviors from LFPs of the STN and GPi from 12 PD and dystonia subjects with DBS by conducting a finger-clicking task with an accuracy of 99.8%. Furthermore, Niketeghad et al. (160) used SVM from 14 patients' LFPs to recognize speech, motor, and random type tasks with 73.2% accuracy. Additionally, several other studies examined the classification of behavioral actions using hidden Markov models (HMMs) in combination with the ML techniques (240; 241; 101). For example, Jiang et al (101) created an HMM in combination with

ML for the classification of motor, language, and motor/language combination tasks. They examined the LFP signals of nine DBS patients using their HMM model, and achieved over 80% accuracy at distinguishing between a combination of any two tasks.

3.5 Surgical Targeting

The surgical implantation of DBS electrodes necessitates accurate real-time identification of a specific brain region. Image guidance with direct magnetic resonance imaging (MRI) based targeting or co-registration of computed tomography (CT) with a 3-Tesla MRI are currently used to localize anatomic targets (i.e, STN and GPi) (14; 72; 158). The selection of targets and surgical approaches varies among DBS centers, with microelectrode recording (MER) and intraoperative stimulation effects used to refine electrode placement (92). ML techniques have been proposed to utilize microelectrode recordings to aid in subthalamic nucleus classification and thereby facilitate electrode placement. Cardona et al. (30; 223) performed a retrospective analysis of MER region localization of 10 patients. They compared traditional ML classifiers (NB, SVM, KNN, etc.) with a multi-output Gaussian approach. The relative performance of the classifiers varied by task, but each achieved an accuracy of over 70%. The classification of other specific brain regions through MER signals has also been explored. For example, Khosravi et al. (110; 111; 112) analyzed the MERs of electrode trajectories with ML techniques over several studies. They retrospectively analyzed 100 DBS subjects' physicians determinations of optimal implantation trajectory's MERs with a deep NN and obtained an accuracy of 92% in localizing specific brain regions. Additionally, they utilized unsupervised machine learning techniques on 50 subjects' MERs and determined the dorsal border of the STN region with 80% accuracy. Finally, a fast Fourier transform was applied to the MERs of 20 DBS surgeries, and a logistic regression algorithm and SVM were applied to localize STN depth. This achieved an accuracy of 85% in detecting the borders of STN. High accuracy region SVM classification models were used by Guillén et

al. (88; 89) with a 99.5% accuracy at stratifying MERs between the STN, thalamus nucleus, and the substantia nigra. Similarly, Lu et al. (136) applied SVM to 16 subjects' MERs guided by 3-Tesla MRIs. This model achieved an AUC of 0.87. It is important to consider that MER measurements can be confounded based on the patient's state of sedation: minimal (awake), light, or deep (general anesthesia).

Alternative approaches, such as the utilization of traditional MRIs (205) and state-of-the-art 7-Tesla MRIs (206; 113) or magnetoencephalography (MEG) (21) in conjunction with ML techniques, have been rarely explored. Ozturk et al. (167) examined ten patients' LFP signals and determined STN localization by an online linear discriminant classifier. Similarly, Darbin et al. (49) used LFP signals in a Tree Bagger algorithm to identify regions in non-human primates. Stuart et al. (211) examined 16 subjects' EEGs with multiple stimulation regions specifically: GPI, VIM, or STN. The fast Fourier transform was used to extract features. They compared the performance of several ML classifiers in determining the stimulation region. A decision tree based gradient booster paired with principal component analysis (PCA) performed the best at region identification (F1 0.62, precision 0.68).

ML has also been used in conjunction with HMMs to model surgical planning. For example, Valsky et al. (220) developed a real-time ML classification of the subthalamic border for accurate detection of the STN in DBS surgery. They examined 58 STN trajectories using an SVM and achieved an accuracy of 97.6%. They used the SVM results to inform an HMM for real-time identification, and this model was tested on an additional 73 trajectories, achieving an accuracy of 94%. Valsky et al. then continued this work in (219) for the identification of the striatopallidal border site to the GPi. They applied a similar SVM/HMM algorithm to 116 trajectories over three patient types: awake PD patients, lightly anesthetized genetic, and non-genetic dystonia subjects. Their algorithm had similar performance to clinical experts at GPi border identification across all patient types considered.

3.6 Insights into DBS Mechanisms

A small body of work exists that aims to better understand the underlying biological and chemical signals that impact DBS performance. Trevathan and colleagues (215) used NN and Volterra kernels to characterize stimulation-evoked neurochemical releases. They compared their proposed frameworks of stimulation-evoked dopamine releases in several animal models. This allowed for a neurochemical signaling comparison of healthy and diseased brains, which theoretically could be used to develop strategies for controlling DBS systems. In another study, Khawaldeh et al. (106) examined the off medication LFPs of 18 STN-DBS patients. They used NB classifiers for movement predictions and determined the band frequency of oscillatory synchronization associated with motor impairment in PD.

Finally, ML techniques have been utilized in the design of DBS systems. Jovanov et al. (103) developed a unique aDBS system hardware platform that allowed for the real-time optimization of DBS treatments by utilizing a genetic algorithm to identify patterns in treatment planning. Other studies that concentrate on hardware limitations include: Zhu et al. (242), which focused on performance, memory, and hardware requirements for seizure detection tasks in PD using EEG/ECoG (electrocorticography) as inputs to a custom decision tree algorithm; and De La Pava et al. (53), which applied hierarchical KNN to reduce the computational load of tissue-activated visualization in DBS modeling.

3.7 Discussion

The application of ML in DBS treatment for PD is a rapidly growing, interdisciplinary field that has unprecedented potential to transform all aspects of DBS. If considered and executed appropriately, ML will prove a tremendously beneficial tool for clinicians. This review examined the role of ML in DBS in the broadest terms, performing an exhaustive review and considering nearly a decade of publications.

We found compelling applications for ML in (1) DBS candidate selection; (2) programming optimization; (3) surgical targeting; and (4) insights into DBS mechanisms. Even within a specific category, there was large heterogeneity among the studies regarding feature type, ML models, and practicality of implementation. Broadly speaking, the clinical features utilized in ML studies benefited from objectiveness. As such, ML research favors LFP, wearable sensors, MERs, and similar unbiased measurements.

The identification of PD patients whose motor control symptoms would be more effectively managed under DBS than medication-based therapies is a clear target for ML. Predicting which patients may have adverse effects from DBS therapies is crucial to ensuring optimal patient outcomes. Non-motor symptoms present further complexity to the PD phenotype, which has implications on disease progression, response to therapy (medication and stimulation), and surgical risk. Recent publications have offered methods to predict neurobehavioral changes preoperatively, hence possibly influencing the DBS decision-making process. Measuring motor and non-motor symptomatology through objective means will permit the expansion and robustness of these ML-based decision models. Additionally, patient phenotyping through unsupervised methods may reveal trends in patient data that have not been incorporated into current predictions. The automatic classification of potential DBS patients could efficiently determine an optimized treatment plan, thus dramatically improving patients' quality of life.

Adaptive DBS has a significant scope of clinical impact on a patient's quality of life. The use of LFPs and/or wearable sensors serve as robust measurement instruments for the classification of specific behavioral tasks, severity, and/or tremor. However, a large portion of these studies utilized assessments obtained in a clinical setting. Further work is needed to identify motor control symptoms in real-world environments. Tremor detection using wearable sensors appears to be one of the nearest methods to application in the clinical field. Additionally, the development of HMMs that accurately describe a DBS would have a great impact on strategic

treatment planning. To that end, dynamic treatment planning would require a further assessment as to the practicality of adaptively tuning DBS parameters.

Surgical targeting utilizing ML to interpret intraoperative signals can be used to enhance the targeting of anatomic substrates, providing immediate clinical relevance as a support tool to assist in surgical planning. However, while many of the studies reviewed and utilized MER signals in classifying or delineating surgical targets to enhance optimal electrode placement, it is important to note that few studies have robustly examined clinically relevant outcomes related to these approaches. Image-based training of surrounding anatomy in ML modeling to enhance surgical target localization is expanding and showing greater precision and adaptability than current direct (image guided) or indirect (Talaraich coordinates) methods (16; 170).

Although ML's application in DBS is relatively early in its development, the preliminary evidence is remarkably promising. Its ability to determine underlying patterns within patient data offers the potential to enhance candidate selection, facilitate surgical targeting, and more efficiently optimize programming approaches to yield rapid and robust clinical outcomes. In general, ML models that have interpretable data structures (explainable AI) should be favored in future studies. This allows experts to translate ML decision-making criteria, and could potentially provide insights into the fundamental nature of DBS. Additional insights may be found in the growing body of literature that proposes using PD-related genetic mutations and genetic profiles to identify DBS therapy effectiveness (131; 54). Furthermore, ML methods generally benefit from large data sets. The majority of studies considered in this review are limited by their small number of patients. Before any of the presented methods can be applied in a clinical setting, large-scale, randomized, multi-center studies are necessary to ensure an expansive data set while minimizing differences in target selection and surgical planning.

Chapter 4

Influence of Deep Brain Stimulation Frequency and Levodopa on Gait Dynamics in Parkinson's Disease

4.1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder resulting from the loss of dopaminergic neurons (52; 12). As the disease advances, gait disorder—manifesting as shuffling, reduction in speed, and freezing of gait, can arise and cause significant disability (162). Early in the disease course, dopaminergic medication can effectively manage PD's cardinal symptoms (i.e., bradykinesia, tremor, rigidity) (77). Levodopa (L-dopa), the most common dopaminergic medication used in the treatment of PD, can further produce spatial gait benefits such as improved stride length, gait velocity, swing velocity, as well as lower leg and arm range of motion (20; 45). However, temporal aspects of gait such as cadence, double limb support, and swing duration tend to be less responsive (20; 45).

In advanced PD patients, deep brain stimulation (DBS) effectively treats motor complications including, wearing offs and L-dopa induced dyskinesia (76; 208; 32; 69). The subthalamic nucleus (STN) is one of the preferred anatomic substrates for the implantation of electrodes for DBS (118; 199). High frequency stimulation (130-185Hz) of the STN engenders an average reduction in dopaminergic medication of $40 - 60\%$ while improving L-dopa on time and lessening dyskinesia (119). Axial symptoms tend not to respond to high frequency STN-DBS, with long-term studies demonstrating a worsening over 5-10 years after STN-DBS related to disease progression (76; 208; 32; 69; 224; 221). However, STN-DBS has been shown to improve similar gait measures as L-dopa,including velocity and stride length, while increasing the amplitude of trunk torsion and flexion along with arm and leg movements (66; 71; 195). Several studies also suggest that the combination of L-dopa and STN-DBS engenders synergistic effects on these gait parameters along with reducing gait variability—a clinical marker associated with increased fall risk among PD individuals (66; 93). Reports of gait benefit with low frequency (60-80Hz) STN-DBS have shown potential benefit, but sustained responses are variable, with little known as to which patients would show improvement (208; 153; 188; 24; 191). Furthermore, there is a lack of understanding as to the interaction of dopaminergic medication and low frequency stimulation, which has limited its widespread use.

In this study, we aim to identify the gait features impacted through the combination of DBS stimulation frequency and medication in a cohort of chronic STN-DBS patients with advanced PD and underlying gait disorder. We apply machine learning (ML) techniques to a data set of gait kinematics acquired from instrumented walking assessments to determine which gait features are impacted by either high frequency (180Hz; HFS) or low frequency stimulation (60Hz; LFS) in both medicated and unmedicated states.

4.2 Methods

4.2.1 Patient cohort and clinical data

The patient cohort consisted of twenty-one chronic PD patients (15 males/6 females), with a mean age of 64 years (range 39-80 years) and a mean disease duration of 15 years (range 3-40 years). The cohort was collected from an ongoing 60Hz STN-DBS gait disorder study (NCT identifier: 04184791) who had bilateral STN-DBS for more than 3 months and had a documented underlying gait disorder as defined as a score of two or more on the Movement Disorder Society's unified Parkinson's disease rating scale (MDS-UPDRS) III gait subscore. Specifically, participants' axial symptoms were measured by the MDS-UPDRS III subscores (3.1, 3.3(a-e), 3.9, 3.11, 3.12, 3.13) for facial masking, speech, limb rigidity, freezing of gait, posture, and postural instability. In addition, this cohort consisted of participants who received 60Hz low frequency (three participants) and 130-186Hz high frequency DBS stimulation (nineteen participants) as their standard-of-care therapy. Exclusion criteria for this study included non-English speaking, cognitive impairment limiting compliance with the study protocol, and vestibular or musculoskeletal problems affecting gait. A summary of the patient cohort is presented in Table 4.1. All data collection and analysis were performed in accordance with the Declaration of Helsinki and approved by the Institutional Review Boards of The Feinstein Medical Research Institutes/Northwell Health and the University of Tennessee (UTK-IRB-19-05559-XP).

4.2.2 Instrumented stand and walk trials

The analysis was performed utilizing sensor data collected from instrumented stand-and- walk (SAW) trials. Participants were evaluated in the following six conditions: OFF DBS/OFF MED, OFF DBS/ON MED, HFS DBS/ON MED,

Table 4.1: Summary of the study's chronic DBS cohort demographics.

	Mean (STD)	**(Min-Max)**
Age (years)	64 (9)	39-80
Sex (Male/Female)	15/6	-
Disease Duration (years)	15 (8)	3-40
LEDD (mg)	624 (310)	150-1317
Standard-of-care Frequency (Hz)	130 (42)	60-186
MDS-UPDRS III Axial Subscore OFF Stimulation/OFF Medication	9.14 (3.81)	2-17
MDS-UPDRS III Axial Subscore OFF Stimulation/ON Medication	6.18 (3.22)	1-12

HFS DBS/OFF MED, LFS DBS/ON MED, and LFS DBS/OFF MED. The medication OFF state was achieved following an overnight withdrawal of dopaminergic medications, while the DBS OFF condition consisted of a 50-minute 'wash-out' period prior to evaluation. To ensure the transition to the ON MED state, participants were administered 1.5 times their usual L-dopa dose (up to 300mg) using carbidopa/levodopa 25-100 tablet(s). SAW trials were performed for each of the STN-DBS electrode's four contact pairs, which may lead to an overall increase in variability of the measurements. Each pair was reprogrammed in the OFF MED state under HFS. The amplitude was increased in 0.1mA increments until sustained side effects were produced, with the final amplitude being 10% lower than the side effect threshold. The LFS amplitude was determined for each contact by using the HFS amplitude and the total electrical energy delivered (TEED), where,

$$\text{TEDD} = \frac{\text{Amplitude}^2 * \text{Frequency} * \text{Pulse Width}}{\text{Impedance}} \tag{4.1}$$

to keep the LFS and HFS conditions equivalent. There were eight stimulation trials per medication state, and participants were blinded to their stimulation changes. During the stimulation trials, contact pairs were randomly stimulated with either HFS or LFS in both the ON MED and OFF MED states. The presence of L-dopa induced dyskinesia was determined by the principal investigator for each of the trials.

Participants' gait dynamics were recorded using full-body Opal sensors (APDM, Portland, OR) on their wrists, feet, sternum, and lumbar spine (L5 level). Each SAW consisted of a 30- second standing period, followed by a 7m walk at a comfortable speed, a 180-degree turn, and a return 7m walk. Both the initial and return 7m walks are considered separate trails for the purposes of this study. Subjects completed two trials per SAW for each stimulation condition. In the OFF stimulation conditions, one SAW (two trials) was conducted compared to four SAWs (eight trials) for each ON stimulation condition, as each contact pair was individually evaluated. Walking aides were permitted for the study and their influence was reduced in the

models by removing upper limb and postural sway parameters from the analysis. The measurements assessed during the instrumented walks included: 1) Gait– cycle duration (seconds), speed (m/s), stance (%GCT), swing (%GCT), heel strike angle (degrees), toe off-angle (degrees), toe out angle (degrees), stride length (cm), cadence (steps/min), step duration (seconds), elevation at midswing (cm) (foot clearance), double support (%GCT), and circumduction (cm); 2) Lumbar – sagittal range of motion (degrees), transverse range of motion (degrees), coronal range of motion (degrees); 3) Trunk – sagittal range of motion (degrees), transverse range of motion (degrees), coronal range of motion (degrees).

4.2.3 Data preprocessing

To preserve data, a standardized approach to recertify missing values (due to partial instrument errors during a completed walk) was applied. Missing values within a single trial were replaced with that participant's mean value over all trails corresponding to the same frequency/medication condition. However, if a participant does not have a measured value under any trial, the missing value was replaced with the mean of all other participants' walks under the corresponding frequency/medication condition.

Additionally, consistent with the literature, stride length, gait speed, and cadence were adjusted to account for patients' height. These gait measurements were adjusted as such (67; 200),

$$\text{Height Adjusted Stride Length} = \frac{\text{Stride Length}}{\frac{\text{Body Height}}{\text{Mean Body Height}}} \tag{4.2}$$

$$\text{Height Adjusted Speed} = \frac{\text{Speed}}{\sqrt{\frac{\text{Body Height}}{\text{Mean Body Height}}}} \tag{4.3}$$

$$\text{Height Adjusted Cadence} = \text{Cadence}\sqrt{\frac{\text{Body Height}}{\text{Mean Body Height}}} \tag{4.4}$$

4.2.4 Random forest

Random forest classifiers were implemented to discriminate between the OFF DBS/OFF MED condition, and each alternative stimulated/medicated condition. Random forests have shown to be robust with high dimensionality data sets and are highly interpretable (23). Features utilized by the classifier can be ordered by their predictive power (175). Thereby, the most impactful features to the model's predictions can be identified. In this analysis, the top ten features were selected to classify between different stimulation frequency and medication conditions, as determined by the Gini index of node impurity (175).

Using leave-one-out cross-validation, the classifiers were trained to predict participants' frequency/medication conditions. Leave-one-out cross-validation was selected to maximize the amount of training data available. In this process, one participant from the training data set was removed during each fold to test the algorithm's performance on that "left out" participant. This process was then repeated for each participant. As each training/testing fold will produce a unique model; these feature sets are compared across models.

As participants conducted a different number of walks under each stimulation/medication condition, data balancing was performed. Specifically, SAW trials from the less common frequency/medication condition were randomly resampled, with replacement, until both conditions were equally represented in the training data. Note this process is not done for the testing data to ensure that each predicted SAW trial is unique.

4.2.5 Statistical analysis

Statistics summarizing the classifier's predictive performance were calculated. These predictive metrics included accuracy, sensitivity, specificity, and F1 score, using the sklearn python library (175). Each performance metric is based on ratios of true positives (TP), true negatives (TF), false positives (FP), and false negatives (FP). The formula for each predictive metric is as follows,

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4.5}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TN} + \text{FN}} \tag{4.6}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4.7}$$

$$\text{F1} = \frac{2\text{TP} + \text{TN}}{2\text{TP} + \text{FP} + \text{FN}} \tag{4.8}$$

Finally, each metric's 95% confidence interval was determined over all folds in cross-validation to provide bounds to the predictive metrics.

