U.PORTO

DOUTORAMENTO

CIÊNCIA DE DADOS DE SAÚDE

Predicting Health-Related Quality of Life in Parkinson's Disease: Interpretable Machine Learning Methods

Daniel Meneses Magano

D

2025

U.PORTO

Predicting Health-Related Quality of Life in Parkinson's Disease: Interpretable Machine Learning Methods

Programa Doutoral em Ciência de Dados de Saúde

Daniel Meneses Magano

Supervisor: Tiago Salgado De Magalhães Taveira-Gomes

Second Supervisor: António de Sousa Barros

2025

Tese apresentada à Faculdade de Medicina da Universidade do Porto para candidatura ao grau de DOUTOR em Ciência de Dados de Saúde.

I declare that this thesis, submitted to the Faculty of Medicine of the University of Porto within the scope of the Doctoral Program in Health Data Science, is my original work. Excluding clearly indicated quotations, no part of it has been reproduced or adapted from scientific publications or any other sources.

To Carolina Sousa, for her support, motivation, and understanding.

To my parents and brother for their unconditional support.

A tree doesn’t question the size of the sky; it grows toward it.

— Anonymous

Biographical Note

Daniel Magano earned his master's degree in Pharmaceutical Sciences from the University of Coimbra in 2017. During his studies, he conducted voluntary research (February–December 2016) at the university's Pharmaceutical Technology Laboratory, investigating topical drug delivery formulations, and completed a short research internship at Inserm U 1045 (France) in July 2016, focusing on the physiopathology of pulmonary hypertension. Daniel gained hands-on experience through two curricular internships: first at São Sebastião Pharmacy (January–April 2017), followed by a Quality Assurance internship at Overpharma (May–July 2017).

In April 2018, he joined BIAL as a Quality Control Trainee, working on procedures, Raman technology implementation, and laboratory compliance. From January 2019 to April 2022, Daniel worked in Compliance and Drug Safety at BIAL, managing the Pharmacovigilance System Master File, deviations, safety agreements, and coordinating audits. In May 2022, he was appointed Corporate Compliance Specialist while simultaneously serving as Medical Digital Strategy Manager (May 2022–April 2024), leading digital initiatives, implementing new systems, and coordinating Customer Relationship Management/ Omnichannel projects. Since May 2024, Daniel has held the position of Medical Excellence and Operations Manager, developing medical platforms, advancing AI initiatives, and ensuring a digital transformation on a global level.

In 2021, Daniel began his PhD in Health Data Science at the Faculty of Medicine, University of Porto, driven by his long-standing passion for digital innovation. He recognized this doctoral path as the ideal way to merge his enthusiasm for technology with his professional dedication to patient care, harnessing data and machine learning to advance health outcomes. His experience in compliance, pharmacovigilance, and digital transformation reflects a steadfast commitment to improving patient care through data-driven, cutting-edge solutions.

Acknowledgments

Pursuing this PhD was both an exploration into the unknown and a daily test of endurance. Countless late nights, minimal hours of sleep, and an ever-present sense of uncertainty. Yet, as I often reminded myself, “The fun is in the challenge”, and that mindset carried me through every twist and obstacle.

First and foremost, I am profoundly grateful to Carolina Sousa, for her unwavering support, motivation, and patience. Her understanding and encouragement provided the steadfast foundation that allowed me to keep pushing forward.

To my parents and brother, thank you for your unconditional support. Their belief in my abilities, and their words of reassurance were essential in keeping me motivated through the most challenging phases of this project.

I also want to acknowledge my friends for their camaraderie and constant motivation. Their presence and support turned even the toughest moments into occasions for laughter, for which I am truly grateful.

I owe special thanks to my supervisors for their guidance, teaching, and patience. Their mentorship shaped every stage of this thesis and enriched both the science and my own growth as a researcher.

To each of you, thank you for empowering me to turn this ambitious challenge into reality.

Abstract

Parkinson's Disease (PD) significantly impairs health-related quality of life (HRQoL), prompting the need for accurate, data-driven methods to identify at-risk patients and guide clinical management. This thesis explores three stages of machine learning approaches to predict HRQoL in PD, with the Parkinson's Disease Questionnaire-39 (PDQ-39) as the primary outcome measure. Widely used in both clinical and research settings, the PDQ-39 evaluates eight domains: Mobility, Activities of Daily Living, Emotional Well-Being, Stigma, Social Support, Cognitions, Communication, and Bodily Discomfort, offering a comprehensive picture of patients' quality of life.

Paper 1 explored whether classification-based models could effectively categorize patients according to the PDQ-39 Summary Index (the questionnaire's overall HRQoL score) by splitting individuals into high- versus low-score groups. As no clinically validated cut-off points exist for the PDQ-39, we employed an expectation-maximization algorithm to identify a threshold that best separated patients into two groups. Leveraging readily available clinical and demographic features, we trained and tested several classification algorithms. While these models demonstrated promising potential for differentiating patients, reliance on an inferred cut-off, rather than a clinically validated reference, ultimately constrained interpretability and practical clinical utility.

Reflecting on this limitation, Paper 2 shifted toward a regression model to predict the continuous PDQ-39 Summary Index rather than imposing artificial categories. By integrating explainable machine learning techniques, we achieved more detailed and clinically relevant predictions of HRQoL outcomes, revealing that features like an earlier age at diagnosis and the presence of dementia are closely linked with worse quality-of-life scores. However, some indications of potential overfitting were observed, leading us to question whether disaggregating the PDQ-39 into its eight domains and incorporating domain-specific predictors, might further refine our models and mitigate overfitting issues. This consideration naturally set the stage for the work undertaken in Paper 3.

Paper 3 built on these insights by individually modeling each of the PDQ-39's eight domains, thereby uncovering the unique predictors pertinent to each one of them. This domain-specific approach, reinforced by the inclusion of additional variables, mitigated overfitting and deepened our understanding of how different clinical features shape the distinct aspects of HRQoL in patients with PD. Notably, non-motor symptoms, measured using the Non-

Motor Symptoms Questionnaire, emerged as the most influential predictor across all domains, underscoring their crucial role in assessing the impact of the disease. Consequently, our results emphasize the importance of predictive models that account for each patient's specific characteristics.

These findings collectively highlight the potential of using machine learning to achieve more personalized care in patients with PD. By identifying key clinical features, especially non-motor symptoms, in both the overall HRQoL score and its individual domains, we enable more refined, patient-specific strategies. Ultimately, this study underscores how data-driven insights can refine clinical decision-making, helping clinicians address the challenges that most profoundly affect each patient's quality of life.

Keywords: Parkinson's Disease; health-related quality of life; machine learning; explainable machine learning

Resumo

A Doença de Parkinson (PD) afeta significativamente a qualidade de vida relacionada com a saúde (HRQoL), evidenciando a necessidade de métodos rigorosos, baseados em dados, para identificar doentes em risco e apoiar a gestão clínica. Esta tese explora três abordagens de machine learning para prever a HRQoL em PD, utilizando os resultados do Parkinson's Disease Questionnaire-39 (PDQ-39) como métrica principal. Amplamente utilizado tanto em contextos clínicos como em investigação, o PDQ-39 avalia oito domínios: mobilidade, atividades do quotidiano, bem-estar emocional, eStigma, apoio social, cognição, comunicação e desconforto corporal, fornecendo uma visão abrangente da qualidade de vida do doente.

O Artigo 1 explora o potencial de modelos de classificação para categorizar os doentes de acordo com o PDQ-39 Summary Index - a pontuação global de HRQoL do questionário - dividindo os indivíduos em grupos de elevada e baixa pontuação. Dado não existirem pontos de corte clinicamente validados para o PDQ-39, utilizámos um algoritmo de *expectation-maximization* para identificar o limiar que melhor separa os doentes em ambos os grupos. Focando-nos em variáveis clínicas e demográficas de fácil obtenção, treinámos e testámos diversos algoritmos de classificação. Embora estes modelos tenham demonstrado um potencial promissor para diferenciar doentes, a dependência de um limiar inferido, em vez de uma referência clinicamente validada, limita a interpretabilidade e a utilidade prática em ambiente clínico.

Considerando esta limitação, o Artigo 2 adota um paradigma de regressão para prever o PDQ-39 Summary Index, tratando-o como uma variável contínua, em contraste com a utilização de categorias artificiais. Integrando técnicas de *machine learning* explicável, obtivemos previsões dos resultados de HRQoL mais detalhadas e clinicamente relevantes, demonstrando que caraterísticas como o surgimento precoce da doença e a presença de demência estão fortemente ligadas a piores pontuações de qualidade de vida. No entanto, observámos sinais de um potencial *overfitting*, o que nos levou a questionar se a desagregação do PDQ-39 nos seus oito domínios e a incorporação de variáveis específicas em cada um destes, poderia melhorar os modelos e mitigar o *overfitting*. Esta reflexão despoletou, naturalmente, o trabalho desenvolvido no Artigo 3.

O artigo 3 baseou-se nestes dados, modelando individualmente cada um dos oito domínios do PDQ-39, identificando assim as variáveis relevantes em cada um destes. Esta

abordagem específica de cada domínio, reforçada pela inclusão de variáveis adicionais, atenuou o overfitting e aprofundou a nossa compreensão de como diferentes caraterísticas clínicas moldam aspectos distintos da HRQoL na PD. Importa notar que os sintomas não motores, avaliados através do Questionário de Sintomas Não Motores (Non-Motor Symptoms Questionnaire), surgiram como a variável preditiva mais influente em todos os domínios, realçando o seu papel crítico na avaliação do impacto da doença. Consequentemente, os nossos resultados destacam a importância de modelos preditivos que considerem o perfil específico de cada doente.

Estes resultados evidenciam o potencial da utilização de machine learning para obter cuidados mais personalizados na PD. Ao identificar as principais variáveis clínicas, particularmente os sintomas não motores, tanto na pontuação global de HRQoL como nos domínios individuais, obtemos estratégias mais adaptadas a cada doente. Em conclusão, este trabalho demonstra a forma como a informação baseada em dados pode auxiliar a tomada de decisão clínica, apoiando os profissionais de saúde nos desafios que afectam mais profundamente a qualidade de vida de cada doente.

Palavras-chave: Doença de Parkinson; qualidade de vida relacionada com a saúde; aprendizagem automática; aprendizagem automática explicável

Table of contents

List of Figures

List of Tables

Appendices

List of Abbreviations

Abbreviation	Full Term / Description
AI	Artificial Intelligence
ALE	Accumulated Local Effects (plot)
AUC	Area Under the Curve (ROC analysis)
CART	Classification and Regression Tree
CI	Confidence Interval
DBS	Deep Brain Stimulation
EM	Expectation-Maximization (algorithm)
FAMD	Factor Analysis of Mixed Data
gaussprPoly	Gaussian Process with Polynomial Kernel.
HCP	Health Care Provider
HRQoL	Health-Related Quality of Life
IQR	Interquartile Range
KNN	K-Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
MAE	Mean Absolute Error
MCCV	Monte Carlo Cross-Validation
MICE	Multivariate Imputation by Chained Equations
ML	Machine Learning
MNAR	Missing Not At Random
NMS	Non-Motor Symptoms
NMSQ	Non-Motor Symptoms Questionnaire

Abbreviation	Full Term / Description
nnet	Feedforward Neural Network
PD	Parkinson’s Disease
PDQ-39	Parkinson’s Disease Questionnaire-39
PDQ-39 SI	Parkinson’s Disease Questionnaire-39 Summary Index
PDQL	Parkinson’s Disease Quality of Life Scale
PRISM	Parkinson’s Real-world Impact assesSMent (study/database)
PROM(s)	Patient-Reported Outcome Measure(s)
R^2	Coefficient of Determination
RF	Random Forest
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristic (curve)
SD	Standard Deviation
SVM	Support Vector Machine
UK	United Kingdom
xgbTree	Extreme Gradient Boosting with Decision Trees

1. Introduction

Parkinson's Disease: A Multisystem Neurodegenerative Disorder

Parkinson's disease (PD) is one of the most prevalent and fastest-growing neurological disorders worldwide, currently affecting over 10 million people globally [1], [2], [3]. Once considered rare, its incidence has more than doubled between 1990 and 2015 and is projected to surpass 17 million by 2040, fueled by aging populations, increasing longevity, declining smoking rates, and the by-products of industrialization [4], [5]. The Global Burden of Disease study has identified PD as the fastest-growing neurological disorder by age-standardized rates of disability and mortality [3], [4], [5].

Clinically, PD is defined by a core set of motor symptoms, including resting tremor, rigidity, bradykinesia, and postural instability, which result from dopaminergic neurodegeneration in the substantia nigra pars compacta. Non-motor symptoms (NMS), including depression, constipation, hyposmia, REM sleep behavior disorder, pain, and cognitive changes, are often present years before motor symptoms appear and contribute significantly to the disease burden [6], [7]. NMS contribute substantially to disability and caregiver burden [3], [8].

The disease also exhibits significant heterogeneity. Whereas some individuals demonstrate relatively slow disease progression with predominant tremor symptoms, others experience symptoms such as early postural instability or cognitive impairment, which have been associated with a more rapid decline [9], [10], [11]. Although subtype classifications based on symptom dominance and disease trajectory are still being refined, they offer valuable insights into prognosis and contribute to a shift toward personalized treatment strategies [11], [12].

PD exhibits notable sex- and age-related variability in both its incidence and clinical manifestations. Men are more frequently affected than women, with incidence ratios consistently reported to be approximately 1.5 to 2 times higher [1], [13]. Female patients with PD are more likely to develop dyskinesias earlier in the disease course, exhibit a slower decline in Activities of Daily Living, and have a reduced risk of developing cognitive impairments than male patients [13]. Age is the most important known risk factor for Parkinson's disease, with its incidence and prevalence increasing steadily as individuals grow older [3]. While most cases occur after the age of 60, a subset of individuals develops

symptoms at a younger age. Parkinson's disease with onset before the age of 50 accounts for approximately 5–10% of cases and is often associated with mutations in genes such as PRKN and PINK1 [3]. These genetic forms tend to follow a distinct clinical profile, often marked by slower progression and sustained responsiveness to dopaminergic treatment [3].

PD imposes substantial socioeconomic and social burden worldwide. This impact stems largely from its progressive and debilitating nature, which significantly affects not only individuals with PD but also their families and broader communities. Caregiver burden, characterized by physical, emotional, social, and economic strain, becomes progressively more demanding as the disease advances, with neuropsychiatric symptoms such as anxiety, depression, apathy, mental fatigue, and psychosis significantly exacerbating this strain [14], [15]. Additionally, the societal costs associated with PD extend beyond healthcare expenditures, encompassing lost productivity and increased dependency, thus contributing substantially to the economic strain on healthcare systems and society as a whole [14], [15]. As populations continue to age and life expectancy rises, this burden is anticipated to increase sharply, highlighting the urgent need for sustainable strategies and supportive measures to mitigate these socioeconomic impacts [4].

The impact and challenges posed by PD on healthcare systems are significant and multifaceted, including escalating healthcare demands and resource allocation difficulties. The disease's chronic and progressive nature necessitates ongoing and often multidisciplinary management, involving neurologists, therapists, and various allied health professionals, which presents considerable resource and organizational challenges for healthcare systems [3]. Furthermore, the heterogeneity of PD symptom presentation and progression complicates patient care, underscoring the need for personalized treatment strategies and a robust infrastructure for long-term management [3]. Future challenges include addressing the anticipated increase in patient numbers, improving early diagnosis to enable timely interventions, developing and integrating disease-modifying therapies, and enhancing caregiver support structures. Leveraging advancements in genetic research, digital health technologies, and telemedicine may offer promising avenues for transforming PD care, making it more personalized, efficient, and accessible in the coming decades [3], [4].

Health-Related Quality of Life in Parkinson's Disease

Chronic diseases often limit an individual's functioning, resulting in reduced Health-Related Quality of Life (HRQoL). Consequently, quality of life has become an important metric for evaluating disease impact and treatment benefits and is now seen as a key outcome in chronic and neurodegenerative disease management. HRQoL captures aspects of the illness impact on daily life that clinical scales may not fully reflect [16], [17]. Over the past few decades, HRQoL has emerged as a standard outcome in PD research and clinical trials [18].

One of the most widely used PD-specific HRQoL instruments is the 39-item Parkinson's Disease Questionnaire (PDQ-39) [19]. This patient-reported metric was developed through in-depth interviews with PD patients and testing, condensing an initial 65-item pool into 39 items spanning eight domains [20], [21]. Validation studies have demonstrated that the PDQ-39 has high internal consistency, test–retest reliability, and construct validity [20], [21], [22], [23]. Shorter tools (e.g. the 8-item PDQ-8 derived from the PDQ-39) and other questionnaires like the Parkinson's Disease Quality of Life scale (PDQL) exist, but PDQ-39 remains one of the most comprehensive and well validated PD HRQoL tools [19].

The PDQ-39 assesses how frequently people with PD experience difficulties across eight dimensions of health [20]:

- Activities of Daily Living (6 items)
- Bodily Discomfort (3 items)
- Cognitions (4 items)
- Communication (3 items)
- Emotional Well-Being (6 items)
- Mobility (10 items)
- Social Support (3 items)
- Stigma (4 items)

Respondents rate each item on a 5-point Likert scale:

- 0 = Never
- 1 = Occasionally
- 2 = Sometimes
- 3 = Often
- 4 = Always

The individual item scores are summed within each domain, and then each domain score is transformed to a 0–100 scale, where 0 indicates no impact from the disease and 100 indicates the worst possible impact on quality of life. Additionally, an overall measure, the PDQ-39 Summary Index (PDQ-39 SI), can be calculated by averaging the scores of all eight domains to provide a global representation of HRQoL in patients with PD [20], [21].

In both clinical practice and research, the PDQ-39 provides valuable insights into patients' overall well-being and treatment outcomes. For example, PDQ-39 scores deteriorate significantly with advancing PD severity, especially in domains directly related to physical impairment such as Mobility and Activities of Daily Living [16]. Such information is vital for guiding holistic patient care, as HRQoL outcomes highlight needs beyond motor symptom control.

Rise of Machine Learning in Healthcare

Artificial intelligence (AI), particularly machine learning (ML), has emerged as a powerful and disruptive tool in healthcare, with the potential to fundamentally transform medical practice and care delivery [24]. In recent years, an explosion of healthcare data and improvement in computational power have driven a surge in ML applications aimed at addressing healthcare challenges. Notably, the integration of ML-driven tools into chronic disease management enhances the efficiency of patient care, optimizing treatment strategies, and even helping reduce costs [25]. This convergence of data and algorithms provides innovative solutions for managing long-term conditions, thereby introducing a new era of data-informed personalized medicine.

Key Applications of ML in Disease Management

ML methods have been applied to multiple aspects of disease management. Key application areas include diagnosis (especially early detection), prediction of disease progression, modeling and predicting treatment responses, and patient monitoring or symptom management:

- **Diagnosis and Early Detection:** ML algorithms can sift through complex biomedical data, from imaging scans to genomic profiles, to detect diseases earlier and more accurately than traditional methods. For example, in oncology, AI-based analysis of pathology and radiology data can identify novel tumor biomarkers and improve cancer screening and diagnosis [26]. In neurodegenerative diseases, ML improves the early detection of conditions that were historically difficult to diagnose in the prodromal stages. A deep learning model analyzing natural smartphone touchscreen typing patterns was able to distinguish early PD patients from healthy individuals, revealing subtle motor impairments that clinicians might miss during a brief exam [27]. Similarly, in Alzheimer's disease, machine learning applied to brain magnetic resonance imaging and other biomarkers can achieve high diagnostic accuracies for early-stage Alzheimer's vs. mild cognitive impairment, underlining how ML can surpass traditional diagnostic constraints in precision [28]. These advances show ML's promise in recognizing the "hidden" patterns of disease onset, enabling interventions at an earlier stage.
- **Disease Progression Prediction:** Another key application of ML is forecasting the progression of a chronic disease in an individual, which is crucial for the management of patients. ML models can analyze baseline clinical and biomarker data to stratify patients according to the likely disease trajectory. In Parkinson's disease, for instance, a data-driven analysis of longitudinal cohorts identified three distinct PD subtypes with different rates of progression (slow, moderate, and fast), and the ML model could predict five-year progression outcomes with relatively high accuracy [29]. Such model-identified subgroups help deconstruct PD's clinical heterogeneity and allow clinicians to anticipate which patients might decline more rapidly [29]. By forecasting disease progression, ML tools enable proactive monitoring and timely interventions tailored to the predicted course of the patient.
- **Treatment Response Modeling:** ML is increasingly employed to model and predict how patients will respond to specific treatments, which supports more personalized

and effective therapy decisions. Predictive models can help identify patients who are likely to benefit from an intervention, or which therapy options may yield the best outcomes. In Parkinson's disease, for example, researchers have developed an ML-based prediction tool to support surgical treatment decisions: a logistic regression model was trained on pre-surgical clinical features to predict which patients with PD would have a poor motor response to deep brain stimulation (DBS). This model identified "weak responders" to DBS, effectively flagging patients who failed to show significant motor improvement one year post-surgery [30]. Such a tool can assist clinicians in patient selection and counseling for DBS by estimating the likely benefits.

- **Symptom Monitoring and Management:** A major promise of ML in chronic disease management lies in continuous patient monitoring and symptom management, often via wearable sensors or smartphone-based tools. By coupling sensor data with ML, healthcare providers can obtain an ongoing, objective picture of a patient's condition outside of clinical visits. Indeed, *mobile health* technologies combined with ML have become a focal point in chronic care, enabling remote monitoring and telemedicine interventions [25]. In Parkinson's disease, this approach has led to the development of smartphone applications that assess symptoms and severity in real time. For example, researchers have developed a tool that uses a phone's microphone, accelerometer, and touchscreen to measure voice, gait, finger tapping, and balance; ML algorithms convert these inputs into a "mobile Parkinson's Disease score" reflecting overall symptom severity [31]. Such continuous tracking is especially valuable given PD's fluctuation in symptoms as it allows doctors to fine-tune medications based on rich data rather than brief snapshots. By empowering patients and clinicians with real-time data, these ML-powered systems facilitate timely adjustments to treatment and support better self-management.

Machine Learning in Parkinson's Disease and the Potential to Predict HRQoL

ML methods have significantly enhanced the management of PD by enabling patient stratification, predictions of disease progression, informed treatment decisions, and real-time symptom tracking through digital technologies [27], [30], [32], [33]. These advances facilitate more personalized and timely interventions, improving patient care and outcomes.

HRQoL has become a central focus in PD management, recognizing that the impact of the disease extends far beyond motor symptoms alone. HRQoL assessments capture comprehensive patient experiences, including emotional, social, cognitive, and physical domains, and provide valuable insights into disease burden and treatment efficacy [16], [18], [19]. Predicting HRQoL through ML has significant potential, as it would enable healthcare providers to proactively identify patients at risk of deterioration, implement timely interventions, and allocate resources more effectively. This predictive capability could ultimately lead to enhanced personalized care, optimized support systems, and improved overall management and quality of life for patients with PD [16], [18].

2. Objectives of the Thesis

The primary aim of this thesis is to explore and develop ML approaches to accurately predict HRQoL in PD, using the PDQ-39 as the key outcome measure. This work addresses three sequential objectives; each reflected in the individual studies comprising this thesis:

- The first study investigated whether classification-based ML models could effectively categorize patients according to their overall HRQoL, as measured by the PDQ-39 Summary Index. This involves identifying an optimal threshold that differentiates patients with higher and lower HRQoL scores, enabling clinicians to quickly recognize individuals at greater risk of experiencing significant disease impact on their quality of life.

- Given the limitations identified in the first study, particularly related to the interpretability and clinical utility of categorical predictions, the second study shifted towards employing regression-based ML models. This approach aims to predict the continuous PDQ-39 Summary Index, providing nuanced and clinically meaningful insights into HRQoL outcomes. Additionally, this study incorporated explainable ML techniques to identify key clinical and demographic variables associated with HRQoL deterioration, such as age at diagnosis and cognitive impairment.

- The third study builds upon the insights gained from the second, aiming to enhance predictive accuracy and clinical relevance by separately modeling each of the eight PDQ-39 domains. This domain-specific approach aims to uncover unique predictors for each aspect of HRQoL, with special emphasis on the impact of non-motor symptoms, thereby refining the model accuracy.

Overall, this thesis seeks to demonstrate the feasibility and clinical utility of ML models in predicting HRQoL, leveraging common clinical variables that are readily available in routine clinical practice. By identifying influential predictors and enhancing the precision of HRQoL assessments, this study aims to support personalized patient management strategies and improve clinical decision-making in Parkinson’s disease.

3. Applying machine learning methods to predict the Parkinson´s Disease Questionnaire-39 Summary Index

Extended Abstract

Applying machine learning methods to predict the Parkinson's Disease Questionnaire-39 Summary Index

Daniel Magano[1,2], Joana B. Moreira[1,3], Pedro P. Rodrigues[4]

[1] Ph.D. Program in Health Data Science, Faculty of Medicine, University of Porto, Porto, Portugal

[2] BIAL – Portela & Ca S.A., Coronado, Portugal

[3] Anaesthesiology, Emergency and Intensive Care Department, Centro Hospitalar Universitário do Porto, Largo do Prof. Abel Salazar, Porto, Portugal

[4] CINTESIS@RISE, MEDCIDS - Faculty of Medicine of the University of Porto, Porto, Portugal

Published in Journal of Statistics on Health Decision (Appendix 1)

https://doi.org/10.34624/jshd.v4i1.28783

3.1. Introduction

Health-related quality of life (HRQoL) gained importance in the last decades and is considered to be an important outcome measure in chronic diseases studies. HRQoL disease specific instruments reflect the consequences of that disease to a particular person and are sensitive to change in perceived HRQoL. In Parkinson's disease (PD) several disease specific HRQoL instruments have become available in the past few years. A 2002 systematic review of HRQoL instruments in PD concluded that in many situations the Parkinson´s Disease Questionnaire -39 (PDQ-39) is the most appropriate HRQoL instrument, because it is the scale that has been tested most thoroughly, has adequate clinimetric characteristics, has been used in the largest number of studies, and is available in many languages[34].

The PQD-39 is a 39-item self-report questionnaire, which assesses PD specific health related quality over the last month. It assesses how often patients experience difficulties across the 8 quality of life dimensions and the impact of PD on specific dimensions of functioning and well-being [21]. Its summary index (PDQ-39 SI) is a widely used patient-reported clinical trial endpoint. Peter Hagell and Maria H. Nilsson assessed the unidimensionality of the PDQ-39 and PDQ-39 SI using Rasch and confirmatory factor analysis and concluded its multidimensional nature [35].

The objective of this study is to do an exploratory analysis on the multidimensional nature of the PDQ-39 SI using machine learning methods to improve the interpretability of the score and obtain a model to predict the perceived quality of life of people with PD.

3.2. Methods

Data analysis was completed with R version 4.1.1. Data used in the preparation of this article were obtained from the Parkinson's Real-World Impact assesSMent (PRISM) database. PRISM study and database was funded by BIAL – Portela & C^{a}, S.A., designed in collaboration with The Cure Parkinson's Trust, an advocacy group based in the United Kingdom (UK), and reviewed by the PRISM steering committee.

The PRISM database contains data from 861 people from 5 European countries with PD collected in the context of an observational study with cross-sectional design [36].

PDQ-39 SI multivariable nature was first assessed using the Expectation-Maximization (EM) algorithm (Figure 1). Based on the latent classes, a transformation of PDQ-39 SI from continuous to binomial outcome was performed: mild vs severe symptoms impacting their quality of life.

Clinically relevant variables for modeling the PDQ-39 SI were initially selected from the PRISM database. Univariate logistic regressions were created with the PDQ-39 SI as a function of each of the previously selected variables. The variables presenting a p-value below 0.05 were ordered according to their impact on the odds ratio and included in the machine learning models by forward selection based on the resulting AUC values. Incomplete observations were then removed resulting in a dataset with 615 patients.

Before training the models, 20% of the dataset was saved to be used as independent validation data. With the remaining 80% of the dataset (training set), several models were trained to predict the classes obtained through EM. A repeated 10-fold cross validation was performed on this training set to estimate the respective ROC curves (Figure 2). Additionally, the performance of each model obtained with the 20% independent validation set was also assessed through the plotting of ROC curves (Figure 3).

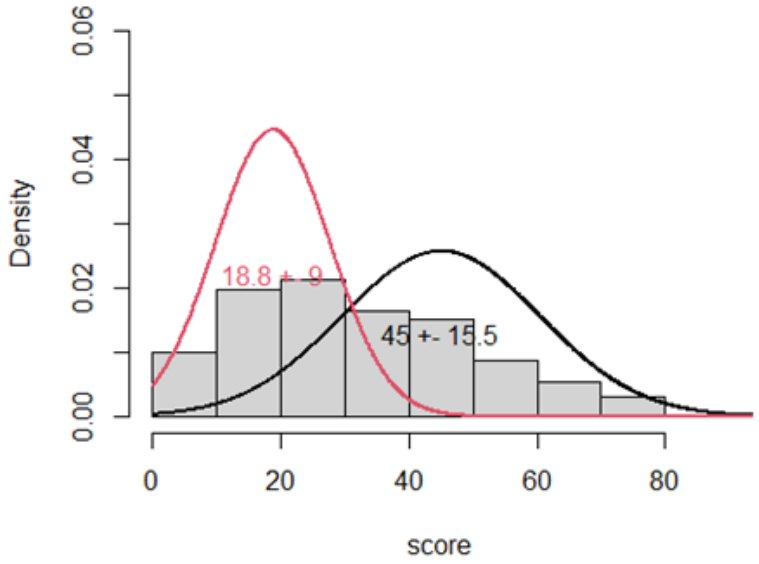

Figure 1: Latent classes on Parkinson´s Disease Summary Index. With the use of Expectation Maximization algorithm (2 kernels) 2 latent classes are described: (i) red distribution with mean of 18.8 and standard deviation of ± 9, and (ii) black distribution with mean of 45 and standard deviation of 15.5. Interception of the two distributions at the 31 value.

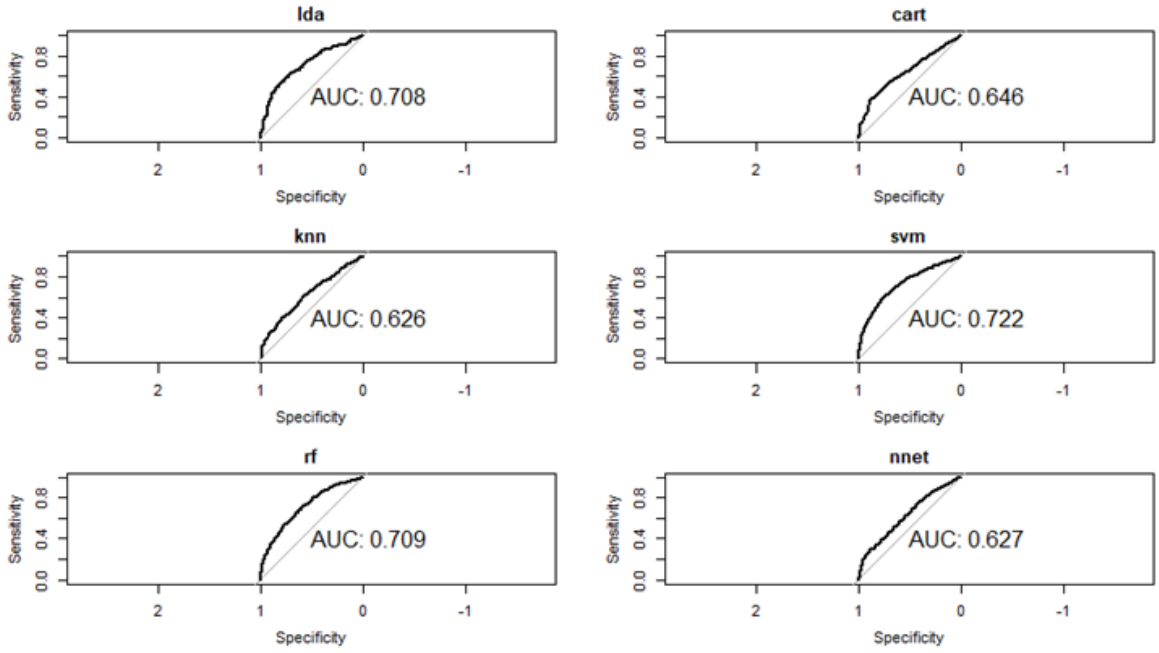

Figure 2: Performance estimation of different machine learning methods for PDQ-39 SI modeling on the training set using repeated 10-fold cross validation. ROC curves (receiver operating characteristic curves) of the repeated 10-fold cross validation training set (80% of the dataset) with (i) lda (Linear Discriminant Analysis), (ii) cart (Classification and Regression Tree, splitting index: gini), (iii) knn (K-nearest Neighbors, K=3), (iv) svm (Support Vector Machine, kernel: radial basis function), (v) rf (Random Forest, number of trees = 500), (vi) nnet (Feedforward Neural Network, 1 hidden layer with 5 neurons). AUC (area under the curve).

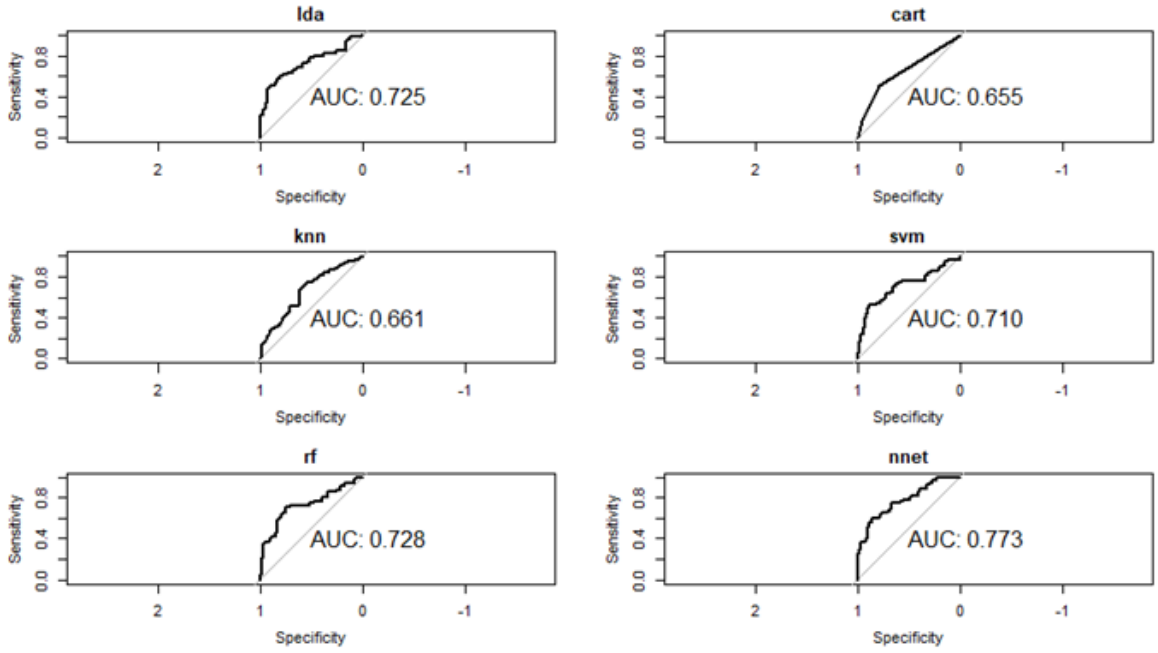

Figure 3: Performance of different machine learning methods for PDQ-39 SI modeling on the independent validation set. ROC curves (receiver operating characteristic curves) of the independent validation set (20% of the dataset) with (i) lda (Linear Discriminant Analysis), (ii) cart (Classification and Regression Tree, splitting index: gini), (iii) knn (K-nearest Neighbors, K=3), (iv) svm (Support Vector Machine, kernel: radial basis function), (v) rf (Random Forest, number of trees = 500), (vi) nnet (Feedforward Neural Network, 1 hidden layer with 5 neurons). AUC (area under the curve).

3.3. Results

The PDQ-39 SI was converted from a continuous outcome into two categories based on the impact on the quality of life of the person with PD: mild (with a PDQ-39 SI below 31) and severe (PDQ-39 SI above 31).

The ROC curves obtained with repeated cross-validation for each of the trained models are presented in Figure 2 while the ROC curves obtained with the independent validation data set are presented in Figure 3.

3.4. Discussion

The obtention of a binomial classification for the PDQ-39 SI can facilitate its interpretation and harmonize its utilization by different health care providers (HCP). Additionally, machine learning models applied to clinical variables can be used to predict to which of these classes a person with Parkinson's disease would belong. This information would allow HCPs to predict which of the patients are expected to have a lower quality of life.

The PDQ-39 SI is based on the patient's interpretation about several dimensions measuring his/her perceived quality of life. The subjectivity of the information collected through a survey is considered a limitation of this study.

The results obtained demonstrate a good performance of Linear Discriminant Analysis, Support Vector Machine and Random Forest methods. Future work will focus on hyperparameters optimization and on studying a wider range of variables to accommodate possible confounding factors.

4. Predicting Quality of Life in Parkinson's Disease: A Machine Learning Approach Employing Common Clinical Variables

Original Article

Predicting Quality of Life in Parkinson's Disease: A Machine *Learning Approach Employing Common Clinical Variables*

Daniel Magano [1,2] , Tiago Taveira-Gomes [3,4,5] , João Massano [6,7] and António S. Barros [8]

[1] Ph.D. Program in Health Data Science, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal

[2] Medical Department, BIAL-Portela & Cª., S.A., 4745-457 São Mamede do Coronado, Portugal

[3] Department of Community Medicine, Information and Decision in Health, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal

[4] Faculty of Health Sciences, University Fernando Pessoa, 4200-150 Porto, Portugal

[5] SIGIL Scientific Enterprises, 4076 Dubai, United Arab Emirates

[6] Department of Clinical Neurosciences and Mental Health, Faculty of Medicine University of Porto, 4200-319 Porto, Portugal

[7] Department of Neurology, Centro Hospitalar Universitário de São João, 4200-319 Porto, Portugal

[8] Department of Surgery and Physiology, Cardiovascular R&D Centre-UnIC@RISE, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal

Published in Journal of Clinical Medicine (Appendix 2)

https://doi.org/10.3390/jcm13175081

4.1. Abstract

Background: Parkinson's Disease significantly impacts health-related quality of life, with the Parkinson's Disease Questionnaire-39 extensively used for its assessment. However, predicting such outcomes remains a challenge due to the subjective nature and variability in patient experiences. This study develops a machine learning model using accessible clinical data to enable predictions of life-quality outcomes in Parkinson's Disease and utilizes explainable machine learning techniques to identify key influencing factors, offering actionable insights for clinicians.

Methods: Data from the Parkinson's Real-world Impact Assessment study (PRISM), involving 861 patients across six European countries, were analyzed. After excluding incomplete data, 627 complete observations were used for the analysis. An ensemble machine learning model was developed with a 90% training and 10% validation split.

Results: The model demonstrated a Mean Absolute Error of 4.82, a Root Mean Squared Error of 8.09, and an R^2 of 0.75 in the training set, indicating a strong model fit. In the validation set, the model achieved a Mean Absolute Error of 11.22, a Root Mean Squared Error of 13.99, and an R^2 of 0.36, showcasing moderate variation. Key predictors such as age at diagnosis, patient's country, dementia, and patient's age were identified, providing insights into the model's decision-making process.

Conclusions: This study presents a robust model capable of predicting the impact of Parkinson's Disease on patients' quality of life using common clinical variables. These results demonstrate the potential of machine learning to enhance clinical decision-making and patient care, suggesting directions for future research to improve model generalizability and applicability.

Keywords: Parkinson's disease; health-related quality of life; predictive modeling; machine learning; Parkinson's disease questionnaire-39; PDQ-39 summary index; explainable machine learning

4.2. Introduction

Parkinson's Disease (PD) global prevalence more than doubled between 1990 and 2015 [5]. The motor symptoms associated with the second most common neurodegenerative disorder include resting tremor, rigidity, bradykinesia, and postural instability. Additionally, non-motor symptoms such as cognitive impairment, mental health disorders, sleep disturbances, pain, and sensory disturbances are also common in PD [5], [37], [38], [39]. All these symptoms decrease the health-related quality of life (HRQoL) of people with PD [34].

In recent years, HRQoL has emerged as a critical outcome measure in the study of chronic diseases. HRQoL instruments, particularly those designed for specific diseases, offer valuable information on the personal effects of disease and any changes that are perceived over time. For example, in the context of PD, several disease-specific HRQoL tools have been developed recently. A comprehensive, systematic review conducted in 2002 determined that the Parkinson's Disease Questionnaire-39 (PDQ-39) was the most suitable instrument for assessing HRQoL in the context of PD. The PDQ-39 was chosen based on several factors, including its extensive testing, robust clinimetric properties, widespread use in numerous studies, and availability in numerous languages [34]. Additionally, recent studies have demonstrated its utility in various contexts, such as assessing the impact of deep brain stimulation of the subthalamic nucleus on the quality of life of PD patients [40] and validating its use in PD patients with cognitive impairment, showcasing its versatility and applicability across different patient populations [22]. Furthermore, the Italian version of the PDQ-39 has been validated, highlighting its cross-cultural applicability [23]. Research suggests that the PDQ-39 is a reliable and valid tool for assessing HRQoL in PD, as it provides a comprehensive view of the disease's impact, extending beyond traditional clinical outcomes, and reflects the patient's perspective, capturing their personal and subjective experience of living with the disease [41].

PDQ-39 is a commonly utilized self-report questionnaire comprised of 39 items that assess the degree of difficulty experienced by individuals diagnosed with PD across eight dimensions of HRQoL, namely mobility, activities of daily living, emotional well-being, stigma, social support, cognitions, communication, and bodily discomfort [21]. The PDQ-39 Summary Index (PDQ-39 SI), which measures the overall impact of PD on a patient's life, is a commonly utilized patient-reported endpoint in clinical trials [21], [42], [43]. Despite its widespread use, accurately predicting the PDQ-39 SI and, consequently, the HRQoL remains a complex and challenging task. This complexity is partly due to the subjective

nature of the questionnaire, which captures individual patient experiences and perceptions that may vary widely based on personal factors, such as disease severity, duration, treatment response, individual resilience, and varying interpretations of quality of life.

Machine and statistical learning have been used in PD to predict disease progression, motor symptoms, and treatment response. For instance, Chandrabhatla et al. explored the co-evolution of machine learning (ML) and digital technologies to enhance the monitoring of Parkinson's Disease motor symptoms, thereby improving real-time symptom tracking and management [44]. Gao et al. employed both model-based and model-free ML techniques for the diagnostic, prediction, and classification of clinical outcomes in PD, highlighting their potential in early diagnosis and personalized treatment planning [45]. Nilashi et al. evaluated the effectiveness of ensemble ML methods for predicting PD progression, demonstrating the improved accuracy and robustness of combined models [46]. Dadu et al. focused on identifying and predicting PD subtypes and progression using ML in two cohorts, which can aid in precise disease classification and management [29]. Additionally, a recent study introduced a Perceiver architecture-based multimodal ML model to estimate PD symptom severity by analyzing raw gait signals and features from Ground Reaction Force sensors [47]. However, its application in predicting HRQoL, as measured by the PDQ-39 Summary Index (SI), remains underexplored. This study sought to develop an ML model capable of predicting PDQ-39 SI in patients with PD using readily available clinical variables, such as the patient's current PD medication, co-existing medical conditions, education level, age, and age at diagnosis. The objective of this research is to develop a ML model equipped with explainable ML techniques to enhance clinicians' ability to predict HRQoL in PD patients. This tool is designed to provide deeper insights into the factors influencing HRQoL, thereby informing treatment decisions and ultimately improving patient care. Additionally, the explainable ML component of this study offers critical information about the variables that have a higher impact on quality of life, which can be valuable for clinicians in understanding disease progression and management.