4.2.6 SHAP analysis

Following the classifiers' identification of impacted gait features, Shapley Additive Explanation (SHAP) analysis is performed to determine the magnitude and directionality of the impact. Shapley values are an explanatory prediction technique that provides a metric to determine how each feature contributes to a model's predictive performance (137). Specifically, SHAP calculates points on a -1.0 to 1.0 scale representing each feature's importance to the model (137). In the context of this classification study, a value of -1.0 would be indicative of a feature's importance in selecting the OFF DBS/OFF MED condition, and 1.0 would be indicative of

the stimulated/medicated condition. Specifically, SHAP analysis, utilizing a random forest regressor, is applied for each stimulation/medication comparison using features which appear in more than five of the models. This analysis is performed over the entire patient cohort.

4.2.7 Comparison to non-PD cohort

To quantify stimulation frequency, and L-dopa impact on gait, four gait features (stride length, gait speed, foot strike angle, and toe-off angle) were compared to age and gender-matched healthy patients. These gait features were selected as a comparable healthy age and gender discretized cohort exists in the literature (67). The difference in each measured gait features between cohorts was calculated. This process was performed for each stimulation/medication condition.

4.3 Results

4.3.1 Random forest performance

This section describes the predictive capabilities of random forest algorithms at discriminating between six stimulation frequency/medication conditions compared to the OFF DBS/OFF MED condition. Differences between the random forest classifiers' performance are due to the impact of stimulation frequency and medication on gait kinematics. Each conditions models' performance is presented in Table 4.2. Overall, the random forest classifier achieves higher performance at discriminating ON MED conditions compared to the OFF DBS/OFF MED conditions. Specifically, the random forest achieves a predictive accuracy of $64 \pm 9\%$ for the HFS/ON MED condition, followed by LFS/ON MED ($64 \pm 10\%$) and OFF DBS/ON MED ($61 \pm 13\%$). For the OFF MED conditions, the random forest achieves a predictive accuracy of $60 \pm 7\%$ for HFS/OFF MED and $53 \pm 7\%$ for LFS/OFF MED, respectively.

Table 4.2: Random forest classifiers' performance at identifying the denoted stimulation frequency/medication condition from OFF DBS/OFF Medication using gait measurements (lower limb, lumbar, trunk) from instrumented SAW trials. Presented are each performance metrics mean and 95% confidence interval over leave-one-out cross validation.

	HFS/ OFF Meds	HFS/ ON Meds	LFS/ OFF Meds	LFS/ ON Meds	OFF DBS/ ON Meds
Accuracy	0.595 ± 0.072	0.637 ± 0.090	0.525 ± 0.070	0.637 ± 0.097	0.607 ± 0.128
Sensitivity	0.848 ± 0.110	0.884 ± 0.083	0.800 ± 0.134	0.849 ± 0.121	0.571 ± 0.207
Specificity	0.343 ± 0.198	0.390 ± 0.205	0.249 ± 0.168	0.425 ± 0.208	0.643 ± 0.205
F1	0.673 ± 0.054	0.719 ± 0.066	0.604 ± 0.089	0.694 ± 0.103	0.510 ± 0.184

As fewer SAW trials were conducted in OFF DBS conditions, the majority of the random forest models must account for imbalance within the training data (i.e., fewer SAW trials under the OFF DBS/OFF MED condition). Therefore, specificity is an important metric. Specificity follows a similar trend as accuracy, with the ON MED conditions having better predictive performance. Specifically, OFF DBS/ON MED has the highest specificity at $64 \pm 21\%$. However, LFS/OFF MED has the lowest specificity at $25 \pm 17\%$, meaning that the model rarely predicts the OFF DBS/OFF MED condition.

4.3.2 Gait feature analysis

Each random forest model used different features in order to discriminate every frequency/medication condition from the OFF DBS/OFF MED condition. The features utilized in each comparison and their direction of influence are present in Figure 4.1-4.3. In Figure 4.1-4.3, each model was trained using cross-validation, meaning a feature with a frequency of 21 appeared in every model. Additionally, in Figure 4.1-4.3, the directionality and magnitude of each feature is shown through SHAP analysis. A negative SHAP value is associated with the OFF DBS/OFF MED condition. In contrast, a positive value is associated with the comparative frequency/medication condition. To assist with the interpretability of these results, a Sankey diagram is provided in Figure 4.4, describing which gait features are impacted by stimulation/medication.

4.3.3 Effects on lower limb kinematics

The effects of frequency/medication on lower limb features are some of the most striking within the data set. All five comparative models used predominately lower limb features. Specifically, stride length is observed under all stimulation (HFS or LFS)/MED (OFF or ON) conditions and gait speed appears in all stimulation (HFS or LFS)/ON MED conditions as well as the HFS/OFF MED condition. Greater

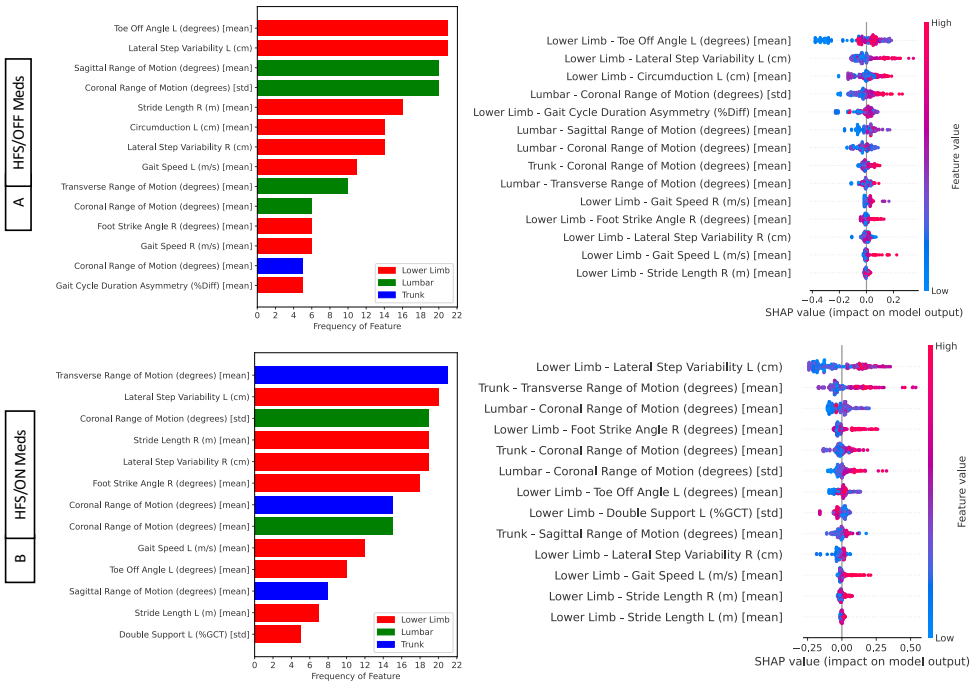

Figure 4.1: Comparison of features in high frequency stimulation conditions by gait category used in each random forest classifier discriminating the OFF DBS/OFF Meds condition. Each graph represents a comparison between OFF DBS/OFF Meds and A) HFS/OFF Meds, or B) HFS/ON Meds. Here A and B represents the frequency of each gait feature inclusion in the top-ten features used in more than five random forest models. Note that as leave-one-out cross-validation is utilized, each fold may produce a model containing different features. As such, the frequency denotes in how many models each feature appears. Additionally, the SHAP values are calculated based on a random forest regressor applied to all participants' top features. The y-axis denotes the relative range in the value of the feature, and the x-axis denotes how the feature impacts the predictive performance. Therefore, a lower SHAP value would be indicative that the feature relates to an OFF DBS/OFF Meds condition prediction. In contrast, a higher SHAP value would be indicative of the non-OFF DBS/OFF Meds condition.

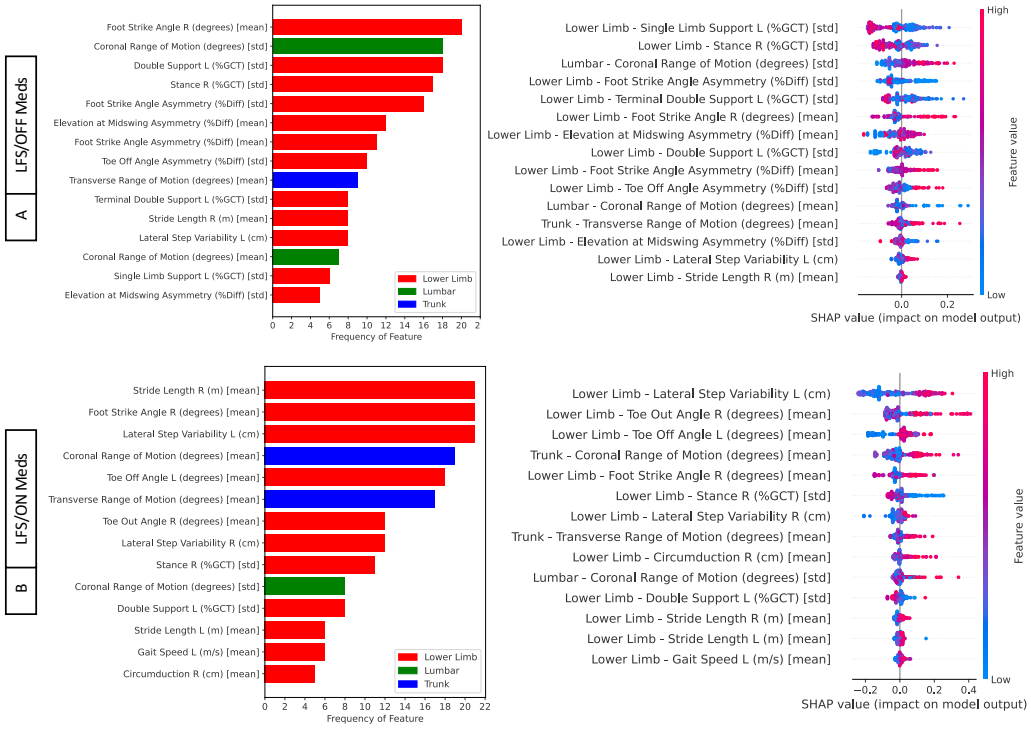

Figure 4.2: Comparison of features in low frequency stimulation conditions by gait category used in each random forest classifier discriminating the OFF DBS/OFF Meds condition. Each graph represents a comparison between OFF DBS/OFF Meds and A) LFS/OFF Meds, or B) LFS/ON Meds. Here A and B represents the frequency of each gait feature inclusion in the top-ten features used in more than five random forest models. Note that as leave-one-out cross-validation is utilized, each fold may produce a model containing different features. As such, the frequency denotes in how many models each feature appears. Additionally, the SHAP values are calculated based on a random forest regressor applied to all participants' top features. The y-axis denotes the relative range in the value of the feature, and the x-axis denotes how the feature impacts the predictive performance. Therefore, a lower SHAP value would be indicative that the feature relates to an OFF DBS/OFF Meds condition prediction. In contrast, a higher SHAP value would be indicative of the non-OFF DBS/OFF Meds condition.

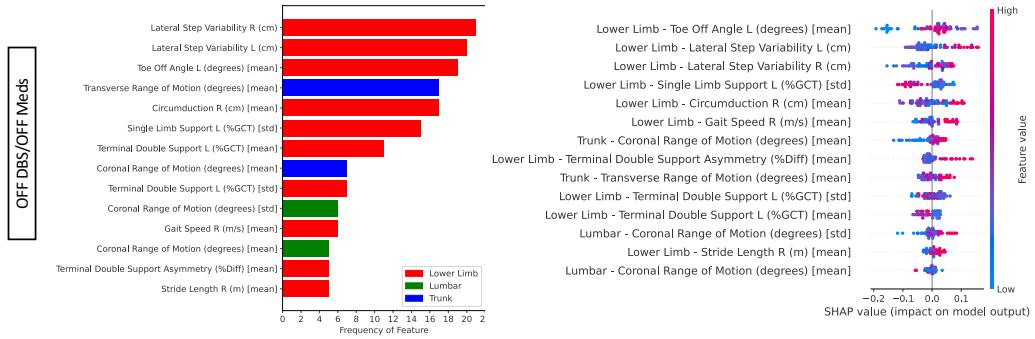

Figure 4.3: Comparison of features in the OFF stimulation condition by gait category used in each random forest classifier discriminating the OFF DBS/OFF Meds condition. Each graph represents a comparison between OFF DBS/OFF Meds and OFF DBS/OFF Meds. This represents the frequency of each gait feature inclusion in the top-ten features used in more than five random forest models. Note that as leave-one-out cross-validation is utilized, each fold may produce a model containing different features. As such, the frequency denotes in how many models each feature appears. Additionally, the SHAP values are calculated based on a random forest regressor applied to all participants' top features. The y-axis denotes the relative range in the value of the feature, and the x-axis denotes how the feature impacts the predictive performance. Therefore, a lower SHAP value would be indicative that the feature relates to an OFF DBS/OFF Meds condition prediction. In contrast, a higher SHAP value would be indicative of the non-OFF DBS/OFF Meds condition.

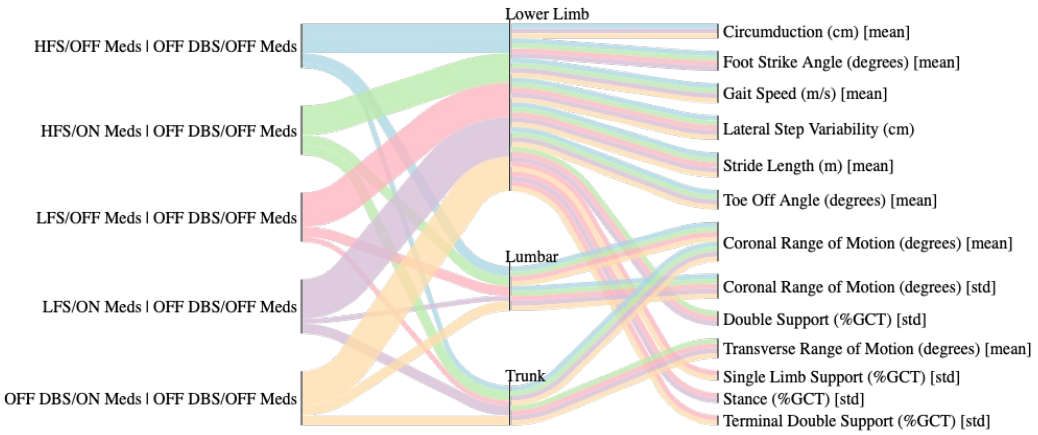

Figure 4.4: Sankey diagram representing features used in multiple random forest classifier discriminating the OFF DBS/OFF Meds condition. Each graph represents a comparison between OFF DBS/OFF Meds and Blue) HFS/OFF Meds, Green) HFS/ON Meds, Red) LFS/OFF Meds, Purple) LFS/ON Meds, and Orange) OFF DBS/ON Meds. Here, following a models colored lines will connect to the features utilized by feature category. Note that only features which are shared between models are included in the Sankey diagram. For example, following HFS/OFF Meds — OFF DBS/OFF Meds (blue) starting from the left one can observe the model uses features from all three gait categories (lower limb, lumbar, trunk). Further, it utilizes a greater number of lower limb features represented by the thicker line. Following the colored lines to the right represents each feature used in the model which is shared with another predictive model. For example, "Foot Strike Angle R" is used in HFS/OFF Meds, HFS/ON Meds, LFS/OFF Meds, and LFS/ON Meds models; however, it is not shared with OFF DBS/ON Meds model.

stride length and gait speed are associated with stimulation (HFS or LFS)/ON MED conditions. Comparing SHAP values, the effect on gait speed appears more prominent than stride length for some participants. Further, foot circumduction is utilized in the HFS/OFF MED condition, LFS/ON MED, and the OFF DBS/OFF MED conditions.

Stimulation and medication additionally impact foot positioning. Participants' toe-off angle measurements are utilized in HFS/OFF and ON MED, LFS/ON MED, and OFF DBS/ON MED conditions. A greater toe-off angle is associated with stimulation/ON MED conditions, whereas a reduced angle associates with the OFF DBS/OFF MED condition. Similarly, foot strike angle was present in all stimulated conditions (both HFS and LFS) with a greater angle associated with Stimulation/ON medication conditions.

Lateral step variability and stride length are the only gait features to be impacted by all frequency and medication conditions. An increased lateral step variability is highly associated with stimulation (HFS or LFS)/ON MED conditions. Further, several gait features' standard deviations (STD) are utilized in the models, albeit inconsistently. Specifically, the STD of the single limb support and terminal double support are present in both the OFF DBS/ON MED and LFS/OFF MED whereas the STD of the stance and double support associate with LFS/ON MED and LFS/OFF MED conditions. A reduced STD in these measurements is generally associated with stimulation (HFS or LFS)/ON MED. Notably, the STD of stance (%GCT) is increased under LFS.

4.3.4 Effects on lumbar kinematics

The effects of frequency/medication on lumbar features are more selective than lower limb features, with the LFS/OFF MED condition using few lumbar features. All stimulation (HFS and LFS)/ON MED conditions, and the HFS/OFF MED condition, utilize coronal range of motion, which also appears to be more disparate amongst participants in the OFF DBS/OFF MED condition. An increased STD in coronal

range of motion is strongly indicative of being in a stimulated/medication condition, compared to the OFF DBS/OFF MED condition.

4.3.5 Effects on trunk kinematics

The effects of frequency/medication on trunk features appear to be associated with medication effects. HFS/ON MED, LFS/ON MED, and OFF DBS/ON MED models consisted of more trunk features than their respective OFF Medication conditions. Coronal range of motion is employed in HFS/OFF MED, HFS/ON MED, LFS/ON MED, and OFF DBS/ON MED conditions. In comparison, transverse range of motion is used in LFS/OFF MED, LFS/ON MED, HFS/ON MED, and OFF DBS/ON MED condition models, while sagittal range of motion is used primarily in the HFS/ON MED condition. Further, both coronal and transverse range of motion increase under stimulated (HFS and LFS)/ON MED conditions.

4.3.6 Gait comparison to a healthy cohort

To consider the overall impact of frequency and medication on gait, features were compared to a reported cohort of gender and age-matched healthy controls (67). These results are presented in Figure 4.5. For most participants, HFS provides improvements across all gait kinematics over LFS and the OFF DBS condition. However, there are participants within the cohort, for example, participant 13, who receive greater benefits from LFS. It is also interesting to note that participants 16 and 18 receive greater benefits under LFS only in combination with medication. Additionally, participants 19, 20, and 21, experience a "narrowing of the gap" of feature correction between stimulation frequencies with the inclusion of medication. Further, a synergistic effect between HFS and medication can be observed in participant 21.

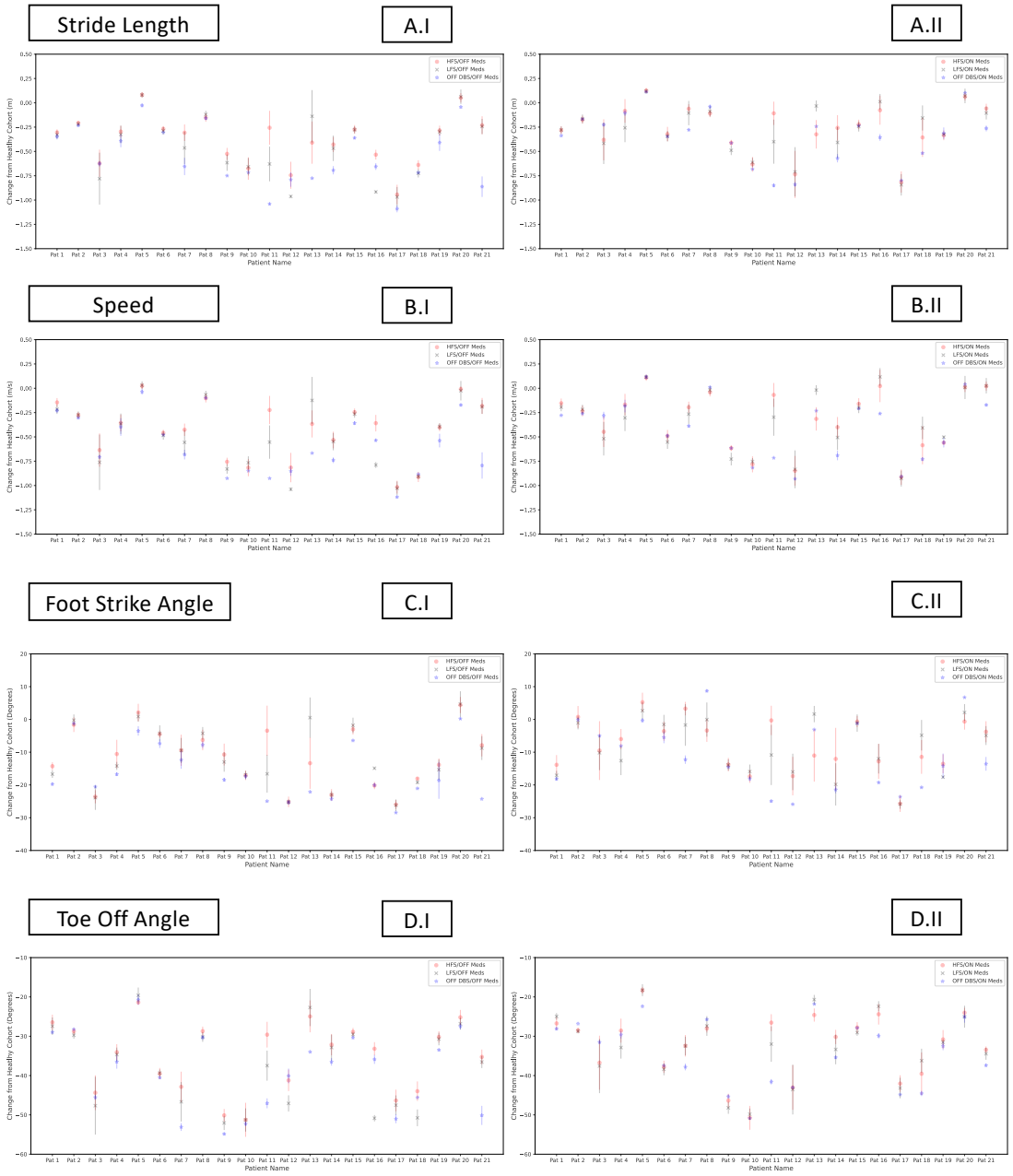

Figure 4.5: Comparison of stimulation and medication conditions' gait features from age and gender matched healthy patients. Each graph represents a comparison between each patient's gait parameters in the I) OFF Medication condition under each stimulation frequency and in the II) the ON Medication condition under each stimulation frequency. Each gait parameter is specific to a patient's initially affected hemibody. Each graph presents a different gait feature where A) Stride Length, B) Speed, C) Foot Strike Angle, and D) Toe Off Angle, which is utilized in one or more random forest models. Here, a value of 0 would represent a patient's gait parameter equaling that of a healthy age and gender matched control.

4.4 Discussion

This study demonstrates the differential effects of stimulation (HFS and LFS) and L-Dopa on gait kinematic domains: lower limb, lumbar, and trunk. We show that by comparing gait features, captured by instrumented walking assessments, that a predictive classification algorithm can be created to infer impact of stimulation and/or medication on specific aspects of gait.

4.4.1 Stimulation frequency and levodopa effects on gait features

The comparison of features in stimulation condition models (either HFS or LFS) that are not shared in the OFF DBS/ON MED models indicate features that are likely impacted primarily by stimulation. The results suggest that foot strike angle increases with stimulation (HFS or LFS) from the OFF DBS/OFF MED condition. Additionally, stride length and lumbar coronal range of motion are shared between stimulation (HFS or LFS)/OFF MED conditions. Prior studies have shown that HFS stimulation improves stride length in PD patients, (71; 195) but the results have been mixed with respect to LFS (153; 177).

Little work has been done to determine the interplay between medication and stimulation. Faist et al. (66) suggested that DBS and L-dopa have a synergistic effect that improves a patient's stride length and gait speed; however, this work did not account for differences in stimulation frequency. Our results show that L-dopa influences lateral step variability, stride length, gait speed, toe off angle, lumbar coronal range of motion, and trunk transverse range of motion. These effects are seen both with and without the presence of stimulation. However, greater effects on lateral step variability, stride length, gait speed, toe off angle, and foot strike angle are seen when either HFS or LFS is combined with L-dopa. This combined effect is also observed with trunk and lumbar features, but since dyskinesia was observed in 11%

of all SAW trials—83% of which were in the stimulation/ON medicated states—we cannot exclude its effects on these specific features.

4.4.2 Treatment synergism and LFS responders

A subset of gait features (stride length, foot strike angle, toe off angle, gait speed), which appear in several classifier models, were adjusted in reference to healthy controls to determine the impact of stimulation frequency solely on these features. We again find that all features improve for several participants with HFS and LFS. The results further suggest that stimulation and L-dopa have a synergistic effect for several participants' stride length (participants 7, 11, 14, 16, 18, 21). Within the cohort, there appears to be participants who are LFS responders. Participant 13 received greater benefits from LFS compared to HFS in the OFF medicated state. For most participants, HFS and LFS gait feature improvements are similar; however, there are three participants (participants 3, 12, and 16), who did demonstrate worsening of certain kinematics under LFS in the OFF MED state. Additionally, there is an added benefit for participants receiving both LFS and medication. This combined effect can be seen in the stride length changes in participants 16 and 18, whose reduced stride length improvement in the LFS/ON MED state exceeds both medication alone and HFS/ON MED. This improvement suggests that responders who would receive additional gait benefits from LFS in combination with medication may exist and requires a larger study cohort to confirm this supposition.

4.4.3 Stimulation frequency and levodopa effects on gait variability

Our results suggest that HFS-DBS, LFS-DBS, and medication, increase lateral step variability. In general, an increase in lateral step variability is seen with normal aging (209), and in chronic PD patients, with STN-DBS, an increased lateral step variability may reflect a trend towards a more age-appropriate gait cycle. Reduced

STD in single limb support, double limb support, terminal double support associates with the medicated and LFS states, while an increase in STD of stance (%GCT) occurs under LFS. These temporal feature changes require additional investigation to determine whether they influence balance dynamics in those with advanced disease. Additionally, all stimulated and medicated conditions increase the STD of the lumbar coronal range of motion, which likely reflects enhanced movement in this axial plane. Such a change may not only underpin potential unmeasured influence on balance but could also result from dyskinesia.

Each condition's model incorporates the STD of one or more gait features, suggesting the importance of feature variability in determining the impact of stimulation and medication. The LFS/OFF MED model utilizes a disproportionate number of gait measurements' STD. While the classifier performs more poorly at discriminating between the LFS/OFF MED and OFF DBS/OFF MED conditions, these features are significant to the models' performance. Typically, these STDs are reduced in the LFS/OFF MED condition. However, the STD of coronal range of motion is increased under all conditions. The reduced variability of gait under LFS conditions could be influenced by the presence of walking aides, which were used more in the LFS/OFF MED condition and likely produced a more consistent walking pattern.

4.4.4 Study limitations

This study is subject to certain limitations: 1) A small cohort size coupled with the fact that subjects were on non-standard of care stimulation/medication conditions for short periods of time reduce our ability to extrapolate any of these results beyond short-term effects. 2) The stimulation of different contact pairs in a randomized manner may have increased some of the variability observed in this study. 3) The use of walking aids could statistically affect gait parameters, particularly stability and variability measurements.

4.5 Conclusions

The results from this study show differential effects of HFS and LFS STN-DBS as well as L-dopa on various spatial and temporal aspects of gait in an advanced PD-DBS cohort. Synergistic benefits of combining stimulation (HFS or LFS) and medication were demonstrated in several gait parameters—speed, stride length, foot strike angle, toe off angle. Additionally, individual LFS responders were identified within the study's cohort. Finally, we demonstrate that leveraging ML modeling can highlight and help elucidate the complexity of the interplay between DBS and medication on gait kinematics, which in time, could lead to patient-specific predictive approaches.

Chapter 5

Adapting Random Forests to Predict Obesity-Associated Gene Expression

5.1 Introduction

Large-scale expression quantitative trait loci (eQTL) analyses have demonstrated that most genes have a genetic component to their expression (135). By combining genotypes from multiple variants proximal to a gene, we can build models that can predict tissue-specific expression of a gene at the patient level. PrediXcan is one such genotype-based prediction method that can predict gene expression from individual-level genotype data (75; 7). While there is empirical evidence that PrediXcan predicts the expression of a wide range of genes well, this is not universally the case for all genes (229; 148).

Machine learning approaches, such as random forests (RFs), have several desirable traits for bioinformatics, including the clear identification of important features, scalability, and robustness against noise (184; 17). Consequently, RFs have been explored to predict gene expression (58; 116). While these studies have shown great

promise, comparative studies between RF regressors and classifiers have been carried out less extensively in the genetics literature.

The literature on how machine learning approaches generalize to many medically relevant genes is limited. One genetic association with far-reaching impact is obesity. Obesity is clinically linked to increased morbidity and mortality (230; 40). Genetic effects on monogenic obesity have been well documented (230) and recent GWAS studies have discovered many novel single nucleotide polymorphisms (SNPs) associated with non-syndromic obesity (230; 40; 96).

In this work, we examine the capability of RFs to predict the gene expression of key obesity genes based on genetic variants. Specifically, each genetic variant is used as a potential feature in our models. Further, we compare these RFs to PrediXcan, a state-of-the-art gene expression prediction method (75). PrediXcan uses predetermined weights, often from PredictDB (https://predictdb.org), making it a popular choice amongst researchers (179; 174). We apply RFs and PrediXcan to subcutaneous adipose tissue samples from the Genotype-Tissue Expression (GTEx) project (135; 87).

5.2 Data

In this study, we utilize the GTEx dataset (135; 87), which contains samples of up to 43 highly genotyped tissue types from more than 700 post-mortem donors. The GTEx database includes eQTL data, which represent gene-regulatory expression across multiple tissues (135; 87). In this study, we examine the normalized expression and genetic variant data in subcutaneous adipose tissue to predict the expression of key obesity genes for 581 donors. The data were analyzed following GTEx's Data Use Certification Agreement. Ethical approval for GTEx was originally obtained from each study site's Institutional Review Board (135).

The role of genetics in obesity has been well-documented leading to the identification of many impactful genes (230; 40; 96). The Centers for Disease Control (CDC) published an article identifying key genes associated with obesity (Centers for Disease Control and Prevention). These genes include *ADIPOQ* (ENSG00000181092), *FTO* (ENSG00000140718), *INSIG2* (ENSG00000125629), *LEP* (ENSG00000174697), *LEPR* (ENSG00000116678), *PCSK1* (ENSG00000175426), and *PPARG* (ENSG00000132170).

5.3 Methods

We apply five RFs (ordinal or one-hot encoded regressors and classifiers and a balanced classifier) to predict gene expression using different feature selection and encoding techniques. RFs are supervised machine learning models that use an ensemble of decision trees, and are generally shown to be very robust when applied to high-dimensional data, and not prone to overfitting (17). Therefore, RFs are well-positioned to predict gene expression from genetic variants, which frequently entail more than $30,000$ features.

We apply RFs to two different versions of the GTEx dataset for comparison with PrediXcan. First, we perform 10-fold cross-validation (CV) on GTEx version 8. Second, each RF is separately trained on GTEx version 7 and then tested on the donors present in GTEx version 8 but not in GTEx version 7. The latter process is visualized in Figure 5.1. Note that we use this process to ensure that we avoid any potential overlap between training and testing data during the analysis. Each RF was implemented using scikit-learn's RF libraries in Python (26). A similar training and testing approach is followed for PrediXcan; weights from PrediXcan defined for GTEx version 7 are used for testing PrediXcan on donors available only in GTEx version 8. This way, PrediXcan's testing and training data do not overlap.

5.3.1 Random Forest Regressor and Classifier

In this study, we compare the predictive performance of RF regressors and classifiers in predicting obesity-associated gene expression. Gene expression is a normalized continuous variable, typically ranging from -3 to 3, which justifies the use of an RF regressor.

Much of clinical value in predicting gene expression comes from stratifying individuals' expression levels (192). Therefore, an RF classifier may prove beneficial at predicting expression values at the extreme ends of the distribution. To apply an RF classifier, we discretize the continuous variable of gene expression. Specifically, we define five classes that group gene expression values. First, we define a class within one standard deviation (SD) from the mean, then between one and two SDs away from the mean (in both positive and negative directions), and finally beyond two SDs away from the mean (in both positive and negative directions). Once an RF predicts a class based on the genetic variants, we map the predicted class to a continuous value at the center of the corresponding class to compute R^2.

5.3.2 Ordinal and One-hot Encoding

The genetic variant data from the GTEx project are provided in the VCF file format. We compare two different encoding methods. First, we apply the ordinal encoding mapping .|. (missing data) $\rightarrow$ 0, 0|0 (no alt alleles) $\rightarrow$ 1, 0|1 and 1|0 (one alt allele) $\rightarrow$ 2, 1|1 (two alt alleles) $\rightarrow$ 3. Such encoding assigns higher importance to SNPs with more copies of the alternate allele.

As an alternative approach, one-hot encoding represents each variant as a binary vector. This encoding method is often a more efficient way to parse and store data. To accomplish this encoding, we use Keras' one-hot encoding library (39).

5.3.3 Oversampling Approach for Data Balancing

One challenge in training a classification model is learning from unbalanced data (60). This is the case in our constructed classes, where the expression values have been pre-normalized by GTEx. One way of ameliorating the class imbalance problem is to apply a domain-specific data oversampling technique in which noise is added to the sampled data. Similar 'oversampling with noise' techniques have been employed in the task of ordinal classification (60). In this study, we randomly sample a genetic profile from an 'under-represented' class. We then randomly alter the ordinal encoding of 100 genetic variants from this profile, acting as noise in the sampling process. This profile maintains the same class representation as before the noise generation process. The oversampling process is repeated for each under-represented class.

5.4 Results

We compare the predictive performance (as determined through R^2, a standard metric in gene expression prediction studies (75; 7)) of five RFs and of PrediXcan at predicting gene expression. We first perform preliminary experiments to tune the RFs' hyper-parameters. Specifically, we determine the optimal number of decision trees by iteratively adding trees from 100 to 1500 with a step size of 100 trees. Further, we compare the optimal number of included features by keeping the most important features ranging from 500 to the maximum number of genetic variants for each gene (approximately $30,000$) with a step size of 500 features, as each gene may be influenced by a unique number of variants. The corresponding CV results on GTEx version 8 are presented in Figures 5.2 and 5.3, respectively. In general, these results indicate that increasing the number of decision trees does not dramatically impact predictive performance. Similarly, a relatively low number of features achieves overall comparative performance. Based on these results, in each of the following

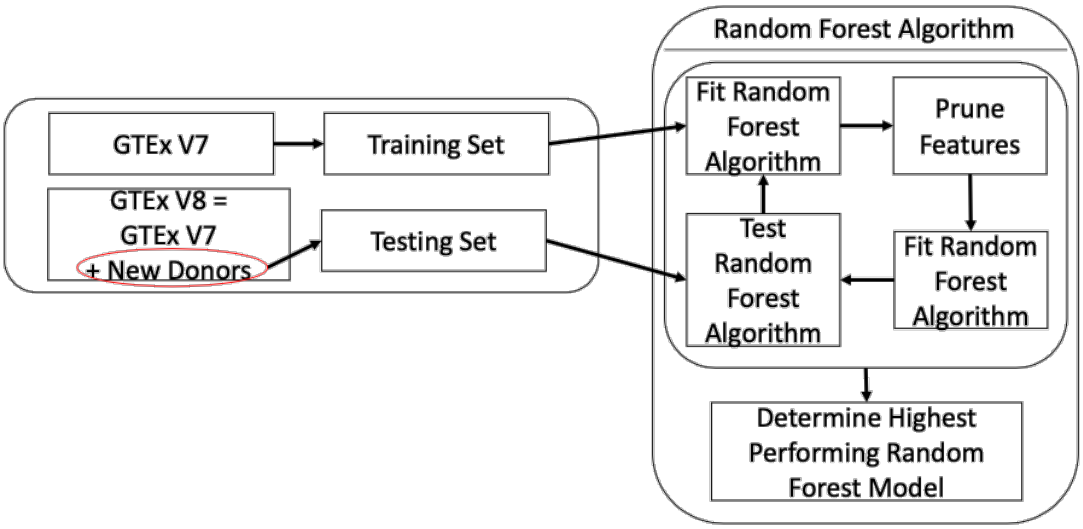

Figure 5.1: Fitting a random forest (RF) to the GTEx data. The RF is trained on GTEx version 7 and tested on the donors present in GTEx version 8 but not in version 7. The same training/testing separation is performed for PrediXcan using the publicly available weights defined for GTEx version 7.

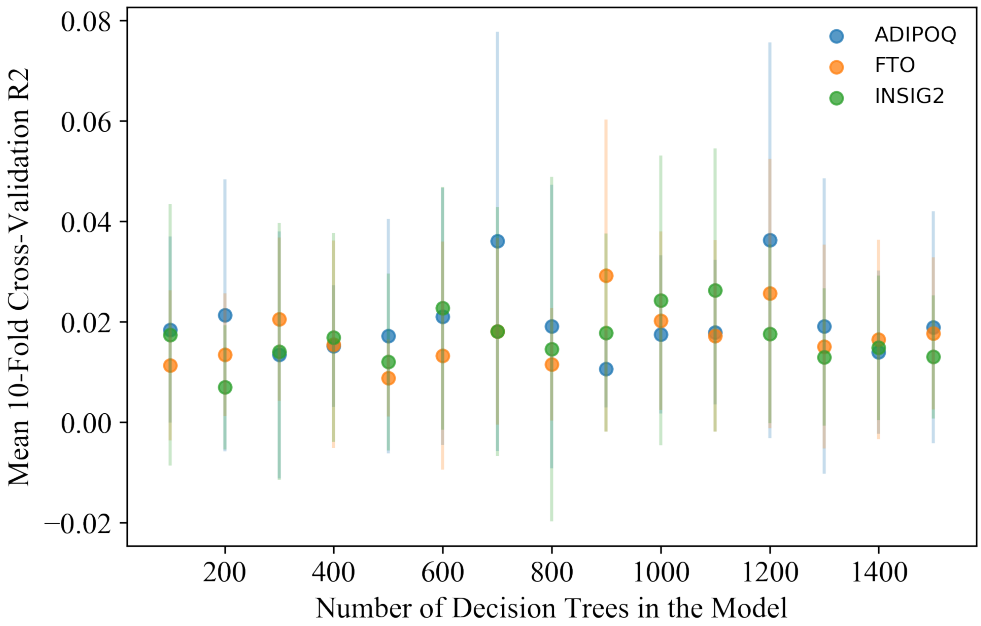

Figure 5.2: Sample of the performance of the random forest regressor for incremental amounts of decision trees for *ADIPOQ*, *FTO*, and *INSIG2*. Each point represents the mean R^2 averaged over 10-fold CV, where the error bars represent the standard deviation.