4.3. Materials and Methods

Database and Data Preparation

The data for this study were sourced from the Parkinson's Real-world Impact assesSMent (PRISM) study, a cross-sectional observational research initiative. A total of 861 patients with PD and 256 caregivers from six European countries participated in the PRISM study, completing an online survey using structured questionnaires including the PDQ-39, Non-Motor Symptoms Questionnaire, and Zarit Burden Interview [36]. This survey captured a broad range of information, including medication use, comorbidities, and healthcare resource utilization [36]. After excluding two observations due to missing PDQ-39 SI scores, 859 valid entries remained for analysis. The demographic profile of the participant pool was nearly evenly split by gender: 48.5% female and 50.3% male, with a mean age of 65.0 years (Standard Deviation (SD) = 10.2 years). The average PDQ-39 SI score was 32.0 (SD = 18.2), with its distribution depicted in Figure 4. Supplementary Table S1 summarizes key variables from the PRISM dataset.

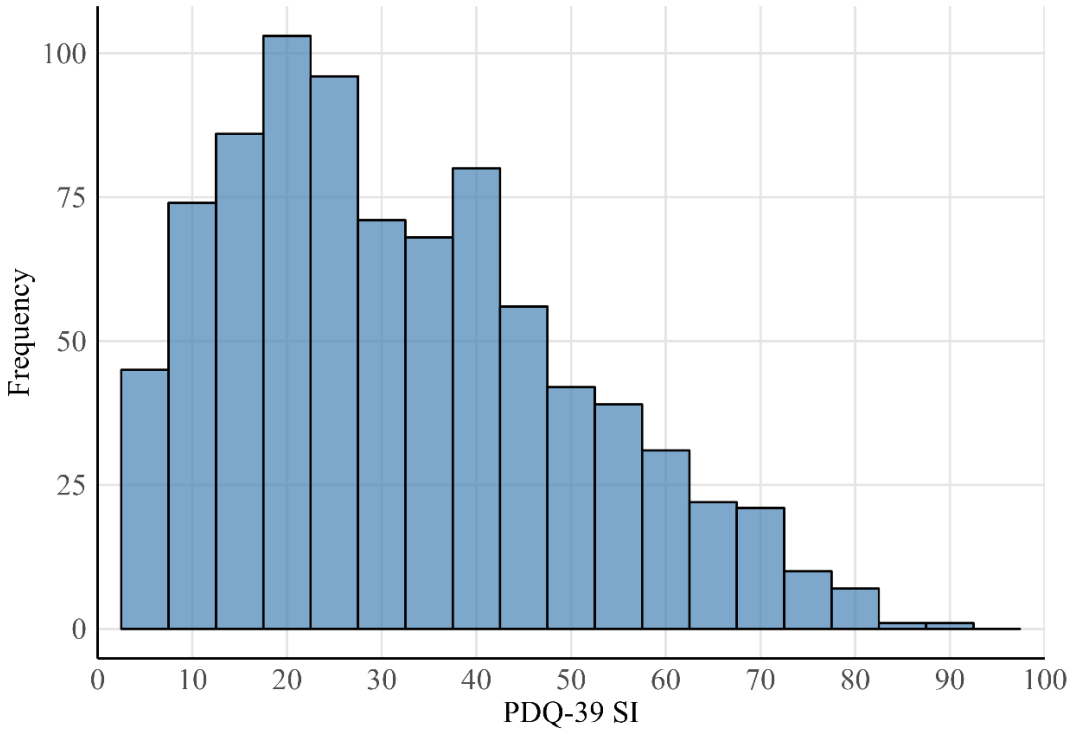

Figure 4: Distribution of PDQ-39 SI across 859 patients from the PRISM study.

Data Analysis

All analyses were performed using the R statistical software (version 4.3.0) [48]. As part of our analytical approach, we utilized Factor Analysis of Mixed Data (FAMD) to explore the underlying relationships between variables within the PRISM dataset, including the PDQ-39 SI. FAMD are particularly relevant for datasets that include both quantitative and qualitative variables, as they extend the principles of principal component analysis to mixed data types. This approach allows us to reduce the complexity of the data by identifying the principal dimensions that capture the most variance. Consequently, FAMD provide a comprehensive understanding of how different variables interact and contribute to the observed patterns in our dataset. The results of the FAMD are visually depicted in Supplementary Figure S1, offering a clear illustration of the interrelations among the dataset where variables that are highly correlated cluster together, including how closely each variable correlates with the outcome variable, the PDQ-39 SI.

After the FAMD analysis, the PRISM dataset was subjected to a data-cleaning process. Categorical variables with responses such as 'prefer not to say' or 'other' were treated as missing values to minimize the noise in the final model. Upon examination of the dataset (Supplementary Table S1), it was observed that it contains minimal missing values. Most of these missing values appear to be missing completely at random, with notable exceptions in the variables age at diagnosis and current medication. These exceptions suggest a missing not at random pattern, likely due to patients not recalling the age of PD diagnosis or being unaware of the therapeutic class of their current medication. Although we initially attempted to address missing values using Multivariate Imputation by Chained Equations (MICE), this approach resulted in a model with worse predictive capabilities [49]. Therefore, we only kept complete observations, resulting in a total of 627 valid entries. Subsequently, the dataset was partitioned into training (90%) and validation (10%) sets to maximize the learning capacity of our model amidst the constraints of limited sample size and high variability inherent in questionnaire-based data. This allocation allows for comprehensive model training on a diverse range of responses, critical for capturing the nuanced patterns of HRQoL assessments in PD, while still reserving a portion of the data for essential validation to assess model performance and mitigate overfitting. Figure 5 illustrates this study's workflow, detailing each step from data acquisition to the data split and comparison of regression metrics in the training and validation sets.

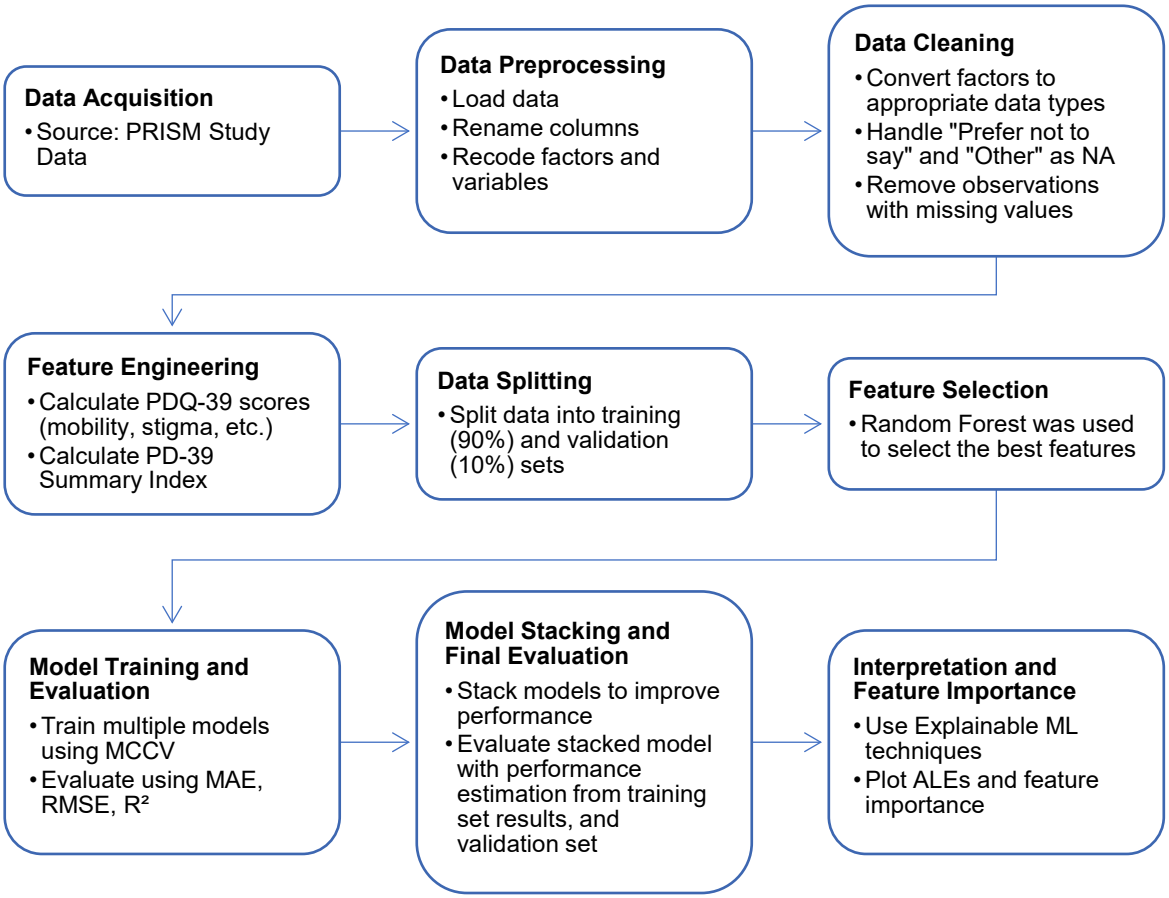

Figure 5: Diagram of the study process from data acquisition to data preparation, feature selection, model training, and performance evaluation.

Model training

The models were trained on a personal computer equipped with an AMD Ryzen 7 7800X3D CPU, 32 GB of RAM, and an AMD Radeon RX 7900 XT GPU (Advanced Micro Devices, Inc., Santa Clara, CA, USA). We evaluated several feature selection methods—specifically, random forest, least absolute shrinkage and selection operator (LASSO), group LASSO, and the option of omitting feature selection—to identify the most suitable technique for our dataset. Feature selection through random forest was selected for its standout performance in enhancing the predictive accuracy of our suite of ML models. We then applied this feature selection approach in training a diverse range of ML models, namely linear regression, boosted generalized linear model, extreme gradient boosting with linear models and decision trees, ridge regression, neural network regression, LASSO regression, random forest, k-nearest neighbors, support vector machines with radial basis function and

polynomial kernel, and Gaussian processes (linear, polynomial, and radial kernels), incorporating essential pre-processing steps such as centering and scaling with the caret package [50]. To accurately train and evaluate the performance of these models within the context of a dataset characterized by limited size and high variability, Monte Carlo Cross-Validation (MCCV) was employed with 500 repetitions for each model. This approach was chosen for its effectiveness in mitigating data variability, thereby ensuring a more reliable and stable assessment of the models' generalizability.

Model Stacking

To further improve the performance of our predictive model, we implemented a stacking approach using an ensemble of the previously trained ML models. Stacking is an advanced ensemble method that combines multiple predictive models to enhance overall accuracy. It utilizes a meta-model that optimally combines the predictions from several base models trained on the same data. This technique leverages the strengths and compensates for the weaknesses of individual models, thereby increasing the robustness and reliability of predictions [51].

We started by assessing the correlation among these models to create an ensemble characterized by both strong individual performance and minimal intermodel correlation. We then trained several ensembles using MCCV and experimented with different numbers of repetitions (100, 250, 500, and 1000) to assess both stability and performance. For the stacking mechanism, we tested both random forest and extreme gradient boosting with decision trees (xgbTree) as techniques. These were selected for their demonstrated capability to effectively handle complex data patterns. This stacked ensemble approach allowed us to enhance the model's ability to generalize across different patient profiles and improve the accuracy of predicting the PDQ-39 SI.

Model Evaluation

The performance of the models was evaluated using multiple metrics. The Mean Absolute Error (MAE) measures the average magnitude of the errors in predictions, providing a straightforward measure of the average error magnitude per prediction. The Root Mean Squared Error (RMSE) offers insight into the accuracy of the model by quantifying the square root of the average squared differences between the predicted values and actual

observations. Lastly, the Coefficient of Determination (R^2) reflects the proportion of variance in the dependent variable that is predictable from the independent variables, thereby indicating the explanatory power of the model. These metrics collectively provide a comprehensive assessment of model performance.

The ensemble model that emerged as the most effective incorporated the pre-trained ML models: Gaussian process with a Radial Basis Function kernel and random forest. This ensemble model was trained using MCCV with 1000 repetitions, with xgbTree serving as the stacking method. MCCV was utilized both to train and to estimate the performance of the ensemble model on the training dataset, while the validation set was reserved for the final assessment to evaluate overfitting susceptibility and ensure an unbiased evaluation of the model's generalizability.

Interpretable machine learning analysis

To unveil the importance and effect of each feature on the predictive model, we employed interpretable ML techniques, particularly focusing on Permutation Feature Importance (PFI). Utilizing the iml package in R, this analysis facilitates a deeper understanding of the behavior of our final ensemble ML model [52]. The core concept of PFI is to measure the increase in the model's prediction error after the permutation of each feature's values, which effectively breaks the relationship between the feature and the true outcome. This process provides a clear indication of a feature's importance; a significant increase in error indicates that the model relied heavily on the feature for prediction, whereas no change in error suggests that the feature was not utilized by the model for prediction [53].

The process of feature importance quantification involves permuting the values of each feature and assessing the change in the model's performance using MAE as the chosen metric, due to its lower sensitivity to outliers and variability within the dataset. This procedure was carried out for each feature individually and repeated 100 times to obtain a more robust estimate, with the results averaged thereafter.

4.4. Results

Performance Metrics

In the training set, the final ensemble model achieved an MAE of 4.82, an RMSE of 8.09, and an R^2 value of 0.75, indicating a strong fit to the data. The distributions of MAE, RMSE, and R^2 metrics from the 1000 MCCV repetitions approximated normal distributions, as shown in Figure 6, Figure 7, and Figure 8. These distributions suggest that the model's performance is consistent and reliable across different subsamples of the training data, reflecting its stability. When applied to the validation set, the model exhibited moderate variations in performance metrics, with an MAE of 11.22, an RMSE of 13.99, and an R^2 value of 0.36, indicating the model's reasonable predictive capability on unseen data. A summary of these results is presented in Table 1.

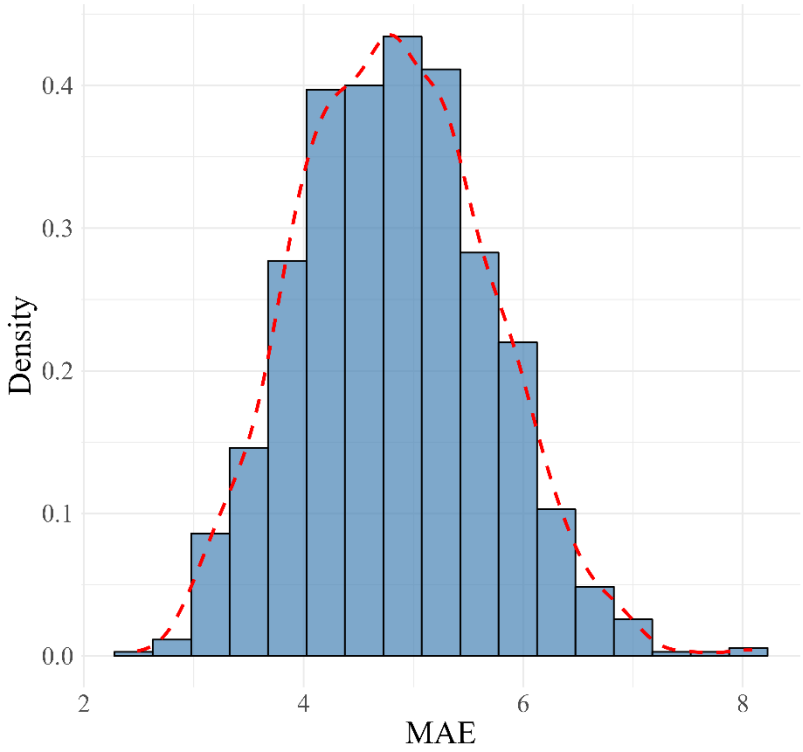

Figure 6: Distribution and density curve of MAE values from 1000 MCCV repetitions, demonstrating the consistency and reliability of the predictive model in training.

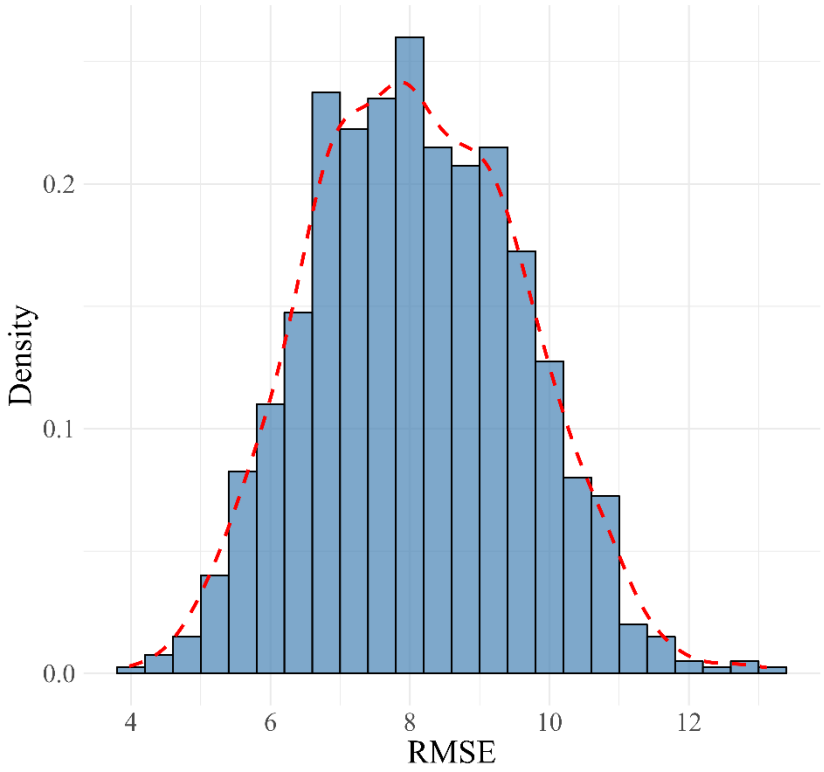

Figure 7: Distribution and density curve of RMSE values from 1000 MCCV repetitions, assessing the model's accuracy and stability during the training phase.

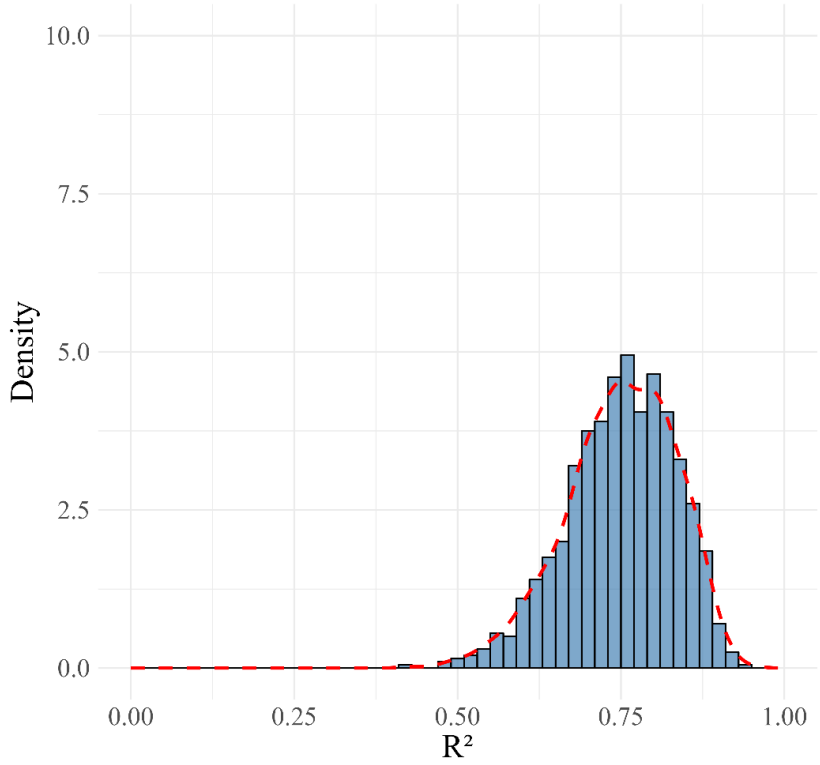

Figure 8: Distribution and density curve of R^2 values from 1000 MCCV repetitions, showcasing the predictive power and fit of the model to the training data.

Table 1: Results.

Metric	Training Set	Validation Set
MAE	4.82	11.22
RMSE	8.09	13.99
R^2	0.75	0.36

The True vs. Predicted Values plot (Figure 9) demonstrates the relationship between the actual PDQ-39 SI scores and the scores predicted by the model. The plot suggests that the model tends to slightly overpredict at lower values and underpredict at higher values, indicating areas for potential refinement. The Residuals vs. Predicted Values plot (Figure 10) shows that while most residuals are clustered around zero, there is a slight spread, suggesting occasional larger prediction errors.

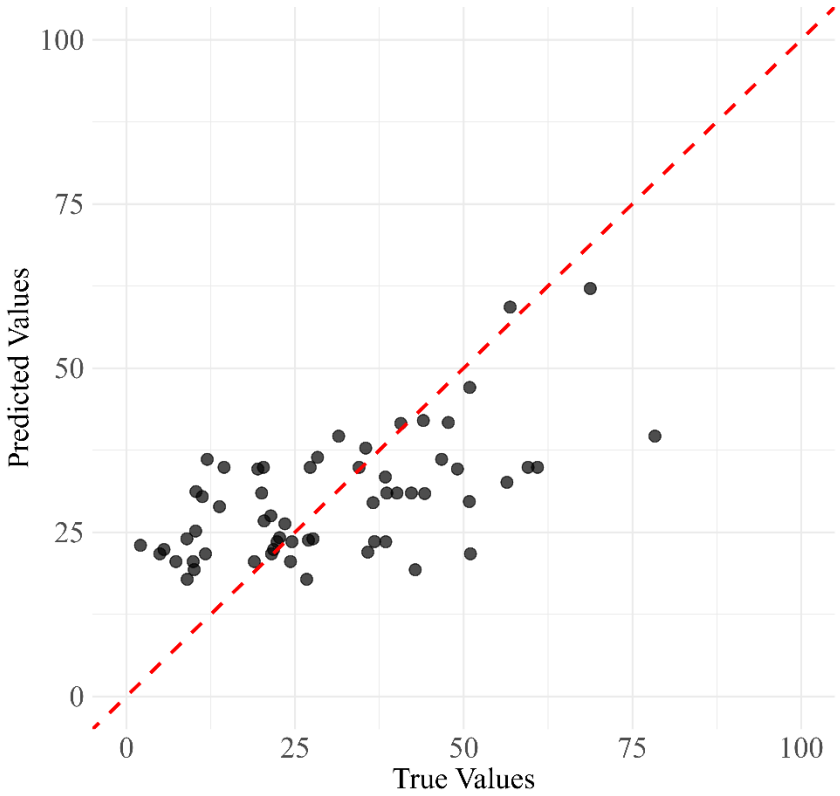

Figure 9: Scatter plot showing the relationship between actual PDQ-39 SI scores (true values) and model predictions (predicted values). The black dots represent individual predictions, and the red dotted line represents the line of perfect prediction where predicted values would equal true values.

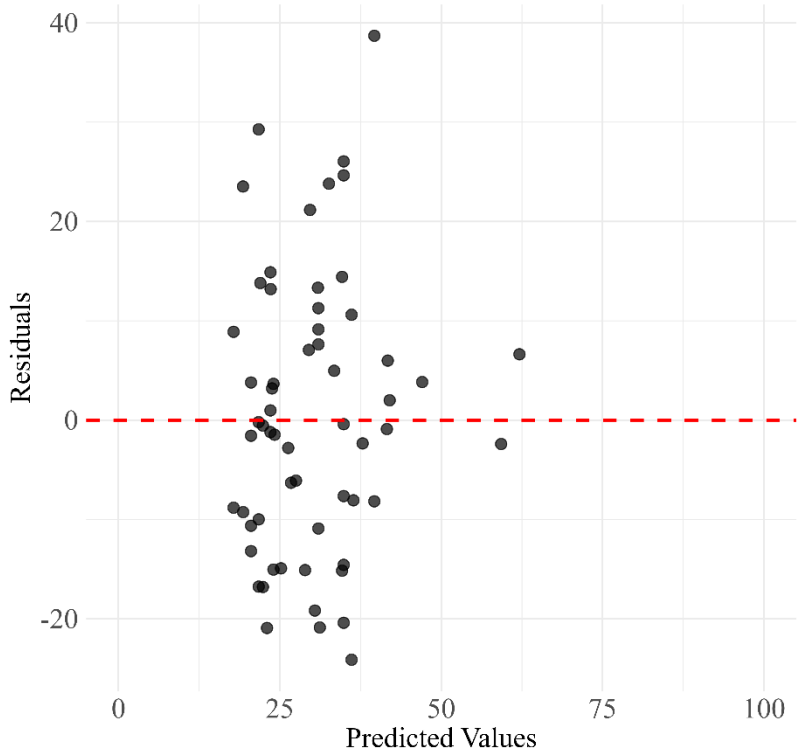

Figure 10: Scatter plot of residuals (errors) against predicted values, indicating prediction accuracy and error distribution. The black dots represent the residuals (actual value − predicted value), and the red dashed line represents the zero-error line.

Interpretation

The ensemble model demonstrated strong predictive capabilities on the training dataset, as indicated by the high R^2 and low MAE values. However, the model performance on the validation set showed moderate variation. Although the model remains a promising tool for estimating PDQ-39 SI scores, these variations suggest areas for future research and model refinement. The observed variations in the validation set may be attributed to the limited size of the dataset and the inherent variability of HRQoL measures, both of which warrant further investigation. The diagnostic plots suggest that while the model generally performs well, addressing the slight overprediction at lower values and underprediction at higher values could enhance its accuracy, aligning with the variability observed in the validation set. The variables selected through feature selection for their significant impact on PDQ-39 SI predictions are listed in Table 2.

Feature importance and effects

The interpretable ML analysis using the iml package in R yielded substantial insights into the significance of the individual features within our predictive model. Table 2 encapsulates the importance of each feature along with the associated 90% confidence intervals, as derived from Permutation Feature Importance, which quantifies the decrease in model accuracy when each feature's data is randomly shuffled. Predominantly, age at diagnosis, patient's country, dementia, and patient's age emerged as the most important features affecting PDQ-39 SI predictions. A graphical representation of feature importance was generated to provide a holistic view of each feature's contribution to the model's predictive performance (Figure 11).

Table 2: Selected Features and their importance in PDQ-39 SI predictions [a-c].

Feature	Importance (0.05)	Importance	Importance (0.95)	Permutation error
Age at diagnosis	1.07	1.16	1.25	13.02
Country	1.05	1.12	1.19	12.53
Dementia	1.02	1.09	1.11	12.26
Age	1.02	1.09	1.19	12.24
Depression	1.03	1.09	1.15	12.19
Time to first levodopa prescription post-diagnosis	1.02	1.08	1.13	12.07
Number of comorbidities	1	1.05	1.11	11.83
Anxiety	1.02	1.05	1.07	11.74
First prescribed oral medication contained levodopa	1	1.02	1.05	11.5
Currently taking Rivastigmine	1	1.02	1.05	11.5
Education	0.98	1.02	1.06	11.46
Peripheral vascular arterial disease	0.99	1.01	1.04	11.37
Diabetes (type 1 or 2)	0.99	1.01	1.02	11.32
Currently taking Catechol-O-methyltransferase inhibitor	0.97	1	1.03	11.21
Currently taking Amantadine	0.98	0.99	1.01	11.07

[a] The importance of a feature is quantified by permuting the values of the feature and measuring the change in the model's performance against a baseline using MAE. The displayed results include the estimated importance values alongside their 90% confidence intervals, represented by the lower (Importance (0.05)) and upper (Importance (0.95)) limits.

[b] A larger permutation error indicates higher feature importance.

[c] The process was executed for each feature and repeated 100 times to obtain a more robust estimate, with the results averaged.

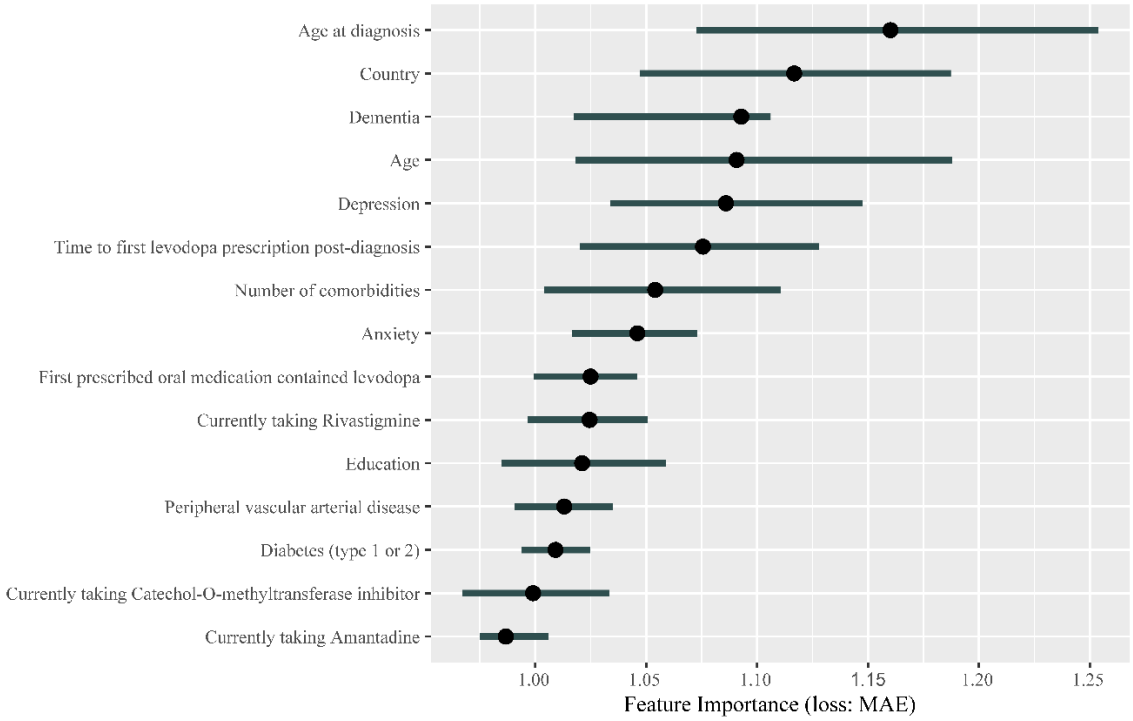

Figure 11: Demonstration of the feature importance of clinical variables in predicting PDQ-39 SI.

Transitioning to a more detailed analysis, the Accumulated Local Effects (ALE) plots facilitate visual exploration of the behavior of the model, underscoring how the prediction shifts locally with variations in each feature. The plots for the features age at diagnosis, dementia, patient's country, and patient's age are represented in Figure 12, Figure 13, Figure 14, and Figure 15. The marks beneath the x-axis represent the distribution of each feature where observations were obtained. In the case of categorical variables such as dementia and the patient's country, the marks were automatically spaced to simplify the visual interpretation. For the features age at diagnosis and patient's age, regions with sparse or no points suggest caution in interpretation due to insufficient data.

The ALE plot for age at diagnosis (Figure 12) indicates that the earlier the disease appears, the higher the scores predicted by the model on the PDQ-39 SI, and therefore the more likely the patients are to have a lower quality of life, which is consistent with disease progression. The ALE plot for the patient's country (Figure 13) indicates a moderate impact on the predictions of PDQ-39 SI, with the United Kingdom showing a slightly more negative impact on the outcome. The ALE plot for dementia (Figure 14) demonstrates the significant negative impact of this comorbidity on the PDQ-39 SI predictions, highlighting its burden on

PD patients. The ALE plot for the patient's age (Figure 15) demonstrates a consistent trend, indicating a direct influence of age on PDQ-39 SI predictions. This interpretive analysis not only enhances the comprehensibility of our predictive model but also extends the foundation for a more informed evaluation of the features affecting the PDQ-39 SI.

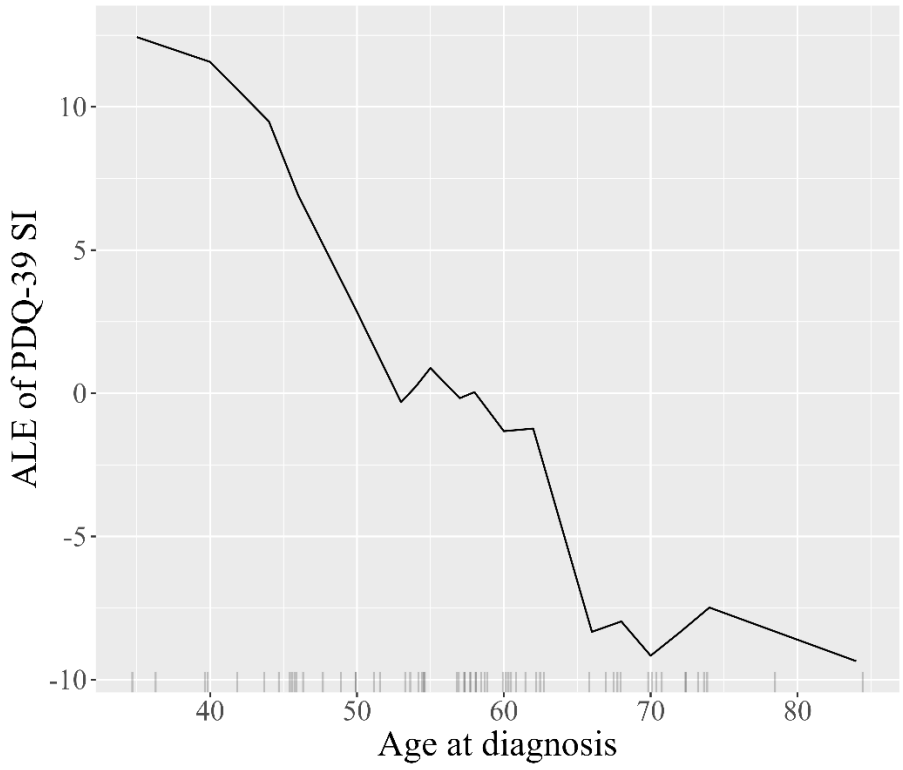

Figure 12: ALE plot illustrating the effect of age at diagnosis on PDQ-39 SI predictions, showing how earlier or later disease appearance impacts the quality of life in PD patients.

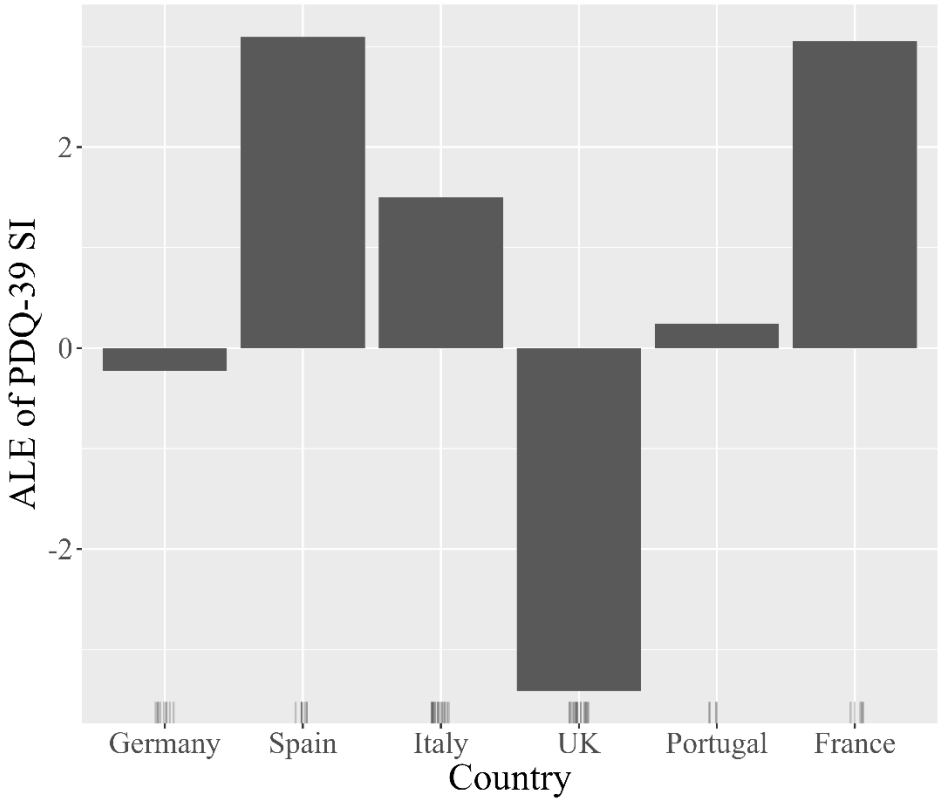

Figure 13: ALE plot illustrating the effect of the patient's country of residence on PDQ-39 SI predictions, highlighting the variations in the quality of life among PD patients across different countries.

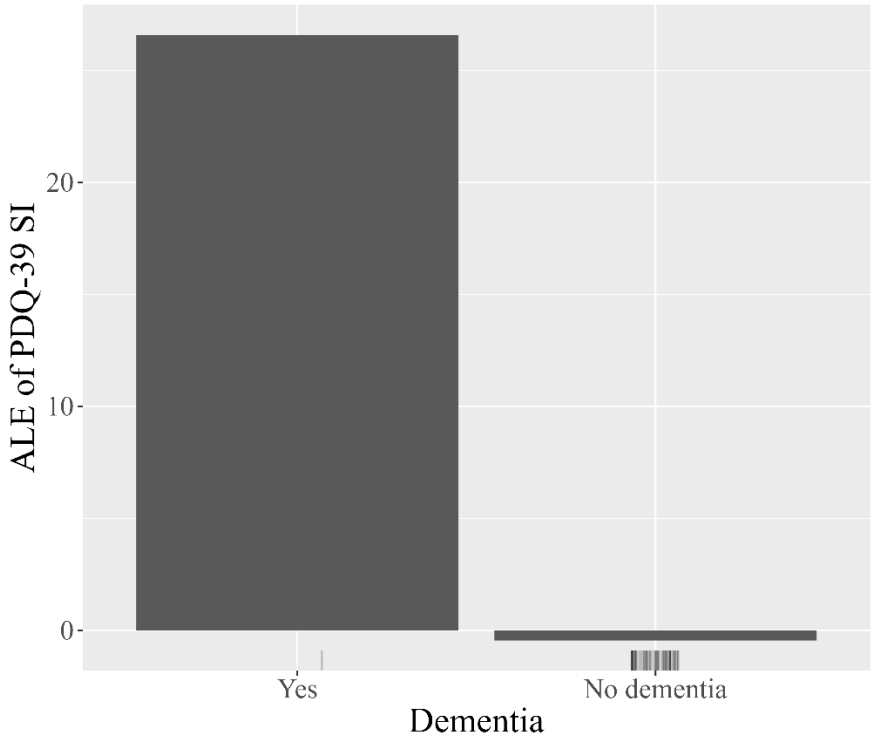

Figure 14: ALE plot demonstrating how the presence of dementia affects PDQ-39 SI scores predicted by the model, emphasizing its considerable negative impact on the quality of life in PD patients.

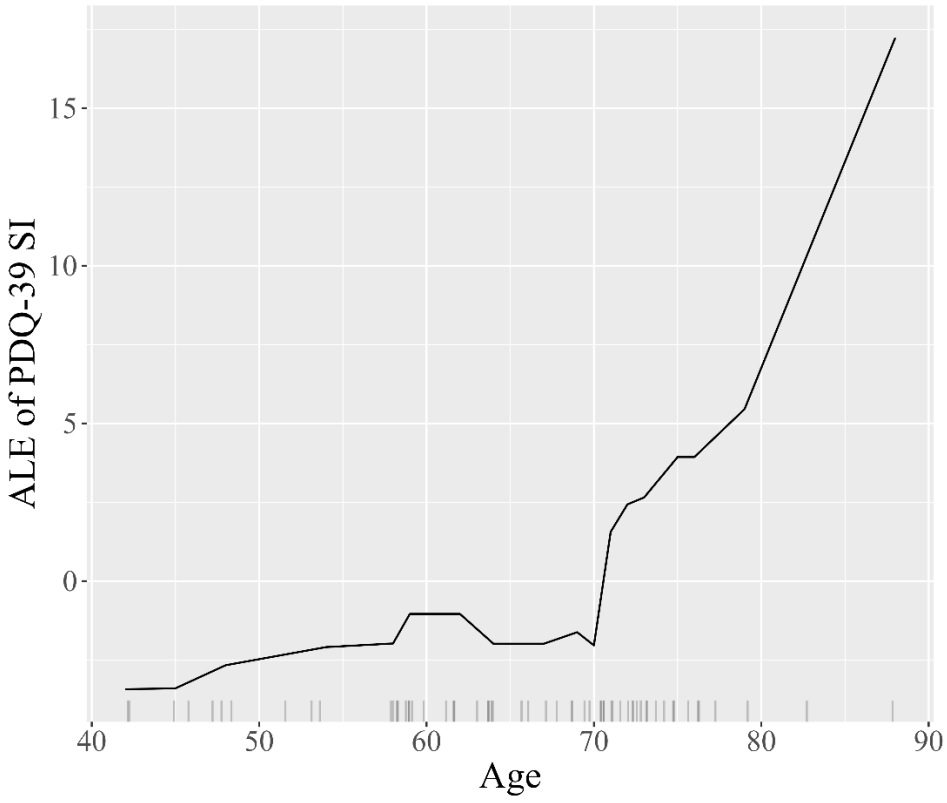

Figure 15: ALE plot detailing the impact of age on PDQ-39 SI predictions, illustrating the direct influence of age increments on the PDQ-39 SI scores predicted by the model.

4.5. Discussion

Principal Results

The primary achievement of this study is the successful development and validation of an ensemble ML model tailored to predict the PDQ-39 SI score in patients with PD. The model's strong performance in the training set, particularly its high R^2 value, underscores its ability to account for a significant proportion of the variance in PDQ-39 SI scores. This is a noteworthy advancement given the complex interplay of factors that can influence the quality of life in patients with PD. However, it is important to note that the significant difference in R^2 values between the training and validation sets indicates a potential issue of overfitting. The R^2 value in the training set was substantially higher than that in the validation set, suggesting that the model may have learned patterns specific to the training data that do not generalize well to unseen data. This is common in ML applications and serves as a valuable indicator of areas that may benefit from further research and optimization. Future work should focus on addressing this overfitting issue by exploring techniques such as

regularization, cross-validation, and the inclusion of more diverse training data. Additionally, model simplification and the use of ensemble methods with a more rigorous selection of base models may help enhance generalizability.

The model's performance in the validation set, although not as strong as that in the training set, still suggests that it holds promise as a tool for estimating PDQ-39 SI scores in a clinical setting. Despite the inherent uncertainty in predictions of patient-reported outcome measures, the model provides valuable insights into disease progression and highlights the potential impact of clinical variables on the quality of life in PD patients. This indicates that the model could become a useful decision-support tool, with further refinement and validation needed to enhance its reliability and applicability.