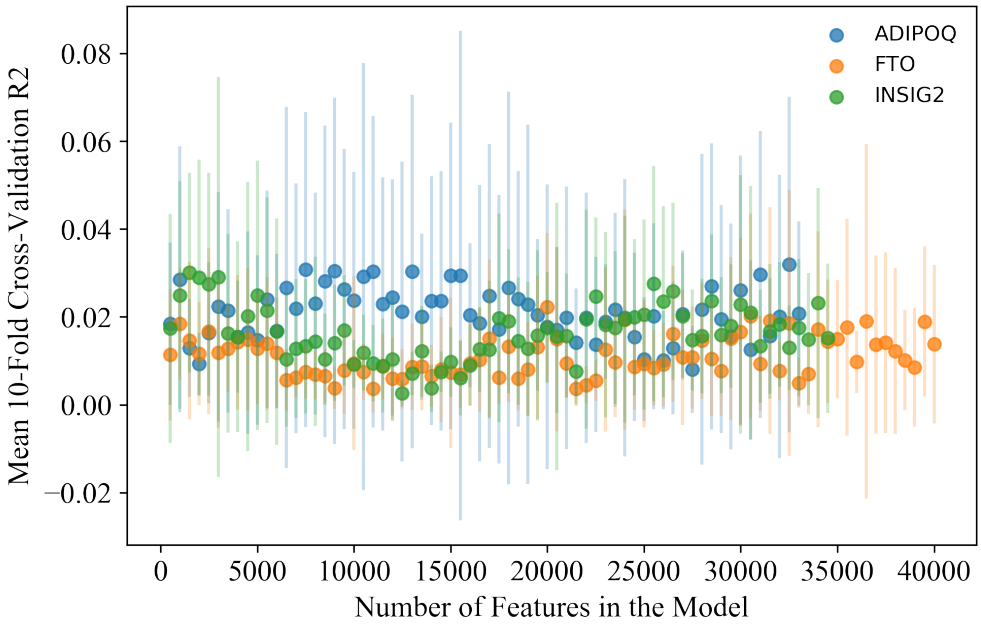

Figure 5.3: Sample of the performance of the random forest regressor for incremental amounts of features for *ADIPOQ*, *FTO*, and *INSIG2*. Each point represents the mean R^2 averaged over 10-fold CV, where the error bars represent the standard deviation.

experiments, we set the number of decision trees to 100, and we perform feature pruning from 25 to 1000 features with a step size of 25 features.

5.4.1 GTEx Version 8

We compare the five implemented RFs using 10-fold CV to predict gene expression in GTEx version 8 subcutaneous adipose tissue. These results are presented in Table 5.1. The RFs are benchmarked against PrediXcan's elastic net and MASHR models, which use predetermined weights previously trained on the entirety of GTEx version 8 (179). Note that the PrediXcan's elastic net model does not have publicly available predetermined weights for the genes *LEP* and *PPARG* in subcutaneous adipose tissue.

These results show that there is no universal winner between RFs and PrediXcan's methods based on paired t-tests on the differences in R^2 between associated model pairs at a 10% level of statistical significance. For *LEPR*, the ordinal RF regressor and the balanced classifier outperform PrediXcan's MASHR model; for *LEP*, the balanced RF classifier outperforms PrediXcan's MASHR model; for *FTO*, PrediXcan's elastic net outperforms PrediXcan's MASHR and the ordinal RF regressor and classifier; for *INSIG2*, PrediXcan's elastic net outperforms PrediXcan's MASHR model and all RFs. Finally, for *LEPR* the ordinal RF regressor outperform both the one-hot encoded regressor and classifier.

5.4.2 GTEx Version 7

We again compare the five implemented RFs and PrediXcan's methods in terms of gene expression predictions on the testing data. As discussed in Section 6.3 and depicted in Figure 5.1, to avoid overlap between training and testing data, we use PrediXcan's predetermined weights from GTEx version 7 to predict expression for a testing set that only contains the new donors added in GTEx version 8. Similarly, the RFs are trained on GTEx version 7 and are tested on the donors added in GTEx version 8 (174). We draw $1,000$ bootstrap samples from the testing set, each with

Table 5.1: Mean R^2 values of random forests and of PrediXcan benchmarks related to the prediction of obesity-associated gene expression (standard deviations are shown in parentheses). Each R^2 is averaged over 10-fold CV on GTEx version 8. Models which exhibit improvement in gene-specific predictive performance at a significance level of 10% are presented in bold (see Section 5.4.1). Significance is determined via paired t-tests in the difference of R^2 values between model pairs.

Gene Name	PrediXcan benchmark		Random forest regressor		Random forest classifier		
	Elastic net	MASHR	Ordinal	One-hot	Ordinal	One-hot	Balanced
ADIPOQ	0.030(0.041)	0.050(0.066)	0.031(0.030)	0.023(0.018)	0.034(0.028)	0.029(0.039)	0.028(0.023)
FTO	**0.050(0.035)**	0.017(0.017)	0.020(0.023)	0.035(0.040)	0.022(0.019)	0.026(0.032)	0.036(0.038)
INSIG2	**0.080(0.048)**	**0.060(0.044)**	0.035(0.024)	0.029(0.034)	0.025(0.046)	0.047(0.061)	0.033(0.030)
LEP	-	0.011(0.007)	0.026(0.037)	0.021(0.027)	0.039(0.064)	0.030(0.045)	**0.040(0.045)**
LEPR	0.035(0.022)	0.016(0.017)	**0.045(0.041)**	0.025(0.029)	0.028(0.026)	0.028(0.030)	0.036(0.031)
PCSK1	0.052(0.062)	0.038(0.051)	0.036(0.056)	0.037(0.060)	0.031(0.033)	0.034(0.042)	0.045(0.036)
PPARG	-	0.028(0.021)	0.037(0.053)	0.020(0.021)	0.034(0.047)	0.041(0.055)	0.039(0.035)

Table 5.2: 90% Bias-corrected and accelerated (BCa) bootstrap confidence intervals for R^2 values of random forests and of PrediXcan related to the prediction of obesity-associated gene expression. The testing data are extracted from GTEx version 8, corresponding to donors not in GTEx version 7. Each BCa confidence interval is estimated from $1,000$ bootstrap samples from the testing set. Models which exhibit improvement in gene-specific predictive performance at a significance level of 10% are presented in bold (see Section 5.4.2). PrediXcan's elastic net model does not have predetermined weights for *LEP* and *PPARG*.

Gene name	PrediXcan benchmark	Random forest regressor		Random forest classifier		
	Elastic net - V7	Ordinal	One-hot	Ordinal	One-hot	Balanced
ADIPOQ	(0.000, 0.052)	(0.014, 0.179)	(0.004, 0.102)	(0.003, 0.143)	(0.004, 0.174)	(0.000, 0.075)
FTO	(0.000, 0.080)	(0.027, 0.257)	(0.011, 0.204)	(0.000, 0.161)	(0.000, 0.139)	(0.000, 0.087)
INSIG2	(0.035, 0.194)	(0.001, 0.102)	(0.003, 0.181)	(0.014, 0.160)	(0.026, 0.258)	(0.000, 0.012)
LEP	-	(0.000, 0.097)	(0.001, 0.083)	(0.014, 0.197)	(0.016, 0.201)	(0.000, 0.000)
LEPR	(0.001, 0.111)	(0.000, 0.079)	(0.006, 0.148)	(0.004, 0.169)	(0.002, 0.146)	(0.009, 0.161)
PCSK1	(0.000, 0.073)	(0.000, 0.052)	(0.016, 0.170)	(0.006, 0.149)	(0.000, 0.192)	(0.000, 0.000)
PPARG	-	(0.002, 0.126)	(0.012, 0.110)	(0.001, 0.129)	(0.000, 0.090)	(0.067, 0.279)

sample size equal to the sample size of the testing set. For each of the $1,000$ bootstrap samples, we compute the R^2 of each model. We then determine bias-corrected and accelerated (BCa) confidence intervals for R^2 over the bootstrap samples from the testing sets as a range of predictive performance (59). These confidence intervals are presented in Table 5.2. To utilize this approach, we translate variants from human genome reference build b37 to b38.

We compute 90% BCa confidence intervals for the differences in R^2 between model pairs as a metric of statistical significance. The predictive performance of RFs is statistically similar to PrediXcan at a significance level of 10%, that is based on the 90% BCa confidence intervals for the differences in R^2. Moreover, at a significance level of 10% the balanced RF classifier is outperformed on *INSIG2* (by PrediXcan, and the one-hot RF classifier), *LEP* (by ordinal and one-hot RF classifiers), and *PCSK1* (by ordinal and one-hot RF classifiers).

5.5 Discussion and Future Work

In this study, we compare RF regressors and classifiers in the prediction of obesity-associated gene expression. The RFs are benchmarked against PrediXcan's state-of-the-art genotype-based prediction algorithms. Our results show, that in general the RFs perform comparably and, in some cases, outperform PrediXcan. Finally, we find that ordinal and one-hot encoding perform similarly for the majority of genes.

This work is subject to certain limitations. Due to the pre-trained nature of PrediXcan, it is not trivial to make a fair and direct comparison. Using the GTEx version 8 dataset, PrediXcan gains an advantage from having previously been trained on the entirety of GTEx, meaning it has had exposure to the testing set. Although GTEx presents one of the most deeply sampled genetic populations, testing on a completely novel/separate obesity genetic dataset would be ideal for benchmarking.

RFs have shown to be effective in predicting gene expression for select gene/tissue pairs (58; 116). This work expands the literature to encompass key obesity genes

while additionally exploring the predictive merits of regressors, classifiers, encoding, and the balancing of training sets. To this end, future work may include developing an RF that trims features (genetic variants) based on their positional proximity relatively to one another, for a biologically-inspired approach to parameter pruning.

Chapter 6

Towards Faster Gene Expression Prediction via Dimensionality Reduction and Feature Selection

6.1 Introduction

The majority of genes' expressions are modulated by multiple proximal variants specific to that gene (135). These variants can be used as inputs to bioinformatics models to predict tissue-specific expression of a gene at the patient level (75; 232). However, the efficiency and effectiveness of such models can be limited due to the large number of input features (genetic variants) relative to samples (donors), as several genetic variants have little to no contribution to gene expression. Several methods exist to reduce input size or dimensionality; however, these methods have not been widely applied to patient-specific gene expression data.

Principal component analysis (PCA) is a dimensionality reduction technique with applications in bioinformatics studies (138; 193). PCA's effectiveness with genetic variants has been demonstrated; however, it is commonly used to summarize genotype data for use alongside other features in phenotypic classification studies (128; 237; 73;

15; 19; 46). PCA's ability to summarize genetic variants for the prediction of gene expression remains an open question.

As an alternative approach, linkage disequilibrium (LD) pruning seeks to remove redundant or highly correlated variants, thus eliciting a reduced but representative set of genetic variants. LD pruning is a commonplace technique in the assessment of polygenic risk scores and for genetic predictions in populations with small effective population sizes and high LD, such as animal studies (28; 225; 98; 125; 229). Several tools exist to apply LD pruning to a variety of genetic data types (183; 28). While LD pruning does not directly relate to gene expression, it facilitates the selection of a representative subset of genetic variants which could be used for the prediction of gene expression.

In this study, we examine the role of PCA and LD pruning to summarize or reduce the number of input genetic variants for the prediction of tissue-specific gene expression. Specifically, the modified genetic variants are used as features in elastic net models. Elastic nets have been shown effective at predicting individual-level gene expression and have been demonstrated to provide accurate predictions (75; 179). We carry out our analysis on skeletal muscle tissue samples from the Genotype-Tissue Expression (GTEx) project (135; 87; 42).

6.2 Data

In this study, we analyze the GTEx data set (version eight, see (135; 87; 42)), which contains individuals' genetic variants and tissue-specific expression data from more than 700 post-mortem donors. Specifically, our models predict the individual's normalized gene expression (dependent variable) using their ordinal encoded genetic variant data as predictors (independent variables). These data are collected from 706 skeletal muscle tissue samples, representing the largest single tissue location in the GTEx database. We predict the expression of 151 genes for which PrediXcan's elastic

net model, a state-of-the-art gene expression prediction model, achieves an $R^2 > 0.5$ (75; 179).

6.3 Methods

We apply an elastic net to predict gene expression utilizing each PCA, LD pruning and a combination of the two techniques to modify the genetic variants as inputs. Elastic nets are used herein, as they have been demonstrated to be highly effective at predicting gene expression (75; 179). An overview of the study's design is presented in Figure 6.1.

We perform 10-fold cross-validation on skeletal muscle samples from the GTEx V8 database (87; 42). For each of the 151 genes, PCA and LD pruning were performed independently. We utilized Scikit-learn's elastic net library in Python (176). In this study, the elastic net's penalty constant multiplier α was set to 1, and the l_1 ratio, which is the mixing parameter defined by $0 \leq l_1 \leq 1$, was optimized over the data set, see Figure 6.2.

6.3.1 PCA

To create a lower-dimensional summary of the genetic variant data, PCA is used. We utilize Scikit-learn's PCA library in Python (176). For each gene and fold of cross-validation, PCs are independently calculated.

To determine the efficiency and accuracy of PCA, different numbers of PCs are analyzed across all genes. Specifically, $1, 3, 5, 10, 25, 50, 100, 200, 300, 400, 500$ PCs are analyzed. For each of these numbers of PCs, the average computational time for PCA and elastic nets across all folds are determined, as well as their corresponding R^2 value.

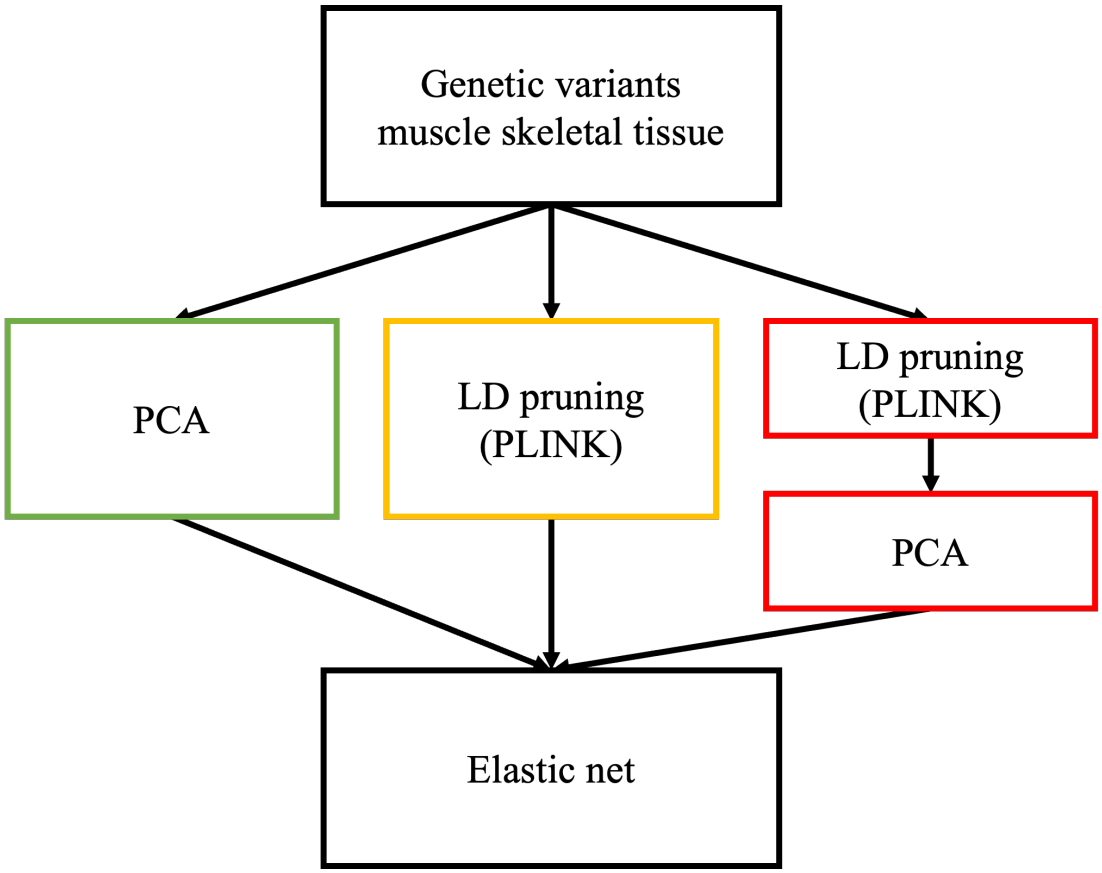

Figure 6.1: An overview of the study. PCA, LD pruning, and a combination of the two are applied to the genetic variant data. LD pruning is performed using the PLINK software package (183). Each reduced data set is analyzed by an elastic net model to predict gene expression. This results in four different elastic net models for comparison.

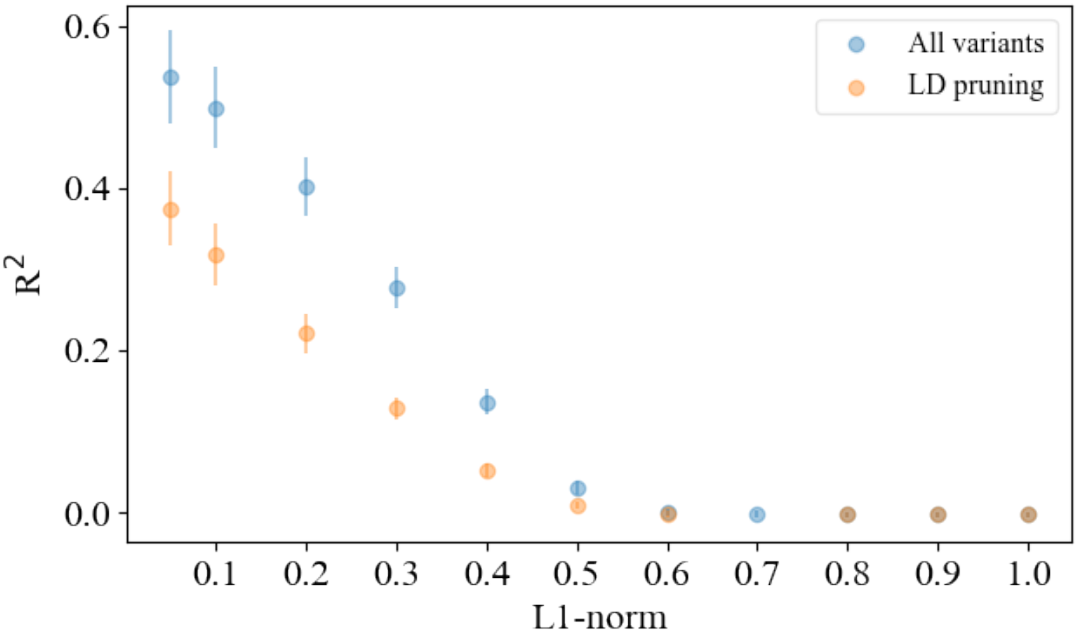

Figure 6.2: Calibration plot of the elastic net over all genes utilizing both all variants and LD-pruned data. Each point represents the mean R^2 value of all 151 genes with standard deviation. The l_1-norm parameter dictates the regularization between ridge (0.0) and lasso (1.0).

6.3.2 LD pruning

LD pruning is a recursive technique that removes highly correlated variants via a sliding window calculating pairwise variants' R^2 values and the variance inflation factor (183). This process seeks to determine a 'representative,' independent variant in regions of high LD. To accomplish this, we use the PLINK software package with a 50 variant window using a 5 variant step and a 0.2 LD threshold (183).

6.3.3 Combined PCA and LD pruning

Finally, a combination of PCA and LD pruning is examined. In this combination, LD pruning is first applied to all variants of each gene as described in Section 6.3.2, and then the pruned subset is summarized by PCA as described in Section 6.3.1. For each gene, 500 PCs are utilized, as such a larger number of PCs provides higher accuracy.

6.4 Results

We compare the performance of elastic net models with PCA and LD-pruned input features to predict gene expression. Each elastic net's l_1-norm is calibrated across all genes; the results of the calibration are provided in Figure 6.2. The elastic net's performance improves for a low l_1-norm, which is descriptive of ridge regression, for elastic nets utilizing both all genetic and LD pruned variants. Therefore, an l_1-norm of 0.05 is selected, as SciKit-learn's elastic net is reported to be unstable for l_1-norms < 0.05 (176).

6.4.1 Dimensionality reduction and feature selection results

We compare the average R^2 for each gene across all elastic net models for each PCA, LD pruning, and a combination of the two. The results of these comparisons are provided in Figure 6.3. For all genes, 10-fold cross-validation is performed for each elastic net model.

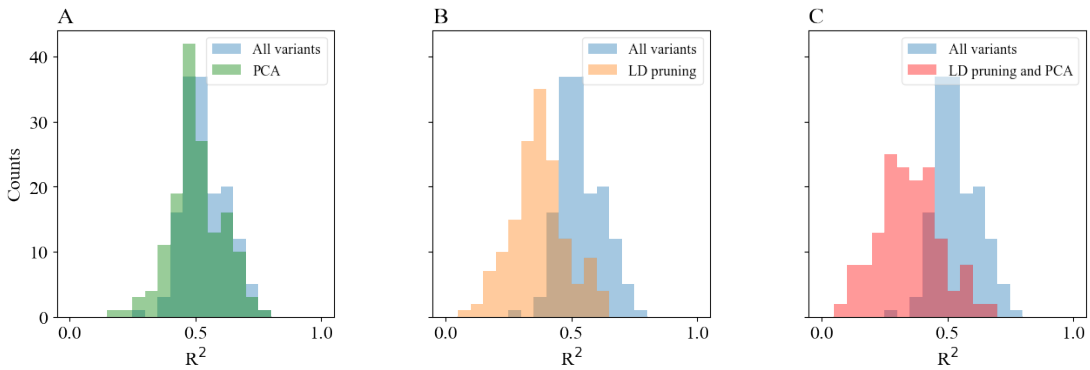

Figure 6.3: Comparison of the elastic net's R^2 performance with each dimensionality reduction or feature selection technique. Each plot displays a comparison between an elastic net utilizing all variants and an elastic net with PCA performed with 500 PCs (left figure), LD pruning via the PLINK software package (middle figure), and a combination of LD pruning and PCA performed with 500 PCs (right figure).

The results indicate that the elastic net utilizing the full genotype data (all variants) has the best performance, followed by PCA with 500 PCs. Specifically, elastic nets utilizing PCA-summarized features have a 7% reduction in overall performance compared to the elastic net model utilizing all variants. However, PCA performs very well at summarizing the genes for which the elastic net has its best performance. These genes' expressions are likely correlated with a few highly significant genetic variants, and therefore, PCA is able to more effectively summarize the information contained in these genetic variants.

Finally, elastic nets utilizing inputs from LD pruning and a combination of LD pruning and PCA have lower average R^2 scores than elastic nets utilizing all variants and PCA. Specifically, an elastic net using LD-pruned variants has an average 30% reduction in overall performance, and the combination of LD pruning and PCA has an average 35% reduction in overall performance. Further, applying PCA to these LD pruned-variants results in a greater number of genes that have an $R^2 < 0.2$.

6.4.2 PCA performance and timing

As PCA outperforms LD pruning, we further compare the number of PCs' impact on predictive accuracy and computational time. The performance results are presented in Figure 6.4-A, and the timing results are presented in Figure 6.4-B. This analysis is performed for $1, 3, 5, 10, 25, 50, 100, 200, 300, 400, 500$ PCs.

These results demonstrate that the elastic net's performance improves with a greater number of PCs as expected, achieving its best performance for 500 PCs. However, the elastic net achieves similar performance with a relatively low number of PCs, such as 100 PCs. Specifically, using $1, 3, 5, 10, 25, 50, 100, 200, 400, 500$ PCs results in an R^2 of $0.06, 0.17, 0.24, 0.42, 0.47, 0.49, 0.50, 0.50, 0.50, 0.50$, respectively.

Similar to the PCA's R^2 values, the computational time required also increases with the number of PCs. Specifically, using $1, 3, 5, 10, 25, 50, 100, 200, 300, 400, 500$ PCs results in an average per gene computation time of $2.03, 2.06, 2.13, 2.28, 2.74, 3.47,$

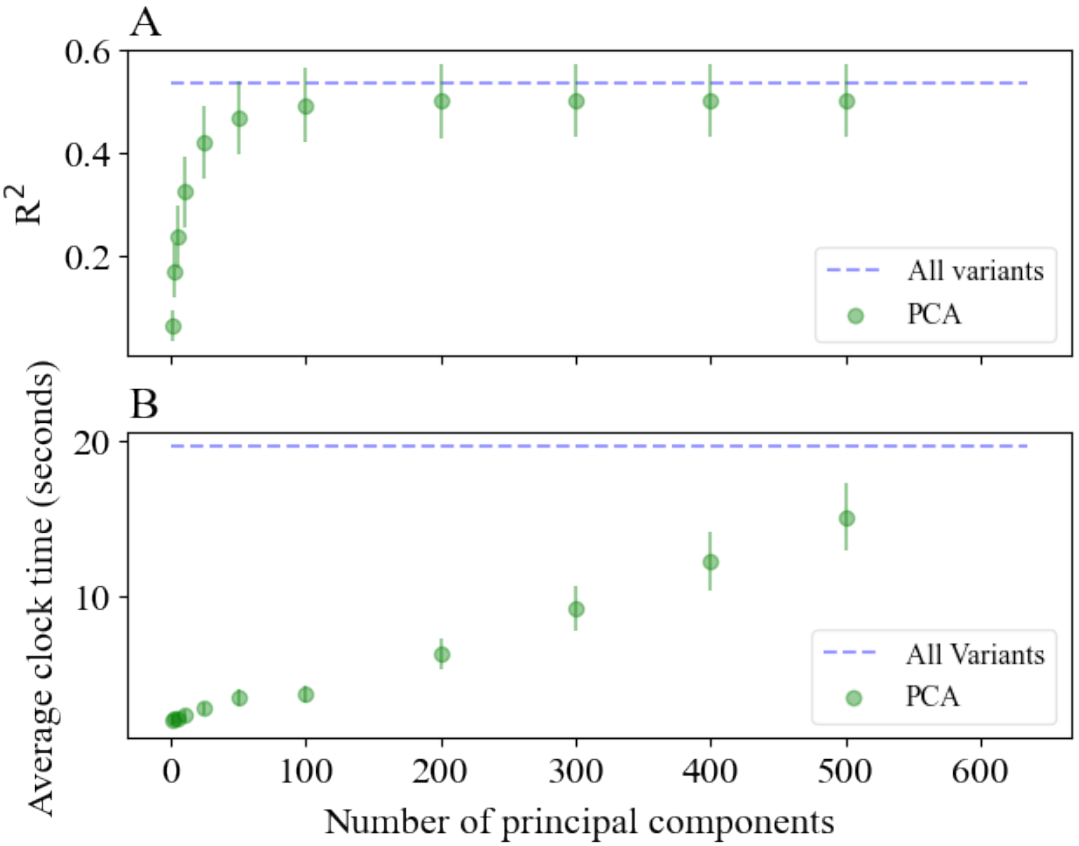

Figure 6.4: Comparison of the effectiveness and efficiency per the number of PCs in PCA. The dashed line displays the performance of the elastic net model when provided all genetic variants. Presented is the average R^2 score of each model (top figure), and the clock-based run time of 10-fold cross-validation averaged over all genes (bottom figure).

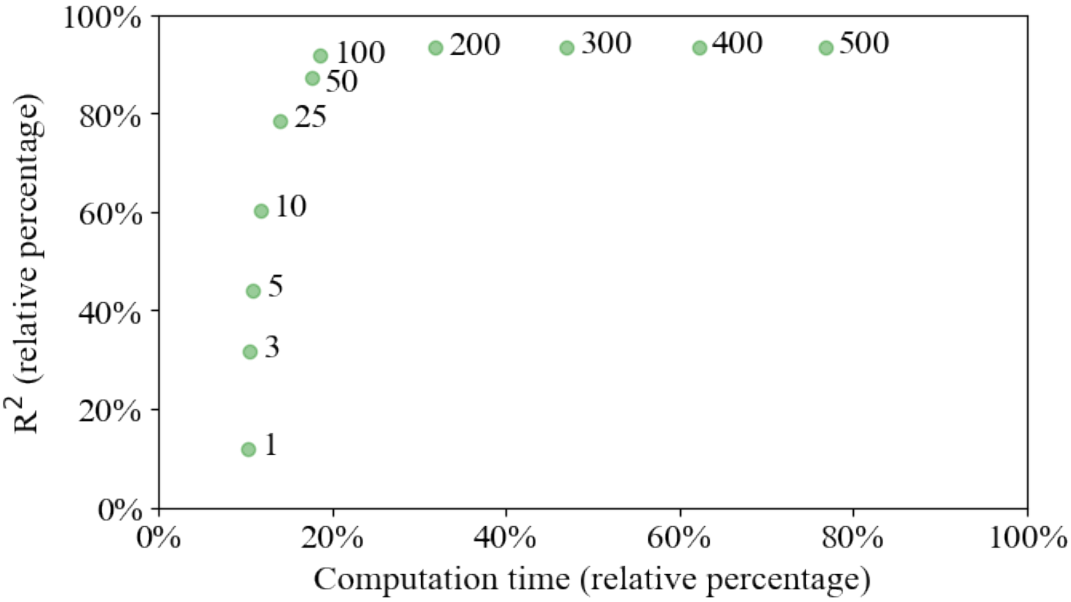

Figure 6.5: The relative effectiveness and efficiency of an elastic net model utilizing lower-dimensional representations of genetic variants based on varying number of PCs in PCA, compared to an elastic net utilizing all genetic variants.

3.66, 6.25, 9.20, 12.19, 15.05 seconds, respectively. The trade-off in computational timing to predictive accuracy relative to the elastic net's performance utilizing all genetic variants is provided in Figure 6.5. For example, when 100 PCs are used, the elastic net achieves 82% of R^2 in only 19% of the time relative to the elastic net utilizing all genetic variants.

6.5 Discussion and Future Work

In this study, we analyze the performance of dimensionality reduction and feature selection techniques to modify genetic variants to be utilized in an elastic net model for the prediction of tissue-specific, individual-level gene expression. The reduction of the input space for large-scale genetic analyses can have a significant impact on the effectiveness and efficiency of genetic models. Our results show that, in general, PCA performs most effectively than LD pruning, and utilizing a relatively low number of PCs (100) can achieve similar performance with significant improvements in computation time. These results are specific to the GTEx database, which was selected due to its high number and quality of samples; however, the trends of this analysis should generalize to any similarly sized gene expression data sets. However, more work is necessary to determine how PCA scales to tissues with a lower number of samples or to genes with a lower correlation between genetic variants and expression.

While previous studies have shown PCA to be effective at creating a lower-dimensional representation of genetic variant data (128; 237; 73; 15; 19), little work has been done to determine the impact of PCA in terms of computational time and predictive accuracy in gene expression prediction. Improvements in computational efficiency in gene expression prediction has significant implications. The GTEx database contains over 27,000 genes for more than 50 tissues; predictive models are routinely made to predict expression for each gene-tissue pair. In our study, on average, an elastic net requires 19.6 seconds to perform 10-fold cross-validation per gene, whereas performing PCA along with an elastic net that takes 100 of the resulting

PCs as input requires only 3.7 seconds per gene. Therefore, PCA may provide an effective method for reducing the computational burden of gene expression prediction.

Further, LD pruning is a variant reduction technique that works to remove highly correlated variants. It is often performed as a 'first step' in many genetic analyses, specifically polygenetic risk scores and genome-wide association studies (GWAS) (28). However, while LD pruning generates a representative subset of the genetic variant data, it does not optimize for specific variants' influence on gene expression. Therefore, as our results show, LD pruning does not as effectively reduce the input variables for the elastic net model, as PCA does. One major consideration is that LD pruning may remove variants that have a very strong biological effect on gene expression, but are not influentially correlated due to their scarcity in the population. Therefore, both LD pruning and PCA, utilizing a low number of PCs, may fail to recognize the influence of several low-level determinate variants that have a large aggregate influence on gene expression. Thus, further work is necessary to determine how PCA and LD pruning generalize to such genes.

Bibliography

[1] Adamiak-Giera, U., Jawień, W., Pierzchlińska, A., Białecka, M., Kobierski, J. D., Janus, T., and Gawrońska-Szklarz, B. (2021). Pharmacokinetics of levodopa and 3-o-methyldopa in Parkinson's ian patients treated with levodopa and ropinirole and in patients with motor complications. *Pharmaceutics*, 13(9):1395. 28, 39, 41

[2] Ahlrichs, C. and Lawo, M. (2013). Parkinson's disease motor symptoms in machine learning: A review. *arXiv preprint arXiv:1312.3825*. 50, 52

[3] Amjad, F., Bhatti, D., Davis, T. L., Oguh, O., Pahwa, R., Kukreja, P., Zamudio, J., and Metman, L. V. (2019). Current practices for outpatient initiation of levodopa-carbidopa intestinal gel for management of advanced Parkinson's disease in the United States. *Advances in Therapy*, 36:2233–2246. 48

[4] Amoozegar, S., Pooyan, M., and Roughani, M. (2019). Toward a closed-loop deep brain stimulation in Parkinson's disease using local field potential in parkinsonian rat model. *Medical Hypotheses*, 132:109360. 59

[5] Angeles, P., Tai, Y., Pavese, N., Wilson, S., and Vaidyanathan, R. (2017). Automated assessment of symptom severity changes during deep brain stimulation (DBS) therapy for Parkinson's disease. In *2017 International Conference on Rehabilitation Robotics (ICORR)*, pages 1512–1517. IEEE. 59

[6] Arav, Y. and Zohar, A. (2023). Model-based optimization of controlled release formulation of levodopa for Parkinson's disease. *medRxiv*, pages 2023–03. 28, 39, 41

[7] Barbeira, A. N., Dickinson, S. P., Bonazzola, R., Zheng, J., Wheeler, H. E., Torres, J. M., Torstenson, E. S., Shah, K. P., Garcia, T., Edwards, T. L., et al. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature communications*, 9(1):1–20. 88, 92

[8] Baston, C., Contin, M., Calandra Buonaura, G., Cortelli, P., and Ursino, M. (2016). A mathematical model of levodopa medication effect on basal ganglia in Parkinson's disease: an application to the alternate finger tapping task. *Frontiers in Human Neuroscience*, 10:280. 28, 39

[9] Baucum, M., Khojandi, A., Vasudevan, R., and Davis, R. (2022). Adapting reinforcement learning treatment policies using limited data to personalize critical care. *INFORMS Journal on Data Science*, 1(1):27–49. 26

[10] Baucum, M., Khojandi, A., Vasudevan, R., and Ramdhani, R. (2021). Optimizing patient-specific medication regimen policies using wearable sensors in Parkinson's disease. 28, 35, 45

[11] Baumgarten, C., Zhao, Y., Sauleau, P., Malrain, C., Jannin, P., and Haegelen, C. (2016). Image-guided preoperative prediction of pyramidal tract side effect in deep brain stimulation: proof of concept and application to the pyramidal tract side effect induced by pallidal stimulation. *Journal of Medical Imaging*, 3(2):025001–025001. 58

[12] Beitz, J. M. (2014). Parkinson's disease: a review. *Frontiers in Bioscience-Scholar*, 6(1):65–74. 23, 24, 66

[13] Belić, M., Bobić, V., Badža, M., Šolaja, N., urić-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. 50, 52

[14] Benabid, A. L. (2003). Deep brain stimulation for Parkinson's disease. *Current opinion in neurobiology*, 13(6):696–706. 52, 61

[15] Berglund, A. E., Welsh, E. A., and Eschrich, S. A. (2017). Characteristics and validation techniques for PCA-based gene-expression signatures. *International Journal of Genomics*, 2017. 101, 110

[16] Bermudez, C., Rodriguez, W., Huo, Y., Hainline, A. E., Li, R., Shults, R., D'Haese, P. D., Konrad, P. E., Dawant, B. M., and Landman, B. A. (2019). Towards machine learning prediction of deep brain stimulation (DBS) intra-operative efficacy maps. In *Medical Imaging 2019: Image Processing*, volume 10949, pages 528–534. SPIE. 65

[17] Biau, G. and Scornet, E. (2016). A random forest guided tour. *Test*, 25(2):197–227. 88, 90

[18] Biba, M., Esposito, F., Ferilli, S., Di Mauro, N., and Basile, T. M. A. (2007). Unsupervised discretization using kernel density estimation. In *IJCAI*, volume 7, pages 696–701. 51

[19] Bicciato, S., Luchini, A., and Di Bello, C. (2003). PCA disjoint models for multiclass cancer analysis using gene expression data. *Bioinformatics*, 19(5):571–578. 101, 110

[20] Blin, O., Ferrandez, A., Pailhous, J., and Serratrice, G. (1991). Dopa-sensitive and dopa-resistant gait parameters in Parkinson's disease. *Journal of the Neurological Sciences*, 103(1):51–54. 66

[21] Boring, M. J., Jessen, Z. F., Wozny, T. A., Ward, M. J., Whiteman, A. C., Richardson, R. M., and Ghuman, A. S. (2019). Quantitatively validating the efficacy of artifact suppression techniques to study the cortical consequences of deep brain stimulation with magnetoencephalography. *Neuroimage*, 199:366–374. 62

[22] Braybrook, M., O'Connor, S., Churchward, P., Perera, T., Farzanehfar, P., and Horne, M. (2016). An ambulatory tremor score for Parkinson's disease. *Journal of Parkinson's Disease*, 6(4):723–731. 3

[23] Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32. 9, 72

[24] Brozova, H., Barnaure, I., Alterman, R. L., and Tagliati, M. (2009). STN-DBS frequency effects on freezing of gait in advanced Parkinson disease. *Neurology*, 72(8):770–771. 67

[25] Buch, E., Boyle, N. G., and Belott, P. H. (2011). Pacemaker and defibrillator lead extraction. *Circulation*, 123(11):e378–e380. 26

[26] Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., and Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122. 90

[27] Burchiel, K. J., McCartney, S., Lee, A., and Raslan, A. M. (2013). Accuracy of deep brain stimulation electrode placement using intraoperative computed tomography without microelectrode recording. *Journal of neurosurgery*, 119(2):301–306. 49