The feature importance analysis conducted in this study provides a granular understanding of the variables that significantly impact the PDQ-39 SI score. This approach not only predicts the PDQ-39 SI score but also elucidates the underlying factors contributing to the model's predictions, offering a more comprehensive insight into the determinants of HRQoL in patients with PD. The notable importance of variables such as age at diagnosis, dementia, and patient's age in our model underlines the well-documented interplay among these factors and PD, as they are highly correlated with disease progression. Age is a recognized risk factor for PD, and both the prevalence and severity of the disease notably increase with age, underscoring the progressive nature of PD. The significant association between age and PD is evident from various epidemiological studies that show an exponential increase in PD prevalence from the age of 60 years onwards, reaching a peak in the age group of 70–79 years [54]. Additionally, an early appearance of the disease (represented by the feature 'age at diagnosis') is associated with a greater impact on the quality of life, highlighting the significant burden of the disease when it manifests early. Furthermore, dementia is a common comorbidity in patients with PD, often exacerbating the severity of the disease and the patient's quality of life. The prevalence of dementia in PD patients ranges between 24% and 31%, with the risk increasing substantially with age [55].

Our model's emphasis on age at diagnosis, dementia, and patient's age as significant predictors of the PDQ-39 SI reflects the clinically observed interactions between these variables and PD symptom severity. This alignment with established clinical knowledge not only enhances the validity of our model but also underscores the potential utility of our model in clinical settings. Utilizing these critical predictors, healthcare providers can anticipate the likely trajectory of PD patients' quality of life, aiding in the development of personalized

treatment plans. It is also worth noting that other variables, such as anxiety, education, and some comorbidities, were identified as relevant by our feature selection method but did not show as high importance in the final model's predictions. These variables might still be important for adjusting the model to control for confounding effects and improve overall accuracy, even though they do not individually have strong predictive power [56].

Although the model does provide important information on the features affecting the PDQ-39 SI, it is also important to highlight that the predictive capabilities still demonstrate some variation, especially in the validation set. This could indicate that while the clinical variables are useful, the features available in the dataset do not entirely capture the PDQ-39 SI. The observed variability suggests that there are other influential factors impacting the PDQ-39 SI that are not represented in the current dataset. As such, future research should explore the inclusion of additional variables and data sources that may better reflect the complexities of HRQoL in PD patients. This would potentially enhance the model's accuracy and provide a more comprehensive tool for clinical application.

Limitations

One of the key limitations of our study was the small sample size. This could have contributed to the observed variations in performance metrics between the training and validation sets as small sample sizes can significantly impact ML performance estimates [57]. Additionally, we decided to exclude incomplete data and only use complete observations, resulting in a total of 627 valid entries. This decision was necessary because data imputation negatively impacted model performance. Future research should aim to address these limitations by incorporating larger and more diverse datasets, which could enhance the model's generalizability and robustness. Exploring different imputation methods or techniques to handle missing data more effectively might also improve model performance.

Comparison with Prior Work

Our study diverges from the approach taken by Candel-Parra et al. for predicting the PDQ-39 SI score in patients with PD. Specifically, Candel-Parra et al. utilized classification models to categorize the PDQ-39 SI score into binary categories (above or below 40), whereas our study employed regression models to predict the actual score value [58]. This distinction is particularly significant given the absence of established cut-off values for the PDQ-39 SI

score. Classification models, such as those used by Candel-Parra et al., require such cut-offs to categorize outcomes, introducing potential arbitrariness and statistical power loss when no clinically validated thresholds exist [58].

In contrast, our regression model predicts the actual PDQ-39 SI score, providing a more nuanced and precise assessment of a patient's quality of life. Treating the PDQ-39 SI scores as a continuous variable aligns our model more closely with the actual behavior of the scores, enhancing its applicability and interpretability in clinical settings. Furthermore, our use of explainable ML techniques allows clinicians to simulate disease progression and understand which variables are influencing the quality of life, providing actionable insights that can inform treatment decisions and patient management. This feature is particularly valuable, as it enables healthcare providers to tailor interventions specifically to the factors that have the greatest impact on a patient's HRQoL, such as early diagnosis, dementia, and patient's age.

4.6. Conclusions

The ensemble model developed in this study is a promising tool for healthcare providers to predict the HRQoL of patients with Parkinson's Disease, thereby aiding in personalized treatment strategies. Given that PDQ-39 SI is commonly used as an endpoint in clinical trials, the ability to predict this score could offer valuable insights for trial design and patient management. By leveraging regression models, our approach offers a nuanced understanding of a patient's quality of life, filling a gap in existing research methodologies.

While the PDQ-39 is a reliable and valid tool for assessing HRQoL in PD patients, it primarily serves as a cross-sectional assessment. In contrast, our predictive model enhances the utility of the PDQ-39 by providing dynamic predictions of the PDQ-39 SI score based on readily available clinical variables. This predictive capability allows for the simulation of disease progression and the identification of key factors influencing HRQoL over time, such as age at diagnosis, dementia, and patient's age. Consequently, clinicians can gain deeper insights into how individual patient characteristics and changes in clinical status impact the quality of life, enabling more personalized and proactive management strategies. The integration of explainable ML techniques further distinguishes our model by offering transparency into the prediction process, thereby empowering healthcare providers to understand and trust the model's recommendations. The practical feature of utilizing clinical

variables that are typically available at the moment of use, as emphasized in a recent critique of predictive models, enhances the usability and real-world applicability of our model, setting it apart from models that depend on variables that are not readily accessible in the clinical context [59].

Although the model's performance on the validation set exhibited moderate variation, it sets the stage for future research in aspects such as feature engineering, hyperparameter optimization, and incorporation of additional clinically relevant variables. Moreover, the model's performance is likely to improve with a larger sample size, given the inherent variability of patient-reported outcomes, such as those measured by the PDQ-39 SI. Future work should also address the potential overfitting suggested by the difference in R^2 values between the training and validation sets.

While the model does provide important insights into the features affecting the PDQ-39 SI, it is essential to recognize that the predictive capabilities still demonstrate some variation, especially in the validation set. This variability suggests that the current dataset's features do not entirely capture the PDQ-39 SI, indicating that additional variables and data sources might be necessary to better reflect the complexities of HRQoL in PD patients. Future research should explore these additional factors to enhance the model's accuracy and provide a more comprehensive tool for clinical application.

The potential impact of the model on clinical practice and clinical trials underscores its value as a significant contribution to the field. Furthermore, the broader implications of this work have the potential to significantly influence how healthcare providers approach PD treatment, thereby contributing to more personalized and effective care.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jcm13175081/s1, Table S1: Descriptive Statistics of the PRISM Database. Figure S2: Factor Analysis of Mixed Data.

Author Contributions: Conceptualization, Daniel Magano, Tiago Taveira-Gomes and António Barros; methodology, Daniel Magano, Tiago Taveira-Gomes and António Barros; software, Daniel Magano; validation, Daniel Magano, Tiago Taveira-Gomes, António Barros, and João Massano; formal analysis, Daniel Magano; data curation, Daniel Magano;

writing—original draft preparation, Daniel Magano; writing—review and editing, Tiago Taveira-Gomes, António Barros, and João Massano; supervision, Tiago Taveira-Gomes and António Barros. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been done under the scope of and funded by the PhD Program in Health Data Science of the Faculty of Medicine of the University of Porto, Portugal heads.med.up.pt.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used in this study were obtained from the Parkinson's Real World Impact assesSMent (PRISM) database (www.prism.bial.com), accessed on 18 February 2022.

Acknowledgments: The PRISM study and database were funded by BIAL – Portela & Cª, S.A., designed in collaboration with The Cure Parkinson's Trust, an advocacy group based in the United Kingdom (UK), and reviewed by the PRISM Steering Committee.

Conflicts of Interest: The authors declare no conflicts of interest.

5. Non-Motor Symptoms as Critical Predictors of Quality of Life in Parkinson's Disease: A Machine Learning Approach

Original Article

Non-Motor Symptoms as Critical Predictors of Quality of Life in Parkinson's Disease: A Machine Learning Approach

Daniel Magano[1, 2], António S. Barros[3], João Massano[4, 5], Laila Alsuwaidi[6], and Tiago Taveira-Gomes[7, 8, 9]

[1] Ph.D. Program in Health Data Science, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal

[2] Medical Department, BIAL-Portela & Cª., S.A., 4745-457 São Mamede do Coronado, Portugal

[3] Department of Surgery and Physiology, Cardiovascular R&D Centre-UnIC@RISE, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal

[4] Department of Clinical Neurosciences and Mental Health, Faculty of Medicine University of Porto, 4200-319 Porto, Portugal

[5] Department of Neurology, Centro Hospitalar Universitário de São João, 4200-319 Porto, Portugal

[6] Department of Community Medicine, Information and Decision in Health, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal

[7] Department of Community Medicine, Information and Decision in Health, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal

[8] Faculty of Health Sciences, University Fernando Pessoa, 4200-150 Porto, Portugal

[9] SIGIL Scientific Enterprises, 4076 Dubai, United Arab Emirates

Published in Health and Quality of Life Outcomes (Appendix 3)

https://doi.org/10.1186/s12955-025-02451-2

5.1. Abstract

Background: Parkinson's disease (PD) considerably impacts health-related quality of life (HRQoL) through motor and non-motor symptoms. The Parkinson's Disease Questionnaire-39 (PDQ-39) is the most widely used tool to assess HRQoL, encompassing eight dimensions and a Summary Index providing an overall score. Despite advances in machine learning (ML) for predicting disease symptoms and progression, its application to predict HRQoL across these dimensions remains underexplored.

Methods: This study uses complete-case data for 478 of 861 patients from PRISM, a cross-sectional observational survey conducted in six European countries in 2018–2019. Participants were adults with PD recruited through advocacy groups and clinical centers who completed online assessments, providing data on demographics, medication, comorbidities, and disease characteristics (Tolosa et al., 2021). ML models were trained to predict PDQ-39 dimensions and Summary Index scores (0–100; higher = worse HRQoL). Features were preselected using the Boruta algorithm on the training data. Model selection was based on the lowest mean RMSE from 100 bootstrap resamples on the training set. Selected models were then retrained using 1000 bootstrap resamples for robust performance estimation. Final performance was evaluated on a held-out 20% validation set using R^2, MAE, and RMSE. Feature importance was assessed using permutation importance with MAE loss (100 permutations) on the held-out validation set. Factor Analysis of Mixed Data (FAMD) was used to explore patterns between non-motor symptoms and PDQ-39.

Results: Selected models: xgbTree (Summary Index; Activities of Daily Living) and gaussprPoly (all other PDQ-39 dimensions). On the validation set, Summary Index/ Cognitions showed the strongest performance with R^2 = 0.56/0.53, MAE = 9.60/12.39, RMSE = 12.66/16.20. Permutation feature importance ranked the Non-Motor Symptoms Questionnaire score (sum of 30 non-motor symptoms, range 0-30) as the most important predictor across all models. FAMD showed clustering of Social Support, Bodily Discomfort, and Stigma dimensions with Anxiety.

Conclusions: Our findings demonstrate the critical role of non-motor symptoms in predicting HRQoL in patients with PD. While ML models effectively predict overall HRQoL and cognitive aspects, achieving comparable performance on other dimensions may require additional variables to reduce error. These insights emphasize comprehensive treatment strategies addressing both motor and non-motor symptoms.

Keywords: Parkinson’s Disease; Health-Related Quality of Life; predictive modeling; machine learning; Parkinson’s Disease Questionnaire-39; explainable machine learning

5.2. Background

Parkinson’s disease (PD) is a progressive neurodegenerative disorder that affects millions of people worldwide, with its global prevalence more than doubling between 1990 and 2015, highlighting the growing burden of this condition [38], [60]. PD is primarily characterized by motor symptoms such as resting tremor, rigidity, bradykinesia, and postural instability, which are often accompanied by a variety of non-motor symptoms (NMS), including cognitive impairment, mental health disorders, sleep disturbances, pain, and sensory disturbances [37], [38], [39]. These symptoms considerably reduce the health-related quality of life (HRQoL) in individuals with PD, making HRQoL an essential outcome measure in both clinical settings and research [34].

HRQoL instruments, particularly disease-specific measures, have become invaluable in capturing the personal impact of PD and the changes experienced by patients over time. Among these instruments, the Parkinson's Disease Questionnaire-39 (PDQ-39) is the most widely used and validated tool for assessing HRQoL in patients with PD [22], [23], [34], [41]. It covers eight dimensions of HRQoL: Activities of Daily Living, Bodily Discomfort, Cognitions, Communication, Emotional Well-Being, Mobility, Social Support, and Stigma, providing a comprehensive overview of the disease's impact on HRQoL [21]. As a patient-reported outcome measure, the PDQ-39 reflects outcomes that matter to patients and aligns with regulatory guidance: both the European Medicines Agency and the U.S. Food and Drug Administration emphasize the use of validated instruments to inform benefit–risk assessment and, when appropriate, to support claims in product labeling [61], [62].

The PDQ-39 Summary Index, hereafter referred to as the Summary Index, is a composite score derived from the individual eight dimensions of the PDQ-39 and is frequently used in clinical trials as a primary outcome measure to evaluate the overall impact of PD on patients' lives [42], [43]. While the Summary Index provides a valuable summary of HRQoL, it can obscure the nuances captured by the individual dimensions, each reflecting a distinct aspect of the disease's impact. Understanding the relationship between these dimensions and clinical variables is crucial for tailoring interventions to address the specific challenges faced by patients with PD. However, this remains challenging due to the complex and multifaceted nature of the disease, where symptoms can vary widely among individuals. Notably, non-

motor symptoms (NMS), which are often among the most disabling yet under-recognized aspects of PD, have been shown to considerably impact the quality of life [41]. The Non-Motor Symptoms Questionnaire (NMSQ) is a self-reported instrument of 30 yes/no items covering common non-motor symptoms in PD. The total NMSQ score is obtained by summing the number of symptoms the patient reports, yielding a score from 0 to 30, with higher scores reflecting a greater burden of non-motor symptoms. This simple sum provides a practical, validated measure of overall non-motor symptom burden in PD [63].

Traditional statistical models often fall short in capturing the nonlinear and high-dimensional relationships between the diverse clinical manifestations of PD and HRQoL. In response, ML methods have gained traction for their ability to model such complexity, with demonstrated success in predicting disease progression, motor subtypes, and treatment response [29], [44], [45], [46], [64], [65], [66], [67], [68], [69]. While these advances highlight the potential of ML in neurodegenerative disease modeling, most applications to date have concentrated on motor symptoms, digital monitoring, and disease classification.

In contrast, ML models targeting HRQoL in PD, particularly using the PDQ-39, remain limited. Most existing studies treat HRQoL as a secondary outcome or focus solely on the global Summary Index, overlooking the predictive modeling of its individual dimensions [58], [70], [71], [72]. Building on our earlier work published in the *Journal of Clinical Medicine* (2024), which focused on predicting the Summary Index, this study addresses a critical gap by extending the analysis to the eight individual PDQ-39 dimensions. By doing so, we aim to enhance the understanding of dimension-specific quality-of-life impacts and support more personalized clinical decision-making in PD care.

This study aims to achieve two critical objectives: first, to explore and clarify the relationship between standard clinical variables and both the PDQ-39 dimensions and the Summary Index; second, to develop and assess the potential of ML models for predicting these outcomes. By incorporating explainable ML techniques, this study seeks to enhance clinicians' understanding of HRQoL outcomes, offering deeper insights into the factors most strongly associated with the different dimensions of HRQoL in PD. This approach is intended to inform more personalized treatment strategies, ultimately improving care and HRQoL for patients with PD.

5.3. Materials and Methods

Study overview and analysis workflow

Figure 16 summarizes the workflow from the PRISM data source and analytic sample selection through model development, evaluation, and interpretability. It is intended as an overview; full details are provided in the subsequent Methods subsections.

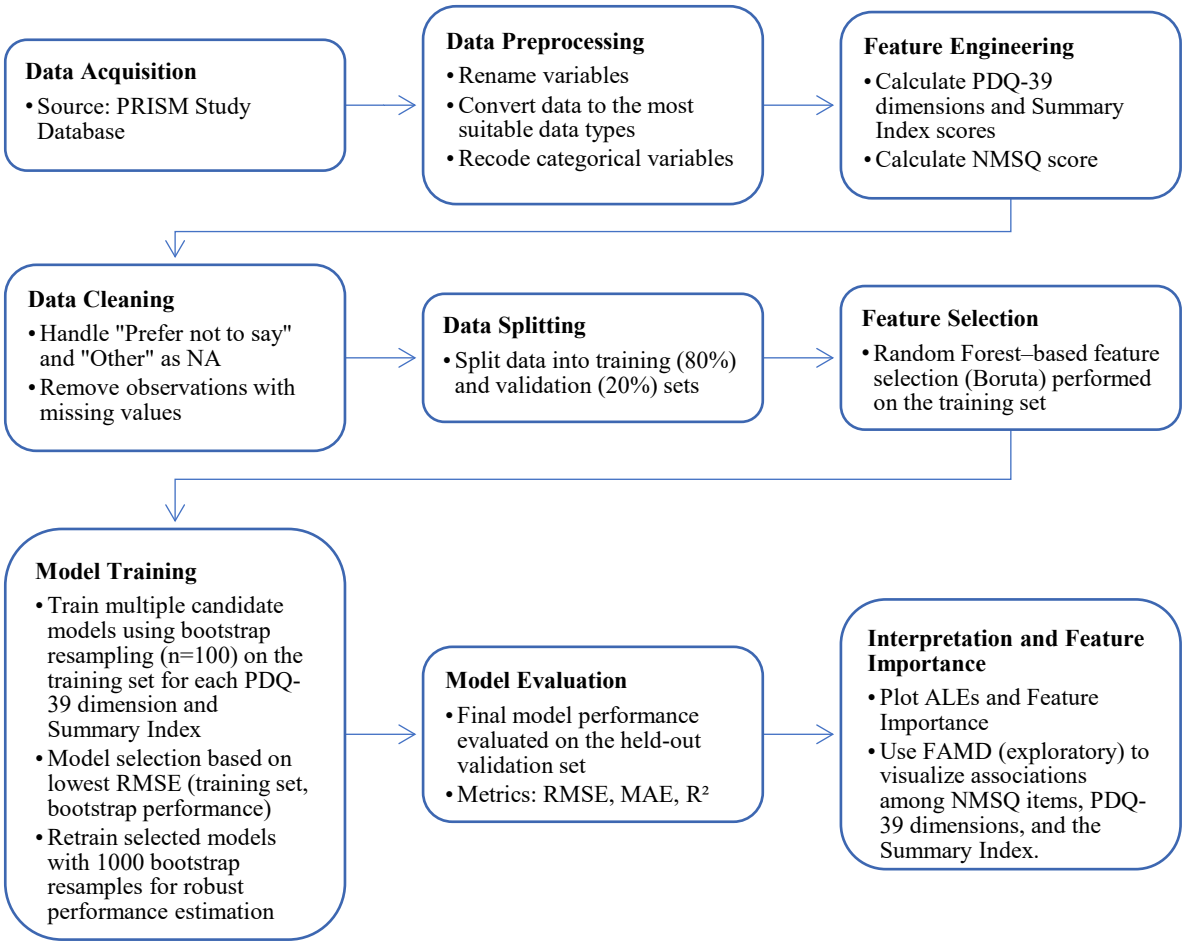

Figure 16. Diagram of the study's workflow. PRISM, Parkinson's Real-world Impact assesSMent; PDQ-39, Parkinson's Disease Questionnaire-39; NMSQ, Non-Motor Symptoms Questionnaire; RMSE, Root Mean Squared Error; MAE, Mean Absolute Error; R^2, Coefficient of Determination; ALE, Accumulated Local Effects; FAMD, Factor Analysis of Mixed Data. Note: "Recode categorical variables" refers to converting categorical variables to suitable numeric/dummy-coded variables for modeling. Model selection was performed using the training set only; the validation set was reserved for final, unbiased performance assessment.

Database

The data for this study were sourced from the Parkinson's Real-world Impact assesSMent (PRISM) study, a cross-sectional observational research initiative conducted across six European countries: France, Germany, Italy, Portugal, Spain, and the United Kingdom. Detailed information on the study design and recruitment can be found in Tolosa et al., 2021 [36]. The PRISM study included 861 patients with PD and 256 caregivers, recruited via patient advocacy groups and clinical centers. Participants completed a comprehensive online survey comprising validated instruments, including the PDQ-39, the NMSQ, as well as questions on demographics, medication, comorbidities, disease characteristics, and healthcare resource utilization. This diverse database supports robust modeling and generalizability to the wider European PD population.

Data preparation and analysis

The PRISM dataset underwent a data cleaning process using R statistical software (version 4.3.2) [48]. Subsequently, all analyses were performed with the same software to ensure consistency and reproducibility. For categorical variables such as gender, medication, and comorbidities, responses like "prefer not to say" and "other" were recoded as missing, as these options do not provide analyzable clinical information and would introduce noise if retained. Most variables in the dataset had low rates of missing data; however, a notable exception was observed for the NMSQ, which had 31.2% missing responses. This suggests a pattern of survey fatigue, where respondents' engagement decreased as the questionnaire progressed, leading to an increased number of missing responses. This indicates that the missing data in the NMSQ likely follow a missing not at random (MNAR) pattern, suggesting an attrition effect in survey completion. Additionally, the variables "age at diagnosis" and "current medication" exhibited an MNAR pattern, likely due to patients not recalling their age at PD diagnosis or being unaware of their current medication.

Although we initially attempted to address missing values using Multivariate Imputation by Chained Equations (MICE), this approach resulted in models with worse predictive capabilities [49]. Therefore, only cases with complete data for all variables required for modeling were retained, resulting in a final analytic sample of 478 patients. No missing data remained in the analytic sample.

The analytic sample (n=478) comprised 46.9% female and 53.1% male participants, with a mean age of 65.6 years (Standard Deviation (SD) = 9.7 years). The distributions of the PDQ-39 dimensions and Summary Index are depicted in Figure 17. The distribution of the NMSQ score is displayed in Figure 18. Table 3 provides an overview of key variables from the 478 complete observations from PRISM database.

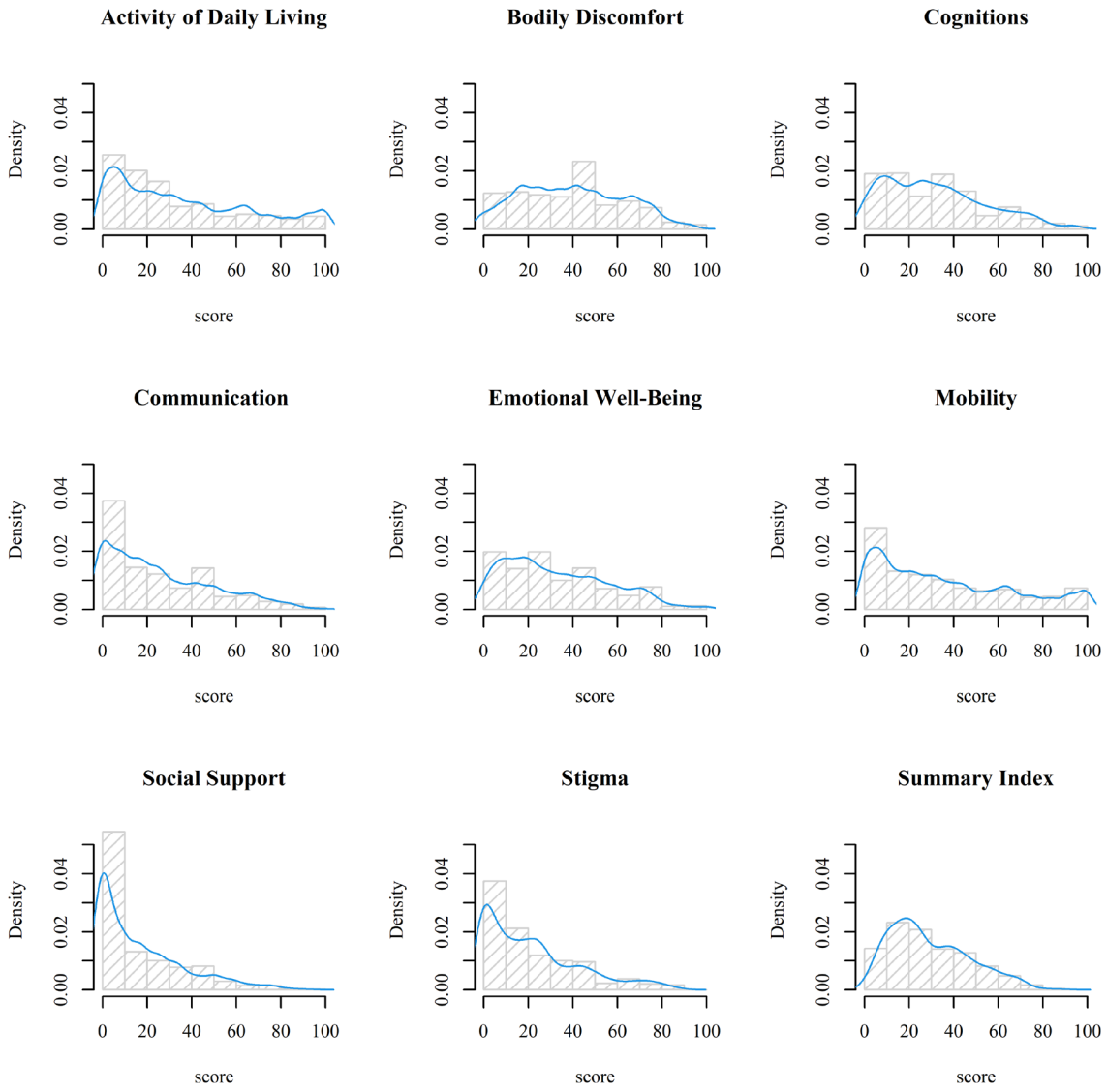

Figure 17: Distribution and density plots of PDQ-39 dimensions and Summary Index scores across 478 patients from the PRISM study.

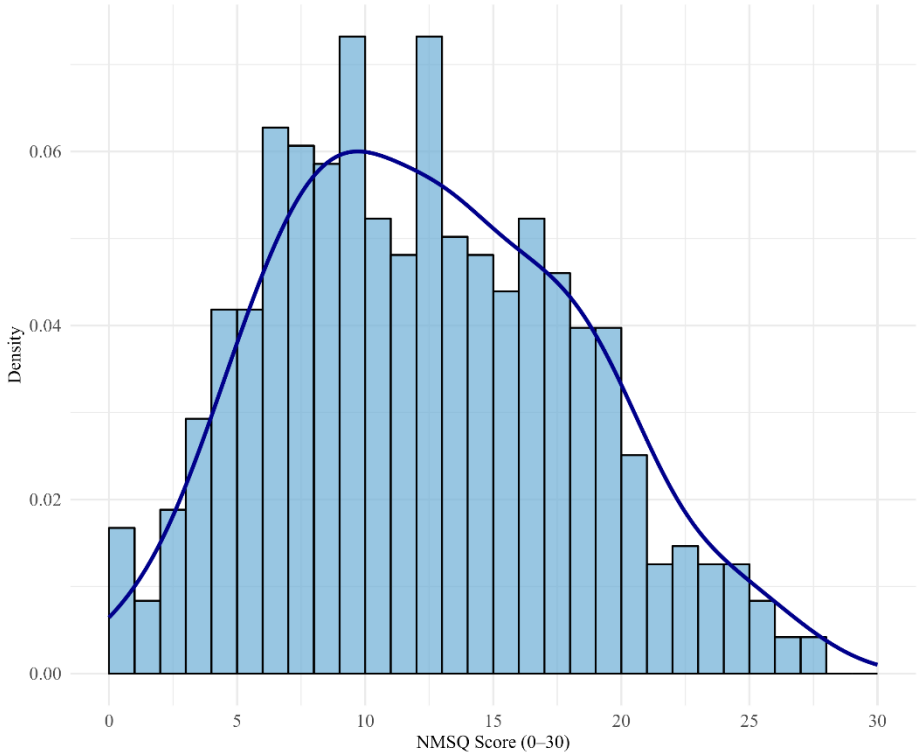

Figure 18: Distribution and density of NMSQ score across 478 patients from the PRISM study.

Table 3: Descriptive Statistics of the complete observations from PRISM Database.

Characteristics	Overall (N=478)
Demographics	
Age, years	
Mean (SD)	65.6 (9.7)
Median [IQR]	66.0 [59.0;72.0]
Min, Max	[38.0;91.0]
Gender	
Female	224 (46.9%)
Male	254 (53.1%)
Country	
France	19 (4.0%)
Germany	50 (10.5%)
Italy	141 (29.5%)
Portugal	45 (9.4%)

Characteristics	Overall (N=478)
Spain	82 (17.2%)
UK	141 (29.5%)
Education	
University	206 (43.1%)
No University	272 (56.9%)
NMS Questionnaire	
Mean (SD)	12.5 (5.9)
Median [IQR]	12.0 [8.0;17.0]
Min, Max	[0.0;28.0]
Comorbidities	
Anxiety	72 (15.1%)
Asthma	25 (5.2%)
Cancer	23 (4.8%)
Chronic Obstructive Pulmonary Disease	10 (2.1%)
Dementia	24 (5.0%)
Depression	97 (20.3%)
Diabetes (type 1 or 2)	29 (6.1%)
Gastric ulcer	22 (4.6%)
Heart issues	36 (7.5%)
High blood pressure	126 (26.4%)
Kidney disease	9 (1.9%)
Liver disease	8 (1.7%)
Peripheral vascular arterial disease	18 (3.8%)
Rheumatic diseases	44 (9.2%)
Stroke	7 (1.5%)
PD Medication (current)	
Amantadine	47 (9.8%)
Anticholinergics	8 (1.7%)
Catechol-O-methyltransferase inhibitor	60 (12.6%)
Dopamine agonists	248 (51.9%)
Levodopa	404 (84.5%)
Monoamine oxidase-B inhibitor	191 (40.0%)
Rivastigmine	17 (3.6%)

Characteristics	Overall (N=478)
PDQ-39	
Activities of Daily Living	
Mean (SD)	30.7 (27.1)
Median [IQR]	20.8 [8.3;45.8]
Min, Max	[0.0;100.0]
Bodily Discomfort	
Mean (SD)	39.9 (23.1)
Median [IQR]	41.7 [18.8;58.3]
Min, Max	[0.0;91.7]
Cognitions	
Mean (SD)	31.7 (22.6)
Median [IQR]	31.2 [12.5;43.8]
Min, Max	[0.0;93.8]
Communication	
Mean (SD)	25.7 (23.5)
Median [IQR]	16.7 [8.3;41.7]
Min, Max	[0.0;100.0]
Emotional Well-Being	
Mean (SD)	33.1 (23.5)
Median [IQR]	29.2 [12.5;50.0]
Min, Max	[0.0;100.0]
Mobility	
Mean (SD)	35.8 (29.9)
Median [IQR]	30.0 [10.0;57.5]
Min, Max	[0.0;100.0]
Social Support	
Mean (SD)	16.7 (19.7)
Median [IQR]	8.3 [0.0;25.0]
Min, Max	[0.0;91.7]
Stigma	
Mean (SD)	21.7 (21.4)
Median [IQR]	18.8 [6.2;31.2]
Min, Max	[0.0;87.5]
Summary Index	
Mean (SD)	29.4 (17.9)

Characteristics	Overall (N=478)
Median [IQR]	25.1 [15.2;41.3]
Min, Max	[1.3;89.3]

IQR, interquartile range; NMS, Non-Motor Symptoms; PD, Parkinson's disease; PDQ-39, Parkinson's Disease Questionnaire-39; SD, standard deviation.

Internal consistency of the PDQ-39 Summary Index and its eight dimensions was assessed using Cronbach's α in the retained sample (n = 478). Scoring followed standard procedures, with all scores ranging from 0 to 100 and higher scores reflecting worse HRQoL. There were no missing items for the PDQ-39 in the analytic sample.

The dataset was then randomly partitioned into training (80%) and validation (20%) sets. This allocation allows for comprehensive model training on a diverse range of responses, critical for capturing the nuanced patterns of HRQoL assessments in PD, while still reserving a portion of the data for essential validation to assess model performance and overfitting.

Model training

We selected machine learning approaches for their scalability and proven ability to capture complex, nonlinear relationships that are often missed by traditional statistical models. This framework supports robust prediction in moderate-sized datasets and can be readily extended to larger cohorts as more data becomes available.

For feature selection, we used the Boruta algorithm. This method is based on Random Forest models and works by checking which variables are truly useful for predicting each PDQ-39 dimension and the Summary Index [73], [74], [75], [76]. Feature selection was performed exclusively on the training set after the initial 80/20 train/validation split, to avoid information leakage. Using this approach, we identified nine distinct sets of features specific to each of the PDQ-39 dimensions and the Summary Index (see Table 6 in the Results section). These nine sets were then utilized in training a diverse range of ML models, namely linear regression, boosted generalized linear model, extreme gradient boosting with linear models and decision trees, ridge regression, neural network regression, LASSO regression, random forest, k-nearest neighbors, support vector machines with radial basis function and polynomial kernels, and Gaussian processes (linear, polynomial, and radial kernels),

incorporating essential pre-processing steps such as centering and scaling with the caret package [50].

To evaluate the performance and stability of these models in a dataset characterized by limited size and high variability, we utilized bootstrap resampling exclusively on the training set, using the standard nonparametric (Efron's) bootstrap method [50], [77]. Bootstrap is a well-established approach that provides a more reliable estimate of the generalizability of machine learning models, particularly in smaller datasets, by generating multiple resampled sets and reducing the bias inherent in model evaluation [78]. Initially, each candidate model was evaluated using 100 bootstrap resamples, and the model with the lowest average Root Mean Squared Error (RMSE) in the training set was selected for each outcome, as RMSE is widely recognized for its effectiveness in assessing predictive performance by penalizing larger errors more than smaller ones [78]. The selected models were then retrained on the training set using 1000 bootstrap resamples to obtain robust estimates of their predictive performance.

The held-out validation set (20% of the data) was reserved strictly for final, unbiased model evaluation and was not subjected to bootstrap resampling. Validation set metrics were calculated as single-point estimates to reflect true out-of-sample predictive performance.

Model evaluation

The performance of the models was evaluated using multiple metrics on both the training set (80%) and the validation set (20%) to assess generalizability and overfitting. The validation set was specifically reserved for the final evaluation to ensure an unbiased assessment of models' performance. Mean Absolute Error (MAE) quantifies the average magnitude of prediction errors, providing a straightforward measure of how much predicted values deviate from true values. RMSE offers insight into the typical magnitude of prediction errors by emphasizing larger errors due to its squared nature, while the Coefficient of Determination (R^2) indicates the proportion of variance in the dependent variable explained by the model, highlighting its explanatory power. In addition to these metrics, scatter plots of predicted versus true values were used to evaluate how well predictions matched the observations, and studentized residual plots were examined for patterns or systematic errors. Together, these quantitative and visual evaluation tools provide a comprehensive assessment of model performance and robustness.

Because the Summary Index captures overall HRQoL and each model targets different aspects of patient experience, we evaluated and compared the predictive performance across all models. This allows us to see whether overall or specific outcomes are predicted with lower error and gives a broader view of which aspects of HRQoL are most predictable in this dataset.

Interpretable machine learning analysis

To elucidate the importance and model reliance on each feature within our predictive models, we employed interpretable ML techniques, with a particular focus on Permutation Feature Importance (PFI). Utilizing the iml package in R, PFI provides deeper insight into the behavior of the ML models [52]. The core concept of PFI is to measure the increase in the model's prediction error after the permutation of each feature's values, which disrupts the relationship between the feature and the true outcome. A considerable increase in error indicates that the model depends heavily on that feature for predictions, while minimal or no change implies that the feature is not as relevant in the predictive process [79]. However, it's crucial to recognize that even features with minimal changes in prediction error may still contribute meaningfully to the overall performance of the model. These variables, while not strongly predictive on their own, can add complementary information and interact with other predictors, which may enhance model generalizability [56]. Feature importance was estimated using the final fitted model (i.e., the best-performing algorithm for each dimension and the Summary Index) applied to the independent validation set. We assessed feature importance in our models using MAE as the error metric for PFI calculations. This process was repeated 100 times to ensure robust estimates, with the results averaged to provide a clearer view of the models' reliance on each variable.

Additionally, to understand the impact of specific variables on our outcomes, we employed Accumulated Local Effects (ALE) plots. ALE plots allow us to visualize the effect of a feature on the predicted outcome while accounting for interactions with other features, particularly in datasets with correlated predictors [79], [80]. By generating ALE plots for the most relevant features identified through PFI, we were able to interpret how changes in these features affect the PDQ-39 dimensions and the Summary Index.

Furthermore, during our analysis, we observed that the NMSQ score (sum of non-motor symptoms reported by the patient) consistently ranked as the most important predictor

across all ML models, when compared to the full set of clinical and demographic variables included in the models (see Table 6 in the Results section).

Exploratory multivariate analysis using FAMD

Given that the NMSQ comprises 30 items assessing various symptoms, we performed a Factor Analysis of Mixed Data (FAMD) to explore the underlying structure and relationships among these items, the eight PDQ-39 dimensions, and the Summary Index. FAMD is particularly suited for mixed data types and simplifies data complexity by identifying principal dimensions that capture the most variance (inertia) in the dataset [81]. This approach elucidates how variables co-vary by grouping associated variables in proximity, thereby revealing underlying patterns and dimensions. FAMD is exploratory and should not be interpreted as indicating predictive importance or causality.

5.4. Results

Psychometric Properties of the PDQ-39

Internal consistency of the PDQ-39 Summary Index and its eight dimensions was assessed using Cronbach's α. Reliability was excellent for the Summary Index (α = 0.96) and generally high for individual dimensions, ranging from α = 0.71 (Bodily Discomfort) to α = 0.96 (Mobility). Detailed results for all dimensions are presented in Table 4.

Table 4: Internal consistency (Cronbach's α) for the PDQ-39 Summary Index and individual dimensions (0–100 scale, n = 478).

Dimensions	Cronbach α
Activities of Daily Living	0.91
Bodily Discomfort	0.71
Cognitions	0.77
Communication	0.83
Emotional Well-Being	0.90
Mobility	0.96
Social Support	0.73
Stigma	0.76
Summary Index	0.96

Cronbach's α ≥ 0.70 indicates acceptable internal consistency; α ≥ 0.80 indicates good consistency; α ≥ 0.90 indicates excellent consistency.

Predictive Model Performance

Performance metrics for the training set are reported as mean (95% CI) [SD] from 1000 bootstrap resamples; validation set values are single point estimates from the independent validation set. Table 5 presents a summary of the performance metrics (MAE, RMSE, and R^2) for both training and validation sets, alongside the selected models for each PDQ-39 dimension and the Summary Index. The Extreme Gradient Boosting with Decision Trees (xgbTree) model was selected as the best-performing model for both the Summary Index and the Activities of Daily Living dimension, while the Gaussian Process with Polynomial Kernel (gausssprPoly) was chosen for all other dimensions.

Table 5: Results.

PDQ-39 dimension / Summary Index	Training			Validation			Model selected
	MAE	RMSE	R^2	MAE	RMSE	R^2	
Activities of Daily Living	15.12 (13.52-16.98) [0.89]	19.40 (17.21-21.80) [1.14]	0.47 (0.35-0.57) [0.06]	18.33	24.41	0.33	xgbTree
Bodily Discomfort	15.27 (13.79-16.78) [0.77]	18.90 (17.09-20.63) [0.91]	0.29 (0.18-0.40) [0.05]	17.82	21.63	0.32	gaussprPoly
Cognitions	12.55 (11.29-13.88) [0.67]	15.64 (13.96-17.14) [0.82]	0.52 (0.41-0.61) [0.05]	12.39	16.20	0.53	gaussprPoly
Communication	14.89 (13.34-16.41) [0.80]	18.88 (16.83-20.91) [1.07]	0.36 (0.25-0.47) [0.06]	13.70	18.44	0.37	gaussprPoly
Emotional Well-Being	13.45 (12.08-14.83) [0.74]	17.24 (15.49-18.99) [0.91]	0.46 (0.36-0.56) [0.05]	14.78	19.50	0.36	gaussprPoly
Mobility	17.96 (16.30-19.64) [0.87]	22.14 (19.99-24.31) [1.12]	0.44 (0.33-0.54) [0.05]	18.02	22.88	0.48	gaussprPoly
Social Support	13.50 (12.16-14.90) [0.70]	17.44 (15.49-19.41) [1.04]	0.17 (0.08-0.28) [0.05]	14.87	19.08	0.26	gaussprPoly
Stigma	15.49 (14.00-17.02) [0.80]	19.40 (17.34-21.54) [1.08]	0.20 (0.09-0.30) [0.05]	14.96	19.34	0.23	gaussprPoly
Summary Index	8.99 (8.04-9.91) [0.49]	11.45 (10.16-12.73) [0.64]	0.58 (0.48-0.67) [0.05]	9.60	12.66	0.56	xgbTree

Values for the training set are reported as mean (95% CI) [SD] from 1000 bootstrap resamples. Validation set metrics are single-point estimates computed on the independent validation set. All model predictors were standardized (centered and scaled) using caret's preprocessing step prior to training. MAE, Mean Absolute Error; RMSE, Root Mean Squared Error; R^2, Coefficient of Determination.

Training and validation results were broadly similar across outcomes, indicating limited overfitting. The Summary Index model outperformed the individual dimensions models, achieving the lowest MAE (9.60) and RMSE (12.66), and the highest R^2 (0.56) in the validation set. The Cognitions dimension model also demonstrated strong predictive performance, with an R^2 of 0.53, MAE of 12.39, and RMSE of 16.20 in the validation set. This indicates that the model explained over half of the variance in Cognitions scores and produced relatively low average prediction errors, suggesting good predictive performance for this dimension.

The Mobility model explained nearly half the variance (R^2=0.48) but had relatively high MAE/RMSE (18.02/22.88), suggesting substantial average prediction errors. This indicates that while the model captures general trends in Mobility, its predictions for individual patients may lack precision due to larger errors. The Emotional Well-Being model had an R^2 of 0.36, with an MAE of 14.78 and RMSE of 19.50. Although the R^2 is lower than that of the Mobility model, the lower MAE and RMSE values indicate smaller average prediction errors, reflecting moderate predictive capability with better precision compared to the Mobility model.

Interestingly, the models for Stigma and Social Support, despite having lower R^2 values (0.23 and 0.26, respectively), exhibited lower MAE and RMSE values compared to models like Mobility. The Stigma model had an MAE of 14.96 and RMSE of 19.34, while the Social Support model had an MAE of 14.87 and RMSE of 19.08. These lower error metrics suggest that, on average, the predictions for Stigma and Social Support were closer to the true values, even though the models explained less variance in the outcomes. This could be attributed to lower variability in the Stigma and Social Support scores, where even small prediction errors can result in lower R^2 values due to the limited variance in these dimensions.

The Activities of Daily Living and Bodily Discomfort models showed lower predictive performance. The Activities of Daily Living model had an R^2 of 0.33, with the highest MAE (18.33) and RMSE (24.41), indicating considerable prediction errors. Similarly, the Bodily Discomfort model had an R^2 of 0.32, an MAE of 17.82, and an RMSE of 21.63. These higher error metrics suggest that these dimensions were more challenging to predict, likely reflecting variability from factors not captured by the available features.

Scatter and Residuals Analysis

In Figure 19, which presents scatter plots of predicted versus true values, the Summary Index model demonstrates the best fit, with predictions closely clustering around the red dashed line, reflecting its superior performance metrics. Cognitions, Emotional Well-Being, and Communication also show an adequate fit, with reasonable alignment between predicted and true values. However, for Mobility, while the model performs well for lower true values, it tends to underpredict higher scores. Similarly, the models for Bodily Discomfort, Social Support, and Stigma slightly overpredict lower values and underpredict higher ones, indicating a regression toward the mean. The Activities of Daily Living model emerges as having the highest prediction error, with substantial deviations from the red dashed line, signaling lower predictive performance.

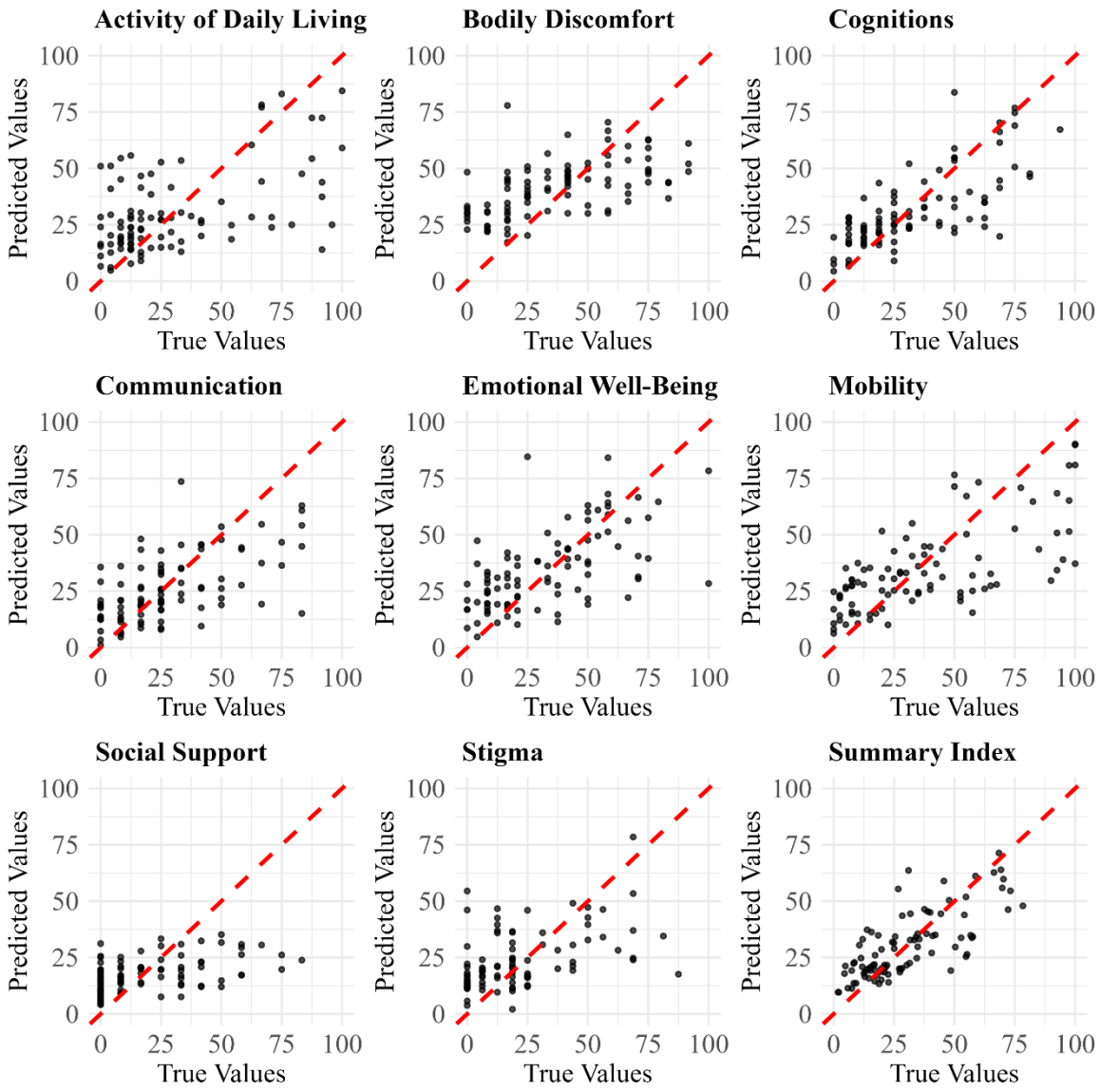

Figure 19. Predicted vs True Values for PDQ-39 Dimensions and Summary Index. The black dots represent individual predictions, and the red dashed line represents the line of perfect prediction where predicted values would equal true values. Predictions were generated on the independent validation set using the following fitted algorithms: xgbTree for Summary Index and Activities of Daily Living; and gaussprPoly for all other dimensions.

In Figure 20, which presents residual plots for each model, we observe distinct patterns in model performance. The residuals for the Summary Index are evenly distributed around zero with low spread, reflecting a well-fitted model with strong predictive performance. Similarly, the Cognitions and Emotional Well-Being models show residuals that are clustered around zero, indicating a solid fit and minimal error variance. Both Communication and Bodily Discomfort exhibit a wider spread of residuals from zero, although their distribution

remains centered, indicating moderate performance. While these models capture some trends in the data, their larger residual spread suggests less precision compared to models like the Summary Index and Cognitions. The Mobility model demonstrates a good fit for lower predicted values, but exhibits an asymmetrical spread at higher values, with a greater number of residuals above zero, suggesting underprediction for larger outcomes. The models for Activities of Daily Living and Stigma exhibit larger residuals across the prediction spectrum. In particular, Activities of Daily Living shows consistently large errors, indicating poor performance, while Stigma and Social Support display heteroscedasticity, with increasing residual variance as predicted values rise, indicating that the models become less reliable when predicting higher values.

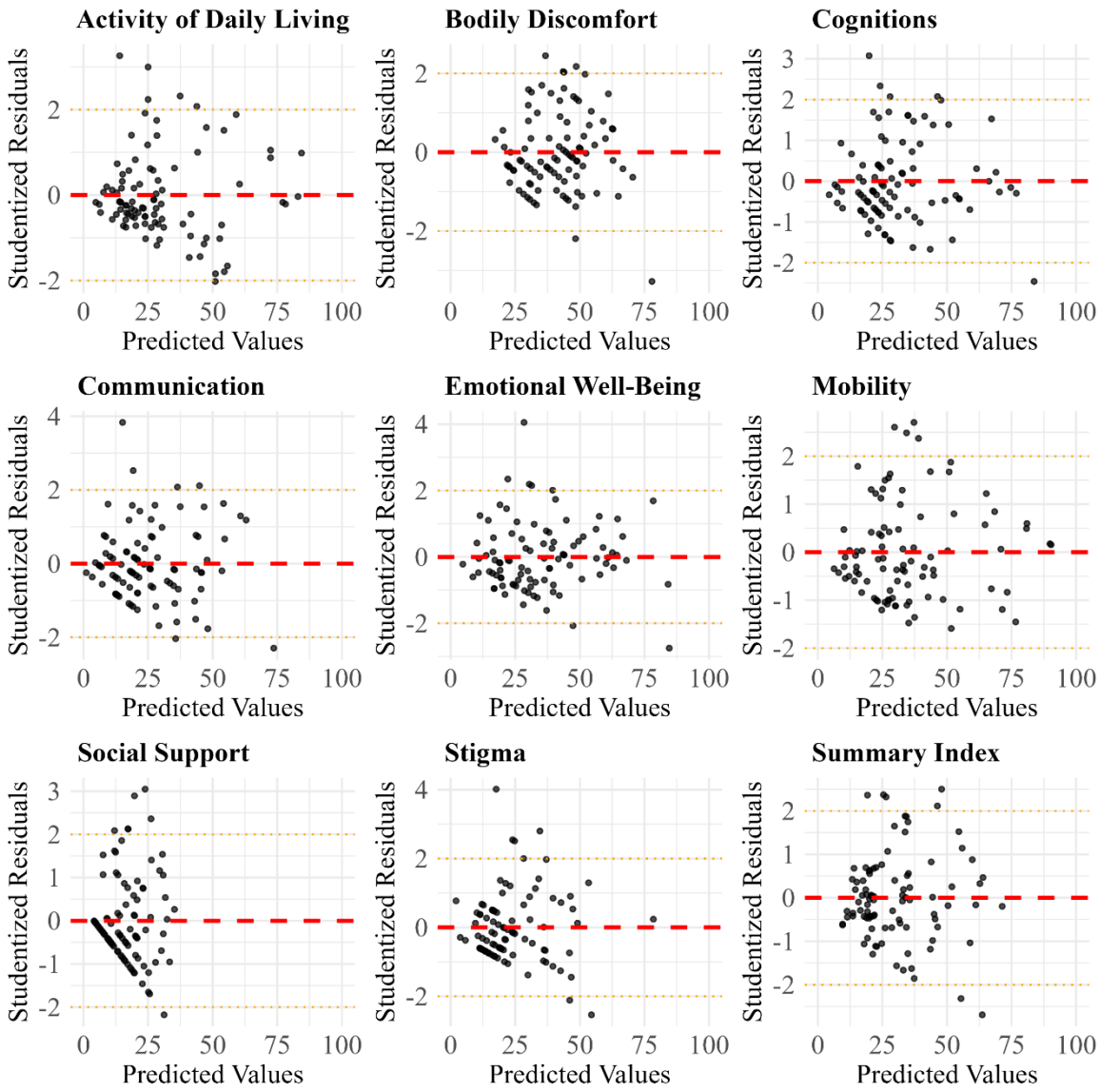

Figure 20. Studentized residuals for each PDQ-39 dimension and Summary Index model, plotted against predicted values from the validation set. The red dashed line indicates zero residual error. The orange dotted lines mark ±2 standard deviations. Residuals were generated on the independent validation set using the following fitted algorithms: xgbTree for Summary Index and Activities of Daily Living; and gaussprPoly for all other dimensions.

Overall, the scatter and residual analyses confirm varying model performance: Summary Index, Cognitions, and Emotional Well-Being demonstrate strong predictive capabilities. Communication and Bodily Discomfort show moderate performance, and Activities of Daily Living exhibits the poorest performance. Four models display specific patterns: Communication, Social Support, and Stigma exhibit heteroscedasticity, while Mobility shows asymmetric residual distribution.

Feature Importance

In Figure 21, the feature importance analysis highlights the universal relevance of the NMSQ score across all models, consistently ranking as the top predictor in every PDQ-39 dimension and the Summary Index. These findings reinforce the central predictive role of NMSQ and align with prior evidence that greater non-motor symptom burden is highly associated with poorer HRQoL [70], [71], [82]. Other clinical variables also showed importance in specific dimensions. Age was particularly relevant in both Mobility and Activities of Daily Living. Notably, for Mobility, age had nearly the same importance as the NMSQ score, indicating its critical role in predicting outcomes in this dimension. Dementia emerged as an important predictor in the Cognitions dimension, aligning with the well-established link between cognitive decline and dementia in patients with PD. Despite these findings, the overall predictive contribution of these clinical variables was less pronounced than that of the NMSQ score. Feature importance values for all selected variables are displayed in Table 6.

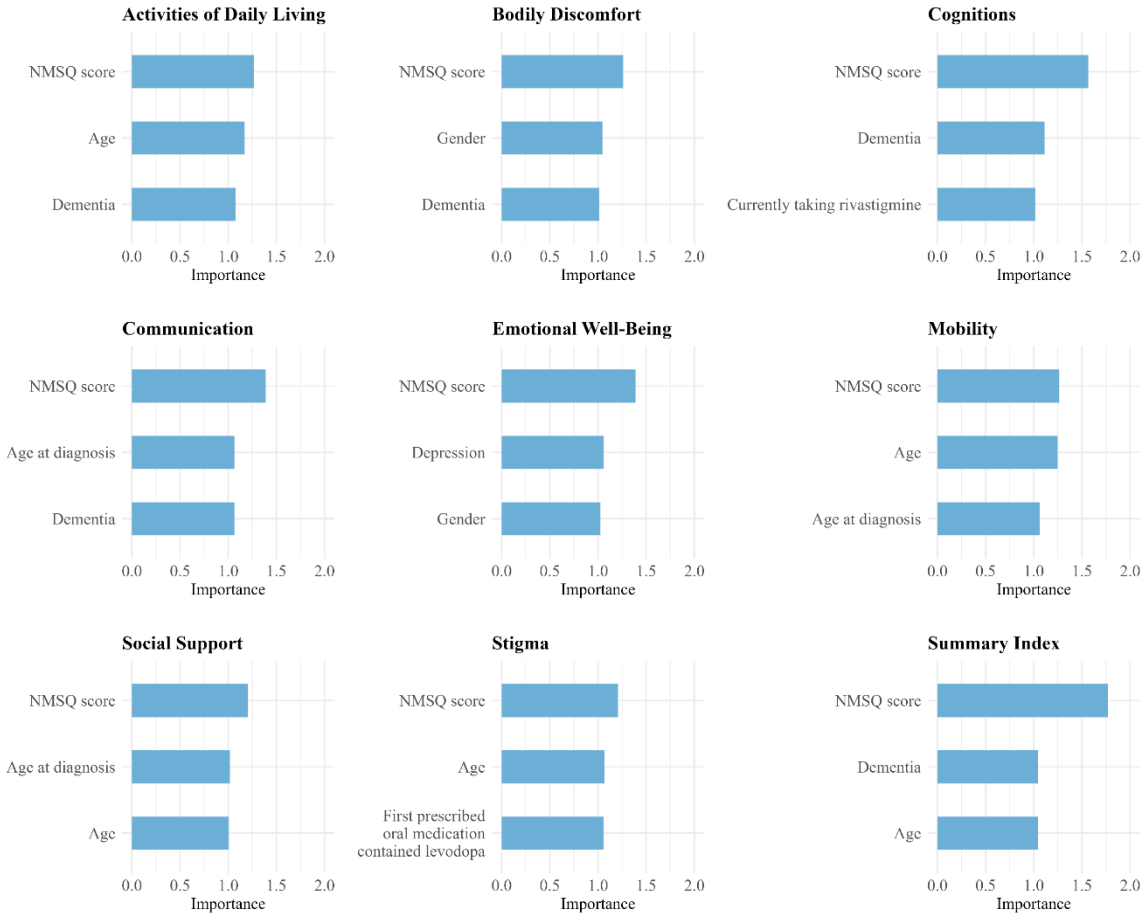

Figure 21. Feature importance for predictive models of PDQ-39 dimensions and Summary Index. Importance was assessed as the mean increase in MAE after permuting each feature (100 permutations), using the final fitted models applied to the independent validation set.

Table 6: Features selected by Boruta algorithm and their permutation-based importance in each PDQ-39 dimension and Summary Index model.

Variables	Importance	Permutation error
Activities of Daily Living model		
NMSQ score	1.27	23.21
Age	1.17	21.47
Dementia	1.08	19.74
Age at diagnosis	1.05	19.27
Currently taking Monoamine oxidase-B inhibitor	1.02	18.61
Cancer	1.01	18.45
Number of comorbidities	1.00	18.33
Peripheral vascular arterial disease	1.00	18.33
Gastric ulcer	1.00	18.33
Anxiety	1.00	18.29
Currently taking rivastigmine	1.00	18.26
Gender	0.99	18.18
Country	0.98	18.00
Bodily Discomfort model		
NMSQ score	1.26	22.48
Gender	1.05	18.62
Dementia	1.01	18.06
Number of comorbidities	1.00	17.90
First prescribed oral medication contained levodopa	1.00	17.85
Age at diagnosis	1.00	17.84
Age	1.00	17.83
Cancer	1.00	17.78
Anxiety	1.00	17.76

Variables	Importance	Permutation error
Cognitions model		
NMSQ score	1.57	19.39
Dementia	1.11	13.78
Currently taking rivastigmine	1.02	12.58
Age	1.02	12.58
Country	1.01	12.50
Depression	1.01	12.48
Heart issues	1.01	12.46
Gastric ulcer	1.01	12.46
Peripheral vascular arterial disease	1.01	12.46
Age at diagnosis	1.00	12.44
Number of comorbidities	1.00	12.44
First prescribed oral medication contained levodopa	1.00	12.40
Communication model		
NMSQ score	1.39	19.04
Age at diagnosis	1.06	14.59
Dementia	1.06	14.57
Age	1.02	13.95
Country	1.01	13.89
Depression	1.00	13.71
Anxiety	1.00	13.69
Currently taking rivastigmine	1.00	13.65
Currently taking Monoamine oxidase B inhibitor	0.99	13.61
First prescribed oral medication contained levodopa	0.99	13.61
Cancer	0.99	13.59
Number of comorbidities	0.99	13.56

Variables	Importance	Permutation error
Emotional Well-Being model		
NMSQ score	1.39	20.53
Depression	1.06	15.70
Gender	1.03	15.16
Number of comorbidities	1.02	15.03
Anxiety	1.01	14.99
Country	1.01	14.96
Dementia	1.00	14.85
Age	1.00	14.74
Peripheral vascular arterial disease	0.99	14.60
Age at diagnosis	0.98	14.47
Mobility model		
NMSQ score	1.27	22.80
Age	1.25	22.48
Age at diagnosis	1.06	19.09
Dementia	1.05	18.91
First prescribed oral medication contained levodopa	1.02	18.42
Currently taking rivastigmine	1.01	18.29
Depression	1.01	18.27
Peripheral vascular arterial disease	1.01	18.21
Currently taking Levodopa	1.01	18.12
Anxiety	1.00	18.09
Rheumatic diseases	1.00	17.99
Number of comorbidities	1.00	17.97

Variables	Importance	Permutation error
Social Support model		
NMSQ score	1.21	17.93
Age at diagnosis	1.02	15.12
Age	1.01	14.95
Anxiety	1.00	14.92
Number of comorbidities	1.00	14.89
Gastric ulcer	1.00	14.87
Depression	1.00	14.87
Stigma model		
NMSQ score	1.21	18.07
Age	1.07	15.96
First prescribed oral medication contained levodopa	1.06	15.82
Currently taking rivastigmine	1.04	15.62
Anxiety	1.04	15.55
Kidney disease	1.04	15.49
Age at diagnosis	1.01	15.14
Summary Index model		
NMSQ score	1.77	17.00
Dementia	1.04	10.03
Age	1.04	10.02
Age at diagnosis	1.04	9.94
Cancer	1.01	9.69
Depression	1.00	9.62
Peripheral vascular arterial disease	1.00	9.62
Anxiety	1.00	9.60
Gastric ulcer	1.00	9.60

Variables	Importance	Permutation error
Currently taking rivastigmine	1.00	9.60
Country	1.00	9.59
Number of comorbidities	1.00	9.59

Features selected through Boruta for each of the models. Feature Permutation Importance: Mean increase in MAE after permuting the feature (100 permutations); permutation error: standard deviation of the MAE increase across permutations.

Accumulated Local Effects

The ALE plots in Figure 22 illustrate the effect of increasing NMSQ scores on the predicted values for each dimension and the Summary Index. ALE plots allow us to visualize the isolated effect of a single feature, in this case, the NMSQ scores, on the models' predictions while accounting for interactions with other features. The x-axis represents the NMSQ score values, and the y-axis represents the accumulated local effect of these scores on the predicted outcomes, centered around zero, where zero indicates no impact on the predicted outcome. An upward trend in ALE values indicates that as NMSQ scores increase, the model predicts worsening HRQoL outcomes, reflecting the detrimental impact of non-motor symptoms.

In general, higher NMSQ scores were associated with worsening HRQoL, as evidenced by increasing ALE values across dimensions. However, the pattern and strength of this relationship varied by dimension. Both the Activities of Daily Living dimension and the Summary Index displayed pronounced stepwise patterns in the ALE plots, which is consistent with the behavior of the tree-based extreme gradient boosting models (xgbTree) selected for these two outcomes. These models create plateaus and jumps in predictions, which appear as steps in the ALE curves. The discrete NMSQ score scale (0–30) may further accentuate this appearance. In contrast, the Stigma dimension showed a steep increase at higher NMSQ scores, highlighting the pronounced impact of non-motor symptoms on patients' perceived social standing and experiences of stigma. For the remaining dimensions, the relationship between NMSQ scores and predicted outcomes appeared fairly linear, indicating a consistent association between higher NMSQ scores and worse predicted HRQoL in these dimensions.

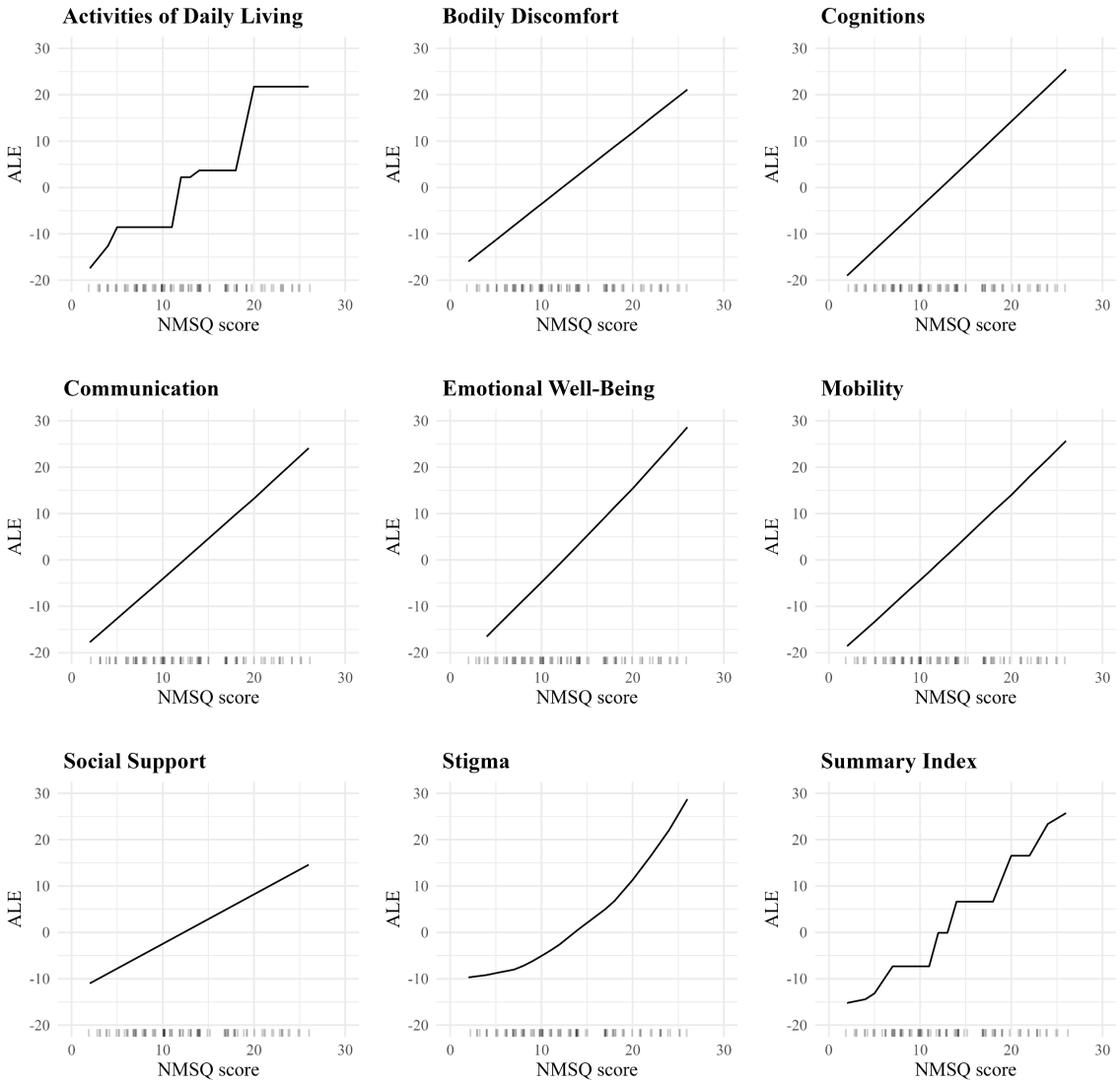

Figure 22. ALE plots of NMSQ scores on PDQ-39 dimensions and Summary Index, computed from the final fitted models and evaluated on the independent validation set.

Factor Analysis of Mixed Data

In Figure 23, the FAMD plot visualizes associations in a low-dimensional factor space: closer proximity indicates stronger associations (co-variation), whereas greater distances reflect weaker associations. Summary Index, Cognitions, and Emotional Well-Being are the most distant from other variables. Their isolation may reflect broader, non-specific associations with multiple NMS, rather than ties to particular items. Activities of Daily Living, Communication, and Mobility appear relatively close to each other. Although they remain distant from individual NMS, their proximity suggests some level of association (co-variation)

among these dimensions. This suggests that patients' daily functioning, communication, and mobility may reflect shared underlying factors, rather than specific non-motor symptoms. Social Support, Bodily Discomfort, and Stigma form a cluster near Anxiety, indicating an association. In particular, Social Support is positioned very close to Anxiety, indicating a strong association between perceived social support and anxiety levels in patients with PD.

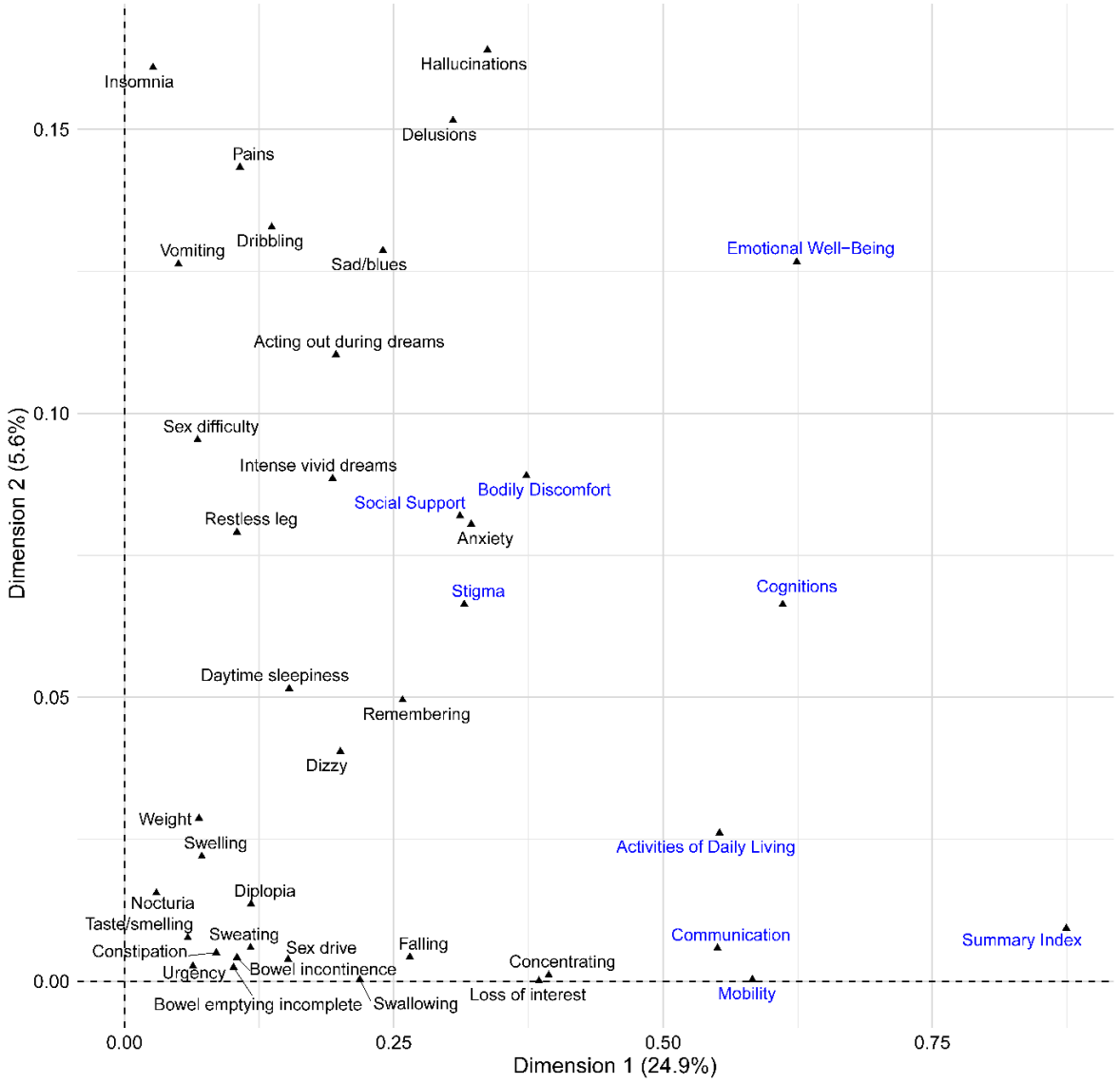

Figure 23. FAMD of NMSQ Items, PDQ-39 dimensions, and Summary Index. PDQ-39 dimensions and Summary Index are colored in blue. Axis percentages show the share of total inertia explained (Dim1 24.9%, Dim2 5.6%; 30.5% combined). Points that are closer indicate stronger associations; this is exploratory and not an evidence of causality or predictive importance.

5.5. Discussion

Principal Results

Our study demonstrates that while the Summary Index and the Cognitions dimension can be predicted with relatively low prediction error using common clinical variables and NMS, other dimensions present more challenges. Models for these two areas showed strong performance, indicating that overall HRQoL and cognitive aspects can be effectively estimated. In contrast, dimensions such as Mobility, Activities of Daily Living, and Bodily Discomfort were more difficult to predict, suggesting that additional variables may be necessary to improve model fit for these aspects of HRQoL.

The NMSQ score emerged as the most important feature across all ML models, underscoring the pervasive association of NMS with HRQoL in patients with PD. This aligns with previous studies highlighting the substantial burden of NMS on patients' daily lives and emphasizes the need for comprehensive management strategies addressing both motor and non-motor symptoms in PD [70], [71], [82]. Apart from the NMSQ score, certain clinical variables were important predictors for specific dimensions. Age was notably important for Mobility and Activities of Daily Living, reflecting the impact of age-related decline on functional abilities. Dementia considerably contributed to the prediction of the Cognitions dimension, aligning with established associations between cognitive decline and dementia in PD [55].

The ALE plots revealed that higher NMSQ scores are associated with worsening HRQoL across all dimensions, with some exhibiting nonlinear relationships. Specifically, the Stigma dimension’s predicted scores increased nonlinearly, with steeper rises at higher NMSQ scores. This indicates that as the burden of NMS becomes more severe, feelings of stigma disproportionately intensify, highlighting the escalating impact of NMS on this aspect of HRQoL.

Furthermore, the FAMD visualization offered a nuanced understanding of the associations between NMS, PDQ-39 dimensions, and the Summary Index. Summary Index, Cognitions, and Emotional Well-Being were positioned distantly from other variables, suggesting broader, non-specific associations with multiple NMS rather than ties to particular symptoms. In contrast, Social Support, Bodily Discomfort, and Stigma clustered near

Anxiety, indicating an association. This suggests that targeted interventions addressing specific NMS may benefit certain HRQoL dimensions.

Overall, these findings reflect the multifaceted nature of HRQoL in PD. The varying model performance suggests that incorporating additional variables may be necessary to improve prediction for certain dimensions of HRQoL.

Comparison to Previous Work

Our findings corroborate prior research that emphasizes the considerable impact of NMS on HRQoL in patients with PD [70], [71], [82]. By employing ML models to predict individual PDQ-39 dimensions, our study extends previous work that predominantly focused on overall HRQoL. This approach provides a more detailed understanding of how specific symptoms and clinical variables are associated with distinct aspects of patients' lives. Moreover, we addressed potential overfitting issues identified in our earlier study [72] by refining our modeling techniques and including additional variables, resulting in more robust and generalizable models. The use of interpretable ML methods offers novel insights into the complex relationships between NMS and HRQoL dimensions, contributing to a deeper comprehension of factors affecting patient well-being.

Limitations

While our study provides important insights, a couple of limitations should be acknowledged. First, there was a considerable amount of missing data, particularly towards the end of the database, suggesting that participants experienced survey fatigue and stopped completing the survey. We attempted data imputation to address the missing values, but this resulted in worse model performance. Consequently, we excluded incomplete observations, resulting in a reduced sample size, which may have affected the generalizability and robustness of our models. Second, the models were developed using data from European patients, which may limit their applicability to other populations.

Future Directions

Future studies should aim to include more diverse populations from different demographic and cultural backgrounds to enhance the generalizability of the models. Incorporating psychosocial factors such as socioeconomic status, social support networks, and environmental influences may reduce prediction error for HRQoL dimensions like Stigma

and Social Support. Additionally, using larger datasets or exploring methods to effectively handle missing data could enhance predictions for challenging dimensions. Longitudinal studies would also be beneficial to model disease progression and capture changes in HRQoL over time.

5.6. Conclusions

Our study confirms that NMS are critical predictors of HRQoL in patients with PD. Utilizing ML models, we successfully predicted the Summary Index and the Cognitions dimension with relatively low prediction error, underscoring the profound impact of NMS on both overall and cognitive aspects of HRQoL. However, predicting other dimensions such as Mobility, Activities of Daily Living, and Bodily Discomfort remains challenging, indicating the need for additional variables to enhance predictive performance. The crucial role of NMS for predicting HRQoL emphasizes the necessity for integrated management approaches that address both motor and non-motor symptoms in PD.

Despite limitations related to missing data and the study's European-centric dataset, our findings offer valuable insights into the complex interplay between NMS and HRQoL. By identifying specific clinical variables associated with certain HRQoL dimensions, our study contributes to the development of more targeted and effective interventions aimed at improving patient outcomes. Future research should focus on incorporating diverse populations, psychosocial factors, and longitudinal data to further refine predictive models and enhance HRQoL outcomes for patients with PD.

In conclusion, this study advances our understanding of how NMS affect various dimensions of HRQoL in PD, demonstrating the potential of ML models in personalizing patient care and informing clinical decision-making. These insights pave the way for more nuanced approaches to managing PD, ultimately aiming to improve the quality of life for those affected by this debilitating disease.

Ethics approval and consent to participate: Not Applicable.

Availability of data and materials: The data used in this study were obtained from the Parkinson's Real-world Impact Assessment (PRISM) database (**www.prism.bial.com**), accessed on 18 February 2022.

Competing interests: The authors declare that they have no competing interests.

Funding: This work has been conducted under the scope of and funded by the PhD Program in Health Data Science of the Faculty of Medicine of the University of Porto, Portugal, **heads.med.up.pt**.

Authors' contributions: DM conceptualized the study, developed the methodology, conducted the analysis, and drafted the manuscript. TTG and ASB contributed to conceptualization, methodology, supervision, validation, and manuscript review. JM participated in validation and reviewed the manuscript. LA reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgements: The PRISM study and database were funded by BIAL—Portela and Cª, S.A., designed in collaboration with the Cure Parkinson's Trust, an advocacy group based in the United Kingdom (UK), and reviewed by the PRISM Steering Committee.

6. Discussion

This thesis explored the utilization of ML methods to predict HRQoL in patients with PD, measured using the PDQ-39 tool. The research progressed from an initial binary classification approach to regression-based general modeling, and ultimately to detailed domain-specific model predictions. Each successive stage of modeling introduced important methodological refinements, leading to a consistent enhancement in predictive power. Given the inherent complexity and clinical heterogeneity of PD symptoms and disease trajectories, this methodological refinement was essential. By advancing from general predictive classifications to precise domain-specific models, this study significantly enhanced the potential clinical utility and interpretability of the findings, facilitating more nuanced insights into patient-specific quality-of-life outcomes and supporting more personalized and informed clinical decision-making.

In Paper 1, we explored the feasibility of applying ML to categorize patients based on their PDQ-39 SI scores by dividing them into high- and low-score groups using a threshold identified by an expectation-maximization algorithm. This threshold, set at a PDQ-39 SI value of 31, enabled binary stratification of individuals. Using common clinical and demographic features as inputs, classification models, such as LDA, SVM, and RF, demonstrated encouraging performance, with all three achieving good discriminative ability in both the training and validation datasets. These findings confirm that routinely available clinical data can be effectively used to distinguish patients based on their HRQoL. Despite these promising outcomes, two key limitations have been identified. First, the absence of clinically validated cut-off values for the PDQ-39 SI reduced the interpretability of the classification and limited its application in clinical practice. Second, the dichotomization of continuous outcomes resulted in a substantial loss of information, restricting the granularity of predictions. Consequently, these issues motivated the transition toward continuous regression modeling in subsequent studies to preserve the richness of the outcome variable and enhance clinical relevance.

To overcome these constraints, Paper 2 implements a predictive model using ensemble regression methods. This approach significantly improved interpretability using explainable ML techniques, such as feature importance analyses and ALE plots. The results highlighted key predictive factors, including age at diagnosis, dementia, patient age, and geographic location (country), reinforcing the established literature on their influence on PD. Although

the ensemble demonstrated robust predictive performance on the training dataset (MAE = 4.82, RMSE = 8.09, R^2 = 0.75), its accuracy declined upon external validation (MAE = 11.22, RMSE = 13.99, R^2 = 0.36), suggesting potential overfitting. The observed discrepancy between the training and validation performances prompted us to reconsider the adequacy of using a single global measure, such as the PDQ-39 SI, to capture the diverse impacts of PD on HRQoL. Therefore, we hypothesized that disaggregating the PDQ-39 into its eight constituent domains and incorporating domain-specific predictors could further improve predictive accuracy, enhance generalizability, and mitigate overfitting.

The research described in Paper 3 advanced the work by developing separate prediction models for each of the eight PDQ-39 domains: Mobility, Activities of Daily Living, Emotional Well-Being, Stigma, Social Support, Cognitions, Communication, and Bodily Discomfort, thereby aligning the model architecture with the multidimensional nature of HRQoL. Extreme Gradient Boosting was identified as the optimal learner for the Activities of Daily Living domain and the Summary Index, whereas the Gaussian Process with Polynomial Kernel performed best for the remaining domains. The PDQ-39 SI demonstrated the strongest overall performance in the validation set, combining the highest R^2 (0.56) with the lowest error (MAE 9.65; RMSE 12.71). Among individual domains, Cognitions performed well, achieving an R^2 of 0.53 with relatively moderate errors (MAE 12.39; RMSE 16.20). Mobility also showed substantial explained variance (R^2 0.48) but presented relatively large prediction errors (MAE 18.02; RMSE 22.88). Emotional Well-Being (R^2 0.36; MAE 14.78; RMSE 19.50) and Communication (R^2 0.35; MAE 13.91; RMSE 18.73) displayed moderate explanatory power with intermediate errors. In contrast, Stigma (R^2 0.23; MAE 14.96; RMSE 19.34) and Social Support (R^2 0.26; MAE 14.87; RMSE 19.08) explained limited variance yet maintained lower average errors, likely due to narrower distributions. Activities of Daily Living (R^2 0.33; MAE 18.22; RMSE 24.28) and Bodily Discomfort (R^2 0.32; MAE 17.82; RMSE 21.63) proved to be the most challenging, showing lower explained variance and higher prediction errors. Across all domains, the NMS Questionnaire consistently emerged as the strongest predictor, emphasizing the critical impact of non-motor symptoms on HRQoL. Age was the second-most important predictor for Mobility and Activities of Daily, reflecting the influence of age-related functional decline, whereas dementia was similarly influential in the Cognition domain. Although domain-specific modeling reduced the overfitting observed in Paper 2, the relatively modest predictive accuracy in motor-focused domains (Mobility, Activities of Daily Living, and Bodily Discomfort) suggests that clinical and demographic variables alone do not fully capture their complexity.

Rigorous methodologies were applied throughout this study. Nevertheless, the cross-sectional and region-specific nature of our dataset inevitably restricted the generalizability of the ML models. Therefore, future studies should prioritize longitudinal data collection and integrate diverse international cohorts to track changes in HRQoL in PD over time. Moreover, incorporating additional psychosocial factors could further refine predictive accuracy, particularly in domains such as Stigma, Social Support, and Emotional Well-Being. Additionally, broadening the predictive scope beyond the PDQ-39 outcomes and including other variables would substantially enhance the practical applicability of ML models. For instance, the integration of objective digital biomarkers derived from wearable technologies capturing metrics such as gait dynamics, tremor severity, and sleep disturbances could significantly improve predictive accuracy in inadequately captured domains, notably Mobility, Activities of Daily Living, and Bodily Discomfort.

Overall, the stepwise improvements from Paper 1 through Paper 3 illustrate the feasibility and incremental benefit of domain-specific ML for personalized HRQoL prediction in PD and lay the groundwork for future multimodal longitudinal modelling frameworks. Clinically, these predictive models hold promises for supporting personalized patient care, facilitating the early identification of individuals at risk for significant quality-of-life deterioration, and guiding targeted therapeutic interventions, such as tailored rehabilitation programs, caregiver support structures, and individualized pharmacological management strategies. The incorporation of PROMs, such as the PDQ-39, not only aligns with evolving clinical standards but also reinforces patient-centered care, ensuring that treatment decisions reflect outcomes that genuinely matter to patients. Indeed, there has been increasing regulatory emphasis on integrating patient perspectives into clinical trial outcomes [83], [84]. Regulatory bodies, such as the European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA) have strongly advocated for the systematic inclusion of PROMs in clinical trials [61], [85], [86], [87], [88]. The FDA also provided explicit guidance on using PROMs to support medical product labeling claims, highlighting their critical role in capturing treatment benefits directly meaningful to patients [89]. Similarly, the EMA continues to emphasize the importance of PROMs in informing clinical and regulatory decision-making processes [62]. This recognition underscores the value and necessity of our research in effectively predicting PROM-based outcomes, which closely aligns with current regulatory and clinical research trends.

At the health-system level, these results demonstrate the potential of ML-based PROMs prediction to inform resource allocation and enable proactive, data-driven care pathways for PD. While promising, the models will require further optimization and prospective validation to ensure consistent, high-quality performance. Prospectively validated algorithms can enable clinicians to direct rehabilitation, caregiver support, and medication reviews toward patients most at risk of HRQoL decline, thereby enhancing outcomes and making more efficient use of healthcare resources.

To translate this potential into routine clinical practice, the next step should involve a pilot phase integrating predictive models into electronic health records, allowing clinicians to evaluate algorithm-generated predictions alongside traditional clinical assessments without immediately influencing decision-making. Additionally, anticipated practical challenges such as clinician acceptance, workflow integration, and technical infrastructure requirements should be proactively addressed to facilitate successful real-world implementation.

Building on insights gained from the initial pilot, future research should prioritize strengthening the generalizability of the model and demonstrating its real-world clinical utility. This entails external validation in multicenter cohorts, prospective testing within routine care workflows, and continuous performance monitoring across diverse demographic and socioeconomic groups to ensure equitable accuracy. Such efforts will translate the current proof-of-concept into a reliable decision-support tool that can be confidently adopted at the point-of-care.

7. Conclusion

This study confirms the feasibility and value of applying ML methods to predict HRQoL outcomes in PD based on PROMs, demonstrating a clear methodological progression from binary classification to domain-specific predictions. A key insight from these models is that, among the available variables, NMS exert the strongest influence on the predicted HRQoL, underscoring their clinical significance and the inherently multidimensional nature of patient well-being.

From a healthcare perspective, predictive models built on PROMs can enhance patient-centered care by enabling the early identification of individuals at risk of quality-of-life deterioration, guiding targeted therapeutic interventions, and informing resource-allocation decisions. In line with regulatory trends, and with PROMs now recognized as key endpoints in clinical trials, these models help embed the patient perspective within routine evaluation and management strategies.

However, achieving full clinical utility will require addressing limitations in data generalizability. Future research should prioritize longitudinal, multicenter studies that include diverse populations and incorporate additional biomarkers to strengthen predictive accuracy and reliability. Ensuring equitable model performance across demographic and socioeconomic groups through rigorous external validation and continuous monitoring is essential.

By overcoming these challenges, ML-driven PROM prediction can mature from proof-of-concept into a reliable decision-support tool that improves the personalization of care and optimizes outcomes for people living with PD.