[28] Calus, M. P. and Vandenplas, J. (2018). SNPrune: an efficient algorithm to prune large SNP array and sequence datasets based on high linkage disequilibrium. *Genetics Selection Evolution*, 50(1):1–11. 101, 111

[29] Camara, C., Subramaniyam, N. P., Warwick, K., Parkkonen, L., Aziz, T., and Pereda, E. (2019). Non-linear dynamical analysis of resting tremor for demand-driven deep brain stimulation. *Sensors*, 19(11):2507. 59

[30] Cardona, H. D. V., Orozco, A. A., and Alvarez, M. A. (2012). Multi-patient learning increases accuracy for subthalamic nucleus identification in deep brain

stimulation. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4341–4344. IEEE. 61

[31] Castaño-Candamil, S., Piroth, T., Reinacher, P., Sajonz, B., Coenen, V. A., and Tangermann, M. (2019). An easy-to-use and fast assessment of patient-specific DBS-induced changes in hand motor control in Parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(10):2155–2163. 58, 59

[32] Castrioto, A., Lozano, A. M., Poon, Y.-Y., Lang, A. E., Fallis, M., and Moro, E. (2011). Ten-year outcome of subthalamic stimulation in Parkinson disease: a blinded evaluation. *Archives of Neurology*, 68(12):1550–1556. 67

[Centers for Disease Control and Prevention] Centers for Disease Control and Prevention. Genes and obesity. https://www.cdc.gov/genomics/resources/diseases/obesity/obesedit.htm. Last accessed 1/4/22. 90

[34] Cerasa, A., Novellino, F., and Quattrone, A. (2016). Connectivity changes in Parkinson's disease. *Current Neurology and Neuroscience Reports*, 16:1–11. 52

[35] Charvin, D., Medori, R., Hauser, R. A., and Rascol, O. (2018). Therapeutic strategies for Parkinson disease: beyond dopaminergic drugs. *Nature Reviews Drug Discovery*, 17(11):804–822. 3

[36] Chaudhuri, K. R. and Schapira, A. H. (2009). Non-motor symptoms of Parkinson's disease: dopaminergic pathophysiology and treatment. *The Lancet Neurology*, 8(5):464–474. 56

[37] Chen, C., Kehtarnavaz, N., and Jafari, R. (2014). A medication adherence monitoring system for pill bottles based on a wearable inertial sensor. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4983–4986. IEEE. 26

[38] Chen, Y., Gong, C., Hao, H., Guo, Y., Xu, S., Zhang, Y., Yin, G., Cao, X., Yang, A., Meng, F., et al. (2019). Automatic sleep stage classification based on subthalamic local field potentials. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(2):118–128. 57

[39] Chollet, F. (2015). Keras. https://github.com/fchollet/keras. 91

[40] Choquet, H. and Meyre, D. (2011). Genetics of obesity: what have we learned? *Current genomics*, 12(3):169–179. 89, 90

[41] Chrabaszcz, A., Neumann, W.-J., Stretcu, O., Lipski, W. J., Bush, A., Dastolfo-Hromack, C. A., Wang, D., Crammond, D. J., Shaiman, S., Dickey, M. W., et al. (2019). Subthalamic nucleus and sensorimotor cortex activity during speech production. *Journal of Neuroscience*, 39(14):2698–2708. 57

[42] Consortium, G. et al. (2020). The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509):1318–1330. 101, 102

[43] Contin, M. and Martinelli, P. (2010). Pharmacokinetics of levodopa. *Journal of Neurology*, 257(2):253–261. 28, 35

[44] Costin, H. and Geman, O. (2013). Parkinson's disease prediction based on multistate Markov models. *International Journal of Computers Communications & Control*, 8(4):525–537. 26

[45] Curtze, C., Nutt, J. G., Carlson-Kuhta, P., Mancini, M., and Horak, F. B. (2015). Levodopa is a double-edged sword for balance and gait in people with Parkinson's disease. *Movement Disorders*, 30(10):1361–1370. 24, 66

[46] Da Costa, J. F. P., Alonso, H., and Roque, L. (2009). A weighted principal component analysis and its application to gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(1):246–252. 101

[47] Dahodwala, N., Karlawish, J., Siderowf, A., Duda, J., and Mandell, D. (2011). Delayed Parkinson's disease diagnosis among African-Americans: The role of reporting of disability. *Neuroepidemiology*, 36(3):150–154. 2

[48] Dams, J., Bornschein, B., Reese, J. P., Conrads-Frank, A., Oertel, W. H., Siebert, U., and Dodel, R. (2011). Modelling the cost effectiveness of treatments for Parkinson's disease: a methodological review. *Pharmacoeconomics*, 29:1025–1049. 26

[49] Darbin, O., Hatanaka, N., Takara, S., Kaneko, M., Chiken, S., Naritoku, D., Martino, A., and Nambu, A. (2020). Local field potential dynamics in the primate cortex in relation to parkinsonism reveled by machine learning: A comparison between the primary motor cortex and the supplementary area. *Neuroscience Research*, 156:66–79. 62

[50] Das, S., Trutoiu, L., Murai, A., Alcindor, D., Oh, M., De la Torre, F., and Hodgins, J. (2011). Quantitative measurement of motor symptoms in Parkinson's disease: A study with full-body motion capture data. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6789–6792. IEEE. 53

[51] Daskalaki, E., Diem, P., and Mougiakakou, S. (2013). Personalized tuning of a reinforcement learning control algorithm for glucose regulation. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3487–3490. IEEE. 27

[52] Davie, C. A. (2008). A review of Parkinson's disease. *British Medical Bulletin*, 86(1):109–127. 23, 24, 66

[53] De La Pava, I., Mejía, J., Álvarez-Meza, A., Álvarez, M., Orozco, A., and Henao, O. (2017). A hierarchical K-nearest neighbor approach for volume of tissue activated estimation. In *Progress in Pattern Recognition, Image Analysis,*

Computer Vision, and Applications: 21st Iberoamerican Congress, CIARP 2016, Lima, Peru, November 8–11, 2016, Proceedings 21, pages 125–133. Springer. 63

[54] de Oliveira, L. M., Barbosa, E. R., Aquino, C. C., Munhoz, R. P., Fasano, A., and Cury, R. G. (2019). Deep brain stimulation in patients with mutations in Parkinson's disease–related genes: a systematic review. *Movement Disorders Clinical Practice*, 6(5):359–368. 65

[55] Dembek, T. A., Reker, P., Visser-Vandewalle, V., Wirths, J., Treuer, H., Klehr, M., Roediger, J., Dafsari, H. S., Barbe, M. T., and Timmermann, L. (2017). Directional DBS increases side-effect thresholds—a prospective, double-blind trial. *Movement Disorders*, 32(10):1380–1388. 50

[56] Deo, R. C. (2015). Machine learning in medicine. *Circulation*, 132(20):1920–1930. 51

[57] Deuschl, G., Schade-Brittinger, C., Krack, P., Volkmann, J., Schäfer, H., Bötzel, K., Daniels, C., Deutschländer, A., Dillmann, U., Eisner, W., et al. (2006). A randomized trial of deep-brain stimulation for Parkinson's disease. *New England Journal of Medicine*, 355(9):896–908. 52

[58] Díaz-Uriarte, R. and De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC bioinformatics*, 7(1):1–13. 88, 98

[59] DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical science*, 11(3):189–228. 98

[60] Domingues, I., Amorim, J. P., Abreu, P. H., Duarte, H., and Santos, J. (2018). Evaluation of oversampling data balancing techniques in the context of ordinal classification. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE. 92

[61] Dorsey, E. R., George, B. P., Leff, B., and Willis, A. W. (2013a). The coming crisis: obtaining care for the growing burden of neurodegenerative conditions. *Neurology*, 80(21):1989–1996. 1, 2, 23

[62] Dorsey, E. R., Venkataraman, V., Grana, M. J., Bull, M. T., George, B. P., Boyd, C. M., Beck, C. A., Rajan, B., Seidmann, A., and Biglan, K. M. (2013b). Randomized controlled clinical trial of "virtual house calls" for Parkinson disease. *JAMA Neurology*, 70(5):565–570. 2, 23

[63] Dudani, S. A. (1976). The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*, (4):325–327. 51

[64] Eggert, K. M., Reese, J. P., Oertel, W. H., and Dodel, R. (2008). Cost effectiveness of pharmacotherapies in early Parkinson's disease. *CNS Drugs*, 22:841–860. 26

[65] Erro, R., Picillo, M., Vitale, C., Palladino, R., Amboni, M., Moccia, M., Pellecchia, M. T., and Barone, P. (2016). Clinical clusters and dopaminergic dysfunction in de-novo Parkinson disease. *Parkinsonism & Related Disorders*, 28:137–140. 20

[66] Faist, M., Xie, J., Kurz, D., Berger, W., Maurer, C., Pollak, P., and Lücking, C. (2001). Effect of bilateral subthalamic nucleus stimulation on gait in Parkinson's disease. *Brain*, 124(8):1590–1600. 67, 84

[67] Fang, X., Liu, C., and Jiang, Z. (2018). Reference values of gait using APDM movement monitoring inertial sensor system. *Royal Society Open Science*, 5(1):170818. 71, 74, 82

[68] Farrokhi, F., Buchlak, Q. D., Sikora, M., Esmaili, N., Marsans, M., McLeod, P., Mark, J., Cox, E., Bennett, C., and Carlson, J. (2020). Investigating risk factors and predicting complications in deep brain stimulation surgery with machine learning algorithms. *World Neurosurgery*, 134:e325–e338. 53

[69] Fasano, A., Daniele, A., and Albanese, A. (2012). Treatment of motor and non-motor features of Parkinson's disease with deep brain stimulation. *The Lancet Neurology*, 11(5):429–442. 56, 67

[70] Feigin, V. L., Abajobir, A. A., Abate, K. H., Abd-Allah, F., Abdulle, A. M., Abera, S. F., Abyu, G. Y., Ahmed, M. B., Aichour, A. N., Aichour, I., et al. (2017). Global, regional, and national burden of neurological disorders during 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Neurology*, 16(11):877–897. 1, 23

[71] Ferrarin, M., Lopiano, L., Rizzone, M., Lanotte, M., Bergamasco, B., Recalcati, M., and Pedotti, A. (2002). Quantitative analysis of gait in Parkinson's disease: a pilot study on the effects of bilateral sub-thalamic stimulation. *Gait & posture*, 16(2):135–148. 67, 84

[72] Foltynie, T., Zrinzo, L., Martinez-Torres, I., Tripoliti, E., Petersen, E., Holl, E., Aviles-Olmos, I., Jahanshahi, M., Hariz, M., and Limousin, P. (2011). MRI-guided STN DBS in Parkinson's disease without microelectrode recording: efficacy and safety. *Journal of Neurology, Neurosurgery & Psychiatry*, 82(4):358–363. 61

[73] Fujisawa, K., Shimo, M., Taguchi, Y.-H., Ikematsu, S., and Miyata, R. (2021). PCA-based unsupervised feature extraction for gene expression analysis of COVID-19 patients. *Scientific Reports*, 11(1):1–11. 100, 110

[74] Gallager, R. (2012). *Discrete stochastic processes*, volume 321. Springer Science & Business Media. 34

[75] Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., Eyler, A. E., Denny, J. C., Nicolae, D. L., Cox, N. J., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics*, 47(9):1091–1098. 88, 89, 92, 100, 101, 102

[76] George, R. S., Nutt, J., Burchiel, K., and Horak, F. (2010). A meta-regression of the long-term effects of deep brain stimulation on balance and gait in PD. *Neurology*, 75(14):1292–1299. 67

[77] Goetz, C., Koller, W., Poewe, W., Rascol, O., Sampaio, C., Brin, M., Lees, A. J., LeWitt, P., Lozano, A., Mizuno, Y., et al. (2002). Levodopa. *Movement Disorders*, 17(SUPPL. 4):S23–S37. 66

[78] Goetz, C. G., Poewe, W., Rascol, O., and Sampaio, C. (2005). Evidence-based medical review update: pharmacological and surgical treatments of Parkinson's disease: 2001 to 2004. *Movement Disorders: Official Journal of the Movement Disorder Society*, 20(5):523–539. 24

[79] Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R., et al. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders: Official Journal of the Movement Disorder Society*, 23(15):2129–2170. 2, 5, 24

[80] Golshan, H. M., Hebb, A. O., Hanrahan, S. J., Nedrud, J., and Mahoor, M. H. (2016). A multiple kernel learning approach for human behavioral task classification using STN-LFP signal. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1030–1033. IEEE. 60

[81] Golshan, H. M., Hebb, A. O., Hanrahan, S. J., Nedrud, J., and Mahoor, M. H. (2018a). A hierarchical structure for human behavior classification using STN local field potentials. *Journal of Neuroscience Methods*, 293:254–263. 60

[82] Golshan, H. M., Hebb, A. O., and Mahoor, M. H. (2020). LFP-Net: A deep learning framework to recognize human behavioral activities using brain STN-LFP signals. *Journal of Neuroscience Methods*, 335:108621. 60

[83] Golshan, H. M., Hebb, A. O., Nedrud, J., and Mahoor, M. H. (2018b). Studying the effects of deep brain stimulation and medication on the dynamics of STN-LFP signals for human behavior analysis. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4720–4723. IEEE. 60

[84] Graupe, D., Khobragade, N., Tuninetti, D., Basu, I., Slavin, K. V., and Verhagen Metman, L. (2018). Who may benefit from on-demand control of deep brain stimulation? Noninvasive evaluation of Parkinson patients. *Neuromodulation: Technology at the Neural Interface*, 21(6):611–616. 59

[85] Griffiths, R. I., Kotschet, K., Arfon, S., Xu, Z. M., Johnson, W., Drago, J., Evans, A., Kempster, P., Raghav, S., and Horne, M. K. (2012). Automated assessment of bradykinesia and dyskinesia in Parkinson's disease. *Journal of Parkinson's Disease*, 2(1):47–55. 3, 19, 29

[86] Grira, N., Crucianu, M., and Boujemaa, N. (2004). Unsupervised and semi-supervised clustering: a brief survey. *A review of machine learning techniques for processing multimedia content*, 1(2004):9–16. 51

[87] GTEx Consortium, Ardlie, K. G., Deluca, D. S., Segrè, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., Trowbridge, C. A., Maller, J. B., Tukiainen, T., et al. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, 348(6235):648–660. 89, 101, 102

[88] Guillén, P., Martinez-de Pison, F., Sanchez, R., Argáez, M., and Velázquez, L. (2011). Characterization of subcortical structures during deep brain stimulation utilizing support vector machines. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 7949–7952. IEEE. 62

[89] Guillén-Rondon, P. and Robinson, M. D. (2016). Deep brain stimulation signal classification using deep belief networks. In *2016 International Conference on*

Computational Science and Computational Intelligence (CSCI), pages 155–158. IEEE. 62

[90] Habets, J. G., Heijmans, M., Kuijf, M. L., Janssen, M. L., Temel, Y., and Kubben, P. L. (2018). An update on adaptive deep brain stimulation in Parkinson's disease. *Movement Disorders*, 33(12):1834–1843. 59

[91] Hansen, L. K. and Salamon, P. (1990). Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, 12(10):993–1001. 51

[92] Hariz, M. I. (2002). Safety and risk of microelectrode recording in surgery for movement disorders. *Stereotactic and functional neurosurgery*, 78(3-4):146–157. 61

[93] Hausdorff, J. M. (2009). Gait dynamics in Parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(2):026113. 56, 67

[94] Hauser, R., Russ, H., Haeger, D., Bruguiere-Fontenille, M., Müller, T., and Wenning, G. (2006). Patient evaluation of a home diary to assess duration and severity of dyskinesia in Parkinson disease. *Clinical Neuropharmacology*, 29(6):322–330. 2, 24

[95] Hebb, A. O., Zhang, J. J., Mahoor, M. H., Tsiokos, C., Matlack, C., Chizeck, H. J., and Pouratian, N. (2014). Creating the feedback loop: closed-loop neurostimulation. *Neurosurgery Clinics*, 25(1):187–204. 52

[96] Herrera, B. M., Keildson, S., and Lindgren, C. M. (2011). Genetics and epigenetics of obesity. *Maturitas*, 69(1):41–49. 89, 90

[97] Herron, J. A., Thompson, M. C., Brown, T., Chizeck, H. J., Ojemann, J. G., and Ko, A. L. (2016). Chronic electrocorticography for sensing movement intention and closed-loop deep brain stimulation with wearable sensors in an essential tremor patient. *Journal of Neurosurgery*, 127(3):580–587. 59

[98] Hill, W. and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, 38(6):226–231. 101

[99] Horne, M. K., McGregor, S., and Bergquist, F. (2015). An objective fluctuation score for Parkinson's disease. *PloS One*, 10(4):e0124522. 3, 24, 29

[100] Huo, W., Angeles, P., Tai, Y. F., Pavese, N., Wilson, S., Hu, M. T., and Vaidyanathan, R. (2020). A heterogeneous sensing suite for multisymptom quantification of Parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(6):1397–1406. 56

[101] Jiang, H., Zhang, J. J., Hebb, A., and Mahoor, M. H. (2013). Time-frequency analysis of brain electrical signals for behavior recognition in patients with Parkinson's disease. In *2013 Asilomar Conference on Signals, Systems and Computers*, pages 1843–1847. IEEE. 60

[102] Jones, H. and Rowland-Yeo, K. (2013). Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT: Pharmacometrics & Systems Pharmacology*, 2(8):1–12. 28, 36

[103] Jovanov, I., Nauman, M., Kumaravelu, K., Lesi, V., Zutshi, A., Grill, W. M., and Pajic, M. (2018). Learning-based control design for deep brain stimulation. In *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*, pages 349–350. IEEE. 63

[104] Kanneganti, K. and Simon, L. (2011). Two-compartment pharmacokinetic models for chemical engineers. *Chemical Engineering Education*, 45(2):101–105. 36

[105] Keijsers, N. L., Horstink, M. W., and Gielen, S. C. (2006). Ambulatory motor assessment in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, 21(1):34–44. 3

[106] Khawaldeh, S., Tinkhauser, G., Shah, S. A., Peterman, K., Debove, I., Nguyen, T. K., Nowacki, A., Lachenmayer, M. L., Schuepbach, M., Pollo, C., et al. (2020). Subthalamic nucleus activity dynamics and limb movement prediction in Parkinson's disease. *Brain*, 143(2):582–596. 63

[107] Khobragade, N., Tuninetti, D., and Graupe, D. (2018). On the need for adaptive learning in on-demand deep brain stimulation for movement disorders. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2190–2193. IEEE. 59

[108] Khodakarami, H., Ricciardi, L., Contarino, M. F., Pahwa, R., Lyons, K. E., Geraedts, V. J., Morgante, F., Leake, A., Paviour, D., De Angelis, A., et al. (2019). Prediction of the levodopa challenge test in Parkinson's disease using data from a wrist-worn sensor. *Sensors*, 19(23):5153. 58

[109] Khojandi, A., Shylo, O., Mannini, L., Kopell, B. H., and Ramdhani, R. A. (2017). Stratifying Parkinson's patients with STN-DBS into high-frequency or 60 hz-frequency modulation using a computational model. *Neuromodulation: Technology at the Neural Interface*, 20(5):450–455. 58

[110] Khosravi, M., Atashzar, S. F., Gilmore, G., Jog, M. S., and Patel, R. V. (2018). Electrophysiological signal processing for intraoperative localization of subthalamic nucleus during deep brain stimulation surgery. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 424–428. IEEE. 61

[111] Khosravi, M., Atashzar, S. F., Gilmore, G., Jog, M. S., and Patel, R. V. (2019). Unsupervised clustering of micro-electrophysiological signals for localization of subthalamic nucleus during DBS surgery. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 17–20. IEEE. 61

[112] Khosravi, M., Atashzar, S. F., Gilmore, G., Jog, M. S., and Patel, R. V. (2020). Intraoperative localization of STN during dbs surgery using a data-driven model. *IEEE Journal of Translational Engineering in Health and Medicine*, 8:1–9. 61

[113] Kim, J., Duchin, Y., Shamir, R. R., Patriat, R., Vitek, J., Harel, N., and Sapiro, G. (2019). Automatic localization of the subthalamic nucleus on patient-specific clinical MRI by incorporating 7 T MRI and machine learning: Application in deep brain stimulation. *Human Brain Mapping*, 40(2):679–698. 62

[114] Kim, Y., Suescun, J., Schiess, M. C., and Jiang, X. (2021). Computational medication regimen for Parkinson's disease using reinforcement learning. *Scientific reports*, 11(1):9313. 26

[115] Koch, M., Geraedts, V., Wang, H., Tannemaat, M., and Bäck, T. (2019). Automated machine learning for EEG-based classification of Parkinson's disease patients. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 4845–4852. IEEE. 57

[116] Kong, Y. and Yu, T. (2018). A deep neural network model using random forest to extract feature representation for gene expression data classification. *Scientific reports*, 8(1):1–9. 88, 98

[117] Kotschet, K., Johnson, W., McGregor, S., Kettlewell, J., Kyoong, A., O'Driscoll, D. M., Turton, A. R., Griffiths, R., and Horne, M. K. (2014). Daytime sleep in Parkinson's disease measured by episodes of immobility. *Parkinsonism & Related Disorders*, 20(6):578–583. 3

[118] Krack, P., Batir, A., Van Blercom, N., Chabardes, S., Fraix, V., Ardouin, C., Koudsie, A., Limousin, P. D., Benazzouz, A., LeBas, J. F., et al. (2003). Five-year follow-up of bilateral stimulation of the subthalamic nucleus in advanced Parkinson's disease. *New England Journal of Medicine*, 349(20):1925–1934. 67

[119] Krack, P., Limousin, P., Benabid, A., and Pollak, P. (1997). Chronic stimulation of subthalamic nucleus improves levodopa-induced dyskinesias in Parkinson's disease. *The Lancet*, 350(9092):1676. 67

[120] Kuhner, A., Schubert, T., Cenciarini, M., Wiesmeier, I. K., Coenen, V. A., Burgard, W., Weiller, C., and Maurer, C. (2017). Correlations between motor symptoms across different motor tasks, quantified via random forest feature classification in Parkinson's disease. *Frontiers in Neurology*, 8:607. 56

[121] Kuhner, A., Wiesmeier, I. K., Cenciarini, M., Maier, T. L., Kammermeier, S., Coenen, V. A., Burgard, W., and Maurer, C. (2020). Motion biomarkers showing maximum contrast between healthy subjects and Parkinson's disease patients treated with deep brain stimulation of the subthalamic nucleus. A pilot study. *Frontiers in Neuroscience*, 13:1450. 56

[122] Kumar, R., Lozano, A., Kim, Y., Hutchison, W., Sime, E., Halket, E., and Lang, A. (1998). Double-blind evaluation of subthalamic nucleus deep brain stimulation in advanced Parkinson's disease. *Neurology*, 51(3):850–855. 52

[123] Kurtzke, J. F. and Goldberg, I. D. (1988). Parkinsonism death rates by race, sex, and geography. *Neurology*, 38(10):1558–1558. 1

[124] Lanska, D. J. (1997). The geographic distribution of Parkinson's disease mortality in the United States. *Journal of the Neurological Sciences*, 150(1):63–70. 1

[125] Larmer, S., Sargolzaei, M., and Schenkel, F. (2014). Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across canadian dairy breeds. *Journal of Dairy Science*, 97(5):3128–3141. 101

[126] Lei, H., Tewari, A., and Murphy, S. (2017). An actor-critic contextual bandit algorithm for personalized mobile health interventions. *arXiv preprint arXiv:1706.09090*. 27

[127] LeMoyne, R., Mastroianni, T., Whiting, D., and Tomycz, N. (2020). Parametric evaluation of deep brain stimulation parameter configurations for Parkinson's disease using a conformal wearable and wireless inertial sensor system and machine learning. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3606–3611. IEEE. 59

[128] Lenz, M., Müller, F.-J., Zenke, M., and Schuppert, A. (2016). Principal components analysis and the reported low intrinsic dimensionality of gene expression microarray data. *Scientific Reports*, 6(1):1–11. 100, 110

[129] LeWitt, P. A. (2015). Levodopa therapy for Parkinson's disease: pharmacokinetics and pharmacodynamics. *Movement Disorders*, 30(1):64–72. 27, 28, 35

[130] Liaw, A., Wiener, M., et al. (2002). Classification and regression by random forest. *R news*, 2(3):18–22. 9, 51

[131] Ligaard, J., Sannæs, J., and Pihlstrøm, L. (2019). Deep brain stimulation and genetic variability in Parkinson's disease: a review of the literature. *NPJ Parkinson's Disease*, 5(1):18. 65

[132] Little, S., Pogosyan, A., Kuhn, A., and Brown, P. (2012). Beta band stability over time correlates with Parkinsonian rigidity and bradykinesia. *Experimental Neurology*, 236(2):383–388. 59

[133] Little, S., Pogosyan, A., Neal, S., Zavala, B., Zrinzo, L., Hariz, M., Foltynie, T., Limousin, P., Ashkan, K., FitzGerald, J., et al. (2013). Adaptive deep brain stimulation in advanced Parkinson disease. *Annals of neurology*, 74(3):449–457. 50, 59

[134] Livi, L., Sadeghian, A., and Sadeghian, H. (2016). Discrimination and characterization of Parkinsonian rest tremors by analyzing long-term correlations

and multifractal signatures. *IEEE Transactions on Biomedical Engineering*, 63(11):2243–2249. 59

[135] Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The genotype-tissue expression (GTEx) project. *Nature genetics*, 45(6):580–585. 88, 89, 100, 101

[136] Lu, C. W., Malaga, K. A., Chou, K. L., Chestek, C. A., and Patil, P. G. (2020). High density microelectrode recording predicts span of therapeutic tissue activation volumes in subthalamic deep brain stimulation for Parkinson disease. *Brain Stimulation*, 13(2):412–419. 62

[137] Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. 73

[138] Ma, S. and Dai, Y. (2011). Principal component analysis based methods in bioinformatics studies. *Briefings in Bioinformatics*, 12(6):714–722. 100

[139] Machado, A., Rezai, A. R., Kopell, B. H., Gross, R. E., Sharan, A. D., and Benabid, A.-L. (2006). Deep brain stimulation for Parkinson's disease: surgical technique and perioperative management. *Movement disorders: official journal of the Movement Disorder Society*, 21(S14):S247–S258. 49

[140] Maetzler, W., Klucken, J., and Horne, M. (2016). A clinical view on the development of technology-based tools in managing Parkinson's disease. *Movement Disorders*, 31(9):1263–1271. 3, 24

[141] Mamun, K., Mace, M., Lutman, M., Stein, J., Liu, X., Aziz, T., Vaidyanathan, R., and Wang, S. (2015). Movement decoding using neural synchronization and inter-hemispheric connectivity from deep brain local field potentials. *Journal of Neural Engineering*, 12(5):056011. 60

[142] Mancy, K., Suresh, G., and Vijayalakshmi, C. (2020). Implementation of a Markov model for the analysis of Parkinson's disease. In *First International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2019*, pages 481–493. Springer. 26

[143] Mariani, B., Jiménez, M. C., Vingerhoets, F. J., and Aminian, K. (2012). On-shoe wearable sensors for gait and turning assessment of patients with Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 60(1):155–158. 27

[144] Marras, C., Chaudhuri, K. R., Titova, N., and Mestre, T. A. (2020). Therapy of Parkinson's disease subtypes. *Neurotherapeutics*, 17:1366–1377. 3

[145] Maurer, A., Hanrahan, S., Nedrud, J., Hebb, A. O., and Papandreou-Suppappola, A. (2016). Suppression of neurostimulation artifacts and adaptive clustering of Parkinson's patients behavioral tasks using EEG. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 851–855. IEEE. 59

[146] Mayeux, R., Marder, K., Cote, L. J., Denaro, J., Hemenegildo, N., Mejia, H., Tang, M.-X., Lantigua, R., Wilder, D., Gurland, B., et al. (1995). The frequency of idiopathic Parkinson's disease by age, ethnic group, and sex in northern Manhattan, 1988–1993. *American Journal of Epidemiology*, 142(8):820–827. 1

[147] McAdams, B. H. and Rizvi, A. A. (2016). An overview of insulin pumps and glucose sensors for the generalist. *Journal of Clinical Medicine*, 5(1):5. 26

[148] Mikhaylova, A. V. and Thornton, T. A. (2019). Accuracy of gene expression prediction from genotype data with PrediXcan varies across and within continental populations. *Frontiers in Genetics*, 10:261. 88

[149] Mizutani, Y., Okada, Y., Ogawa, M., Hasegawa, T., and Nabeshima, T. (1995). Pharmacokinetic model of oral levodopa and role of carbidopa in Parkinson's patients. *Biological and Pharmaceutical Bulletin*, 18(12):1729–1737. 28, 39

[150] Mohammed, A., Bayford, R., and Demosthenous, A. (2020). A framework for adapting deep brain stimulation using Parkinsonian state estimates. *Frontiers in Neuroscience*, 14:499. 59

[151] Mohammed, A., Zamani, M., Bayford, R., and Demosthenous, A. (2015). Patient specific Parkinson's disease detection for adaptive deep brain stimulation. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1528–1531. IEEE. 60

[152] Mohammed, A., Zamani, M., Bayford, R., and Demosthenous, A. (2017). Toward on-demand deep brain stimulation using online Parkinson's disease prediction driven by dynamic detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(12):2441–2452. 60

[153] Moreau, C., Defebvre, L., Destée, A., Bleuse, S., Clement, F., Blatt, J.-L., Krystkowiak, P., and Devos, D. (2008). STN-DBS frequency effects on freezing of gait in advanced Parkinson disease. *Neurology*, 71(2):80–84. 67, 84

[154] Moro, E., Esselink, R., Xie, J., Hommel, M., Benabid, A., and Pollak, P. (2002). The impact on Parkinson's disease of electrical parameter settings in STN stimulation. *Neurology*, 59(5):706–713. 50

[155] Mu, J., Chaudhuri, K. R., Bielza, C., de Pedro-Cuesta, J., Larrañaga, P., and Martinez-Martin, P. (2017). Parkinson's disease subtypes identified from cluster analysis of motor and non-motor symptoms. *Frontiers in Aging Neuroscience*, 9:301. 20

[156] Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press. 51

[157] Nahab, F., Abu-Hussain, H., and Moreno, L. (2019). Evaluation of clinical utility of the Personal KinetiGraph® in the management of Parkinson disease. *Advances in Parkinson's Disease*, 8(03):42. xv, xvii, 4, 5, 11, 19, 27, 28, 30

[158] Neumann, J.-O., Giese, H., Biller, A., Nagel, A. M., and Kiening, K. (2015). Spatial distortion in MRI-guided stereotactic procedures: evaluation in 1.5-, 3-and 7-Tesla MRI scanners. *Stereotactic and Functional Neurosurgery*, 93(6):380–386. 61

[159] Neumann, W.-J., Turner, R. S., Blankertz, B., Mitchell, T., Kühn, A. A., and Richardson, R. M. (2019). Toward electrophysiology-based intelligent adaptive deep brain stimulation for movement disorders. *Neurotherapeutics*, 16:105–118. 59

[160] Niketeghad, S., Hebb, A. O., Nedrud, J., Hanrahan, S. J., and Mahoor, M. H. (2014). Single trial behavioral task classification using subthalamic nucleus local field potential signals. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3793–3796. IEEE. 60

[161] Nutt, J. G. (2008). Pharmacokinetics and pharmacodynamics of levodopa. *Movement Disorders: Official Journal of the Movement Disorder Society*, 23(S3):S580–S584. 24, 27, 28

[162] Okuma, Y. (2014). Freezing of gait and falls in Parkinson's disease. *Journal of Parkinson's disease*, 4(2):255–260. 66

[163] Olanow, C. W., Obeso, J. A., and Stocchi, F. (2006). Continuous dopamine-receptor treatment of Parkinson's disease: scientific rationale and clinical implications. *The Lancet Neurology*, 5(8):677–687. 2

[164] Oliveira, F. H. M., Machado, A. R., Andrade, A. O., et al. (2018). On the use of t-distributed stochastic neighbor embedding for data visualization and classification of individuals with Parkinson's disease. *Computational and Mathematical Methods in Medicine*, 2018. 56

[165] Ossig, C., Gandor, F., Bosredon, C., Fauser, M., Reichmann, H., Horne, M., et al. (2015). Correlation of objective measurement of motor states using a kinetograph and patient diaries in advanced Parkinson's disease. *PLoS One*. 29

[166] Othman, A. A. and Dutta, S. (2014). Population pharmacokinetics of levodopa in subjects with advanced Parkinson's disease: levodopa-carbidopa intestinal gel infusion vs. oral tablets. *British Journal of Clinical Pharmacology*, 78(1):94–105. xiii, 28, 39, 41, 42

[167] Ozturk, M., Telkes, I., Jimenez-Shahed, J., Viswanathan, A., Tarakad, A., Kumar, S., Sheth, S. A., and Ince, N. F. (2020). Randomized, double-blind assessment of LFP versus SUA guidance in STN-DBS lead implantation: a pilot study. *Frontiers in Neuroscience*, 14:611. 62

[168] Pahwa, R., Isaacson, S. H., Torres-Russotto, D., Nahab, F. B., Lynch, P. M., and Kotschet, K. E. (2018). Role of the Personal KinetiGraph in the routine clinical assessment of Parkinson's disease: recommendations from an expert panel. *Expert Review of Neurotherapeutics*, 18(8):669–680. 19, 21, 29

[169] Paliwal, S., Mosley, P. E., Breakspear, M., Coyne, T., Silburn, P., Aponte, E., Mathys, C., and Stephan, K. E. (2019). Subjective estimates of uncertainty during gambling and impulsivity after subthalamic deep brain stimulation for Parkinson's disease. *Scientific Reports*, 9(1):14795. 57

[170] Park, S.-C., Cha, J. H., Lee, S., Jang, W., Lee, C. S., and Lee, J. K. (2019). Deep learning-based deep brain stimulation targeting and clinical applications. *Frontiers in Neuroscience*, 13:1128. 65

[171] Patel, N., Plaha, P., O'sullivan, K., McCarter, R., Heywood, P., and Gill, S. (2003). MRI directed bilateral stimulation of the subthalamic nucleus in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(12):1631–1637. 49

[172] Patel, S., Lorincz, K., Hughes, R., Huggins, N., Growdon, J., Standaert, D., Akay, M., Dy, J., Welsh, M., and Bonato, P. (2009). Monitoring motor fluctuations

in patients with Parkinson's disease using wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*, 13(6):864–873. 27

[173] Paula, M. D., Acosta, G., and Martínez, E. (2015). On-line policy learning and adaptation for real-time personalization of an artificial pancreas. *Expert Systems with Applications*, 42(4):2234–2255. 27

[174] PedictDB Team (2017). GTEx v7 expression models. https://predictdb.org/post/2017/11/29/gtex-v7-expression-models/. 89, 95

[175] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011a). Scikit-learn: Machine learning in python. *the Journal of Machine Learning Research*, 12:2825–2830. 9, 72, 73, 146

[176] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011b). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830. 102, 105

[177] Phibbs, F. T., Arbogast, P. G., and Davis, T. L. (2014). 60-Hz frequency effect on gait in Parkinson's disease with subthalamic nucleus deep brain stimulation. *Neuromodulation: Technology at the Neural Interface*, 17(8):717–720. 84

[178] Prange, S., Metereau, E., and Thobois, S. (2019). Structural imaging in Parkinson's disease: new developments. *Current Neurology and Neuroscience Reports*, 19:1–13. 52

[179] PredictDB Team (2021). GTEx v8 models on eQTL and sQTL. https://predictdb.org/post/2021/07/21/gtex-v8-models-on-eqtl-and-sqtl/. 89, 95, 101, 102

[180] Przybyszewski, A. W., Szlufik, S., Habela, P., and Koziorowski, D. M. (2017a). Rough set rules determine disease progressions in different groups of Parkinson's patients. In *Pattern Recognition and Machine Intelligence: 7th International Conference, PReMI 2017, Kolkata, India, December 5-8, 2017, Proceedings 7*, pages 270–275. Springer. 58

[181] Przybyszewski, A. W., Szlufik, S., Habela, P., and Koziorowski, D. M. (2017b). Rules found by multimodal learning in one group of patients help to determine optimal treatment to other group of Parkinson's patients. In *Intelligent Information and Database Systems: 9th Asian Conference, ACIIDS 2017, Kanazawa, Japan, April 3–5, 2017, Proceedings, Part II 9*, pages 359–367. Springer. 58

[182] Pulliam, C. L., Heldman, D. A., Brokaw, E. B., Mera, T. O., Mari, Z. K., and Burack, M. A. (2017). Continuous assessment of levodopa response in Parkinson's disease using wearable motion sensors. *IEEE Transactions on Biomedical Engineering*, 65(1):159–164. 3

[183] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3):559–575. xxiii, 101, 103, 105

[184] Qi, Y. (2012). Random forest for bioinformatics. In *Ensemble machine learning*, pages 307–323. Springer. 88

[185] Quisel, T., Foschini, L., Zbikowski, S., and Juusola, J. (2019). The association between medication adherence for chronic conditions and digital health activity tracking: Retrospective analysis. *Journal of Medical Internet Research*, 21(3):e11486. 26

[186] Rajkomar, A., Dean, J., and Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14):1347–1358. 50

[187] Ramdhani, R. A., Khojandi, A., Shylo, O., and Kopell, B. H. (2018). Optimizing clinical assessments in Parkinson's disease through the use of wearable sensors and data driven modeling. *Frontiers in computational neuroscience*, 12:72. 3, 24, 27, 50, 52

[188] Ramdhani, R. A., Patel, A., Swope, D., and Kopell, B. H. (2015). Early use of 60 Hz frequency subthalamic stimulation in Parkinson's disease: a case series and review. *Neuromodulation: Technology at the Neural Interface*, 18(8):664–669. 67

[189] Rao, S. S., Hofmann, L. A., and Shakil, A. (2006). Parkinson's disease: Diagnosis and treatment. *American Family Physician*, 74(12):2046–2054. 28

[190] Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., and Yang, G.-Z. (2016). Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 21(1):4–21. 50