8. References

[1] “Parkinson’s Foundation.” Accessed: Apr. 11, 2025. [Online]. Available: https://www.parkinson.org/

[2] “Parkinson’s Europe.” Accessed: Apr. 11, 2025. [Online]. Available: https://parkinsonseurope.org/

[3] B. R. Bloem, M. S. Okun, and C. Klein, “Parkinson’s disease,” *The Lancet*, vol. 397, no. 10291, pp. 2284–2303, Jun. 2021, doi: 10.1016/S0140-6736(21)00218-X.

[4] E. R. Dorsey, T. Sherer, M. S. Okun, and B. R. Bloemd, “The Emerging Evidence of the Parkinson Pandemic,” *J. Parkinsons Dis.*, vol. 8, no. Suppl 1, p. S3, 2018, doi: 10.3233/JPD-181474.

[5] V. L. Feigin *et al.*, “Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015,” *Lancet Neurol.*, vol. 16, no. 11, pp. 877–897, Nov. 2017, doi: 10.1016/S1474-4422(17)30299-5/ATTACHMENT/CD9DB039-439A-4F44-9963-3AA014FE39DC/MMC1.PDF.

[6] M. Ruiz-Lopez *et al.*, “Diagnostic delay in Parkinson’s disease caused by PRKN mutations,” *Parkinsonism Relat. Disord.*, vol. 63, pp. 217–220, Jun. 2019, doi: 10.1016/J.PARKRELDIS.2019.01.010.

[7] M. J. Armstrong and M. S. Okun, “Diagnosis and Treatment of Parkinson Disease: A Review,” *JAMA*, vol. 323, no. 6, pp. 548–560, Feb. 2020, doi: 10.1001/JAMA.2019.22360.

[8] Z. A. Macchi *et al.*, “Patient and caregiver characteristics associated with caregiver burden in Parkinson’s disease: a palliative care approach,” *Ann. Palliat. Med.*, vol. 9, no. Suppl 1, pp. S24–S33, Feb. 2020, doi: 10.21037/APM.2019.10.01.

[9] S. M. Fereshtehnejad and R. B. Postuma, “Subtypes of Parkinson’s Disease: What Do They Tell Us About Disease Progression?,” *Curr. Neurol. Neurosci. Rep.*, vol. 17, no. 4, pp. 1–10, Apr. 2017, doi: 10.1007/S11910-017-0738-X/FIGURES/1.

[10] F. Albrecht *et al.*, “Unraveling Parkinson’s disease heterogeneity using subtypes based on multimodal data,” *Parkinsonism Relat. Disord.*, vol. 102, pp. 19–29, Sep. 2022, doi: 10.1016/J.PARKRELDIS.2022.07.014.

[11] T. Petrijan and M. Menih, “Demographic and Clinical Characteristics in Different Motor Subtypes of Parkinson’s Disease: How Well Do the Findings Fit Within the Framework of Existing Hypotheses?,” *Neurology International 2025, Vol. 17, Page 51*, vol. 17, no. 4, p. 51, Mar. 2025, doi: 10.3390/NEUROLINT17040051.

[12] M. Filidei, L. Marsili, and C. Colosimo, “Do Parkinson’s Disease clinical subtypes really exist?,” *Neurol. Neurochir. Pol.*, Mar. 2025, doi: 10.5603/PJNNS.103572.

[13] H. Iwaki *et al.*, “Differences in the Presentation and Progression of Parkinson’s Disease by Sex,” *Mov. Disord.*, vol. 36, no. 1, p. 106, Jan. 2020, doi: 10.1002/MDS.28312.

[14] P. Martinez-Martin *et al.*, “Neuropsychiatric symptoms and caregiver’s burden in Parkinson’s disease,” *Parkinsonism Relat. Disord.*, vol. 21, no. 6, pp. 629–634, Jun. 2015, doi: 10.1016/J.PARKRELDIS.2015.03.024.

[15] P. Martínez-Martín *et al.*, “Caregiver burden in Parkinson’s disease,” *Mov. Disord.*, vol. 22, no. 7, pp. 924–931, May 2007, doi: 10.1002/MDS.21355.

[16] A. Schrag, M. Jahanshahi, and N. Quinn, “How Does Parkinson’s Disease Affect Quality of Life? A Comparison With Quality of Life in the General Population,” 2000. doi: https://doi.org/10.1002/1531-8257(200011)15:6<1112::AID-MDS1008>3.0.CO;2-A.

[17] C. Jenkinson, R. Fitzpatrick, and V. Peto, “Health-related quality-of-life measurement in patients with Parkinson’s disease,” *Pharmacoeconomics*, vol. 15, no. 2, pp. 157–165, 1999, doi: 10.2165/00019053-199915020-00004.

[18] P. Martinez Martin and M. M. Kurtis, “Health-related quality of life as an outcome variable in Parkinson’s disease,” *Ther. Adv. Neurol. Disord.*, vol. 5, no. 2, p. 105, 2012, doi: 10.1177/1756285611431974.

[19] M. L. Maiuolo, R. Giorgini, M. G. Vaccaro, A. Facchin, A. Quattrone, and A. Quattrone, “Assessments scales for the evaluation of health-related quality of life in Parkinson’s

disease, progressive supranuclear palsy, and multiple system atrophy: a systematic review," *Front. Psychol.*, vol. 15, Sep. 2024, doi: 10.3389/FPSYG.2024.1438830.

[20] C. Jenkinson, R. Fitzpatrick, V. Peto, R. Greenhall, and N. Hyman, "The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score," *Age Ageing*, vol. 26, no. 5, pp. 353–357, Sep. 1997, doi: 10.1093/AGEING/26.5.353.

[21] C. Jenkinson, R. Fitzpatrick, V. Peto, R. Greenhall, and N. Hyman, "The PDQ-8: Development and validation of a short-form Parkinson's disease questionnaire," *Psychol. Health*, vol. 12, no. 6, pp. 805–814, 1997, doi: 10.1080/08870449708406741.

[22] A. Schönenberg and T. Prell, "Measuring quality of life with the Parkinson's Disease Questionnaire-39 in people with cognitive impairment," *PLoS One*, vol. 17, no. 4, p. e0266140, Apr. 2022, doi: 10.1371/JOURNAL.PONE.0266140.

[23] G. Galeoto *et al.*, "Quality of life in Parkinson's disease: Italian validation of the Parkinson's Disease Questionnaire (PDQ-39-IT)," *Neurological Sciences*, vol. 39, no. 11, pp. 1903–1909, Nov. 2018, doi: 10.1007/S10072-018-3524-X/METRICS.

[24] J. Bajwa, U. Munir, A. Nori, and B. Williams, "Artificial intelligence in healthcare: transforming the practice of medicine," *Future Healthc. J.*, vol. 8, no. 2, p. e188, Jul. 2021, doi: 10.7861/FHJ.2021-0095.

[25] M. Pan *et al.*, "Application of artificial intelligence in the health management of chronic disease: bibliometric analysis.," *Front. Med. (Lausanne).*, vol. 11, p. 1506641, Jan. 2024, doi: 10.3389/FMED.2024.1506641/BIBTEX.

[26] J. Liao *et al.*, "Artificial intelligence assists precision medicine in cancer treatment," *Front. Oncol.*, vol. 12, p. 998222, Jan. 2023, doi: 10.3389/FONC.2022.998222.

[27] D. Iakovakis *et al.*, "Screening of Parkinsonian subtle fine-motor impairment from touchscreen typing via deep learning," *Scientific Reports 2020 10:1*, vol. 10, no. 1, pp. 1–13, Jul. 2020, doi: 10.1038/s41598-020-69369-1.

[28] M. Yousefi *et al.*, "Machine learning based algorithms for virtual early detection and screening of neurodegenerative and neurocognitive disorders: a systematic-review,"

Front. Neurol., vol. 15, p. 1413071, Dec. 2024, doi: 10.3389/FNEUR.2024.1413071/XML/NLM.

[29] A. Dadu *et al.*, “Identification and prediction of Parkinson’s disease subtypes and progression using machine learning in two cohorts,” *NPJ Parkinsons Dis.*, vol. 8, no. 1, p. 172, Dec. 2022, doi: 10.1038/s41531-022-00439-z.

[30] J. G. V. Habets *et al.*, “Machine learning prediction of motor response after deep brain stimulation in Parkinson’s disease—proof of principle in a retrospective cohort,” *PeerJ*, vol. 8, p. e10317, Nov. 2020, doi: 10.7717/PEERJ.10317/SUPP-2.

[31] A. Zhan *et al.*, “Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity: The Mobile Parkinson Disease Score,” *JAMA Neurol.*, vol. 75, no. 7, pp. 876–880, Jul. 2018, doi: 10.1001/JAMANEUROL.2018.0809.

[32] R. T. Gerraty, A. Provost, L. Li, E. Wagner, M. Haas, and L. Lancashire, “Machine learning within the Parkinson’s progression markers initiative: Review of the current state of affairs,” *Front. Aging Neurosci.*, vol. 15, p. 1076657, Feb. 2023, doi: 10.3389/FNAGI.2023.1076657/BIBTEX.

[33] M. Aborageh, P. Krawitz, and H. Fröhlich, “Genetics in parkinson’s disease: From better disease understanding to machine learning based precision medicine,” *Frontiers in Molecular Medicine*, vol. 2, p. 933383, Oct. 2022, doi: 10.3389/FMMED.2022.933383.

[34] J. Marinus, “Health related quality of life in Parkinson’s disease: a systematic review of disease specific instruments,” *J. Neurol. Neurosurg. Psychiatry*, vol. 72, no. 2, pp. 241–248, Feb. 2002, doi: 10.1136/jnnp.72.2.241.

[35] P. Hagell and M. H. Nilsson, “The 39-item Parkinson’s Disease Questionnaire (PDQ-39): is it a unidimensional construct?,” *Ther. Adv. Neurol. Disord.*, vol. 2, no. 4, pp. 205–214, 2009, doi: 10.1177/1756285609103726.

[36] E. Tolosa *et al.*, “The Parkinson’s Real-World Impact Assessment (PRISM) Study: A European Survey of the Burden of Parkinson’s Disease in Patients and their Carers,” *J. Parkinsons Dis.*, vol. 11, no. 3, pp. 1309–1323, Aug. 2021, doi: 10.3233/JPD-212611.

[37] J. Jankovic, “Parkinson’s disease: clinical features and diagnosis,” *J. Neurol. Neurosurg. Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008, doi: 10.1136/JNNP.2007.131045.

[38] World Health Organization, “Parkinson Disease: A public health approach,” *World Health Organization*, vol. 291, no. 3, p. 390, 2022.

[39] J. Massano and K. P. Bhatia, “Clinical approach to Parkinson’s disease: features, diagnosis, and principles of management,” *Cold Spring Harb. Perspect. Med.*, vol. 2, no. 6, 2012, doi: 10.1101/CSHPERSPECT.A008870.

[40] A. S. Heluani *et al.*, “Neuropsychological and quality of life assessment in patients with Parkinson’s disease submitted to bilateral deep brain stimulation in the subthalamic nucleus,” *Dement. Neuropsychol.*, vol. 6, no. 4, pp. 260–265, 2012, doi: 10.1590/S1980-57642012DN06040010.

[41] P. Martinez-Martin, “The importance of non-motor disturbances to quality of life in Parkinson’s disease,” *J. Neurol. Sci.*, vol. 310, no. 1–2, pp. 12–16, Nov. 2011, doi: 10.1016/j.jns.2011.05.006.

[42] F. Rocha *et al.*, “Effect of opicapone in Parkinson’s disease patients with ‘early’ motor fluctuations: Parkinson’s disease questionnaire (PDQ-39) analysis from the BIPARK-I double-blind experience,” *J. Neurol. Sci.*, vol. 405, pp. 190–191, Oct. 2019, doi: 10.1016/j.jns.2019.10.1150.

[43] N. Hattori *et al.*, “Effects of rasagiline on Parkinson’s Disease Questionnaire (PDQ-39) emotional well-being domain in patients with Parkinson’s disease: A post-hoc analysis of clinical trials in Japan,” *PLoS One*, vol. 17, no. 1, p. e0262796, Jan. 2022, doi: 10.1371/journal.pone.0262796.

[44] A. S. Chandrabhatla, I. J. Pomeraniec, and A. Ksendzovsky, “Co-evolution of machine learning and digital technologies to improve monitoring of Parkinson’s disease motor symptoms,” *npj Digital Medicine 2022 5:1*, vol. 5, no. 1, pp. 1–18, Mar. 2022, doi: 10.1038/s41746-022-00568-y.

[45] C. Gao *et al.*, “Model-based and Model-free Machine Learning Techniques for Diagnostic Prediction and Classification of Clinical Outcomes in Parkinson’s Disease,” *Sci. Rep.*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/S41598-018-24783-4.

[46] M. Nilashi *et al.*, “Predicting Parkinson’s Disease Progression: Evaluation of Ensemble Methods in Machine Learning,” *J. Healthc. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/2793361.

[47] N. Faiem, T. Asuroglu, K. Acici, A. Kallonen, and M. van Gils, “Assessment of Parkinson’s Disease Severity Using Gait Data: A Deep Learning-Based Multimodal Approach,” *Communications in Computer and Information Science*, vol. 2084 CCIS, pp. 29–48, 2024, doi: 10.1007/978-3-031-59091-7_3/TABLES/6.

[48] “R: The R Project for Statistical Computing.” Accessed: Jul. 12, 2023. [Online]. Available: https://www.r-project.org/

[49] S. van Buuren and K. Groothuis-Oudshoorn, “mice: Multivariate Imputation by Chained Equations in R,” *J. Stat. Softw.*, vol. 45, no. 3, pp. 1–67, Dec. 2011, doi: 10.18637/JSS.V045.I03.

[50] M. Kuhn, “Building Predictive Models in R Using the caret Package,” *J. Stat. Softw.*, vol. 28, no. 5, 2008, doi: 10.18637/jss.v028.i05.

[51] B. Pavlyshenko, “Using Stacking Approaches for Machine Learning Models,” in *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, IEEE, Aug. 2018, pp. 255–258. doi: 10.1109/DSMP.2018.8478522.

[52] C. Molnar, G. Casalicchio, and B. Bischl, “iml: An R package for Interpretable Machine Learning,” *J. Open Source Softw.*, vol. 3, no. 26, p. 786, Jun. 2018, doi: 10.21105/JOSS.00786.

[53] Christoph Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed. 2022.

[54] T. Pringsheim, N. Jette, A. Frolkis, and T. D. L. Steeves, “The prevalence of Parkinson’s disease: A systematic review and meta-analysis,” *Movement Disorders*, vol. 29, no. 13, pp. 1583–1590, Nov. 2014, doi: 10.1002/MDS.25945.

[55] D. Aarsland, K. Andersen, J. P. Larsen, A. Lolk, H. Nielsen, and P. Kragh-Sørensen, “Risk of dementia in Parkinson’s disease: A community-based, prospective study,” *Neurology*, vol. 56, no. 6, pp. 730–736, Mar. 2001, doi: 10.1212/WNL.56.6.730.

[56] G. James, D. Witten, T. Hastie, and R. Tibshirani, “An Introduction to Statistical Learning,” 2021, doi: 10.1007/978-1-0716-1418-1.

[57] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, “Machine learning algorithm validation with a limited sample size,” *PLoS One*, vol. 14, no. 11, p. e0224365, Nov. 2019, doi: 10.1371/journal.pone.0224365.

[58] E. Candel-Parra, M. P. Córcoles-Jiménez, V. Delicado-Useros, M. C. Ruiz-Grao, A. Hernández-Martínez, and M. Molina-Alarcón, “Predictive Model of Quality of Life in Patients with Parkinson’s Disease,” *Int. J. Environ. Res. Public Health*, vol. 19, no. 2, p. 672, Jan. 2022, doi: 10.3390/ijerph19020672.

[59] G. S. Collins and P. Dhiman, “Prediction models should contain predictors known at the moment of intended use,” *Aging Clin. Exp. Res.*, pp. 1–2, Oct. 2023, doi: 10.1007/S40520-023-02560-2/METRICS.

[60] V. L. Feigin *et al.*, “Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015,” *Lancet Neurol.*, vol. 16, no. 11, pp. 877–897, Nov. 2017, doi: 10.1016/S1474-4422(17)30299-5.

[61] “Focus Area: Patient-Reported Outcomes and other Clinical Outcome Assessments | FDA.” Accessed: Jun. 01, 2025. [Online]. Available: https://www.fda.gov/science-research/focus-areas-regulatory-science-report/focus-area-patient-reported-outcomes-and-other-clinical-outcome-assessments

[62] “EMA and European Organisation for Research and Treatment of Cancer (EORTC) workshop: How can patient-reported outcomes (PRO) and health-related quality of life (HRQoL) data inform regulatory decisions? | European Medicines Agency (EMA).” Accessed: Jun. 01, 2025. [Online]. Available: https://www.ema.europa.eu/en/events/ema-european-organisation-research-treatment-cancer-eortc-workshop-how-can-patient-reported-outcomes-pro-health-related-quality-life-hrqol-data-inform-regulatory-decisions

[63] K. R. Chaudhuri *et al.*, “International Multicenter Pilot Study of the First Comprehensive Self-Completed Nonmotor Symptoms Questionnaire for Parkinson’s Disease: The NMSQuest Study,” *Movement Disorders*, vol. 21, no. 7, pp. 916–923, 2006, doi: 10.1002/mds.20844.

[64] K. Cho, J. Yoon, N.-Y. Shin, and S. Han, “A Machine Learning Approach to Identify Whole-Brain Functional Connectivity Signatures of Cognitive Subtypes in Parkinson’s Disease,” 2025, doi: 10.2139/SSRN.5477430.

[65] J. Lu, D. K. Tran, X. L. Bui, N. Q. K. Le, and M. C. H. Chua, “Multimodal machine learning for Parkinson’s disease diagnosis and genetic subtyping,” *Health and Technology 2025*, pp. 1–7, Oct. 2025, doi: 10.1007/S12553-025-01021-2.

[66] M. S. Hosseini, S. M. R. Aghamiri, and M. Panahi, “Cross-regional radiomics: a novel framework for relationship-based feature extraction with validation in Parkinson’s disease motor subtyping,” *BioData Min.*, vol. 18, no. 1, p. 67, Sep. 2025, doi: 10.1186/S13040-025-00483-4.

[67] G. E. Hey, R. Eisinger, D. Guarin, D. Safarpour, M. Rodríguez-Violante, and A. Ramirez-Zamora, “Developing personalized treatment strategies for Parkinson’s disease based on disease subtypes,” *Expert Rev. Neurother.*, vol. 25, no. 10, 2025, doi: 10.1080/14737175.2025.2552784.

[68] W. Zhang, Z. Xu, R. Yu, M. Jiang, and Q. Dai, “DualGCN-GE: integration of spatiotemporal representations from whole-blood expression data with dual-view graph convolution network to identify Parkinson’s disease subtypes,” *BMC Bioinformatics*, vol. 26, no. 1, pp. 1–18, Dec. 2025, doi: 10.1186/S12859-025-06181-6/FIGURES/5.

[69] X. Xu *et al.*, “Identifying subtypes of longitudinal motor symptom severity trajectories in early Parkinson’s disease patients,” *Front. Neurol.*, vol. 16, p. 1597132, Aug. 2025, doi: 10.3389/FNEUR.2025.1597132/BIBTEX.

[70] W. Song *et al.*, “The impact of non-motor symptoms on the Health-Related Quality of Life of Parkinson’s disease patients from Southwest China,” *Parkinsonism Relat. Disord.*, vol. 20, no. 2, pp. 149–152, Feb. 2014, doi: 10.1016/J.PARKRELDIS.2013.10.005.

[71] P. Martinez-Martin, C. Rodriguez-Blazquez, M. M. Kurtis, and K. R. Chaudhuri, “The impact of non-motor symptoms on health-related quality of life of patients with Parkinson’s disease,” *Movement Disorders*, vol. 26, no. 3, pp. 399–406, Feb. 2011, doi: 10.1002/MDS.23462.

[72] D. Magano, T. Taveira-Gomes, J. Massano, and A. S. Barros, “Predicting Quality of Life in Parkinson’s Disease: A Machine Learning Approach Employing Common Clinical Variables,” *J. Clin. Med.*, vol. 13, no. 17, p. 5081, Aug. 2024, doi: 10.3390/jcm13175081.

[73] A. Verikas, A. Gelzinis, and M. Bacauskiene, “Mining data with random forests: A survey and results of new tests,” *Pattern Recognit.*, vol. 44, no. 2, pp. 330–349, Feb. 2011, doi: 10.1016/J.PATCOG.2010.08.011.

[74] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, “Selecting critical features for data classification based on machine learning methods,” *J. Big Data*, vol. 7, no. 1, pp. 1–26, Dec. 2020, doi: 10.1186/S40537-020-00327-4/FIGURES/13.

[75] R. Iranzad and X. Liu, “A review of random forest-based feature selection methods for data science education and applications,” *Int. J. Data Sci. Anal.*, pp. 1–15, Feb. 2024, doi: 10.1007/S41060-024-00509-W/METRICS.

[76] M. B. Kursa and W. R. Rudnicki, “Feature Selection with the Boruta Package,” *J. Stat. Softw.*, vol. 36, no. 11, pp. 1–13, Sep. 2010, doi: 10.18637/JSS.V036.I11.

[77] B. Efron and R. J. Tibshirani, “An Introduction to the Bootstrap,” *An Introduction to the Bootstrap*, May 1994, doi: 10.1201/9780429246593/INTRODUCTION-BOOTSTRAP-BRADLEY-EFRON-TIBSHIRANI/ACCESSIBILITY-INFORMATION.

[78] R. Nisbet, G. Miner, and K. Yale, “Handbook of statistical analysis and data mining applications,” *Handbook of Statistical Analysis and Data Mining Applications*, pp. 1–792, Nov. 2017, doi: 10.1016/C2012-0-06451-4.

[79] Christoph Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.)*, 2nd ed. 2022.

[80] D. W. Apley and J. Zhu, “Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models,” *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 82, no. 4, pp. 1059–1086, Sep. 2020, doi: 10.1111/rssb.12377.

[81] S. Lê, J. Josse, and F. Husson, “FactoMineR: An R package for multivariate analysis,” *J. Stat. Softw.*, vol. 25, no. 1, pp. 1–18, 2008, doi: 10.18637/JSS.V025.I01.

[82] P. Barone *et al.*, “The PRIAMO study: A multicenter assessment of nonmotor symptoms and their impact on quality of life in Parkinson’s disease,” *Mov. Disord.*, vol. 24, no. 11, pp. 1641–1649, Aug. 2009, doi: 10.1002/MDS.22643.

[83] J. M. Bonsel, A. J. Itiola, A. S. Huberts, G. J. Bonsel, and H. Penton, “The use of patient-reported outcome measures to improve patient-related outcomes – a systematic review,” *Health Qual. Life Outcomes*, vol. 22, no. 1, pp. 1–19, Dec. 2024, doi: 10.1186/S12955-024-02312-4/TABLES/1.

[84] N. Jeyaraman, M. Jeyaraman, S. Ramasubramanian, S. Balaji, and S. Muthu, “Voices that matter: The impact of patient-reported outcome measures on clinical decision-making,” *http://www.wjgnet.com/*, vol. 15, no. 2, Jun. 2025, doi: 10.5662/WJM.V15.I2.98066.

[85] M. Pe *et al.*, “Using patient-reported outcomes and health-related quality of life data in regulatory decisions on cancer treatment: highlights from an EMA-EORTC workshop,” *Lancet Oncol.*, vol. 26, no. 6, pp. 687–690, Jun. 2025, doi: 10.1016/S1470-2045(25)00150-0.

[86] “FDA Patient-Focused Drug Development Guidance Series for Enhancing the Incorporation of the Patient’s Voice in Medical Product Development and Regulatory Decision Making | FDA.” Accessed: Jun. 01, 2025. [Online]. Available: https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical

[87] “ISPOR - Requirements for Patient-Reported Outcome and Data Analytics in Health Technology Assessment in England, France and Germany, and Opportunities for Methods Harmonization across European Markets.” Accessed: Jun. 01, 2025. [Online]. Available: https://www.ispor.org/heor-resources/presentations-database/presentation-cti/ispor-2025/the-role-of-patient-reported-outcomes-and-patient-preference-information-in-regulatory-and-hta-decision-making/requirements-for-patient-reported-outcome-and-data-analytics-in-health-technology-assessment-in-england-france-and-germany-and-opportunities-for-methods-harmonization-across-european-markets

[88] C. Yap *et al.*, “Advancing patient-centric care: integrating patient reported outcomes for tolerability assessment in early phase clinical trials – insights from an expert virtual

roundtable," *EClinicalMedicine*, vol. 76, p. 102838, Oct. 2024, doi: 10.1016/J.ECLINM.2024.102838/ATTACHMENT/820F354C-22A6-420B-8D20-B0C20C267E24/MMC1.DOCX.

[89] "Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims | FDA." Accessed: Jun. 01, 2025. [Online]. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims

9. Appendices

Appendix 1

Applying machine learning methods to predict the Parkinson's Disease Questionnaire-39 Summary Index

Journal of Statistics on Health Decision

EXTENDED ABSTRACT

Applying machine learning methods to predict the Parkinson´s Disease Questionnaire-39 Summary Index

Daniel Magano[1,2], Joana B. Moreira[1,3], Pedro P. Rodrigues[4]

[1]Ph.D. Program in Health Data Science, Faculty of Medicine, University of Porto, Porto, Portugal;
[2]BIAL – Portela & Ca S.A., Coronado, Portugal
[3]Anaesthesiology, Emergency and Intensive Care Department, Centro Hospitalar Universitário do Porto, Largo do Prof. Abel Salazar, Porto, Portugal
[4]CINTESIS@RISE, MEDCIDS - Faculty of Medicine of the University of Porto, Porto, Portugal

Introduction:

Health-related quality of life (HRQoL) gained importance in the last decades and is considered to be an important outcome measure in chronic diseases studies. HRQoL disease specific instruments reflect the consequences of that disease to a particular person and are sensitive to change in perceived HRQoL. In Parkinson's disease (PD) several disease specific HRQoL instruments have become available in the past few years. A 2002 systematic review of HRQoL instruments in PD concluded that in many situations the Parkinson´s Disease Questionnaire -39 (PDQ-39) is the most appropriate HRQoL instrument, because it is the scale that has been tested most thoroughly, has adequate clinimetric characteristics, has been used in the largest number of studies, and is available in many languages [1].

The PQD-39 is a 39-item self-report questionnaire, which assesses PD specific health related quality over the last month. It assesses how often patients experience difficulties across the 8 quality of life dimensions and the impact of PD on specific dimensions of functioning and well-being [2]. Its summary index (PDQ-39SI) is a widely used patient-reported clinical trial endpoint. Peter Hagell and Maria H. Nilsson assessed the unidimensionality of the PDQ-39 and PDQ-39SI using Rasch and confirmatory factor analysis and concluded its multidimensional nature [3].

The objective of this study is to do an exploratory analysis on the multidimensional nature of the PDQ-39SI using machine learning methods to improve the interpretability of the score and obtain a model to predict the perceived quality of life of people with PD.

Methods:

Data analysis was completed with R version 4.1.1. Data used in the preparation of this article were obtained from the Parkinson's Real-World Impact assesSMent (PRISM) database. PRISM study and database was funded by BIAL – Portela & C[a], S.A., designed in collaboration with The Cure Parkinson's Trust, an advocacy group based in the United Kingdom (UK), and reviewed by the PRISM steering committee.

The PRISM database contains data from 861 people from 5 European countries with PD collected in the context of an observational study with cross-sectional design [4].

PDQ-39SI multivariable nature was first assessed using the Expectation-Maximization (EM) algorithm (Figure 1). Based on the latent classes, a transformation of PDQ-39SI from continuous to binomial outcome was performed: mild vs severe symptoms impacting their quality of life.

Clinically relevant variables for modeling the PDQ-39SI were initially selected from the PRISM database. Univariate logistic regressions were created with the PDQ-39SI as a function of each of the previously selected variables. The variables presenting a p-value below 0.05 were ordered according to their impact on the odds ratio and included in the machine learning models by forward selection based on the resulting AUC values. Incomplete observations were then removed resulting in a dataset with 615 patients

Before training the models, 20% of the dataset was saved to be used as independent validation data. With the remaining 80% of the dataset (training set), several models were trained to predict the classes obtained through EM. A repeated 10-fold cross validation was performed on this training set to estimate the respective ROC curves (Figure 2). Additionally, the performance of each model obtained with the 20% independent validation set was also assessed through the plotting of ROC curves (Figure 3).

Keywords:
Health-Related Quality of Life, Machine Learning, Parkinson Disease, Parkinson´s Disease Questionnaire -39, Parkinson´s Disease Questionnaire -39 Summary Index

Corresponding author:
Daniel Magano
danielmagano@outlook.com

Conflict of interest:
the authors declare no conflict of interests

First published: 20JUL2022

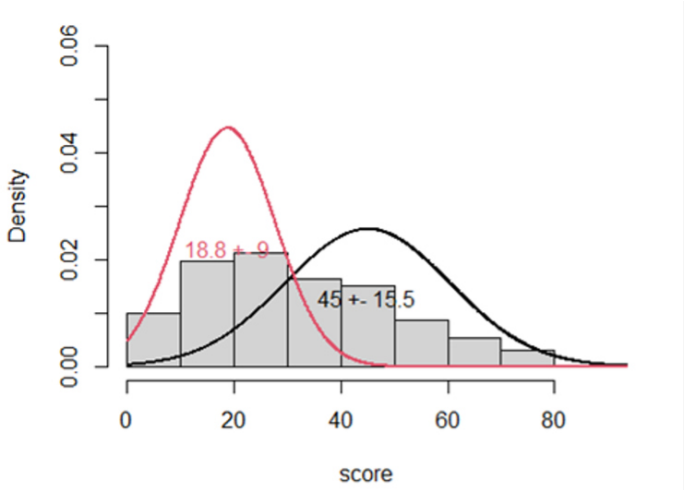

Figure 1 - Latent classes on Parkinson´s Disease Summary Index. ***Legend:*** With the use of Expectation Maximization algorithm (2 kernels) 2 latent classes are described: **(i)** red distribution with mean of 18.8 and standard deviation of ± 9, and **(ii)** black distribution with mean of 45 and standard deviation of 15.5. Interception of the two distributions at the 31 value.

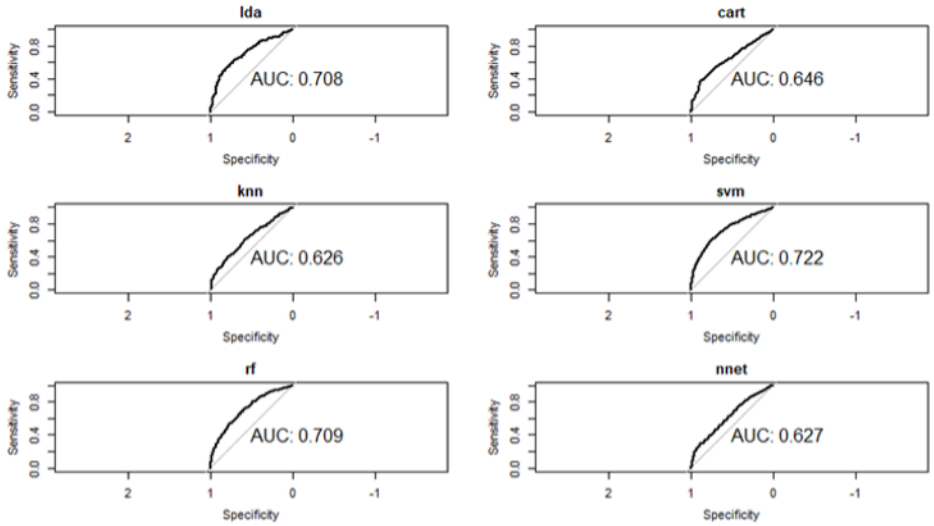

Figure 2 - Performance estimation of different machine learning methods for PDQ-39SI modeling on the training set using repeated 10-fold cross validation. ***Legend:*** ROC curves (receiver operating characteristic curves) of the repeated 10-fold cross validation training set (80% of the dataset) with **(i)** lda (Linear Discriminant Analysis), **(ii)** cart (Classification and Regression Tree, splitting index: gini), **(iii)** knn (K-nearest Neighbors, K=3), **(iv)** svm (Support Vector Machine, kernel: radial basis function), **(v)** rf (Random Forest, number of trees = 500), **(vi)** nnet (Feedforward Neural Network, 1 hidden layer with 5 neurons). AUC (area under the curve)

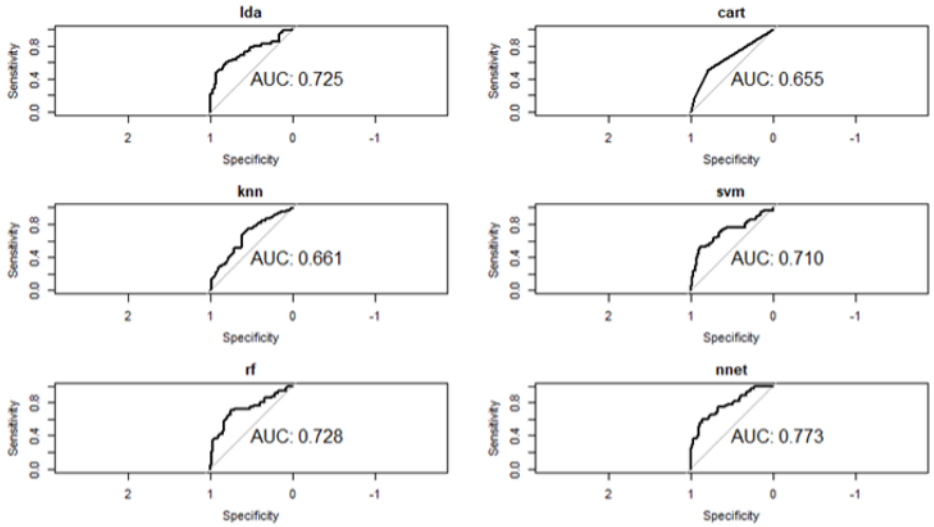

Figure 3 - Performance of different machine learning methods for PDQ-39SI modeling on the independent validation set. ***Legend:*** ROC curves (receiver operating characteristic curves) of the independent validation set (20% of the dataset) with **(i)** lda (Linear Discriminant Analysis), **(ii)** cart (Classification and Regression Tree, splitting index: gini), **(iii)** knn (K-nearest Neighbors, K=3), **(iv)** svm (Support Vector Machine, kernel: radial basis function), **(v)** rf (Random Forest, number of trees = 500), **(vi)** nnet (Feedforward Neural Network, 1 hidden layer with 5 neurons). AUC (area under the curve)

Results:

The PDQ-39SI was converted from a continuous outcome into two categories based on the impact on the quality of life of the person with PD: mild (with a PDQ-39SI below 31) and severe (PDQ-39SI above 31).

The ROC curves obtained with repeated cross-validation for each of the trained models are presented in figure 2 while the ROC curves obtained with the independent validation data set are presented in figure 3.

Discussion:

The obtention of a binomial classification for the PDQ-39SI can facilitate its interpretation and harmonize its utilization by different health care providers (HCP). Additionally, machine learning models applied to clinical variables can be used to predict to which of these classes a person with Parkinson's disease would belong. This information would allow HCPs to predict which of the patients are expected to have a lower quality of life.

The PDQ-39SI is based on the patient's interpretation about several dimensions measuring his/her perceived quality of life. The subjectivity of the information collected through a survey is considered a limitation of this study.

The results obtained demonstrate a good performance of Linear Discriminant Analysis, Support Vector Machine and Random Forest methods. Future work will focus on hyperparameters optimization and on studying a wider range of variables to accommodate possible confounding factors.

References:

1. Marinus J, Ramaker C, van Hilten JJ, Stiggelbout AM. Health related quality of life in Parkinson´s disease: a systematic review of disease specific instruments. J Neurol Neurosurg Psychiatry. 2002 Feb 1;72(2):241 LP – 248. https://doi.org/10.1136/jnnp.72.2.241
2. parkinsons-disease-questionnaire-pdq-39-pdq-8. Available from: https://innovation.ox.ac.uk/outcome-measures/parkinsons-disease-questionnaire-pdq-39-pdq-8/
3. Hagell P, Nilsson MH. The 39-item Parkinson's Disease Questionnaire (PDQ-39): is it a unidimensional construct? Ther Adv Neurol Disord. 2009 May 28;2(4):205–14. https://doi.org/10.1177/1756285609103726
4. Parkinson's Real World Impact assesSMent (PRISM). Available from: www.prism.bial.com

Appendix 2

Predicting Quality of Life in Parkinson's Disease: A Machine Learning Approach Employing Common Clinical Variables

Journal of
Clinical Medicine

MDPI

Article

Predicting Quality of Life in Parkinson's Disease: A Machine Learning Approach Employing Common Clinical Variables

Daniel Magano [1,2,*], Tiago Taveira-Gomes [3,4,5], João Massano [6,7] and António S. Barros [8]

1 Ph.D. Program in Health Data Science, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal
2 Medical Department, BIAL-Portela & Cª., S.A., 4745-457 São Mamede do Coronado, Portugal
3 Department of Community Medicine, Information and Decision in Health, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal; tiagogomes@med.up.pt
4 Faculty of Health Sciences, University Fernando Pessoa, 4200-150 Porto, Portugal
5 SIGIL Scientific Enterprises, 4076 Dubai, United Arab Emirates
6 Department of Clinical Neurosciences and Mental Health, Faculty of Medicine University of Porto, 4200-319 Porto, Portugal; jmassano@med.up.pt
7 Department of Neurology, Centro Hospitalar Universitário de São João, 4200-319 Porto, Portugal
8 Department of Surgery and Physiology, Cardiovascular R&D Centre-UnIC@RISE, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal; asbarros@med.up.pt
* Correspondence: danielmagano@outlook.com

Abstract: Background: Parkinson's Disease significantly impacts health-related quality of life, with the Parkinson's Disease Questionnaire-39 extensively used for its assessment. However, predicting such outcomes remains a challenge due to the subjective nature and variability in patient experiences. This study develops a machine learning model using accessible clinical data to enable predictions of life-quality outcomes in Parkinson's Disease and utilizes explainable machine learning techniques to identify key influencing factors, offering actionable insights for clinicians. **Methods**: Data from the Parkinson's Real-world Impact Assessment study (PRISM), involving 861 patients across six European countries, were analyzed. After excluding incomplete data, 627 complete observations were used for the analysis. An ensemble machine learning model was developed with a 90% training and 10% validation split. **Results**: The model demonstrated a Mean Absolute Error of 4.82, a Root Mean Squared Error of 8.09, and an R^2 of 0.75 in the training set, indicating a strong model fit. In the validation set, the model achieved a Mean Absolute Error of 11.22, a Root Mean Squared Error of 13.99, and an R^2 of 0.36, showcasing moderate variation. Key predictors such as age at diagnosis, patient's country, dementia, and patient's age were identified, providing insights into the model's decision-making process. **Conclusions**: This study presents a robust model capable of predicting the impact of Parkinson's Disease on patients' quality of life using common clinical variables. These results demonstrate the potential of machine learning to enhance clinical decision-making and patient care, suggesting directions for future research to improve model generalizability and applicability.

Keywords: Parkinson's disease; health-related quality of life; predictive modeling; machine learning; Parkinson's disease questionnaire—39; PDQ-39 summary index; explainable machine learning

Citation: Magano, D.; Taveira-Gomes, T.; Massano, J.; Barros, A.S. Predicting Quality of Life in Parkinson's Disease: A Machine Learning Approach Employing Common Clinical Variables. *J. Clin. Med.* **2024**, *13*, 5081. https://doi.org/10.3390/jcm13175081

Academic Editors: Lindsay A. Farrer and Frank Klawonn

Received: 30 July 2024
Accepted: 23 August 2024
Published: 27 August 2024

1. Introduction

Parkinson's Disease (PD)'s global prevalence more than doubled between 1990 and 2015 [1]. The motor symptoms associated with the second most common neurodegenerative disorder include resting tremor, rigidity, bradykinesia, and postural instability. Additionally, non-motor symptoms such as cognitive impairment, mental health disorders, sleep disturbances, pain, and sensory disturbances are also common in PD [2–4]. All these symptoms decrease the health-related quality of life (HRQoL) of people with PD [5].

In recent years, HRQoL has emerged as a critical outcome measure in the study of chronic diseases. HRQoL instruments, particularly those designed for specific diseases,

offer valuable information on the personal effects of disease and any changes that are perceived over time. For example, in the context of PD, several disease-specific HRQoL tools have been developed recently. A comprehensive, systematic review conducted in 2002 determined that the Parkinson's Disease Questionnaire-39 (PDQ-39) was the most suitable instrument for assessing HRQoL in the context of PD. The PDQ-39 was chosen based on several factors, including its extensive testing, robust clinimetric properties, widespread use in numerous studies, and availability in numerous languages [5]. Additionally, recent studies have demonstrated its utility in various contexts, such as assessing the impact of deep brain stimulation of the subthalamic nucleus on the quality of life of PD patients [6] and validating its use in PD patients with cognitive impairment, showcasing its versatility and applicability across different patient populations [7]. Furthermore, the Italian version of the PDQ-39 has been validated, highlighting its cross-cultural applicability [8]. Research suggests that the PDQ-39 is a reliable and valid tool for assessing HRQoL in PD, as it provides a comprehensive view of the disease's impact, extending beyond traditional clinical outcomes, and reflects the patient's perspective, capturing their personal and subjective experience of living with the disease [9].

PDQ-39 is a commonly utilized self-report questionnaire comprised of 39 items that assess the degree of difficulty experienced by individuals diagnosed with PD across eight dimensions of HRQoL, namely mobility, activities of daily living, emotional well-being, stigma, social support, cognition, communication, and bodily discomfort [10]. The PDQ-39 Summary Index (PDQ-39 SI), which measures the overall impact of PD on a patient's life, is a commonly utilized patient-reported endpoint in clinical trials [10–12]. Despite its widespread use, accurately predicting the PDQ-39 SI and, consequently, the HRQoL remains a complex and challenging task. This complexity is partly due to the subjective nature of the questionnaire, which captures individual patient experiences and perceptions that may vary widely based on personal factors, such as disease severity, duration, treatment response, individual resilience, and varying interpretations of quality of life.

Machine and statistical learning have been used in PD to predict disease progression, motor symptoms, and treatment response. For instance, Chandrabhatla et al. explored the co-evolution of machine learning (ML) and digital technologies to enhance the monitoring of Parkinson's Disease motor symptoms, thereby improving real-time symptom tracking and management [13]. Gao et al. employed both model-based and model-free ML techniques for the diagnostic, prediction, and classification of clinical outcomes in PD, highlighting their potential in early diagnosis and personalized treatment planning [14]. Nilashi et al. evaluated the effectiveness of ensemble ML methods for predicting PD progression, demonstrating the improved accuracy and robustness of combined models [15]. Dadu et al. focused on identifying and predicting PD subtypes and progression using ML in two cohorts, which can aid in precise disease classification and management [16]. Additionally, a recent study introduced a Perceiver architecture-based multimodal ML model to estimate PD symptom severity by analyzing raw gait signals and features from Ground Reaction Force sensors [17]. However, its application in predicting HRQoL, as measured by the PDQ-39 Summary Index (SI), remains underexplored. This study sought to develop an ML model capable of predicting PDQ-39 SI in patients with PD using readily available clinical variables, such as the patient's current PD medication, co-existing medical conditions, education level, age, and age at diagnosis. The objective of this research is to develop an ML model equipped with explainable ML techniques to enhance clinicians' ability to predict HRQoL in PD patients. This tool is designed to provide deeper insights into the factors influencing HRQoL, thereby informing treatment decisions and ultimately improving patient care. Additionally, the explainable ML component of this study offers critical information about the variables that have a higher impact on quality of life, which can be valuable for clinicians in understanding disease progression and management.