[191] Ricchi, V., Zibetti, M., Angrisano, S., Merola, A., Arduino, N., Artusi, C. A., Rizzone, M., Lopiano, L., and Lanotte, M. (2012). Transient effects of 80 Hz stimulation on gait in STN DBS treated PD patients: a 15 months follow-up study. *Brain Stimulation*, 5(3):388–392. 67

[192] Richardson, T. G., O'Nunain, K., Relton, C. L., and Smith, G. D. (2021). Harnessing whole genome polygenic risk scores to stratify individuals based on cardiometabolic risk factors and biomarkers at age 10 in the lifecourse. *Arteriosclerosis, Thrombosis, and Vascular Biology*, pages ATVBAHA–121. 91

[193] Ringnér, M. (2008). What is principal component analysis? *Nature Biotechnology*, 26(3):303–304. 100

[194] Rish, I. et al. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46. 51

[195] Rizzone, M., Ferrarin, M., Pedotti, A., Bergamasco, B., Bosticco, E., Lanotte, M., Perozzo, P., Tavella, A., Torre, E., Recalcati, M., et al. (2002). High-frequency electrical stimulation of the subthalamic nucleus in Parkinson's disease: kinetic and kinematic gait analysis. *Neurological Sciences*, 23:s103–s104. 67, 84

[196] Santiago, A., Langston, J. W., Gandhy, R., Dhall, R., Brillman, S., Rees, L., and Barlow, C. (2019). Qualitative evaluation of the Personal Kinetigraph TM movement recording system in a Parkinson's clinic. *Journal of Parkinson's Disease*, 9(1):207–219. 27

[197] Saunders-Pullman, R., Wang, C., Stanley, K., and Bressman, S. (2011). Diagnosis and referral delay in women with Parkinson's disease. *Gender Medicine*, 8(3):209–217. 2

[198] Schapira, A. H., Chaudhuri, K. R., and Jenner, P. (2017). Non-motor features of Parkinson disease. *Nature Reviews Neuroscience*, 18(7):435–450. 1

[199] Schüpbach, W., Chastan, N., Welter, M., Houeto, J., Mesnage, V., Bonnet, A., Czernecki, V., Maltete, D., Hartmann, A., Mallet, L., et al. (2005). Stimulation of the subthalamic nucleus in Parkinson's disease: a 5 year follow up. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(12):1640–1644. 67

[200] Schwesig, R., Leuchte, S., Fischer, D., Ullmann, R., and Kluttig, A. (2011). Inertial sensor based reference gait data for healthy subjects. *Gait & posture*, 33(4):673–678. 71

[201] Seppi, K., Weintraub, D., Coelho, M., Perez-Lloret, S., Fox, S. H., Katzenschlager, R., Hametner, E.-M., Poewe, W., Rascol, O., Goetz, C. G., et al. (2011). The Movement Disorder Society evidence-based medicine review update: treatments for the non-motor symptoms of Parkinson's disease. *Movement Disorders*, 26(S3):S42–S80. 56

[202] Severson, K. A., Chahine, L. M., Smolensky, L., Ng, K., Hu, J., and Ghosh, S. (2020). Personalized input-output hidden Markov models for disease progression modeling. In *Machine Learning for Healthcare Conference*, pages 309–330. PMLR. 26

[203] Shah, S. A., Tinkhauser, G., Chen, C. C., Little, S., and Brown, P. (2018). Parkinsonian tremor detection from subthalamic nucleus local field potentials for closed-loop deep brain stimulation. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2320–2324. IEEE. 58, 59, 60

[204] Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press. 51

[205] Shamir, R. R., Dolber, T., Noecker, A. M., Walter, B. L., and McIntyre, C. C. (2015). Machine learning approach to optimizing combined stimulation and medication therapies for Parkinson's disease. *Brain Stimulation*, 8(6):1025–1032. 58, 62

[206] Shamir, R. R., Duchin, Y., Kim, J., Patriat, R., Marmor, O., Bergman, H., Vitek, J. L., Sapiro, G., Bick, A., Eliahou, R., et al. (2019). Microelectrode recordings validate the clinical visualization of subthalamic-nucleus based on 7T magnetic resonance imaging and machine learning for deep brain stimulation surgery. *Neurosurgery*, 84(3):749–757. 62

[207] Shen, D., Wu, G., and Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248. 50

[208] Sidiropoulos, C., Walsh, R., Meaney, C., Poon, Y., Fallis, M., and Moro, E. (2013). Low-frequency subthalamic nucleus deep brain stimulation for axial symptoms in advanced Parkinson's disease. *Journal of Neurology*, 260:2306–2311. 67

[209] Skiadopoulos, A., Moore, E. E., Sayles, H. R., Schmid, K. K., and Stergiou, N. (2020). Step width variability as a discriminator of age-related gait changes. *Journal of Neuroengineering and Rehabilitation*, 17:1–13. 85

[210] Steigerwald, F., Matthies, C., and Volkmann, J. (2019). Directional deep brain stimulation. *Neurotherapeutics*, 16:100–104. 50

[211] Stuart, M., Wickramasinghe, C. S., Marino, D. L., Kumbhare, D., Holloway, K., and Manic, M. (2019). Machine learning for deep brain stimulation efficacy using dense array EEG. In *2019 12th International Conference on Human System Interaction (HSI)*, pages 143–150. IEEE. 62

[212] Suykens, J. A. and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9:293–300. 51

[213] Tahafchi, P., Molina, R., Roper, J. A., Sowalsky, K., Hass, C. J., Gunduz, A., Okun, M. S., and Judy, J. W. (2017). Freezing-of-gait detection using temporal, spatial, and physiological features with a support-vector-machine classifier. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2867–2870. IEEE. 59, 60

[214] Tan, P.-N., Steinbach, M., and Kumar, V. (2006). Introduction to data mining, Pearson education. *Inc., New Delhi*. 9

[215] Trevathan, J. K., Yousefi, A., Park, H. O., Bartoletta, J. J., Ludwig, K. A., Lee, K. H., and Lujan, J. L. (2017). Computational modeling of neurotransmitter release evoked by electrical stimulation: nonlinear approaches to predicting stimulation-evoked dopamine release. *ACS Chemical Neuroscience*, 8(2):394–410. 63

[216] Tseng, Z. H., Hayward, R. M., Clark, N. M., Christopher, G., Colburn, B. J., Ursell, P. C., Olgin, J. E., Hart, A. P., and Moffatt, E. (2015). Sudden death in patients with cardiac implantable electronic devices. *JAMA Internal Medicine*, 175(8):1342–1350. 26

[217] Tsipouras, M. G., Tzallas, A. T., Rigas, G., Bougia, P., Fotiadis, D. I., and Konitsiotis, S. (2010). Automated levodopa-induced dyskinesia assessment. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 2411–2414. IEEE. 3

[218] Ursino, M., Magosso, E., Lopane, G., Calandra-Buonaura, G., Cortelli, P., and Contin, M. (2020). Mathematical modeling and parameter estimation of levodopa motor response in patients with Parkinson's disease. *PloS One*, 15(3):e0229729. 28, 39

[219] Valsky, D., Blackwell, K. T., Tamir, I., Eitan, R., Bergman, H., and Israel, Z. (2020). Real-time machine learning classification of pallidal borders during deep brain stimulation surgery. *Journal of Neural Engineering*, 17(1):016021. 62

[220] Valsky, D., Marmor-Levin, O., Deffains, M., Eitan, R., Blackwell, K. T., Bergman, H., and Israel, Z. (2017). Stop! border ahead: A utomatic detection of subthalamic exit during deep brain stimulation surgery. *Movement Disorders*, 32(1):70–79. 62

[221] van Nuenen, B. F., Esselink, R. A., Munneke, M., Speelman, J. D., van Laar, T., and Bloem, B. R. (2008). Postoperative gait deterioration after bilateral subthalamic nucleus stimulation in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, 23(16):2404–2406. 67

[222] Van Rooden, S. M., Heiser, W. J., Kok, J. N., Verbaan, D., Van Hilten, J. J., and Marinus, J. (2010). The identification of Parkinson's disease subtypes using cluster analysis: a systematic review. *Movement Disorders*, 25(8):969–978. 8, 20

[223] Vargas Cardona, H. D., Álvarez, M. A., and Orozco, Á. A. (2018). Multi-task learning for subthalamic nucleus identification in deep brain stimulation. *International Journal of Machine Learning and Cybernetics*, 9:1181–1192. 61

[224] Vercruysse, S., Vandenberghe, W., Münks, L., Nuttin, B., Devos, H., and Nieuwboer, A. (2014). Effects of deep brain stimulation of the subthalamic nucleus on freezing of gait in Parkinson's disease: a prospective controlled study. *Journal of Neurology, Neurosurgery & Psychiatry*, 85(8):871–877. 67

[225] Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G., Do, R., et al. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The American Journal of Human Genetics*, 97(4):576–592. 101

[226] Volkmann, J. (2004). Deep brain stimulation for the treatment of Parkinson's disease. *Journal of clinical neurophysiology*, 21(1):6–17. 52

[227] Volkmann, J., Herzog, J., Kopper, F., and Deuschl, G. (2002). Introduction to the programming of deep brain stimulators. *Movement disorders: official journal of the Movement Disorder Society*, 17(S3):S181–S187. 50

[228] Vozikis, A. and Goulionis, J. E. (2009). Medical decision making for patients with Parkinson's disease under average cost criterion. *Australia and New Zealand health policy*, 6(1). 26

[229] Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A. N., Knowles, D. A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. *Nature genetics*, 51(4):592–599. 88, 101

[230] Walley, A. J., Asher, J. E., and Froguel, P. (2009). The genetic contribution to non-syndromic human obesity. *Nature Reviews Genetics*, 10(7):431–442. 89, 90

[231] Wan, K. R., Maszczyk, T., See, A. A. Q., Dauwels, J., and King, N. K. K. (2019). A review on microelectrode recording selection of features for machine learning in deep brain stimulation surgery for Parkinson's disease. *Clinical Neurophysiology*, 130(1):145–154. 52

[232] Watts, J., Allen, E., Mitoubsi, A., Khojandi, A., Eales, J., Jalali-Najafabadi, F., and Papamarkou, T. (2022). Adapting random forests to predict obesity-associated gene expression. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 4407–4410. IEEE. 100

[233] Watts, J., Khojandi, A., Vasudevan, R., and Ramdhani, R. (2020). Optimizing individualized treatment planning for Parkinson's disease using deep reinforcement learning. In *2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*, pages 5406–5409. IEEE. 26, 50

[234] Willis, A., Schootman, M., Evanoff, B., Perlmutter, J., and Racette, B. (2011). Neurologist care in Parkinson disease: A utilization, outcomes, and survival study. *Neurology*, 77(9):851–857. 2, 23

[235] Yao, L., Brown, P., and Shoaran, M. (2018). Resting tremor detection in Parkinson's disease with machine learning and Kalman filtering. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4. IEEE. 59, 60

[236] Yao, L., Brown, P., and Shoaran, M. (2020). Improved detection of parkinsonian resting tremor with feature engineering and kalman filtering. *Clinical Neurophysiology*, 131(1):274–284. 59, 60

[237] Yeung, K. Y. and Ruzzo, W. L. (2001). Principal component analysis for clustering gene expression data. *Bioinformatics*, 17(9):763–774. 100, 110

[238] Yock, A. D. and Kim, G.-Y. (2017). Using k-means clustering to determine the number and position of isocenters in MLC-based multiple target intracranial radiosurgery. *Journal of Applied Clinical Medical Physics*, 18(5):351–357. 9

[239] Yohanandan, S. A., Jones, M., Peppard, R., Tan, J. L., McDermott, H. J., and Perera, T. (2016). Evaluating machine learning algorithms estimating tremor

severity ratings on the Bain–Findley scale. *Measurement Science and Technology*, 27(12):125702. 56

[240] Zaker, N., Dutta, A., Maurer, A., Zhang, J. J., Hanrahan, S., Hebb, A. O., Kovvali, N., and Papandreou-Suppappola, A. (2014). Adaptive learning of behavioral tasks for patients with Parkinson's disease using signals from deep brain stimulation. In *2014 48th Asilomar Conference on Signals, Systems and Computers*, pages 208–212. IEEE. 60

[241] Zaker, N., Zhang, J. J., Hanrahan, S., Nedrud, J., and Hebb, A. O. (2015). A new approach for automated detection of behavioral task onset for patients with Parkinson's disease using subthalamic nucleus local field potentials. In *2015 49th Asilomar Conference on Signals, Systems and Computers*, pages 780–784. IEEE. 60

[242] Zhu, B., Farivar, M., and Shoaran, M. (2020). Resot: Resource-efficient oblique trees for neural signal classification. *IEEE Transactions on Biomedical Circuits and Systems*, 14(4):692–704. 63

[243] Zou, H., Banerjee, P., Leung, S. S. Y., and Yan, X. (2020). Application of pharmacokinetic-pharmacodynamic modeling in drug delivery: development and challenges. *Frontiers in Pharmacology*, 11:997. 28, 36, 39

Appendix

Appendix A

Clustering Performance

A.1 Cohort Analysis

A total of 1309 features were extracted and engineered from dyskinesia and bradykinesia time-series collected by PKG sensors using TSFresh (175). Demographic information (patient's age at visit 1, age at diagnosis, number of years experiencing PD symptoms, and gender) was also provided as features. The top ten most important features were included in the random forest analysis based on preliminary experiments. These parameters are presented in Table A.1. The Within Cluster Sum of Squares (WCSS) measurement is presented in Figure A.1. The WCSS measurement seeks to minimize the within-cluster variance applied in k-means clusters for the number of cluster selections.

Table A.1: Features included in the random forest analysis based on preliminary experiments.

Calculated Features	Description
Dyskinesia	
Dyskinesia ar coefficient k 10 coeff 2	Unconditional maximum likelihood of an autoregressive AR(k) process—coeff 2—k10
Dyskinesia fft coefficient coeff 39 attr "angle"	One-dimensional discrete fast Fourier transform —coeff 39—angle
Dyskinesia spkt welch density coeff 5	Cross power spectral density —coeff 5
Bradykinesia	
Bradykinesia agg autocorrelation f agg "mean" maxlag 40	Aggregation function fagg —mean maxlag—40
Bradykinesia agg linear trend f agg "mean"chunk len 50 att "rvalue"	Linear least-squares regression aggregated over chunks versus the sequence from 0 up to the number of chunks minus one—mean—chunk length 50—attribute rvalue
Bradykinesia agg linear trend f agg "min" chunk len 50 attr "slope"	Linear least-squares regression aggregated over chunks versus the sequence from 0 up to the number of chunks minus one—mean—chunk length 50—attribute slope
Bradykinesia fft coefficient coeff 23 attr "abs"	One-dimensional discrete fast Fourier transform—coeff 23—abs
Bradykinesia fft coefficient coeff 94 attr "angle"	One-dimensional discrete fast Fourier transform—coeff 94—angle
Bradykinesia fft coefficient coeff 76 attr "imag"	One-dimensional discrete fast Fourier transform—coeff 76—imag
Bradykinesia index mass quantile q 0.9	Relative index i where q% of the mass of the time series x lie left ofI-quantile 0.9

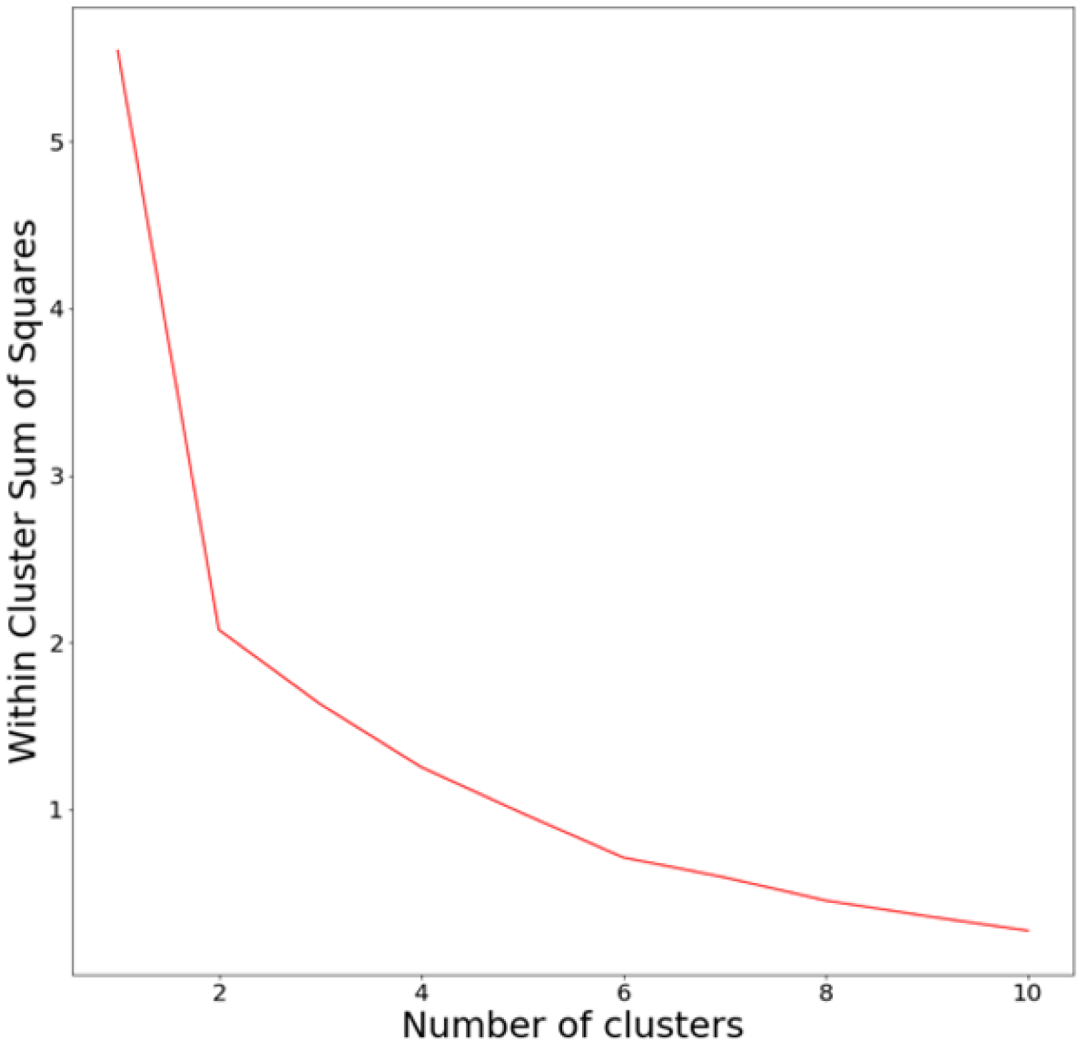

Figure A.1: WCSS as a function of the number of clusters when clustering patients based on their best symptom control medication regimens between the two visits according to PKG's summary dyskinesia and bradykinesia scores. The number of clusters is set to four, as a number towards the bottom of the "elbow" is considered best, but further increasing the number of clusters has been shown to reduce the model's robustness.

Vita

Jeremy Watts graduated from the University of Tennessee at Chattanooga in 2016 with a Bachelor of Science in Physics. He then received a Master of Science in Nuclear Engineering from the University of Tennessee in 2019, with a concentration in Radiological Engineering. He then began his PhD in Industrial Engineering at the University of Tennessee, graduating in 2023. His doctoral research focuses on the application of machine learning and decision sciences in the treatment of chronic disease, primarily Parkinson's disease.