2. Materials and Methods

2.1. Database and Data Preparation

The data for this study were sourced from the Parkinson's Real-world Impact Assessment (PRISM) study, a cross-sectional observational research initiative. A total of 861 patients with PD and 256 caregivers from six European countries participated in the PRISM study, completing an online survey using structured questionnaires including the PDQ-39, Non-Motor Symptoms Questionnaire, and Zarit Burden Interview [18]. This survey captured a broad range of information, including medication use, comorbidities, and healthcare resource utilization [18]. After excluding two observations due to missing PDQ-39 SI scores, 859 valid entries remained for analysis. The demographic profile of the participant pool was nearly evenly split by gender: 48.5% female and 50.3% male, with a mean age of 65.0 years (Standard Deviation (SD) = 10.2 years). The average PDQ-39 SI score was 32.0 (SD = 18.2), with its distribution depicted in Figure 1. Supplementary Table S1 summarizes key variables from the PRISM dataset.

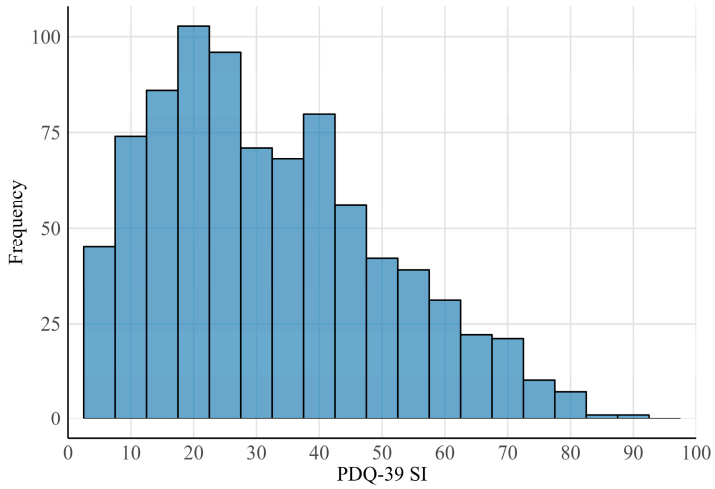

Figure 1. Distribution of PDQ-39 SI across 859 patients from the PRISM study.

2.2. Data Analysis

All analyses were performed using the R statistical software (version 4.3.0) [19]. As part of our analytical approach, we utilized Factor Analysis of Mixed Data (FAMD) to explore the underlying relationships between variables within the PRISM dataset, including the PDQ-39 SI. FAMD are particularly relevant for datasets that include both quantitative and qualitative variables, as they extend the principles of principal component analysis to mixed data types. This approach allows us to reduce the complexity of the data by identifying the principal dimensions that capture the most variance. Consequently, FAMD provide a comprehensive understanding of how different variables interact and contribute to the observed patterns in our dataset. The results of the FAMD are visually depicted in Supplementary Figure S1, offering a clear illustration of the interrelations among the dataset where variables that are highly correlated cluster together, including how closely each variable correlates with the outcome variable, the PDQ-39 SI.

After the FAMD analysis, the PRISM dataset was subjected to a data-cleaning process. Categorical variables with responses such as 'prefer not to say' or 'other' were treated as missing values to minimize the noise in the final model. Upon examination of the dataset (Supplementary Table S1), it was observed that it contains minimal missing values. Most of these missing values appear to be missing completely at random, with notable exceptions in the variables age at diagnosis and current medication. These exceptions suggest a missing not at random pattern, likely due to patients not recalling the age of PD diagnosis or being unaware of the therapeutic class of their current medication. Although we initially attempted to address missing values using Multivariate Imputation by Chained

Equations (MICE), this approach resulted in a model with worse predictive capabilities [20]. Therefore, we only kept complete observations, resulting in a total of 627 valid entries. Subsequently, the dataset was partitioned into training (90%) and validation (10%) sets to maximize the learning capacity of our model amidst the constraints of limited sample size and high variability inherent in questionnaire-based data. This allocation allows for comprehensive model training on a diverse range of responses, critical for capturing the nuanced patterns of HRQoL assessments in PD, while still reserving a portion of the data for essential validation to assess model performance and mitigate overfitting. Figure 2 illustrates this study's workflow, detailing each step from data acquisition to the data split and comparison of regression metrics in the training and validation sets.

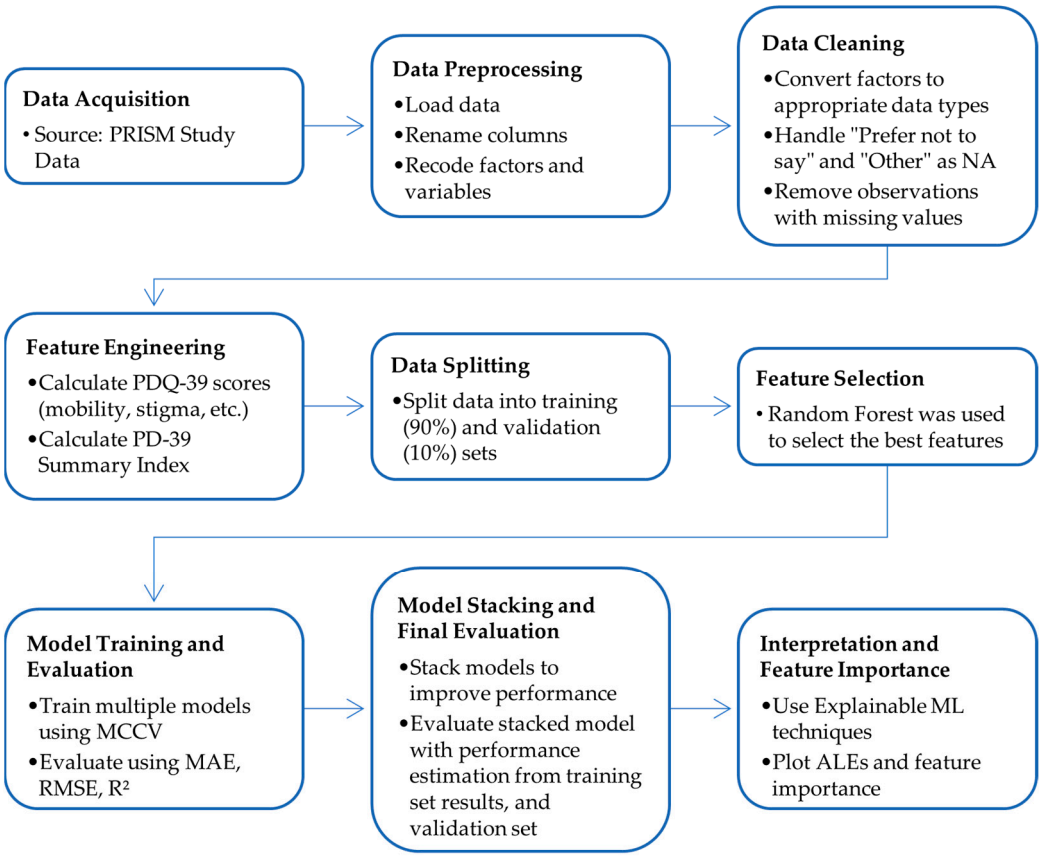

Figure 2. Diagram of the study process from data acquisition to data preparation, feature selection, model training, and performance evaluation.

2.3. Model Training

The models were trained on a personal computer equipped with an AMD Ryzen 7 7800X3D CPU, 32 GB of RAM, and an AMD Radeon RX 7900 XT GPU (Advanced Micro Devices, Inc., Santa Clara, CA, USA). We evaluated several feature selection methods—specifically, random forest, least absolute shrinkage and selection operator (LASSO), group LASSO, and the option of omitting feature selection—to identify the most suitable technique for our dataset. Feature selection through random forest was selected for its standout performance in enhancing the predictive accuracy of our suite of ML models. We then applied this feature selection approach in training a diverse range of ML models, namely linear regression, boosted generalized linear model, extreme gradient boosting with linear models and decision trees, ridge regression, neural network regression, LASSO regression, random forest, k-nearest neighbors, support vector machines with radial basis function and polynomial kernel, and Gaussian processes (linear, polynomial, and radial kernels), incorporating essential pre-processing steps such as centering and scaling with the caret package [21]. To accurately train and evaluate the performance of these models within the

context of a dataset characterized by limited size and high variability, Monte Carlo Cross-Validation (MCCV) was employed with 500 repetitions for each model. This approach was chosen for its effectiveness in mitigating data variability, thereby ensuring a more reliable and stable assessment of the models' generalizability.

2.4. Model Stacking

To further improve the performance of our predictive model, we implemented a stacking approach using an ensemble of the previously trained ML models. Stacking is an advanced ensemble method that combines multiple predictive models to enhance overall accuracy. It utilizes a meta-model that optimally combines the predictions from several base models trained on the same data. This technique leverages the strengths and compensates for the weaknesses of individual models, thereby increasing the robustness and reliability of predictions [22].

We started by assessing the correlation among these models to create an ensemble characterized by both strong individual performance and minimal intermodel correlation. We then trained several ensembles using MCCV and experimented with different numbers of repetitions (100, 250, 500, and 1000) to assess both stability and performance. For the stacking mechanism, we tested both random forest and extreme gradient boosting with decision trees (XGBTree) as techniques. These were selected for their demonstrated capability to effectively handle complex data patterns. This stacked ensemble approach allowed us to enhance the model's ability to generalize across different patient profiles and improve the accuracy of predicting the PDQ-39 SI.

2.5. Model Evaluation

The performance of the models was evaluated using multiple metrics. The Mean Absolute Error (MAE) measures the average magnitude of the errors in predictions, providing a straightforward measure of the average error magnitude per prediction. The Root Mean Squared Error (RMSE) offers insight into the accuracy of the model by quantifying the square root of the average squared differences between the predicted values and actual observations. Lastly, the Coefficient of Determination (R^2) reflects the proportion of variance in the dependent variable that is predictable from the independent variables, thereby indicating the explanatory power of the model. These metrics collectively provide a comprehensive assessment of model performance.

The ensemble model that emerged as the most effective incorporated the pre-trained ML models: Gaussian process with a Radial Basis Function kernel and random forest. This ensemble model was trained using MCCV with 1000 repetitions, with XGBTree serving as the stacking method. MCCV was utilized both to train and to estimate the performance of the ensemble model on the training dataset, while the validation set was reserved for the final assessment to evaluate overfitting susceptibility and ensure an unbiased evaluation of the model's generalizability.

2.6. Interpretable Machine Learning Analysis

To unveil the importance and effect of each feature on the predictive model, we employed interpretable ML techniques, particularly focusing on Permutation Feature Importance (PFI). Utilizing the iml package in R, this analysis facilitates a deeper understanding of the behavior of our final ensemble ML model [23]. The core concept of PFI is to measure the increase in the model's prediction error after the permutation of each feature's values, which effectively breaks the relationship between the feature and the true outcome. This process provides a clear indication of a feature's importance; a significant increase in error indicates that the model relied heavily on the feature for prediction, whereas no change in error suggests that the feature was not utilized by the model for prediction [24].

The process of feature importance quantification involves permuting the values of each feature and assessing the change in the model's performance using MAE as the chosen metric, due to its lower sensitivity to outliers and variability within the dataset. This

procedure was carried out for each feature individually and repeated 100 times to obtain a more robust estimate, with the results averaged thereafter.

3. Results

3.1. Performance Metrics

In the training set, the final ensemble model achieved an MAE of 4.82, an RMSE of 8.09, and an R^2 value of 0.75, indicating a strong fit to the data. The distributions of MAE, RMSE, and R^2 metrics from the 1000 MCCV repetitions approximated normal distributions, as shown in Figures 3–5. These distributions suggest that the model's performance is consistent and reliable across different subsamples of the training data, reflecting its stability. When applied to the validation set, the model exhibited moderate variations in performance metrics, with an MAE of 11.22, an RMSE of 13.99, and an R^2 value of 0.36, indicating the model's reasonable predictive capability on unseen data. A summary of these results is presented in (Table 1).

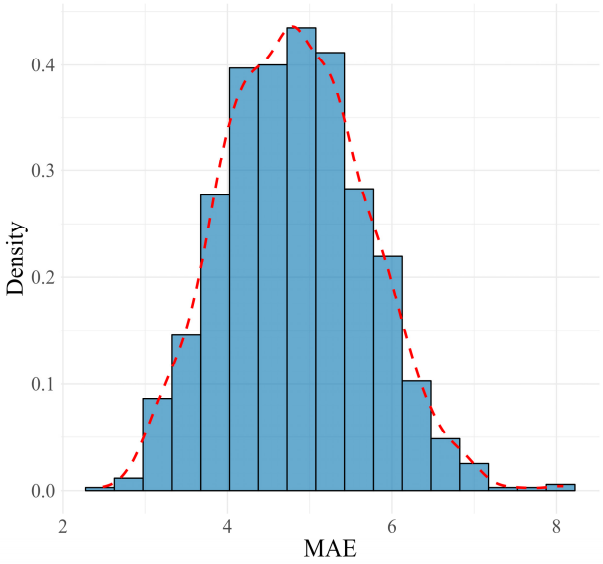

Figure 3. Distribution and density curve of MAE values from 1000 MCCV repetitions, demonstrating the consistency and reliability of the predictive model in training.

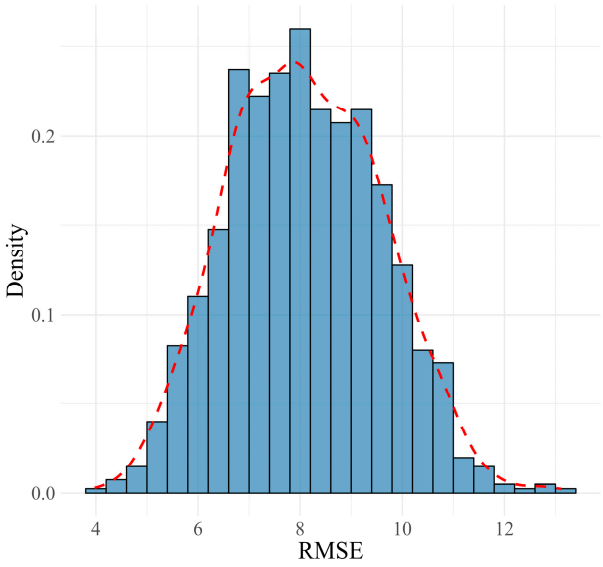

Figure 4. Distribution and density curve of RMSE values from 1000 MCCV repetitions, assessing the model's accuracy and stability during the training phase.

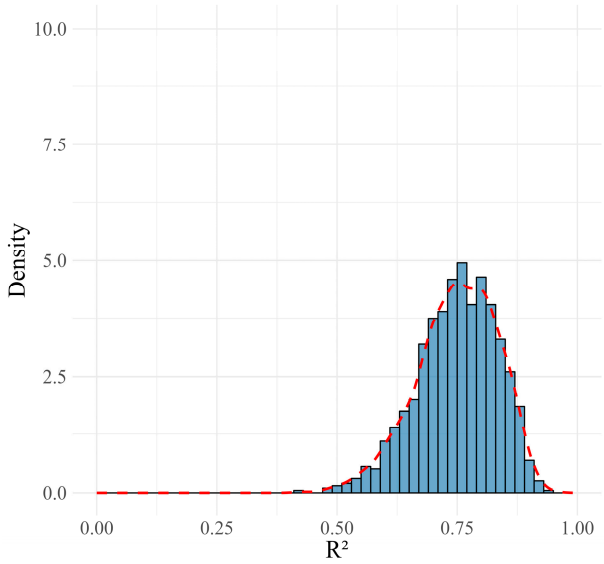

Figure 5. Distribution and density curve of R^2 values from 1000 MCCV repetitions, showcasing the predictive power and fit of the model to the training data.

Table 1. Results.

Metric	Training Set	Validation Set
MAE	4.82	11.22
RMSE	8.09	13.99
R^2	0.75	0.36

The True vs. Predicted Values plot (Figure 6) demonstrates the relationship between the actual PDQ-39 SI scores and the scores predicted by the model. The plot suggests that the model tends to slightly overpredict at lower values and underpredict at higher values, indicating areas for potential refinement. The Residuals vs. Predicted Values plot (Figure 7) shows that while most residuals are clustered around zero, there is a slight spread, suggesting occasional larger prediction errors.

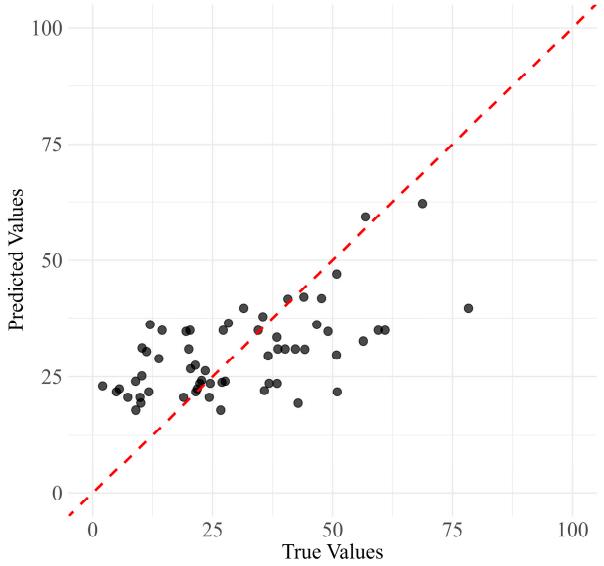

Figure 6. Scatter plot showing the relationship between actual PDQ-39 SI scores (true values) and model predictions (predicted values). The black dots represent individual predictions, and the red dotted line represents the line of perfect prediction where predicted values would equal true values.

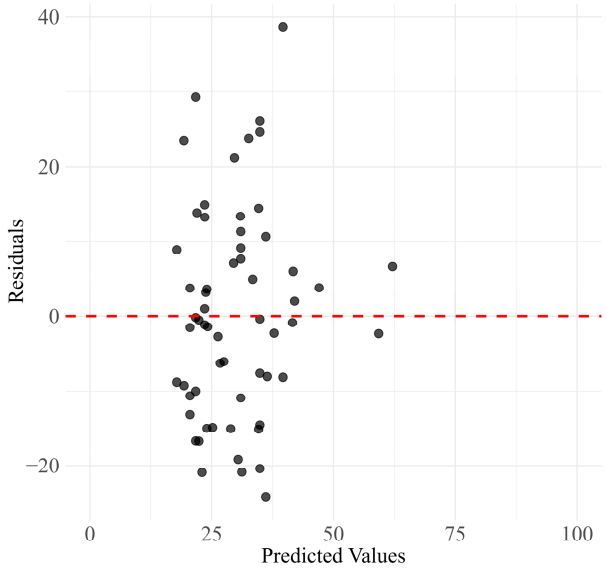

Figure 7. Scatter plot of residuals (errors) against predicted values, indicating prediction accuracy and error distribution. The black dots represent the residuals (actual value − predicted value), and the red dashed line represents the zero-error line.

3.2. Interpretation

The ensemble model demonstrated strong predictive capabilities on the training dataset, as indicated by the high R^2 and low MAE values. However, the model performance on the validation set showed moderate variation. Although the model remains a promising tool for estimating PDQ-39 SI scores, these variations suggest areas for future research and model refinement. The observed variations in the validation set may be attributed to the limited size of the dataset and the inherent variability of HRQoL measures, both of which warrant further investigation. The diagnostic plots suggest that while the model generally performs well, addressing the slight overprediction at lower values and underprediction at higher values could enhance its accuracy, aligning with the variability observed in the validation set. The variables selected through feature selection for their significant impact on PDQ-39 SI predictions are listed in Table 2. Feature importance and effects .

The interpretable ML analysis using the iml package in R yielded substantial insights into the significance of the individual features within our predictive model. Table 2 encapsulates the importance of each feature along with the associated 90% confidence intervals, as derived from Permutation Feature Importance, which quantifies the decrease in model accuracy when each feature's data is randomly shuffled. Predominantly, age at diagnosis, patient's country, dementia, and patient's age emerged as the most important features affecting PDQ-39 SI predictions. A graphical representation of feature importance was generated to provide a holistic view of each feature's contribution to the model's predictive performance (Figure 8).

Table 2. Selected Features and their importance in PDQ-39 SI predictions [a–c].

Feature	Importance (0.05)	Importance	Importance (0.95)	Permutation Error
Age at diagnosis	1.07	1.16	1.25	13.02
Country	1.05	1.12	1.19	12.53
Dementia	1.02	1.09	1.11	12.26
Age	1.02	1.09	1.19	12.24
Depression	1.03	1.09	1.15	12.19
Time to first levodopa prescription post-diagnosis	1.02	1.08	1.13	12.07

Table 2. *Cont.*

Feature	Importance (0.05)	Importance	Importance (0.95)	Permutation Error
Number of comorbidities	1	1.05	1.11	11.83
Anxiety	1.02	1.05	1.07	11.74
First prescribed oral medication contained levodopa	1	1.02	1.05	11.5
Currently taking Rivastigmine	1	1.02	1.05	11.5
Education	0.98	1.02	1.06	11.46
Peripheral vascular arterial disease	0.99	1.01	1.04	11.37
Diabetes (type 1 or 2)	0.99	1.01	1.02	11.32
Currently taking Catechol-O-methyltransferase inhibitor	0.97	1	1.03	11.21
Currently taking Amantadine	0.98	0.99	1.01	11.07

[a] The importance of a feature is quantified by permuting the values of the feature and measuring the change in the model's performance against a baseline using MAE. The displayed results include the estimated importance values alongside their 90% confidence intervals, represented by the lower (Importance (0.05)) and upper (Importance (0.95)) limits. [b] A larger permutation error indicates higher feature importance. [c] The process was executed for each feature and repeated 100 times to obtain a more robust estimate, with the results averaged.

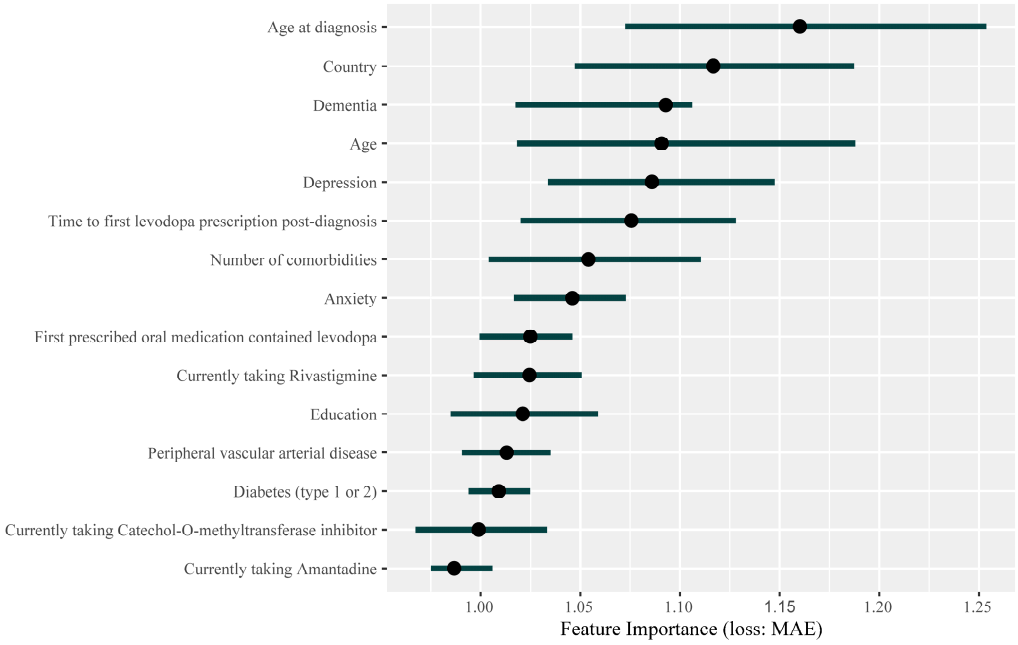

Figure 8. Demonstration of the feature importance of clinical variables in predicting PDQ-39 SI.

Transitioning to a more detailed analysis, the Accumulated Local Effects (ALE) plots facilitate visual exploration of the behavior of the model, underscoring how the prediction shifts locally with variations in each feature. The plots for the features age at diagnosis, dementia, patient's country, and patient's age are represented in Figures 9–12. The marks beneath the *x*-axis represent the distribution of each feature where observations were obtained. In the case of categorical variables such as dementia and the patient's country, the marks were automatically spaced to simplify the visual interpretation. For the features age at diagnosis and patient's age, regions with sparse or no points suggest caution in interpretation due to insufficient data.

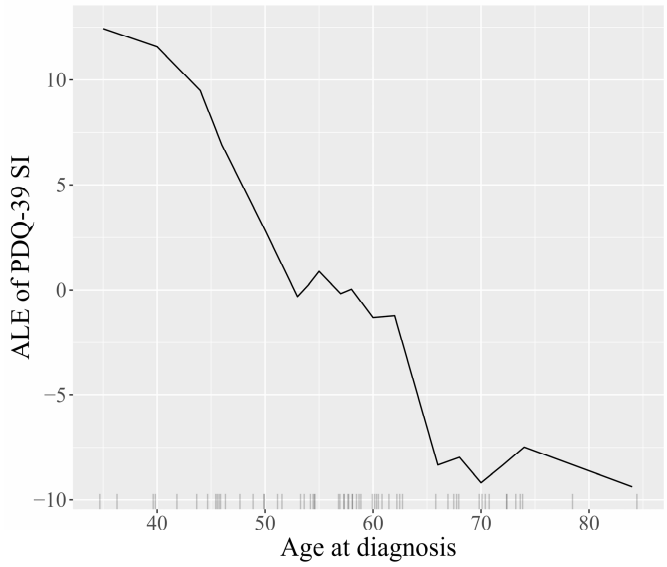

Figure 9. ALE plot illustrating the effect of age at diagnosis on PDQ-39 SI predictions, showing how earlier or later disease appearance impacts the quality of life in PD patients.

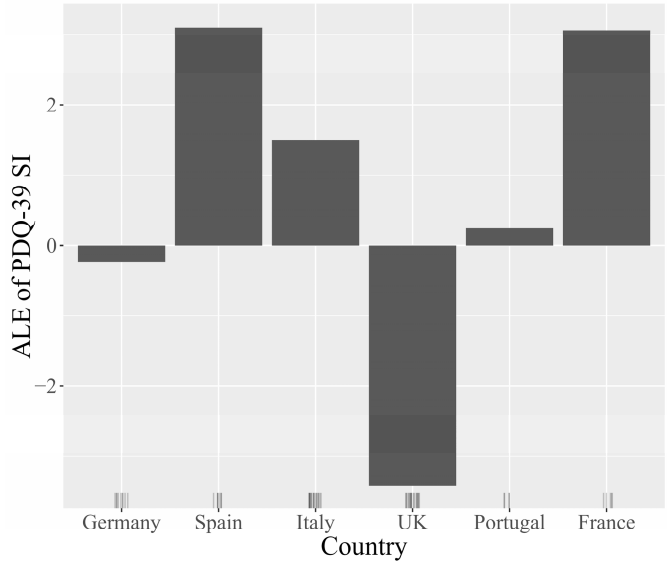

Figure 10. ALE plot illustrating the effect of the patient's country of residence on PDQ-39 SI predictions, highlighting the variations in the quality of life among PD patients across different countries.

The ALE plot for age at diagnosis (Figure 9) indicates that the earlier the disease appears, the higher the scores predicted by the model on the PDQ-39 SI, and therefore the more likely the patients are to have a lower quality of life, which is consistent with disease progression. The ALE plot for the patient's country (Figure 10) indicates a moderate impact on the predictions of PDQ-39 SI, with the United Kingdom showing a slightly more negative impact on the outcome. The ALE plot for dementia (Figure 11) demonstrates the significant negative impact of this comorbidity on the PDQ-39 SI predictions, highlighting its burden on PD patients. The ALE plot for the patient's age (Figure 12) demonstrates a consistent trend, indicating a direct influence of age on PDQ-39 SI predictions. This interpretive analysis not only enhances the comprehensibility of our predictive model but also extends the foundation for a more informed evaluation of the features affecting the PDQ-39 SI.

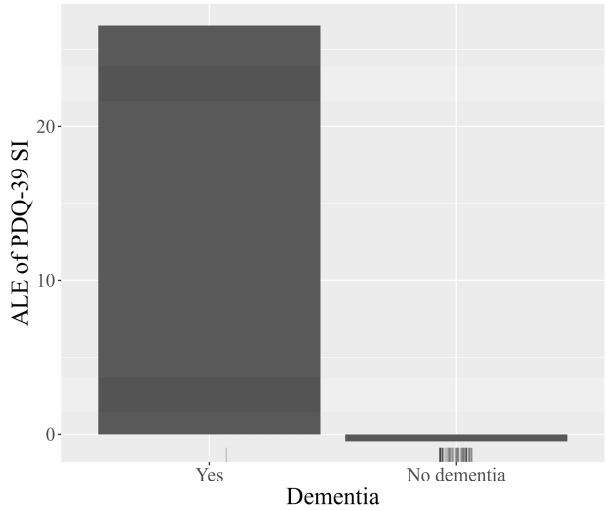

Figure 11. ALE plot demonstrating how the presence of dementia affects PDQ-39 SI scores predicted by the model, emphasizing its considerable negative impact on the quality of life in PD patients.

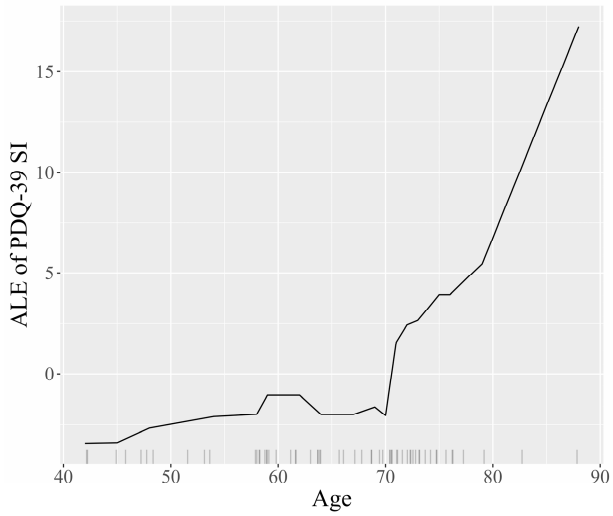

Figure 12. ALE plot detailing the impact of age on PDQ-39 SI predictions, illustrating the direct influence of age increments on the PDQ-39 SI scores predicted by the model.

4. Discussion

4.1. Principal Results

The primary achievement of this study is the successful development and validation of an ensemble ML model tailored to predict the PDQ-39 SI score in patients with PD. The model's strong performance in the training set, particularly its high R^2 value, underscores its ability to account for a significant proportion of the variance in PDQ-39 SI scores. This is a noteworthy advancement given the complex interplay of factors that can influence the quality of life in patients with PD. However, it is important to note that the significant difference in R^2 values between the training and validation sets indicates a potential issue of overfitting. The R^2 value in the training set was substantially higher than that in the validation set, suggesting that the model may have learned patterns specific to the training data that do not generalize well to unseen data. This is common in ML applications and serves as a valuable indicator of areas that may benefit from further research and optimization. Future work should focus on addressing this overfitting issue by exploring techniques such as regularization, cross-validation, and the inclusion of more diverse training data. Additionally, model simplification and the use of ensemble methods with a more rigorous selection of base models may help enhance generalizability.

The model's performance in the validation set, although not as strong as that in the training set, still suggests that it holds promise as a tool for estimating PDQ-39 SI scores in a clinical setting. Despite the inherent uncertainty in predictions of patient-reported outcome measures, the model provides valuable insights into disease progression and highlights the potential impact of clinical variables on the quality of life in PD patients. This indicates that the model could become a useful decision-support tool, with further refinement and validation needed to enhance its reliability and applicability.

The feature importance analysis conducted in this study provides a granular understanding of the variables that significantly impact the PDQ-39 SI score. This approach not only predicts the PDQ-39 SI score but also elucidates the underlying factors contributing to the model's predictions, offering a more comprehensive insight into the determinants of HRQoL in patients with PD. The notable importance of variables such as age at diagnosis, dementia, and patient's age in our model underlines the well-documented interplay among these factors and PD, as they are highly correlated with disease progression. Age is a recognized risk factor for PD, and both the prevalence and severity of the disease notably increase with age, underscoring the progressive nature of PD. The significant association between age and PD is evident from various epidemiological studies that show an exponential increase in PD prevalence from the age of 60 years onwards, reaching a peak in the age group of 70–79 years [25]. Additionally, an early appearance of the disease (represented by the feature 'age at diagnosis') is associated with a greater impact on the quality of life, highlighting the significant burden of the disease when it manifests early. Furthermore, dementia is a common comorbidity in patients with PD, often exacerbating the severity of the disease and the patient's quality of life. The prevalence of dementia in PD patients ranges between 24% and 31%, with the risk increasing substantially with age [26].

Our model's emphasis on age at diagnosis, dementia, and patient's age as significant predictors of the PDQ-39 SI reflects the clinically observed interactions between these variables and PD symptom severity. This alignment with established clinical knowledge not only enhances the validity of our model but also underscores the potential utility of our model in clinical settings. Utilizing these critical predictors, healthcare providers can anticipate the likely trajectory of PD patients' quality of life, aiding in the development of personalized treatment plans. It is also worth noting that other variables, such as anxiety, education, and some comorbidities, were identified as relevant by our feature selection method but did not show as high importance in the final model's predictions. These variables might still be important for adjusting the model to control for confounding effects and improve overall accuracy, even though they do not individually have strong predictive power [27].

Although the model does provide important information on the features affecting the PDQ-39 SI, it is also important to highlight that the predictive capabilities still demonstrate some variation, especially in the validation set. This could indicate that while the clinical variables are useful, the features available in the dataset do not entirely capture the PDQ-39 SI. The observed variability suggests that there are other influential factors impacting the PDQ-39 SI that are not represented in the current dataset. As such, future research should explore the inclusion of additional variables and data sources that may better reflect the complexities of HRQoL in PD patients. This would potentially enhance the model's accuracy and provide a more comprehensive tool for clinical application.

4.2. Limitations

One of the key limitations of our study was the small sample size. This could have contributed to the observed variations in performance metrics between the training and validation sets as small sample sizes can significantly impact ML performance estimates [28]. Additionally, we decided to exclude incomplete data and only use complete observations, resulting in a total of 627 valid entries. This decision was necessary because data imputation negatively impacted model performance. Future research should aim to address these limitations by incorporating larger and more diverse datasets, which could enhance

the model's generalizability and robustness. Exploring different imputation methods or techniques to handle missing data more effectively might also improve model performance.

4.3. Comparison with Prior Work

Our study diverges from the approach taken by Candel-Parra et al. for predicting the PDQ-39 SI score in patients with PD. Specifically, Candel-Parra et al. utilized classification models to categorize the PDQ-39 SI score into binary categories (above or below 40), whereas our study employed regression models to predict the actual score value [29]. This distinction is particularly significant given the absence of established cut-off values for the PDQ-39 SI score. Classification models, such as those used by Candel-Parra et al., require such cut-offs to categorize outcomes, introducing potential arbitrariness and statistical power loss when no clinically validated thresholds exist [29].

In contrast, our regression model predicts the actual PDQ-39 SI score, providing a more nuanced and precise assessment of a patient's quality of life. Treating the PDQ-39 SI scores as a continuous variable aligns our model more closely with the actual behavior of the scores, enhancing its applicability and interpretability in clinical settings. Furthermore, our use of explainable ML techniques allows clinicians to simulate disease progression and understand which variables are influencing the quality of life, providing actionable insights that can inform treatment decisions and patient management. This feature is particularly valuable, as it enables healthcare providers to tailor interventions specifically to the factors that have the greatest impact on a patient's HRQoL, such as early diagnosis, dementia, and patient's age.

5. Conclusions

The ensemble model developed in this study is a promising tool for healthcare providers to predict the HRQoL of patients with Parkinson's Disease, thereby aiding in personalized treatment strategies. Given that PDQ-39 SI is commonly used as an endpoint in clinical trials, the ability to predict this score could offer valuable insights for trial design and patient management. By leveraging regression models, our approach offers a nuanced understanding of a patient's quality of life, filling a gap in existing research methodologies.

While the PDQ-39 is a reliable and valid tool for assessing HRQoL in PD patients, it primarily serves as a cross-sectional assessment. In contrast, our predictive model enhances the utility of the PDQ-39 by providing dynamic predictions of the PDQ-39 SI score based on readily available clinical variables. This predictive capability allows for the simulation of disease progression and the identification of key factors influencing HRQoL over time, such as age at diagnosis, dementia, and patient's age. Consequently, clinicians can gain deeper insights into how individual patient characteristics and changes in clinical status impact the quality of life, enabling more personalized and proactive management strategies. The integration of explainable ML techniques further distinguishes our model by offering transparency into the prediction process, thereby empowering healthcare providers to understand and trust the model's recommendations. The practical feature of utilizing clinical variables that are typically available at the moment of use, as emphasized in a recent critique of predictive models, enhances the usability and real-world applicability of our model, setting it apart from models that depend on variables that are not readily accessible in the clinical context [30].

Although the model's performance on the validation set exhibited moderate variation, it sets the stage for future research in aspects such as feature engineering, hyperparameter optimization, and incorporation of additional clinically relevant variables. Moreover, the model's performance is likely to improve with a larger sample size, given the inherent variability of patient-reported outcomes, such as those measured by the PDQ-39 SI. Future work should also address the potential overfitting suggested by the difference in R^2 values between the training and validation sets.

While the model does provide important insights into the features affecting the PDQ-39 SI, it is essential to recognize that the predictive capabilities still demonstrate some

variation, especially in the validation set. This variability suggests that the current dataset's features do not entirely capture the PDQ-39 SI, indicating that additional variables and data sources might be necessary to better reflect the complexities of HRQoL in PD patients. Future research should explore these additional factors to enhance the model's accuracy and provide a more comprehensive tool for clinical application.

The potential impact of the model on clinical practice and clinical trials underscores its value as a significant contribution to the field. Furthermore, the broader implications of this work have the potential to significantly influence how healthcare providers approach PD treatment, thereby contributing to more personalized and effective care.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jcm13175081/s1, Table S1: Descriptive Statistics of the PRISM Database. Figure S1: Factor Analysis of Mixed Data.

Author Contributions: Conceptualization, D.M., T.T.-G. and A.S.B.; methodology, D.M., T.T.-G. and A.S.B.; software, D.M.; validation, D.M., T.T.-G., A.S.B. and J.M.; formal analysis, D.M.; data curation, D.M.; writing—original draft preparation, D.M.; writing—review and editing, T.T.-G., A.S.B. and J.M.; supervision, T.T.-G. and A.S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been conducted under the scope of and funded by the PhD Program in Health Data Science of the Faculty of Medicine of the University of Porto, Portugal, heads.med.up.pt.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the use of a public database where the original study confirmed that no NHS Research Ethics Committee (REC) approval was re-quired. The study involved voluntary participation, with no identifying information collected, and complied with General Data Protection Regulation (GDPR).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study were obtained from the Parkinson's Real-world Impact Assessment (PRISM) database (www.prism.bial.com), accessed on 18 February 2022.

Acknowledgments: The PRISM study and database were funded by BIAL—Portela and C[a], S.A., designed in collaboration with the Cure Parkinson's Trust, an advocacy group based in the United Kingdom (UK), and reviewed by the PRISM Steering Committee.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Feigin, V.L.; Krishnamurthi, R.V.; Theadom, A.M.; Abajobir, A.A.; Mishra, S.R.; Ahmed, M.B.; Abate, K.H.; Mengistie, M.A.; Wakayo, T.; Abd-Allah, F.; et al. Global, Regional, and National Burden of Neurological Disorders during 1990–2015: A Systematic Analysis for the Global Burden of Disease Study 2015. *Lancet Neurol.* **2017**, *16*, 877–897. [CrossRef]
2. Jankovic, J. Parkinson's Disease: Clinical Features and Diagnosis. *J. Neurol. Neurosurg. Psychiatry* **2008**, *79*, 368–376. [CrossRef] [PubMed]
3. World Health Organization. *Parkinson Disease: A Public Health Approach*; World Health Organization: Geneva, Switzerland, 2022; Volume 291, p. 390.
4. Massano, J.; Bhatia, K.P. Clinical Approach to Parkinson's Disease: Features, Diagnosis, and Principles of Management. *Cold Spring Harb. Perspect. Med.* **2012**, *2*, a008870. [CrossRef]
5. Marinus, J. Health Related Quality of Life in Parkinson's Disease: A Systematic Review of Disease Specific Instruments. *J. Neurol. Neurosurg. Psychiatry* **2002**, *72*, 241–248. [CrossRef] [PubMed]
6. Heluani, A.S.; de Porto, F.H.G.; Listik, S.; de Campos, A.W.; Machado, A.A.C.; Cukiert, A.; de Oliveira, J.O. Neuropsychological and Quality of Life Assessment in Patients with Parkinson's Disease Submitted to Bilateral Deep Brain Stimulation in the Subthalamic Nucleus. *Dement. Neuropsychol.* **2012**, *6*, 260–265. [CrossRef]
7. Schönenberg, A.; Prell, T. Measuring Quality of Life with the Parkinson's Disease Questionnaire-39 in People with Cognitive Impairment. *PLoS ONE* **2022**, *17*, e0266140. [CrossRef] [PubMed]
8. Galeoto, G.; Colalelli, F.; Massai, P.; Berardi, A.; Tofani, M.; Pierantozzi, M.; Servadio, A.; Fabbrini, A.; Fabbrini, G. Quality of Life in Parkinson's Disease: Italian Validation of the Parkinson's Disease Questionnaire (PDQ-39-IT). *Neurol. Sci.* **2018**, *39*, 1903–1909. [CrossRef]
9. Martinez-Martin, P. The Importance of Non-Motor Disturbances to Quality of Life in Parkinson's Disease. *J. Neurol. Sci.* **2011**, *310*, 12–16. [CrossRef]

10. Jenkinson, C.; Fitzpatrick, R.; Peto, V.; Greenhall, R.; Hyman, N. The PDQ-8: Development and Validation of a Short-Form Parkinson's Disease Questionnaire. *Psychol. Health* **1997**, *12*, 805–814. [CrossRef]
11. Rocha, F.; Tolosa, E.; Ferreira, J.; Rascol, O.; Poewe, W.; Santos, A.; Magalhaes, D.; Soares-da-Silva, P. Effect of Opicapone in Parkinson's Disease Patients with 'Early' Motor Fluctuations: Parkinson's Disease Questionnaire (PDQ-39) Analysis from the BIPARK-I Double-Blind Experience. *J. Neurol. Sci.* **2019**, *405*, 190–191. [CrossRef]
12. Hattori, N.; Takeda, A.; Hanya, Y.; Kitagawa, T.; Arai, M.; Furusawa, Y.; Mochizuki, H.; Nagai, M.; Takahashi, R. Effects of Rasagiline on Parkinson's Disease Questionnaire (PDQ-39) Emotional Well-Being Domain in Patients with Parkinson's Disease: A Post-Hoc Analysis of Clinical Trials in Japan. *PLoS ONE* **2022**, *17*, e0262796. [CrossRef] [PubMed]
13. Chandrabhatla, A.S.; Pomeraniec, I.J.; Ksendzovsky, A. Co-Evolution of Machine Learning and Digital Technologies to Improve Monitoring of Parkinson's Disease Motor Symptoms. *NPJ Digit. Med.* **2022**, *5*, 32. [CrossRef] [PubMed]
14. Gao, C.; Sun, H.; Wang, T.; Tang, M.; Bohnen, N.I.; Müller, M.L.T.M.; Herman, T.; Giladi, N.; Kalinin, A.; Spino, C.; et al. Model-Based and Model-Free Machine Learning Techniques for Diagnostic Prediction and Classification of Clinical Outcomes in Parkinson's Disease. *Sci. Rep.* **2018**, *8*, 7129. [CrossRef] [PubMed]
15. Nilashi, M.; Abumalloh, R.A.; Minaei-Bidgoli, B.; Samad, S.; Yousoof Ismail, M.; Alhargan, A.; Abdu Zogaan, W. Predicting Parkinson's Disease Progression: Evaluation of Ensemble Methods in Machine Learning. *J. Healthc. Eng.* **2022**, *2022*, 2793361. [CrossRef]
16. Dadu, A.; Satone, V.; Kaur, R.; Hashemi, S.H.; Leonard, H.; Iwaki, H.; Makarious, M.B.; Billingsley, K.J.; Bandres-Ciga, S.; Sargent, L.J.; et al. Identification and Prediction of Parkinson's Disease Subtypes and Progression Using Machine Learning in Two Cohorts. *NPJ Parkinsons Dis.* **2022**, *8*, 172. [CrossRef]
17. Faiem, N.; Asuroglu, T.; Acici, K.; Kallonen, A.; van Gils, M. Assessment of Parkinson's Disease Severity Using Gait Data: A Deep Learning-Based Multimodal Approach. *Commun. Comput. Inf. Sci.* **2024**, *2084*, 29–48. [CrossRef]
18. Tolosa, E.; Ebersbach, G.; Ferreira, J.J.; Rascol, O.; Antonini, A.; Foltynie, T.; Gibson, R.; Magalhaes, D.; Rocha, J.F.; Lees, A. The Parkinson's Real-World Impact Assessment (PRISM) Study: A European Survey of the Burden of Parkinson's Disease in Patients and Their Carers. *J. Parkinsons Dis.* **2021**, *11*, 1309–1323. [CrossRef]
19. R: The R Project for Statistical Computing. Available online: https://www.r-project.org/ (accessed on 12 July 2023).
20. van Buuren, S.; Groothuis-Oudshoorn, K. Mice: Multivariate Imputation by Chained Equations in R. *J. Stat Softw.* **2011**, *45*, 1–67. [CrossRef]
21. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat Softw.* **2008**, *28*, 1–26. [CrossRef]
22. Pavlyshenko, B. Using Stacking Approaches for Machine Learning Models. In Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 21–25 August 2018; pp. 255–258.
23. Molnar, C.; Casalicchio, G.; Bischl, B. Iml: An R Package for Interpretable Machine Learning. *J. Open Source Softw.* **2018**, *3*, 786. [CrossRef]
24. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed.; Leanpub: Victoria, BC, Canada, 2022.
25. Pringsheim, T.; Jette, N.; Frolkis, A.; Steeves, T.D.L. The Prevalence of Parkinson's Disease: A Systematic Review and Meta-Analysis. *Mov. Disord.* **2014**, *29*, 1583–1590. [CrossRef] [PubMed]
26. Aarsland, D.; Andersen, K.; Larsen, J.P.; Lolk, A.; Nielsen, H.; Kragh-Sørensen, P. Risk of Dementia in Parkinson's Disease. *Neurology* **2001**, *56*, 730–736. [CrossRef] [PubMed]
27. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2021. [CrossRef]
28. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine Learning Algorithm Validation with a Limited Sample Size. *PLoS ONE* **2019**, *14*, e0224365. [CrossRef] [PubMed]
29. Candel-Parra, E.; Córcoles-Jiménez, M.P.; Delicado-Useros, V.; Ruiz-Grao, M.C.; Hernández-Martínez, A.; Molina-Alarcón, M. Predictive Model of Quality of Life in Patients with Parkinson's Disease. *Int. J. Environ. Res. Public Health* **2022**, *19*, 672. [CrossRef]
30. Collins, G.S.; Dhiman, P. Prediction Models Should Contain Predictors Known at the Moment of Intended Use. *Aging Clin. Exp. Res.* **2023**, *35*, 3243–3244. [CrossRef] [PubMed]

Appendix 3

Non-Motor Symptoms as Critical Predictors of Quality of Life in Parkinson's Disease: A Machine Learning Approach

Health and Quality of Life Outcomes

RESEARCH

Open Access

Non-motor symptoms as critical predictors of quality of life in Parkinson's disease: a machine learning approach

Daniel Magano[1,2*], António S. Barros[3], João Massano[4,5], Laila Alsuwaidi[6] and Tiago Taveira-Gomes[6,7,8]

Abstract

Background Parkinson's disease (PD) considerably impacts health-related quality of life (HRQoL) through motor and non-motor symptoms. The Parkinson's Disease Questionnaire-39 (PDQ-39) is the most widely used tool to assess HRQoL, encompassing eight dimensions and a Summary Index providing an overall score. Despite advances in machine learning (ML) for predicting disease symptoms and progression, its application to predict HRQoL across these dimensions remains underexplored.

Methods This study uses complete-case data for 478 of 861 patients from PRISM, a cross-sectional observational survey conducted in six European countries in 2018–2019. Participants were adults with PD recruited through advocacy groups and clinical centers who completed online assessments, providing data on demographics, medication, comorbidities, and disease characteristics (Tolosa et al., 2021). ML models were trained to predict PDQ-39 dimensions and Summary Index scores (0–100; higher = worse HRQoL). Features were preselected using the Boruta algorithm on the training data. Model selection was based on the lowest mean RMSE from 100 bootstrap resamples on the training set. Selected models were then retrained using 1000 bootstrap resamples for robust performance estimation. Final performance was evaluated on a held-out 20% validation set using R^2, MAE, and RMSE. Feature importance was assessed using permutation importance with MAE loss (100 permutations) on the held-out validation set. Factor Analysis of Mixed Data (FAMD) was used to explore patterns between non-motor symptoms and PDQ-39.

Results Selected models: xgbTree (Summary Index; Activities of Daily Living) and gausssprPoly (all other PDQ-39 dimensions). On the validation set, Summary Index/ Cognitions showed the strongest performance with R^2 = 0.56/0.53, MAE = 9.60/12.39, RMSE = 12.66/16.20. Permutation feature importance ranked the Non-Motor Symptoms Questionnaire score (sum of 30 non-motor symptoms, range 0–30) as the most important predictor across all models. FAMD showed clustering of Social Support, Bodily Discomfort, and Stigma dimensions with Anxiety.

Conclusions Our findings demonstrate the critical role of non-motor symptoms in predicting HRQoL in patients with PD. While ML models effectively predict overall HRQoL and cognitive aspects, achieving comparable performance on other dimensions may require additional variables to reduce error. These insights emphasize comprehensive treatment strategies addressing both motor and non-motor symptoms.

*Correspondence:
Daniel Magano
danielmagano@outlook.com

Full list of author information is available at the end of the article

Keywords Parkinson's disease, Health-related quality of life, Predictive modeling, Machine learning, Parkinson's disease questionnaire-39, Explainable machine learning

Background

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects millions of people worldwide, with its global prevalence more than doubling between 1990 and 2015, highlighting the growing burden of this condition [1, 2]. PD is primarily characterized by motor symptoms such as resting tremor, rigidity, bradykinesia, and postural instability, which are often accompanied by a variety of non-motor symptoms (NMS), including cognitive impairment, mental health disorders, sleep disturbances, pain, and sensory disturbances [2–4]. These symptoms considerably reduce the health-related quality of life (HRQoL) in individuals with PD, making HRQoL an essential outcome measure in both clinical settings and research [5].

HRQoL instruments, particularly disease-specific measures, have become invaluable in capturing the personal impact of PD and the changes experienced by patients over time. Among these instruments, the Parkinson's Disease Questionnaire-39 (PDQ-39) is the most widely used and validated tool for assessing HRQoL in patients with PD [5–8]. It covers eight dimensions of HRQoL: Activities of Daily Living, Bodily Discomfort, Cognitions, Communication, Emotional Well-Being, Mobility, Social Support, and Stigma, providing a comprehensive overview of the disease's impact on HRQoL [9]. As a patient-reported outcome measure, the PDQ-39 reflects outcomes that matter to patients and aligns with regulatory guidance: both the European Medicines Agency and the U.S. Food and Drug Administration emphasize the use of validated instruments to inform benefit–risk assessment and, when appropriate, to support claims in product labeling [10, 11].

The PDQ-39 Summary Index, hereafter referred to as the Summary Index, is a composite score derived from the individual eight dimensions of the PDQ-39 and is frequently used in clinical trials as a primary outcome measure to evaluate the overall impact of PD on patients' lives [12, 13]. While the Summary Index provides a valuable summary of HRQoL, it can obscure the nuances captured by the individual dimensions, each reflecting a distinct aspect of the disease's impact. Understanding the relationship between these dimensions and clinical variables is crucial for tailoring interventions to address the specific challenges faced by patients with PD. However, this remains challenging due to the complex and multifaceted nature of the disease, where symptoms can vary widely among individuals. Notably, non-motor symptoms (NMS), which are often among the most disabling yet under-recognized aspects of PD, have been shown to considerably impact the quality of life [8]. The Non-Motor Symptoms Questionnaire (NMSQ) is a self-reported instrument of 30 yes/no items covering common non-motor symptoms in PD. The total NMSQ score is obtained by summing the number of symptoms the patient reports, yielding a score from 0 to 30, with higher scores reflecting a greater burden of non-motor symptoms. This simple sum provides a practical, validated measure of overall non-motor symptom burden in PD [14].

Traditional statistical models often fall short in capturing the nonlinear and high-dimensional relationships between the diverse clinical manifestations of PD and HRQoL. In response, ML methods have gained traction for their ability to model such complexity, with demonstrated success in predicting disease progression, motor subtypes, and treatment response [15–24]. While these advances highlight the potential of ML in neurodegenerative disease modeling, most applications to date have concentrated on motor symptoms, digital monitoring, and disease classification.

In contrast, ML models targeting HRQoL in PD, particularly using the PDQ-39, remain limited. Most existing studies treat HRQoL as a secondary outcome or focus solely on the global Summary Index, overlooking the predictive modeling of its individual dimensions [25–28]. Building on our earlier work published in the *Journal of Clinical Medicine* (2024), which focused on predicting the Summary Index, this study addresses a critical gap by extending the analysis to the eight individual PDQ-39 dimensions. By doing so, we aim to enhance the understanding of dimension-specific quality-of-life impacts and support more personalized clinical decision-making in PD care.

This study aims to achieve two critical objectives: first, to explore and clarify the relationship between standard clinical variables and both the PDQ-39 dimensions and the Summary Index; second, to develop and assess the potential of ML models for predicting these outcomes. By incorporating explainable ML techniques, this study seeks to enhance clinicians' understanding of HRQoL outcomes, offering deeper insights into the factors most strongly associated with the different dimensions of HRQoL in PD. This approach is intended to inform more personalized treatment strategies, ultimately improving care and HRQoL for patients with PD.

Materials and methods

Study overview and analysis workflow

Figure 1 summarizes the workflow from the PRISM data source and analytic sample selection through model development, evaluation, and interpretability. It is intended as an overview; full details are provided in the subsequent Methods subsections.

Database

The data for this study were sourced from the Parkinson's Real-world Impact assesSMent (PRISM) study, a cross-sectional observational research initiative conducted across six European countries: France, Germany, Italy, Portugal, Spain, and the United Kingdom. Detailed information on the study design and recruitment can be found in Tolosa et al., 2021 [29]. The PRISM study included 861 patients with PD and 256 caregivers, recruited via patient advocacy groups and clinical centers. Participants completed a comprehensive online survey comprising validated instruments, including the PDQ-39, the NMSQ, as well as questions on demographics, medication, comorbidities, disease characteristics, and healthcare resource utilization. This diverse database supports robust modeling and generalizability to the wider European PD population.

Data preparation and analysis

The PRISM dataset underwent a data cleaning process using R statistical software (version 4.3.2) [30]. Subsequently, all analyses were performed with the same software to ensure consistency and reproducibility. For categorical variables such as gender, medication, and

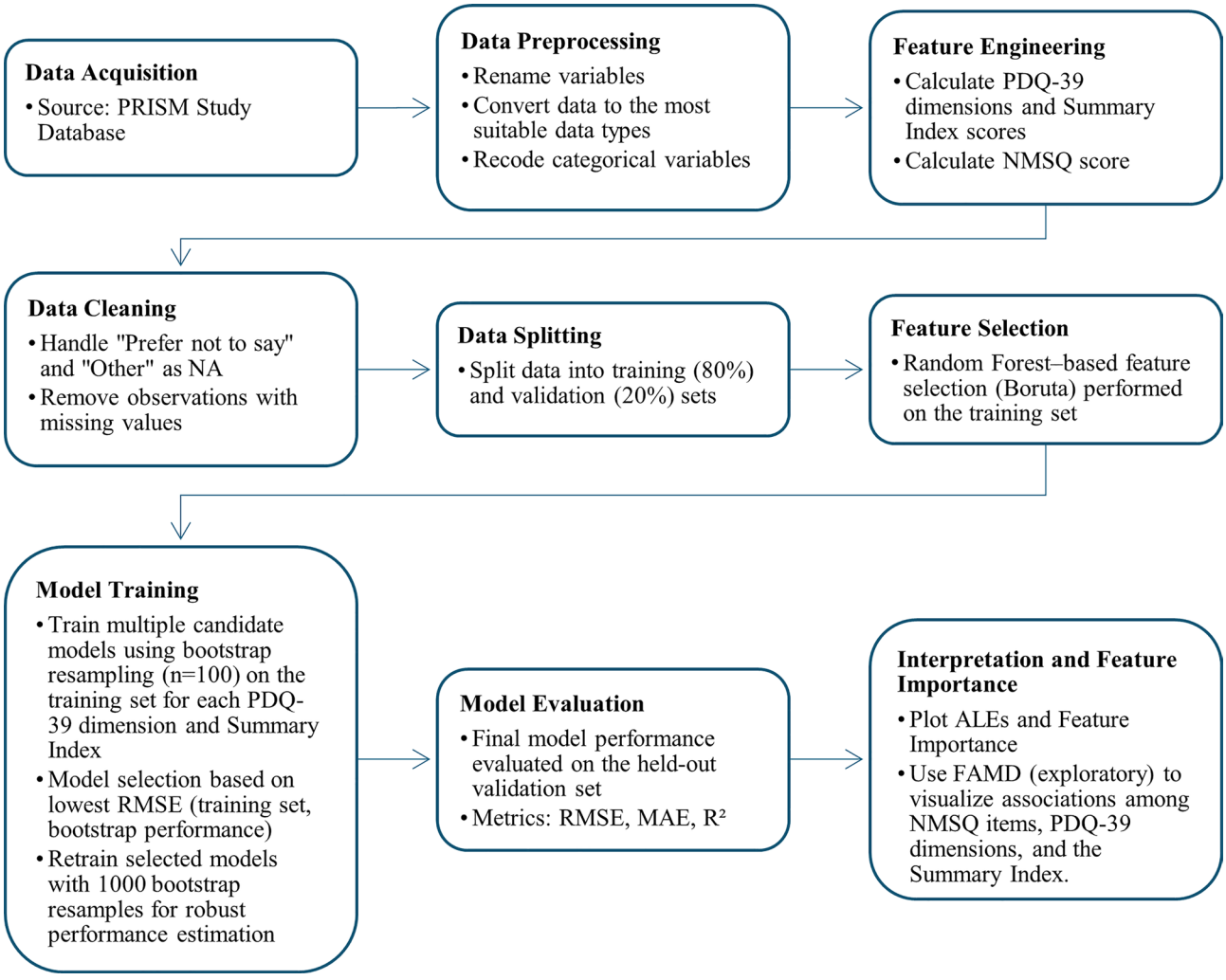

Fig. 1 Diagram of the study's workflow. PRISM, Parkinson's Real-world Impact assesSMent; PDQ-39, Parkinson's Disease Questionnaire-39; NMSQ, Non-Motor Symptoms Questionnaire; RMSE, Root Mean Squared Error; MAE, Mean Absolute Error; R^2, Coefficient of Determination; ALE, Accumulated Local Effects; FAMD, Factor Analysis of Mixed Data. Note: "Recode categorical variables" refers to converting categorical variables to suitable numeric/dummy-coded variables for modeling. Model selection was performed using the training set only; the validation set was reserved for final, unbiased performance assessment

comorbidities, responses like "prefer not to say" and "other" were recoded as missing, as these options do not provide analyzable clinical information and would introduce noise if retained. Most variables in the dataset had low rates of missing data; however, a notable exception was observed for the NMSQ, which had 31.2% missing responses. This suggests a pattern of survey fatigue, where respondents' engagement decreased as the questionnaire progressed, leading to an increased number of missing responses. This indicates that the missing data in the NMSQ likely follow a missing not at random (MNAR) pattern, suggesting an attrition effect in survey completion. Additionally, the variables "age at diagnosis" and "current medication" exhibited an MNAR pattern, likely due to patients not recalling their age at PD diagnosis or being unaware of their current medication.

Although we initially attempted to address missing values using Multivariate Imputation by Chained Equations (MICE), this approach resulted in models with worse predictive capabilities [31]. Therefore, only cases with complete data for all variables required for modeling were retained, resulting in a final analytic sample of 478 patients. No missing data remained in the analytic sample.

The analytic sample ($n = 478$) comprised 46.9% female and 53.1% male participants, with a mean age of 65.6 years (Standard Deviation (SD) = 9.7 years). The distributions of the PDQ-39 dimensions and Summary Index are depicted in Fig. 2. The distribution of the NMSQ score is displayed in Fig. 3. Table 1 provides an overview of key variables from the 478 complete observations from PRISM database.

Internal consistency of the PDQ-39 Summary Index and its eight dimensions was assessed using Cronbach's α in the retained sample ($n = 478$). Scoring followed standard procedures, with all scores ranging from 0 to 100 and higher scores reflecting worse HRQoL. There were no missing items for the PDQ-39 in the analytic sample.

The dataset was then randomly partitioned into training (80%) and validation (20%) sets. This allocation allows for comprehensive model training on a diverse range of responses, critical for capturing the nuanced patterns of HRQoL assessments in PD, while still reserving a portion of the data for essential validation to assess model performance and overfitting.

Model training

We selected machine learning approaches for their scalability and proven ability to capture complex, nonlinear relationships that are often missed by traditional statistical models. This framework supports robust prediction in moderate-sized datasets and can be readily extended to larger cohorts as more data becomes available.

For feature selection, we used the Boruta algorithm. This method is based on Random Forest models and works by checking which variables are truly useful for predicting each PDQ-39 dimension and the Summary Index [32–35]. Feature selection was performed exclusively on the training set after the initial 80/20 train/validation split, to avoid information leakage. Using this approach, we identified nine distinct sets of features specific to each of the PDQ-39 dimensions and the Summary Index (see Table 4 in the Results section). These nine sets were then utilized in training a diverse range of ML models, namely linear regression, boosted generalized linear model, extreme gradient boosting with linear models and decision trees, ridge regression, neural network regression, LASSO regression, random forest, k-nearest neighbors, support vector machines with radial basis function and polynomial kernels, and Gaussian processes (linear, polynomial, and radial kernels), incorporating essential pre-processing steps such as centering and scaling with the caret package [36].

To evaluate the performance and stability of these models in a dataset characterized by limited size and high variability, we utilized bootstrap resampling exclusively on the training set, using the standard nonparametric (Efron's) bootstrap method [36, 37]. Bootstrap is a well-established approach that provides a more reliable estimate of the generalizability of machine learning models, particularly in smaller datasets, by generating multiple resampled sets and reducing the bias inherent in model evaluation [38]. Initially, each candidate model was evaluated using 100 bootstrap resamples, and the model with the lowest average Root Mean Squared Error (RMSE) in the training set was selected for each outcome, as RMSE is widely recognized for its effectiveness in assessing predictive performance by penalizing larger errors more than smaller ones [38]. The selected models were then retrained on the training set using 1000 bootstrap resamples to obtain robust estimates of their predictive performance.

The held-out validation set (20% of the data) was reserved strictly for final, unbiased model evaluation and was not subjected to bootstrap resampling. Validation set metrics were calculated as single-point estimates to reflect true out-of-sample predictive performance.

Model evaluation

The performance of the models was evaluated using multiple metrics on both the training set (80%) and the validation set (20%) to assess generalizability and overfitting. The validation set was specifically reserved for the final evaluation to ensure an unbiased assessment of models' performance. Mean Absolute Error (MAE) quantifies the average magnitude of prediction errors, providing a straightforward measure of how much predicted values

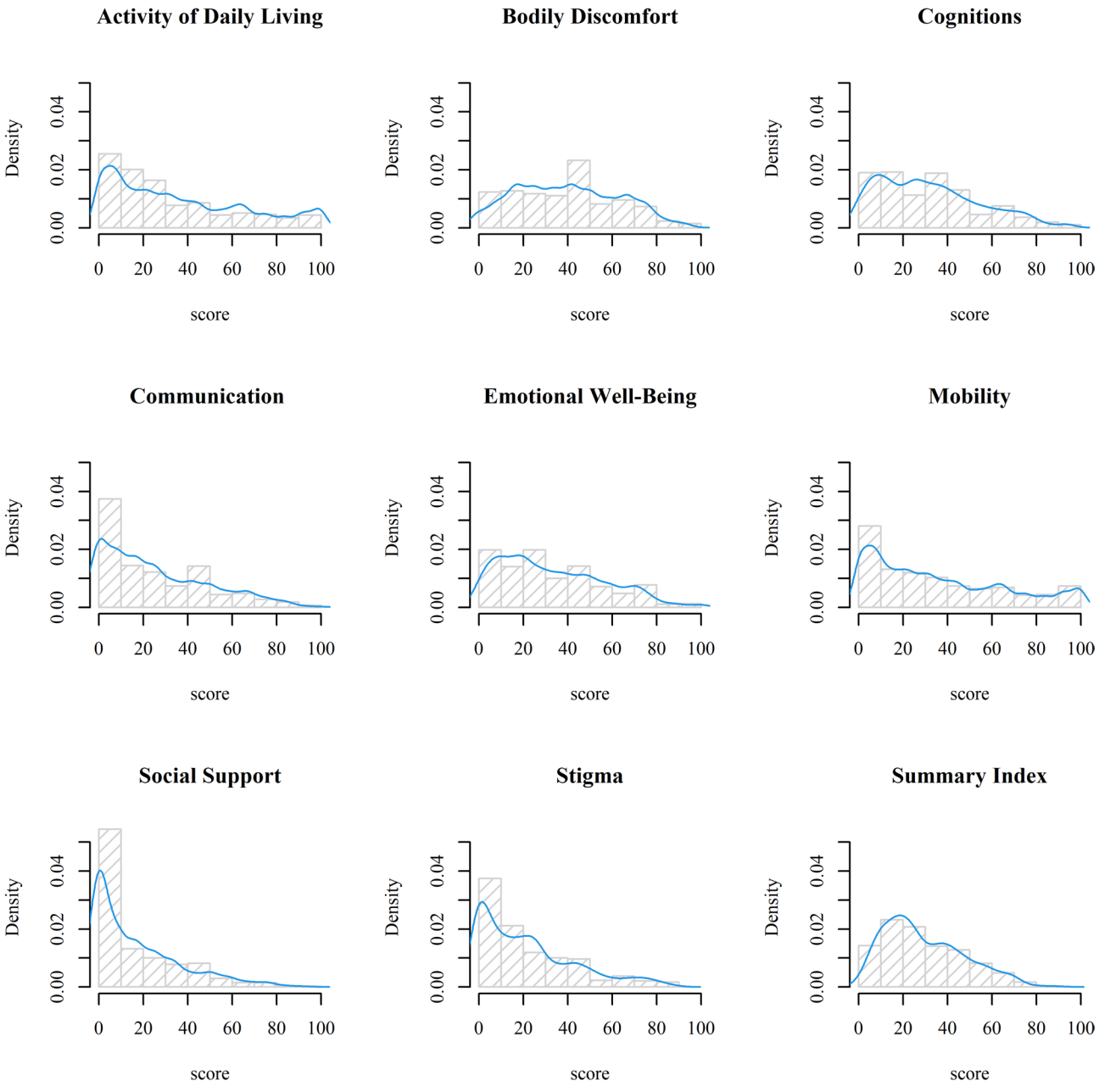

Fig. 2 Distribution and density plots of PDQ-39 dimensions and Summary Index scores across 478 patients from the PRISM study

deviate from true values. RMSE offers insight into the typical magnitude of prediction errors by emphasizing larger errors due to its squared nature, while the Coefficient of Determination (R^2) indicates the proportion of variance in the dependent variable explained by the model, highlighting its explanatory power. In addition to these metrics, scatter plots of predicted versus true values were used to evaluate how well predictions matched the observations, and studentized residual plots were examined for patterns or systematic errors. Together, these quantitative and visual evaluation tools provide a comprehensive assessment of model performance and robustness.

Because the Summary Index captures overall HRQoL and each model targets different aspects of patient experience, we evaluated and compared the predictive performance across all models. This allows us to see whether overall or specific outcomes are predicted with lower error and gives a broader view of which aspects of HRQoL are most predictable in this dataset.

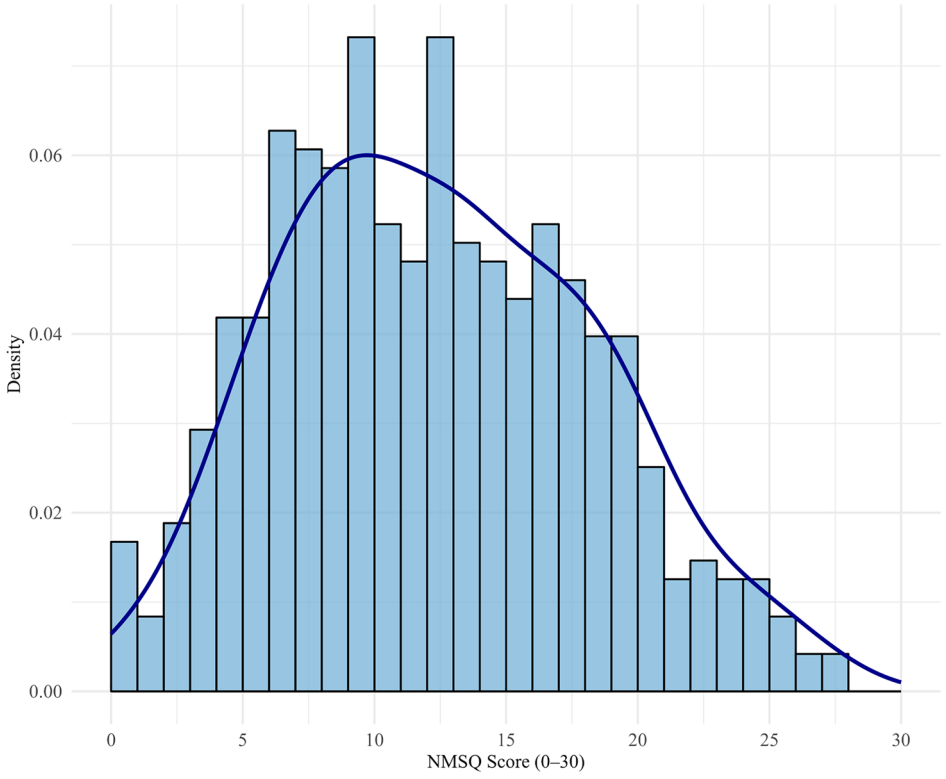

Fig. 3 Distribution and density of NMSQ score across 478 patients from the PRISM study

Interpretable machine learning analysis

To elucidate the importance and model reliance on each feature within our predictive models, we employed interpretable ML techniques, with a particular focus on Permutation Feature Importance (PFI). Utilizing the iml package in R, PFI provides deeper insight into the behavior of the ML models [39]. The core concept of PFI is to measure the increase in the model's prediction error after the permutation of each feature's values, which disrupts the relationship between the feature and the true outcome. A considerable increase in error indicates that the model depends heavily on that feature for predictions, while minimal or no change implies that the feature is not as relevant in the predictive process [40]. However, it's crucial to recognize that even features with minimal changes in prediction error may still contribute meaningfully to the overall performance of the model. These variables, while not strongly predictive on their own, can add complementary information and interact with other predictors, which may enhance model generalizability [41]. Feature importance was estimated using the final fitted model (i.e., the best-performing algorithm for each dimension and the Summary Index) applied to the independent validation set. We assessed feature importance in our models using MAE as the error metric for PFI calculations. This process was repeated 100 times to ensure robust estimates, with the results averaged to provide a clearer view of the models' reliance on each variable.

Additionally, to understand the impact of specific variables on our outcomes, we employed Accumulated Local Effects (ALE) plots. ALE plots allow us to visualize the effect of a feature on the predicted outcome while accounting for interactions with other features, particularly in datasets with correlated predictors [40, 42]. By generating ALE plots for the most relevant features identified through PFI, we were able to interpret how changes in these features affect the PDQ-39 dimensions and the Summary Index.

Furthermore, during our analysis, we observed that the NMSQ score (sum of non-motor symptoms reported by the patient) consistently ranked as the most important predictor across all ML models, when compared to the full set of clinical and demographic variables included in the models (see Table 4 in the Results section).

Exploratory multivariate analysis using FAMD

Given that the NMSQ comprises 30 items assessing various symptoms, we performed a Factor Analysis of Mixed Data (FAMD) to explore the underlying structure and relationships among these items, the eight PDQ-39 dimensions, and the Summary Index. FAMD is particularly suited for mixed data types and simplifies data complexity by identifying principal dimensions that

Table 1 Descriptive statistics of the complete observations from PRISM database

Characteristics	Overall (*N* = 478)
Demographics	
Age, years	
Mean (SD)	65.6 (9.7)
Median [IQR]	66.0 [59.0;72.0]
Min, Max	[38.0;91.0]
Gender	
Female	224 (46.9%)
Male	254 (53.1%)
Country	
France	19 (4.0%)
Germany	50 (10.5%)
Italy	141 (29.5%)
Portugal	45 (9.4%)
Spain	82 (17.2%)
UK	141 (29.5%)
Education	
University	206 (43.1%)
No University	272 (56.9%)
NMS Questionnaire	
Mean (SD)	12.5 (5.9)
Median [IQR]	12.0 [8.0;17.0]
Min, Max	[0.0;28.0]
Comorbidities	
Anxiety	72 (15.1%)
Asthma	25 (5.2%)
Cancer	23 (4.8%)
Chronic Obstructive Pulmonary Disease	10 (2.1%)
Dementia	24 (5.0%)
Depression	97 (20.3%)
Diabetes (type 1 or 2)	29 (6.1%)
Gastric ulcer	22 (4.6%)
Heart issues	36 (7.5%)
High blood pressure	126 (26.4%)
Kidney disease	9 (1.9%)
Liver disease	8 (1.7%)
Peripheral vascular arterial disease	18 (3.8%)
Rheumatic diseases	44 (9.2%)
Stroke	7 (1.5%)
PD Medication (current)	
Amantadine	47 (9.8%)
Anticholinergics	8 (1.7%)
Catechol-O-methyltransferase inhibitor	60 (12.6%)
Dopamine agonists	248 (51.9%)
Levodopa	404 (84.5%)
Monoamine oxidase-B inhibitor	191 (40.0%)
Rivastigmine	17 (3.6%)
PDQ-39	
Activities of Daily Living	
Mean (SD)	30.7 (27.1)
Median [IQR]	20.8 [8.3;45.8]
Min, Max	[0.0;100.0]
Bodily Discomfort	
Mean (SD)	39.9 (23.1)
Median [IQR]	41.7 [18.8;58.3]
Min, Max	[0.0;91.7]
Cognitions	
Mean (SD)	31.7 (22.6)
Median [IQR]	31.2 [12.5;43.8]
Min, Max	[0.0;93.8]
Communication	
Mean (SD)	25.7 (23.5)
Median [IQR]	16.7 [8.3;41.7]
Min, Max	[0.0;100.0]
Emotional Well-Being	
Mean (SD)	33.1 (23.5)
Median [IQR]	29.2 [12.5;50.0]
Min, Max	[0.0;100.0]
Mobility	
Mean (SD)	35.8 (29.9)
Median [IQR]	30.0 [10.0;57.5]
Min, Max	[0.0;100.0]
Social Support	
Mean (SD)	16.7 (19.7)
Median [IQR]	8.3 [0.0;25.0]
Min, Max	[0.0;91.7]
Stigma	
Mean (SD)	21.7 (21.4)
Median [IQR]	18.8 [6.2;31.2]
Min, Max	[0.0;87.5]
Summary Index	
Mean (SD)	29.4 (17.9)
Median [IQR]	25.1 [15.2;41.3]
Min, Max	[1.3;89.3]

IQR, interquartile range; NMS, Non-Motor Symptoms; PD, Parkinson's disease; PDQ-39, Parkinson's Disease Questionnaire-39; SD, standard deviation

capture the most variance (inertia) in the dataset [43]. This approach elucidates how variables co-vary by grouping associated variables in proximity, thereby revealing underlying patterns and dimensions. FAMD is exploratory and should not be interpreted as indicating predictive importance or causality.

Results

Psychometric Properties of the PDQ-39

Internal consistency of the PDQ-39 Summary Index and its eight dimensions was assessed using Cronbach's α. Reliability was excellent for the Summary Index ($\alpha = 0.96$) and generally high for individual dimensions, ranging from $\alpha = 0.71$ (Bodily Discomfort) to $\alpha = 0.96$ (Mobility). Detailed results for all dimensions are presented in Table 2.

Table 2 Internal consistency (Cronbach's α) for the PDQ-39 summary index and individual dimensions (0–100 scale, $n = 478$)

Dimensions	Cronbach α
Activities of Daily Living	0.91
Bodily Discomfort	0.71
Cognitions	0.77
Communication	0.83
Emotional Well-Being	0.90
Mobility	0.96
Social Support	0.73
Stigma	0.76
Summary Index	0.96

Cronbach's $\alpha \geq 0.70$ indicates acceptable internal consistency; $\alpha \geq 0.80$ indicates good consistency; $\alpha \geq 0.90$ indicates excellent consistency

Predictive model performance

Performance metrics for the training set are reported as mean (95% CI) [SD] from 1000 bootstrap resamples; validation set values are single point estimates from the independent validation set. Table 3 presents a summary of the performance metrics (MAE, RMSE, and R^2) for both training and validation sets, alongside the selected models for each PDQ-39 dimension and the Summary Index. The Extreme Gradient Boosting with Decision Trees (xgbTree) model was selected as the best-performing model for both the Summary Index and the Activities of Daily Living dimension, while the Gaussian Process with Polynomial Kernel (gaussprPoly) was chosen for all other dimensions.

Training and validation results were broadly similar across outcomes, indicating limited overfitting. The Summary Index model outperformed the individual dimensions models, achieving the lowest MAE (9.60) and RMSE (12.66), and the highest R^2 (0.56) in the validation set. The Cognitions dimension model also demonstrated strong predictive performance, with an R^2 of 0.53, MAE of 12.39, and RMSE of 16.20 in the validation set. This indicates that the model explained over half of the variance in Cognitions scores and produced relatively low average prediction errors, suggesting good predictive performance for this dimension.

The Mobility model explained nearly half the variance (R^2=0.48) but had relatively high MAE/RMSE (18.02/22.88), suggesting substantial average prediction errors. This indicates that while the model captures general trends in Mobility, its predictions for individual patients may lack precision due to larger errors. The Emotional Well-Being model had an R^2 of 0.36, with an MAE of 14.78 and RMSE of 19.50. Although the R^2 is

Table 3 Results

PDQ-39 dimension / Summary Index	Training			Validation			Model selected
	MAE	RMSE	R^2	MAE	RMSE	R^2	
Activities of Daily Living	15.12 (13.52–16.98) [0.89]	19.40 (17.21–21.80) [1.14]	0.47 (0.35–0.57) [0.06]	18.33	24.41	0.33	xgbTree
Bodily Discomfort	15.27 (13.79–16.78) [0.77]	18.90 (17.09–20.63) [0.91]	0.29 (0.18–0.40) [0.05]	17.82	21.63	0.32	gaussprPoly
Cognitions	12.55 (11.29–13.88) [0.67]	15.64 (13.96–17.14) [0.82]	0.52 (0.41–0.61) [0.05]	12.39	16.20	0.53	gaussprPoly
Communication	14.89 (13.34–16.41) [0.80]	18.88 (16.83–20.91) [1.07]	0.36 (0.25–0.47) [0.06]	13.70	18.44	0.37	gaussprPoly
Emotional Well-Being	13.45 (12.08–14.83) [0.74]	17.24 (15.49–18.99) [0.91]	0.46 (0.36–0.56) [0.05]	14.78	19.50	0.36	gaussprPoly
Mobility	17.96 (16.30-19.64) [0.87]	22.14 (19.99–24.31) [1.12]	0.44 (0.33–0.54) [0.05]	18.02	22.88	0.48	gaussprPoly
Social Support	13.50 (12.16–14.90) [0.70]	17.44 (15.49–19.41) [1.04]	0.17 (0.08–0.28) [0.05]	14.87	19.08	0.26	gaussprPoly
Stigma	15.49 (14.00-17.02) [0.80]	19.40 (17.34–21.54) [1.08]	0.20 (0.09–0.30) [0.05]	14.96	19.34	0.23	gaussprPoly
Summary Index	8.99 (8.04–9.91) [0.49]	11.45 (10.16–12.73) [0.64]	0.58 (0.48–0.67) [0.05]	9.60	12.66	0.56	xgbTree

Values for the training set are reported as mean (95% CI) [SD] from 1000 bootstrap resamples. Validation set metrics are single-point estimates computed on the independent validation set. All model predictors were standardized (centered and scaled) using caret's preprocessing step prior to training. MAE, Mean Absolute Error; RMSE, Root Mean Squared Error; R^2, Coefficient of Determination

lower than that of the Mobility model, the lower MAE and RMSE values indicate smaller average prediction errors, reflecting moderate predictive capability with better precision compared to the Mobility model.

Interestingly, the models for Stigma and Social Support, despite having lower R^2 values (0.23 and 0.26, respectively), exhibited lower MAE and RMSE values compared to models like Mobility. The Stigma model had an MAE of 14.96 and RMSE of 19.34, while the Social Support model had an MAE of 14.87 and RMSE of 19.08. These lower error metrics suggest that, on average, the predictions for Stigma and Social Support were closer to the true values, even though the models explained less variance in the outcomes. This could be attributed to lower variability in the Stigma and Social Support scores, where even small prediction errors can result in lower R^2 values due to the limited variance in these dimensions.

The Activities of Daily Living and Bodily Discomfort models showed lower predictive performance. The Activities of Daily Living model had an R^2 of 0.33, with the highest MAE (18.33) and RMSE (24.41), indicating considerable prediction errors. Similarly, the Bodily Discomfort model had an R^2 of 0.32, an MAE of 17.82, and an RMSE of 21.63. These higher error metrics suggest that these dimensions were more challenging to predict, likely reflecting variability from factors not captured by the available features.

Scatter and residuals analysis

In Fig. 4, which presents scatter plots of predicted versus true values, the Summary Index model demonstrates the best fit, with predictions closely clustering around the red dashed line, reflecting its superior performance metrics. Cognitions, Emotional Well-Being, and Communication also show an adequate fit, with reasonable alignment between predicted and true values. However, for Mobility, while the model performs well for lower true values, it tends to underpredict higher scores. Similarly, the models for Bodily Discomfort, Social Support, and Stigma slightly overpredict lower values and underpredict higher ones, indicating a regression toward the mean. The Activities of Daily Living model emerges as having the highest prediction error, with substantial deviations from the red dashed line, signaling lower predictive performance.

In Fig. 5, which presents residual plots for each model, we observe distinct patterns in model performance. The residuals for the Summary Index are evenly distributed around zero with low spread, reflecting a well-fitted model with strong predictive performance. Similarly, the Cognitions and Emotional Well-Being models show residuals that are clustered around zero, indicating a solid fit and minimal error variance. Both Communication and Bodily Discomfort exhibit a wider spread of residuals from zero, although their distribution remains centered, indicating moderate performance. While these models capture some trends in the data, their larger residual spread suggests less precision compared to models like the Summary Index and Cognitions. The Mobility model demonstrates a good fit for lower predicted values, but exhibits an asymmetrical spread at higher values, with a greater number of residuals above zero, suggesting underprediction for larger outcomes. The models for Activities of Daily Living and Stigma exhibit larger residuals across the prediction spectrum. In particular, Activities of Daily Living shows consistently large errors, indicating poor performance, while Stigma and Social Support display heteroscedasticity, with increasing residual variance as predicted values rise, indicating that the models become less reliable when predicting higher values.

Overall, the scatter and residual analyses confirm varying model performance: Summary Index, Cognitions, and Emotional Well-Being demonstrate strong predictive capabilities. Communication and Bodily Discomfort show moderate performance, and Activities of Daily Living exhibits the poorest performance. Four models display specific patterns: Communication, Social Support, and Stigma exhibit heteroscedasticity, while Mobility shows asymmetric residual distribution.

Feature importance

In Fig. 6, the feature importance analysis highlights the universal relevance of the NMSQ score across all models, consistently ranking as the top predictor in every PDQ-39 dimension and the Summary Index. These findings reinforce the central predictive role of NMSQ and align with prior evidence that greater non-motor symptom burden is highly associated with poorer HRQoL [25, 26, 44]. Other clinical variables also showed importance in specific dimensions. Age was particularly relevant in both Mobility and Activities of Daily Living. Notably, for Mobility, age had nearly the same importance as the NMSQ score, indicating its critical role in predicting outcomes in this dimension. Dementia emerged as an important predictor in the Cognitions dimension, aligning with the well-established link between cognitive decline and dementia in patients with PD. Despite these findings, the overall predictive contribution of these clinical variables was less pronounced than that of the NMSQ score. Feature importance values for all selected variables are displayed in Table 4.

Accumulated local effects

The ALE plots in Fig. 7 illustrate the effect of increasing NMSQ scores on the predicted values for each dimension and the Summary Index. ALE plots allow us to visualize the isolated effect of a single feature, in this case, the NMSQ scores, on the models' predictions while

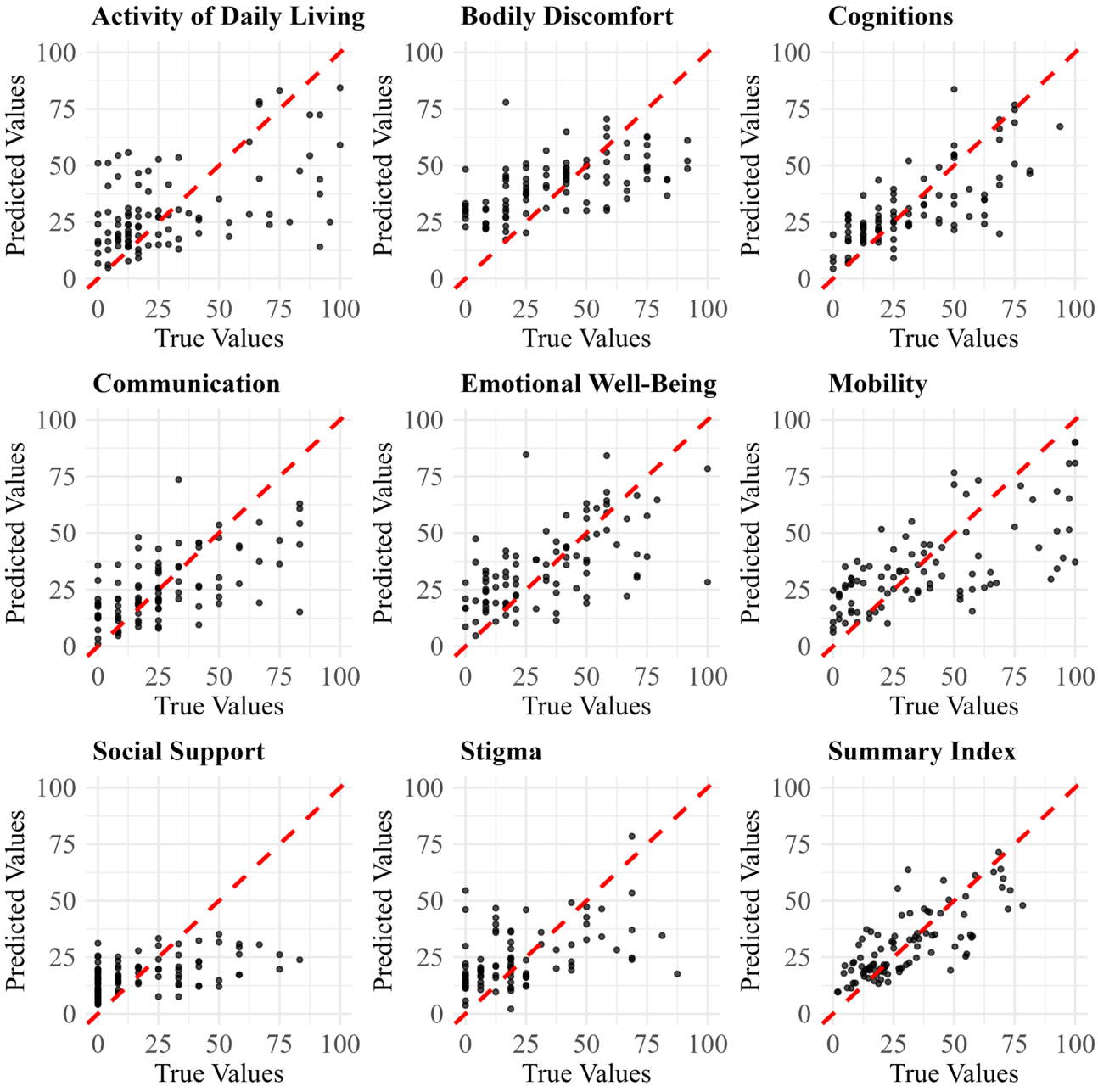

Fig. 4 Predicted vs. true values for PDQ-39 dimensions and summary index. The black dots represent individual predictions, and the red dashed line represents the line of perfect prediction where predicted values would equal true values. Predictions were generated on the independent validation set using the following fitted algorithms: xgbTree for Summary Index and Activities of Daily Living; and gausssprPoly for all other dimensions

accounting for interactions with other features. The x-axis represents the NMSQ score values, and the y-axis represents the accumulated local effect of these scores on the predicted outcomes, centered around zero, where zero indicates no impact on the predicted outcome. An upward trend in ALE values indicates that as NMSQ scores increase, the model predicts worsening HRQoL outcomes, reflecting the detrimental impact of non-motor symptoms.

In general, higher NMSQ scores were associated with worsening HRQoL, as evidenced by increasing ALE values across dimensions. However, the pattern and strength of this relationship varied by dimension. Both the Activities of Daily Living dimension and the Summary Index displayed pronounced stepwise patterns in the ALE plots, which is consistent with the behavior of the tree-based extreme gradient boosting models (xgbTree) selected for these two outcomes. These models create plateaus and jumps in predictions, which appear as steps in the ALE

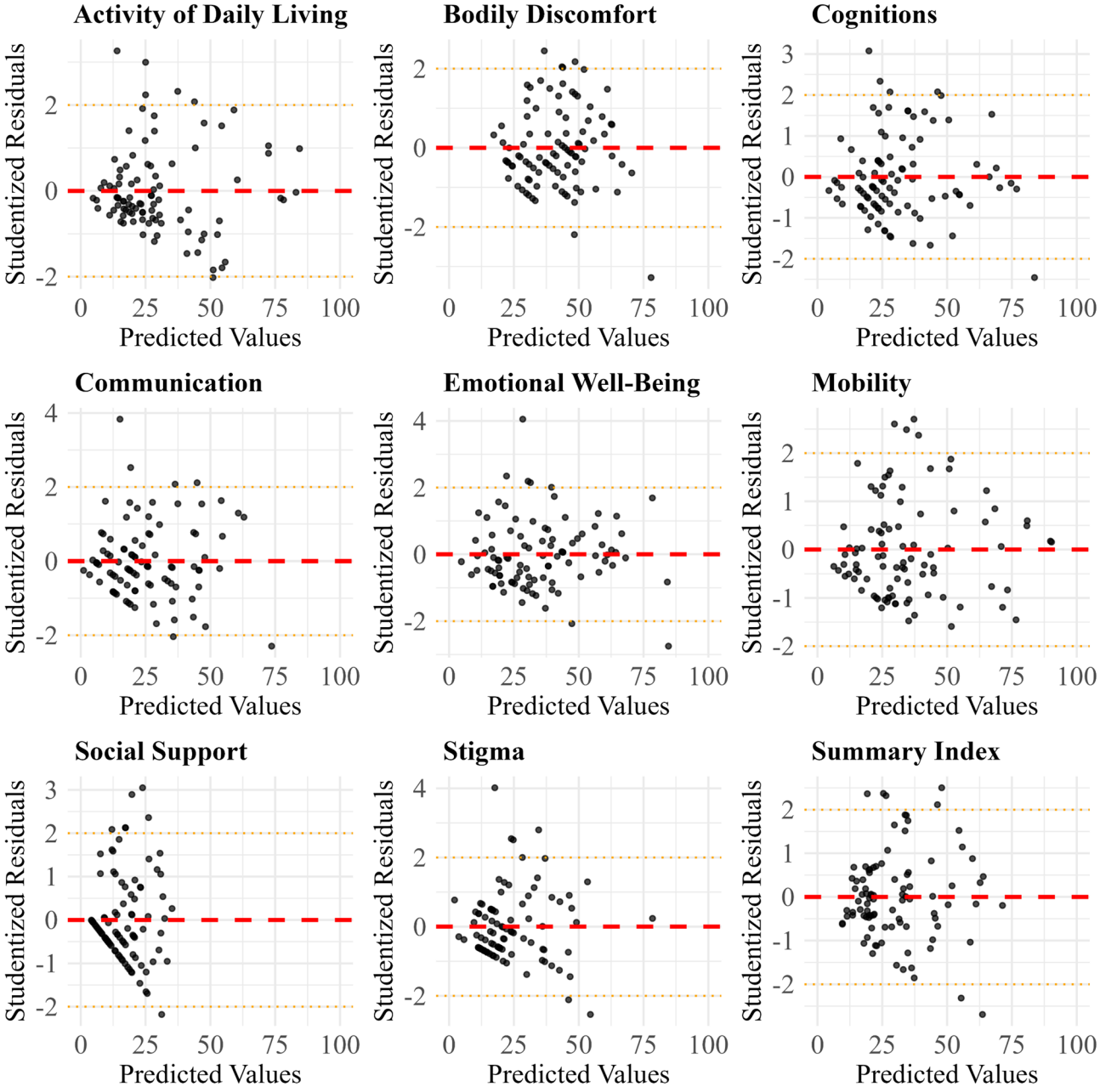

Fig. 5 Studentized residuals for each PDQ-39 dimension and summary index model, plotted against predicted values from the validation set. The red dashed line indicates zero residual error. The orange dotted lines mark ± 2 standard deviations. Residuals were generated on the independent validation set using the following fitted algorithms: xgbTree for Summary Index and Activities of Daily Living; and gausssprPoly for all other dimensions

curves. The discrete NMSQ score scale (0–30) may further accentuate this appearance. In contrast, the Stigma dimension showed a steep increase at higher NMSQ scores, highlighting the pronounced impact of non-motor symptoms on patients' perceived social standing and experiences of stigma. For the remaining dimensions, the relationship between NMSQ scores and predicted outcomes appeared fairly linear, indicating a consistent association between higher NMSQ scores and worse predicted HRQoL in these dimensions.

Factor analysis of mixed data

In Fig. 8, the FAMD plot visualizes associations in a low-dimensional factor space: closer proximity indicates stronger associations (co-variation), whereas greater distances reflect weaker associations. Summary Index, Cognitions, and Emotional Well-Being are the most distant from other variables. Their isolation may reflect broader, non-specific associations with multiple NMS, rather than ties to particular items. Activities of Daily Living, Communication, and Mobility appear relatively

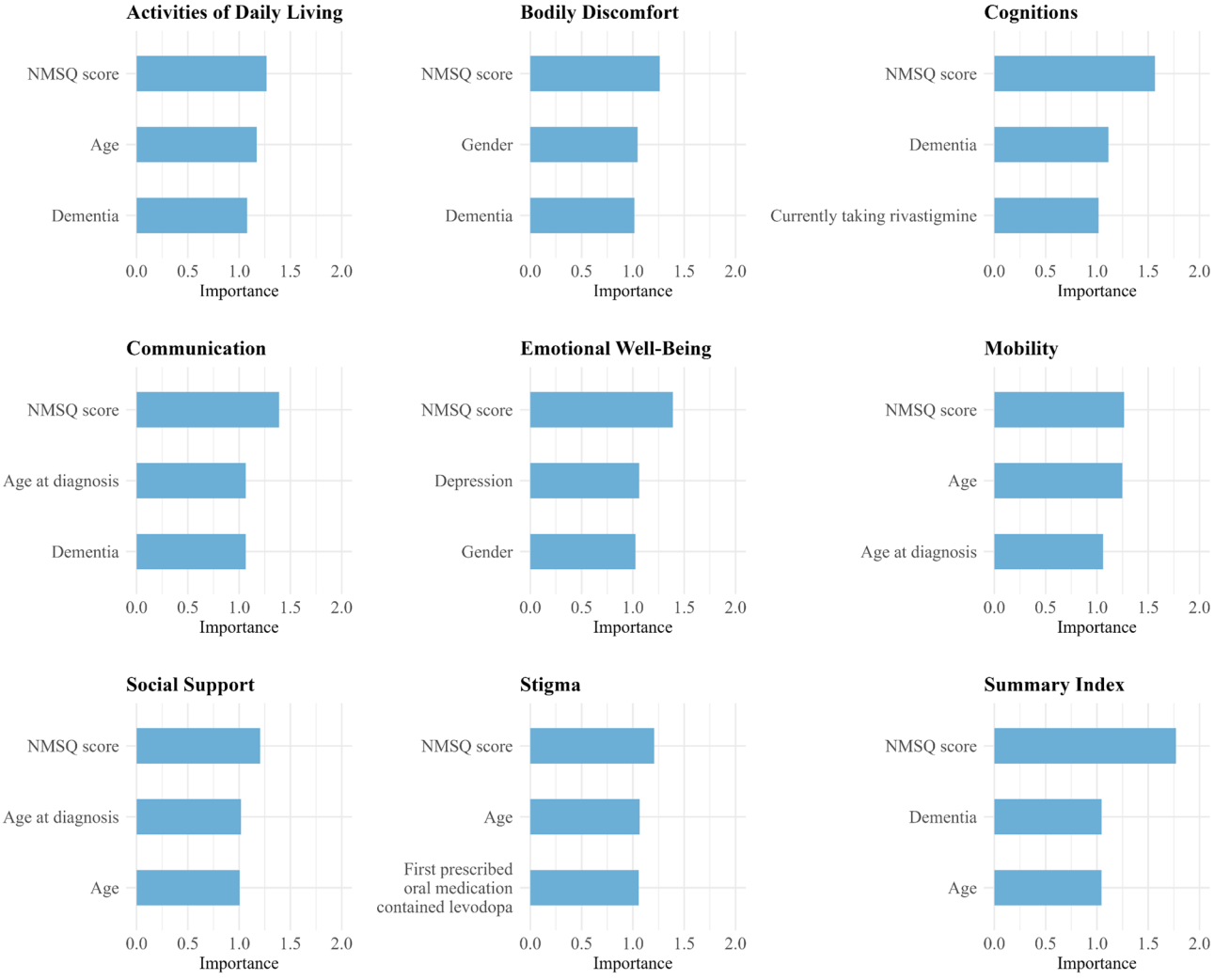

Fig. 6 Feature importance for predictive models of PDQ-39 dimensions and summary index. Importance was assessed as the mean increase in MAE after permuting each feature (100 permutations), using the final fitted models applied to the independent validation set

close to each other. Although they remain distant from individual NMS, their proximity suggests some level of association (co-variation) among these dimensions. This suggests that patients' daily functioning, communication, and mobility may reflect shared underlying factors, rather than specific non-motor symptoms. Social Support, Bodily Discomfort, and Stigma form a cluster near Anxiety, indicating an association. In particular, Social Support is positioned very close to Anxiety, indicating a strong association between perceived social support and anxiety levels in patients with PD.

Discussion

Principal results

Our study demonstrates that while the Summary Index and the Cognitions dimension can be predicted with relatively low prediction error using common clinical variables and NMS, other dimensions present more challenges. Models for these two areas showed strong performance, indicating that overall HRQoL and cognitive aspects can be effectively estimated. In contrast, dimensions such as Mobility, Activities of Daily Living, and Bodily Discomfort were more difficult to predict, suggesting that additional variables may be necessary to improve model fit for these aspects of HRQoL.

The NMSQ score emerged as the most important feature across all ML models, underscoring the pervasive association of NMS with HRQoL in patients with PD. This aligns with previous studies highlighting the substantial burden of NMS on patients' daily lives and emphasizes the need for comprehensive management strategies addressing both motor and non-motor symptoms in PD [25, 26, 44]. Apart from the NMSQ score, certain clinical variables were important predictors for specific dimensions. Age was notably important for Mobility and Activities of Daily Living, reflecting the

Table 4 Features selected by Boruta algorithm and their permutation-based importance in each PDQ-39 dimension and summary index model

Variables	Importance	Permutation error
Activities of Daily Living model		
NMSQ score	1.27	23.21
Age	1.17	21.47
Dementia	1.08	19.74
Age at diagnosis	1.05	19.27
Currently taking Monoamine oxidase-B inhibitor	1.02	18.61
Cancer	1.01	18.45
Number of comorbidities	1.00	18.33
Peripheral vascular arterial disease	1.00	18.33
Gastric ulcer	1.00	18.33
Anxiety	1.00	18.29
Currently taking rivastigmine	1.00	18.26
Gender	0.99	18.18
Country	0.98	18.00
Bodily Discomfort model		
NMSQ score	1.26	22.48
Gender	1.05	18.62
Dementia	1.01	18.06
Number of comorbidities	1.00	17.90
First prescribed oral medication contained levodopa	1.00	17.85
Age at diagnosis	1.00	17.84
Age	1.00	17.83
Cancer	1.00	17.78
Anxiety	1.00	17.76
Cognitions model		
NMSQ score	1.57	19.39
Dementia	1.11	13.78
Currently taking rivastigmine	1.02	12.58
Age	1.02	12.58
Country	1.01	12.50
Depression	1.01	12.48
Heart issues	1.01	12.46
Gastric ulcer	1.01	12.46
Peripheral vascular arterial disease	1.01	12.46
Age at diagnosis	1.00	12.44
Number of comorbidities	1.00	12.44
First prescribed oral medication contained levodopa	1.00	12.40
Communication model		
NMSQ score	1.39	19.04
Age at diagnosis	1.06	14.59
Dementia	1.06	14.57
Age	1.02	13.95
Country	1.01	13.89
Depression	1.00	13.71
Anxiety	1.00	13.69
Currently taking rivastigmine	1.00	13.65
Currently taking Monoamine oxidase B inhibitor	0.99	13.61
First prescribed oral medication contained levodopa	0.99	13.61
Cancer	0.99	13.59
Number of comorbidities	0.99	13.56
Emotional Well-Being model		
NMSQ score	1.39	20.53
Depression	1.06	15.70
Gender	1.03	15.16
Number of comorbidities	1.02	15.03
Anxiety	1.01	14.99
Country	1.01	14.96
Dementia	1.00	14.85
Age	1.00	14.74
Peripheral vascular arterial disease	0.99	14.60
Age at diagnosis	0.98	14.47
Mobility model		
NMSQ score	1.27	22.80
Age	1.25	22.48
Age at diagnosis	1.06	19.09
Dementia	1.05	18.91
First prescribed oral medication contained levodopa	1.02	18.42
Currently taking rivastigmine	1.01	18.29
Depression	1.01	18.27
Peripheral vascular arterial disease	1.01	18.21
Currently taking Levodopa	1.01	18.12
Anxiety	1.00	18.09
Rheumatic diseases	1.00	17.99
Number of comorbidities	1.00	17.97
Social Support model		
NMSQ score	1.21	17.93
Age at diagnosis	1.02	15.12
Age	1.01	14.95
Anxiety	1.00	14.92
Number of comorbidities	1.00	14.89
Gastric ulcer	1.00	14.87
Depression	1.00	14.87
Stigma model		
NMSQ score	1.21	18.07
Age	1.07	15.96
First prescribed oral medication contained levodopa	1.06	15.82
Currently taking rivastigmine	1.04	15.62
Anxiety	1.04	15.55
Kidney disease	1.04	15.49
Age at diagnosis	1.01	15.14
Summary Index model		
NMSQ score	1.77	17.00
Dementia	1.04	10.03

Table 4 (continued)

Variables	Importance	Permutation error
Age	1.04	10.02
Age at diagnosis	1.04	9.94
Cancer	1.01	9.69
Depression	1.00	9.62
Peripheral vascular arterial disease	1.00	9.62
Anxiety	1.00	9.60
Gastric ulcer	1.00	9.60
Currently taking rivastigmine	1.00	9.60
Country	1.00	9.59
Number of comorbidities	1.00	9.59

Features selected through Boruta for each of the models. Feature Permutation Importance: Mean increase in MAE after permuting the feature (100 permutations); permutation error: standard deviation of the MAE increase across permutations

impact of age-related decline on functional abilities. Dementia considerably contributed to the prediction of the Cognitions dimension, aligning with established associations between cognitive decline and dementia in PD [45].

The ALE plots revealed that higher NMSQ scores are associated with worsening HRQoL across all dimensions, with some exhibiting nonlinear relationships. Specifically, the Stigma dimension's predicted scores increased nonlinearly, with steeper rises at higher NMSQ scores. This indicates that as the burden of NMS becomes more severe, feelings of stigma disproportionately intensify, highlighting the escalating impact of NMS on this aspect of HRQoL.

Furthermore, the FAMD visualization offered a nuanced understanding of the associations between NMS, PDQ-39 dimensions, and the Summary Index. Summary Index, Cognitions, and Emotional Well-Being were positioned distantly from other variables, suggesting broader, non-specific associations with multiple NMS rather than ties to particular symptoms. In contrast, Social Support, Bodily Discomfort, and Stigma clustered near Anxiety, indicating an association. This suggests that targeted interventions addressing specific NMS may benefit certain HRQoL dimensions.

Overall, these findings reflect the multifaceted nature of HRQoL in PD. The varying model performance suggests that incorporating additional variables may be necessary to improve prediction for certain dimensions of HRQoL.

Comparison to previous work

Our findings corroborate prior research that emphasizes the considerable impact of NMS on HRQoL in patients with PD [25, 26, 44]. By employing ML models to predict individual PDQ-39 dimensions, our study extends previous work that predominantly focused on overall HRQoL. This approach provides a more detailed understanding of how specific symptoms and clinical variables are associated with distinct aspects of patients' lives. Moreover, we addressed potential overfitting issues identified in our earlier study [27] by refining our modeling techniques and including additional variables, resulting in more robust and generalizable models. The use of interpretable ML methods offers novel insights into the complex relationships between NMS and HRQoL dimensions, contributing to a deeper comprehension of factors affecting patient well-being.

Limitations

While our study provides important insights, a couple of limitations should be acknowledged. First, there was a considerable amount of missing data, particularly towards the end of the database, suggesting that participants experienced survey fatigue and stopped completing the survey. We attempted data imputation to address the missing values, but this resulted in worse model performance. Consequently, we excluded incomplete observations, resulting in a reduced sample size, which may have affected the generalizability and robustness of our models. Second, the models were developed using data from European patients, which may limit their applicability to other populations.

Future directions

Future studies should aim to include more diverse populations from different demographic and cultural backgrounds to enhance the generalizability of the models. Incorporating psychosocial factors such as socioeconomic status, social support networks, and environmental influences may reduce prediction error for HRQoL dimensions like Stigma and Social Support. Additionally, using larger datasets or exploring methods to effectively handle missing data could enhance predictions for challenging dimensions. Longitudinal studies would also be beneficial to model disease progression and capture changes in HRQoL over time.

Conclusions

Our study confirms that NMS are critical predictors of HRQoL in patients with PD. Utilizing ML models, we successfully predicted the Summary Index and the Cognitions dimension with relatively low prediction error, underscoring the profound impact of NMS on both overall and cognitive aspects of HRQoL. However, predicting other dimensions such as Mobility, Activities of Daily Living, and Bodily Discomfort remains challenging, indicating the need for additional variables to enhance predictive performance. The crucial role of NMS for predicting HRQoL emphasizes the necessity for integrated

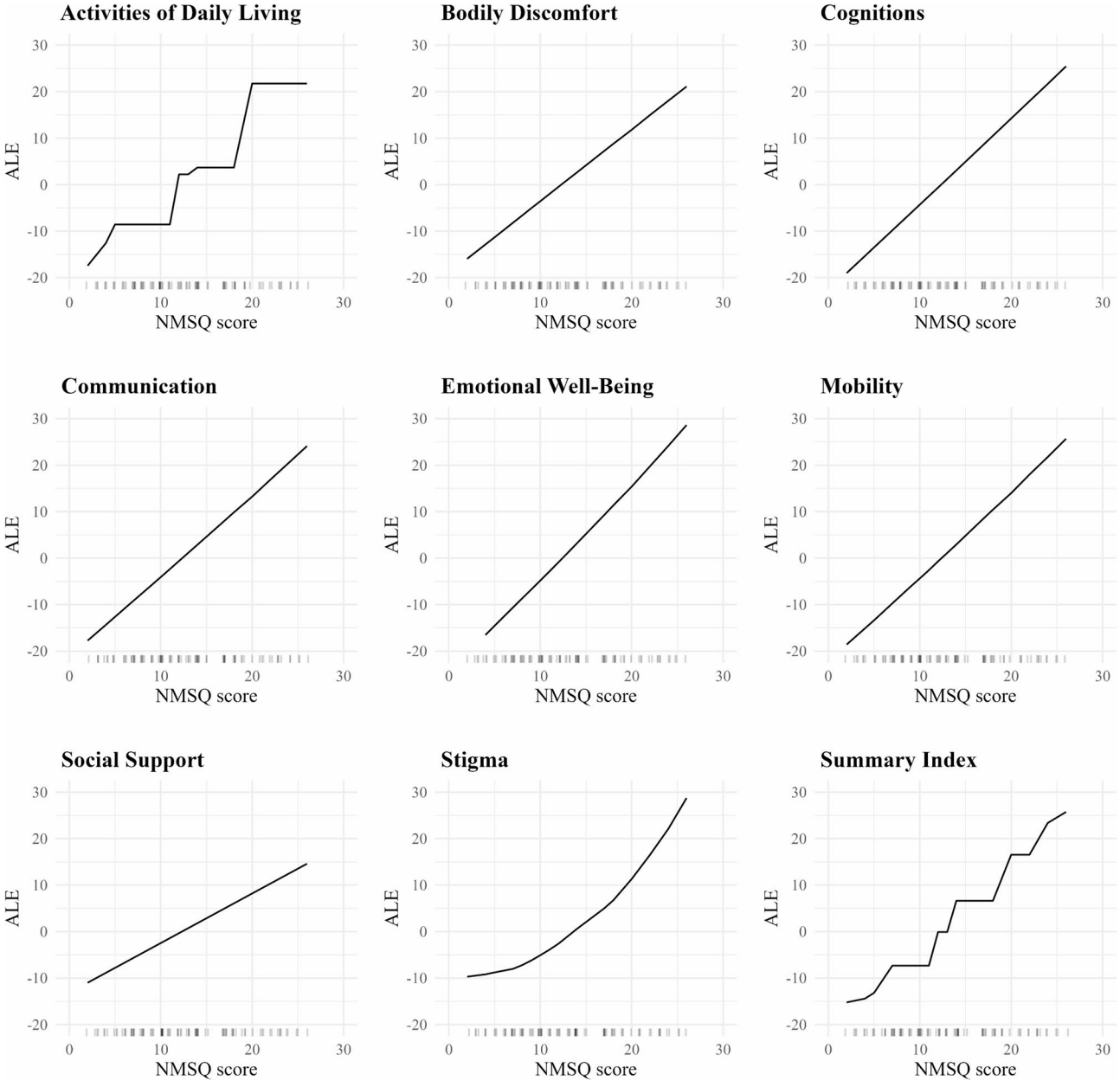

Fig. 7 ALE plots of NMSQ scores on PDQ-39 dimensions and Summary Index, computed from the final fitted models and evaluated on the independent validation set

management approaches that address both motor and non-motor symptoms in PD.

Despite limitations related to missing data and the study's European-centric dataset, our findings offer valuable insights into the complex interplay between NMS and HRQoL. By identifying specific clinical variables associated with certain HRQoL dimensions, our study contributes to the development of more targeted and effective interventions aimed at improving patient outcomes. Future research should focus on incorporating diverse populations, psychosocial factors, and longitudinal data to further refine predictive models and enhance HRQoL outcomes for patients with PD.

In conclusion, this study advances our understanding of how NMS affect various dimensions of HRQoL in PD, demonstrating the potential of ML models in personalizing patient care and informing clinical decision-making. These insights pave the way for more nuanced approaches to managing PD, ultimately aiming to improve the quality of life for those affected by this debilitating disease.

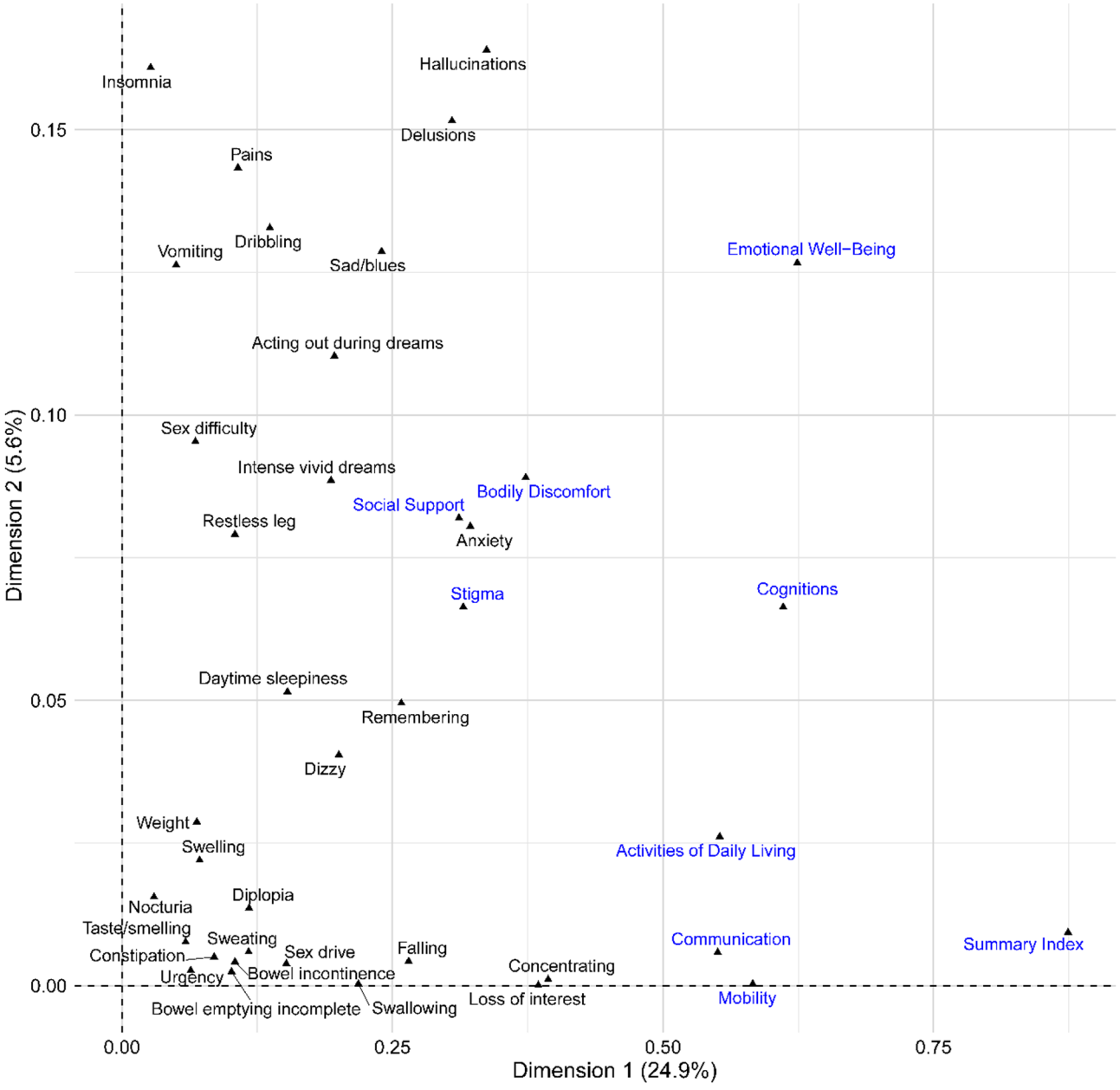

Fig. 8 FAMD of NMSQ Items, PDQ-39 dimensions, and Summary Index. PDQ-39 dimensions and summary index are colored in blue. Axis percentages show the share of total inertia explained (Dim1 24.9%, Dim2 5.6%; 30.5% combined). Points that are closer indicate stronger associations; this is exploratory and not an evidence of causality or predictive importance

Abbreviations

ALE	Accumulated Local Effects
CI	Confidence Interval
FAMD	Factor Analysis of Mixed Data
gaussprPoly	Gaussian Process with Polynomial Kernel
HRQoL	Health—Related Quality of Life
IQR	Interquartile Range
MAE	Mean Absolute Error
MICE	Multivariate Imputation by Chained Equations
ML	Machine Learning
MNAR	Missing Not At Random
NMS	Non—Motor Symptoms
NMSQ	Non—Motor Symptoms Questionnaire
PD	Parkinson's Disease
PDQ	39—Parkinson's Disease Questionnaire—39
PRISM	Parkinson's Real—world Impact Assessment
R^2	Coefficient of Determination
RMSE	Root Mean Squared Error
SD	Standard Deviation
xgbTree	Extreme Gradient Boosting with Decision Trees

Acknowledgements
The PRISM study and database were funded by BIAL—Portela and C[a], S.A., designed in collaboration with the Cure Parkinson's Trust, an advocacy group based in the United Kingdom (UK), and reviewed by the PRISM Steering Committee.

Author contributions
DM conceptualized the study, developed the methodology, conducted the analysis, and drafted the manuscript. TTG and ASB contributed to conceptualization, methodology, supervision, validation, and manuscript review. JM participated in validation and reviewed the manuscript. LA reviewed the manuscript. All authors read and approved the final manuscript.

Funding
This work has been conducted under the scope of and funded by the PhD Program in Health Data Science of the Faculty of Medicine of the University of Porto, Portugal, heads.med.up.pt.

Data availability
The data used in this study were obtained from the Parkinson's Real-world Impact Assessment (PRISM) database (https://next.prism.bial.com/), accessed on 18 February 2022.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Ph.D. Program in Health Data Science, Faculty of Medicine, University of Porto, Porto 4200-319, Portugal
[2]Medical Department, BIAL-Portela & Cª., S.A, São Mamede do Coronado 4745-457, Portugal
[3]RISE-Health, Department of Surgery and Physiology, Faculty of Medicine, University of Porto, Porto 4200-319, Portugal
[4]Department of Clinical Neurosciences and Mental Health, Faculty of Medicine, University of Porto, Porto 4200-319, Portugal
[5]Department of Neurology, Centro Hospitalar Universitário de São João, Porto 4200-319, Portugal
[6]Department of Community Medicine, Information and Decision in Health, Faculty of Medicine, University of Porto, Porto 4200-319, Portugal
[7]Faculty of Health Sciences, University Fernando Pessoa, Porto 4200-150, Portugal
[8]SIGIL Scientific Enterprises, Dubai 4076, United Arab Emirates

Received: 28 December 2024 / Accepted: 28 October 2025
Published online: 06 December 2025

References

1. Feigin VL, Krishnamurthi RV, Theadom AM, Abajobir AA, Mishra SR, Ahmed MB, et al. Global, regional, and National burden of neurological disorders during 1990–2015: a systematic analysis for the global burden of disease study 2015. Lancet Neurol. 2017;16:877–97. https://doi.org/10.1016/S1474-4422(17)30299-5.
2. World Health Organization. Parkinson disease: A public health approach. World Health Organ. 2022;291:390.
3. Jankovic J. Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry. 2008;79:368–76. https://doi.org/10.1136/JNNP.2007.131045.
4. Massano J, Bhatia KP. Clinical approach to parkinson's disease: features, diagnosis, and principles of management. Cold Spring Harb Perspect Med. 2012. https://doi.org/10.1101/CSHPERSPECT.A008870. 2.
5. Marinus J. Health related quality of life in parkinson's disease: a systematic review of disease specific instruments. J Neurol Neurosurg Psychiatry. 2002;72:241–8. https://doi.org/10.1136/jnnp.72.2.241.
6. Galeoto G, Colalelli F, Massai P, Berardi A, Tofani M, Pierantozzi M, et al. Quality of life in parkinson's disease: Italian validation of the parkinson's disease questionnaire (PDQ-39-IT). Neurol Sci. 2018;39:1903–9. https://doi.org/10.1007/S10072-018-3524-X/METRICS.
7. Schönenberg A, Prell T. Measuring quality of life with the parkinson's disease Questionnaire-39 in people with cognitive impairment. PLoS ONE. 2022;17:e0266140. https://doi.org/10.1371/JOURNAL.PONE.0266140.
8. Martinez-Martin P. The importance of non-motor disturbances to quality of life in parkinson's disease. J Neurol Sci. 2011;310:12–6. https://doi.org/10.1016/j.jns.2011.05.006.
9. Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The PDQ-8: development and validation of a short-form parkinson's disease questionnaire. Psychol Health. 1997;12:805–14. https://doi.org/10.1080/08870449708406741.
10. Focus Area. Patient-Reported Outcomes and other Clinical Outcome Assessments | FDA. https://www.fda.gov/science-research/focus-areas-regulatory-science-report/focus-area-patient-reported-outcomes-and-other-clinical-outcome-assessments. Accessed 1 Jun 2025.
11. EMA and European Organisation for Research and Treatment of Cancer (EORTC) workshop. How can patient-reported outcomes (PRO) and health-related quality of life (HRQoL) data inform regulatory decisions? | European Medicines Agency (EMA). https://www.ema.europa.eu/en/events/ema-european-organisation-research-treatment-cancer-eortc-workshop-how-can-patient-reported-outcomes-pro-health-related-quality-life-hrqol-data-inform-regulatory-decisions. Accessed 1 Jun 2025.
12. Rocha F, Tolosa E, Ferreira J, Rascol O, Poewe W, Santos A, et al. Effect of Opicapone in parkinson's disease patients with 'early' motor fluctuations: parkinson's disease questionnaire (PDQ-39) analysis from the BIPARK-I double-blind experience. J Neurol Sci. 2019;405:190–1. https://doi.org/10.1016/j.jns.2019.10.1150.
13. Hattori N, Takeda A, Hanya Y, Kitagawa T, Arai M, Furusawa Y, et al. Effects of Rasagiline on parkinson's disease questionnaire (PDQ-39) emotional well-being domain in patients with parkinson's disease: A post-hoc analysis of clinical trials in Japan. PLoS ONE. 2022;17:e0262796. https://doi.org/10.1371/journal.pone.0262796.
14. Chaudhuri KR, Martinez-Martin P, Schapira AHV, Stocchi F, Sethi K, Odin P, et al. International multicenter pilot study of the first comprehensive Self-Completed nonmotor symptoms questionnaire for parkinson's disease: the NMSQuest study. Mov Disord. 2006;21:916–23. https://doi.org/10.1002/mds.20844.
15. Chandrabhatla AS, Pomeraniec IJ, Ksendzovsky A. Co-evolution of machine learning and digital technologies to improve monitoring of parkinson's disease motor symptoms. Npj Digit Med 2022. 2022;5:1. https://doi.org/10.1038/s41746-022-00568-y.
16. Gao C, Sun H, Wang T, Tang M, Bohnen NI, Müller MLTM, et al. Model-based and Model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. Sci Rep. 2018;8. https://doi.org/10.1038/S41598-018-24783-4.
17. Nilashi M, Abumalloh RA, Minaei-Bidgoli B, Samad S, Yousoof Ismail M, Alhargan A, et al. Predicting Parkinson's Disease progression: evaluation of ensemble methods in machine learning. J Healthc Eng. 2022;2022. https://doi.org/10.1155/2022/2793361
18. Dadu A, Satone V, Kaur R, Hashemi SH, Leonard H, Iwaki H, et al. Identification and prediction of parkinson's disease subtypes and progression using machine learning in two cohorts. NPJ Parkinsons Dis. 2022;8:172. https://doi.org/10.1038/s41531-022-00439-z.
19. Cho K, Yoon J, Shin N-Y, Han S. A machine learning approach to identify Whole-Brain functional connectivity signatures of cognitive subtypes in parkinson's disease. 2025. https://doi.org/10.2139/SSRN.5477430
20. Lu J, Tran DK, Bui XL, Le NQK, Chua MCH. Multimodal machine learning for parkinson's disease diagnosis and genetic subtyping. Health Technol 2025. 2025;1–7. https://doi.org/10.1007/S12553-025-01021-2.
21. Hosseini MS, Aghamiri SMR, Panahi M. Cross-regional radiomics: a novel framework for relationship-based feature extraction with validation in parkinson's disease motor subtyping. BioData Min. 2025;18:67. https://doi.org/10.1186/S13040-025-00483-4.
22. Hey GE, Eisinger R, Guarin D, Safarpour D, Rodríguez-Violante M, Ramirez-Zamora A. Developing personalized treatment strategies for Parkinson's disease based on disease subtypes. Expert Rev Neurother. 2025;25. https://doi.org/10.1080/14737175.2025.2552784.
23. Zhang W, Xu Z, Yu R, Jiang M, Dai Q. DualGCN-GE: integration of Spatiotemporal representations from whole-blood expression data with dual-view graph Convolution network to identify parkinson's disease subtypes. BMC Bioinformatics. 2025;26:1–18. https://doi.org/10.1186/S12859-025-06181-6/FIGURES/5.

24. Xu X, Zhang S, Xu C, Zhang W, Zhao H, Liu Y, et al. Identifying subtypes of longitudinal motor symptom severity trajectories in early parkinson's disease patients. Front Neurol. 2025;16:1597132. https://doi.org/10.3389/FNEUR.2025.1597132/BIBTEX.
25. Song W, Guo X, Chen K, Chen X, Cao B, Wei Q, et al. The impact of non-motor symptoms on the Health-Related quality of life of parkinson's disease patients from Southwest China. Parkinsonism Relat Disord. 2014;20:149–52. https://doi.org/10.1016/J.PARKRELDIS.2013.10.005.
26. Martinez-Martin P, Rodriguez-Blazquez C, Kurtis MM, Chaudhuri KR. The impact of non-motor symptoms on health-related quality of life of patients with parkinson's disease. Mov Disord. 2011;26:399–406. https://doi.org/10.1002/MDS.23462.
27. Magano D, Taveira-Gomes T, Massano J, Barros AS. Predicting quality of life in parkinson's disease: A machine learning approach employing common clinical variables. J Clin Med. 2024;13:5081. https://doi.org/10.3390/jcm13175081.
28. Candel-Parra E, Córcoles-Jiménez MP, Delicado-Useros V, Ruiz-Grao MC, Hernández-Martínez A, Molina-Alarcón M. Predictive model of quality of life in patients with parkinson's disease. Int J Environ Res Public Health. 2022;19:672. https://doi.org/10.3390/ijerph19020672.
29. Tolosa E, Ebersbach G, Ferreira JJ, Rascol O, Antonini A, Foltynie T, et al. The parkinson's Real-World impact assessment (PRISM) study: A European survey of the burden of parkinson's disease in patients and their carers. J Parkinsons Dis. 2021;11:1309–23. https://doi.org/10.3233/JPD-212611.
30. R: The R Project for Statistical Computing. https://www.r-project.org/. Accessed 12 Jul 2023.
31. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Softw. 2011;45:1–67. https://doi.org/10.18637/JSS.V045.I03.
32. Verikas A, Gelzinis A, Bacauskiene M. Mining data with random forests: A survey and results of new tests. Pattern Recognit. 2011;44:330–49. https://doi.org/10.1016/J.PATCOG.2010.08.011.
33. Chen RC, Dewi C, Huang SW, Caraka RE. Selecting critical features for data classification based on machine learning methods. J Big Data. 2020;7:1–26. https://doi.org/10.1186/S40537-020-00327-4/FIGURES/13.
34. Iranzad R, Liu X. A review of random forest-based feature selection methods for data science education and applications. Int J Data Sci Anal. 2024;1–15. https://doi.org/10.1007/S41060-024-00509-W/METRICS.
35. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. 2010;36:1–13. https://doi.org/10.18637/JSS.V036.I11.
36. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28. https://doi.org/10.18637/jss.v028.i05.
37. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. An Introduction to the Bootstrap. 1994. https://doi.org/10.1201/9780429246593/INTRODUCTION-BOOTSTRAP-BRADLEY-EFRON-TIBSHIRANI/ACCESSIBILITY-INFORMATION
38. Nisbet R, Miner G, Yale K. Handbook of statistical analysis and data mining applications. Handbook of Statistical Analysis and Data Mining. Applications. 2017;1–792. https://doi.org/10.1016/C2012-0-06451-4.
39. Molnar C, Casalicchio G, Bischl B. Iml: an R package for interpretable machine learning. J Open Source Softw. 2018;3:786. https://doi.org/10.21105/JOSS.00786.
40. Christoph Molnar. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.). 2nd edition. 2022.
41. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. 2021. https://doi.org/10.1007/978-1-0716-1418-1
42. Apley DW, Zhu J. Visualizing the effects of predictor variables in black box supervised learning models. J R Stat Soc Ser B Stat Methodol. 2020;82:1059–86. https://doi.org/10.1111/rssb.12377.
43. Lê S, Josse J, Husson F, FactoMineR. An R package for multivariate analysis. J Stat Softw. 2008;25:1–18. https://doi.org/10.18637/JSS.V025.I01.
44. Barone P, Antonini A, Colosimo C, Marconi R, Morgante L, Avarello TP, et al. The PRIAMO study: A multicenter assessment of nonmotor symptoms and their impact on quality of life in parkinson's disease. Mov Disord. 2009;24:1641–9. https://doi.org/10.1002/MDS.22643.
45. Aarsland D, Andersen K, Larsen JP, Lolk A, Nielsen H, Kragh-Sørensen P. Risk of dementia in parkinson's disease: A community-based, prospective study. Neurology. 2001;56:730–6. https://doi.org/10.1212/WNL.56.6.730.

Publisher's note