Group 17 Project Proposal

Leaksmy Heng

Khoury College of Computer Science, Northeastern University

March 3, 2023

Abstract

Parkinson's disease (PD) is a condition that impacts both physical and non-physical functions of the body, and is caused by a disruption in the nervous system. The effects ranging from slow movement, muscle rigidity, walking difficulties to cognitive impairment, mental health disorders, sleep problems, and sensory disturbances. These motor and non-motor symptoms can limit speech and mobility and affect various aspects of a person's life. As the disease progresses, disability rates increase, and care becomes necessary. Furthermore, many individuals with PD may also develop dementia as a result of the disease. The rate of disability and mortality related to Parkinson's disease (PD) is growing at a faster pace than any other neurological disorder worldwide. Over the past 25 years, the incidence of PD has doubled, with global estimates in 2019 indicating that more than 8.5 million people are affected. In the year 2019, Parkinson's disease (PD) led to 5.8 million disability-adjusted life years, which represents a significant rise of 81% compared to the year 2000. It also was responsible for causing 329,000 deaths, representing a rise of more than 100% since 2000 [1]. Despite the seriousness of Parkinson's disease, there is currently no known cure. Nevertheless, medical practitioners utilize the Unified Parkinson's Disease Rating Scale (UPDRS) to keep track of the advancement of the condition among their patients. The maximum total UPDRS score is 199, which indicates the most severe level of disability from PD [2].

Basic Idea

This paper aims to highlight the significance of predicting the Unified Parkinson's Disease Rating Scale (UPDRS) in the assessment of the severity and progression of Parkinson's disease among patients. By accurately tracking UPDRS scores, healthcare professionals can better inform treatment decisions and adjust care plans to optimize patient outcomes. Predictive models for UPDRS scores can provide valuable insights for clinicians and researchers, aiding in the monitoring of treatment effectiveness and facilitating the improvement of new therapies for Parkinson's disease. Predictive models may also help in clinical trials, allowing researchers to monitor the effectiveness of new treatments and therapies for Parkinson's disease.

Approach to Solution

We are going to use the dataset published in Kaggle for AMP Parkinson's Disease Progression Prediction [3]. To ensure accurate and effective machine learning models, it is crucial to start with a high-quality and well-labeled dataset. This involves performing data cleaning to remove any errors, inconsistencies, or missing values that may affect the accuracy of the model. Once the data is cleaned and labeled, exploratory data analysis (EDA) can be conducted to gain insights into the data, identify any patterns or trends, and develop an understanding of the relationships between the input variables and the target variable. To further enhance the quality of the analysis and ensure clear communication of findings, a comprehensive report will be generated after the EDA. This report will provide an insightful overview of the data, the insights gained during the analysis, and any recommendations for the subsequent steps in the machine learning process.

We are currently considering a range of machine learning methods for our project, including Support Vector Regression (SVM), and Gradient Boosting. These are all powerful machine learning algorithms that are widely used for supervised learning tasks. Their ability to solve both regression and classification problems makes them highly versatile and applicable across a range of industries and domains.

By training on labeled data, these algorithms can effectively learn patterns and relationships within the data, enabling them to make accurate predictions on new, unseen data. This makes them valuable tools for a variety of applications, from predicting customer behavior to identifying disease risk factors.

Related Work

Paper called "An improved approach for prediction of Parkinson's disease using machine learning techniques" published in IEEE has conducted an extensive study and provide an effective algorithm for predicting the Parkinson's disease. In the paper, an extension of previous work that utilized non-motor

Group 17 Project Proposal

Leaksmy Heng

Khoury College of Computer Science, Northeastern University

March 3, 2023

features such as RBD and olfactory loss, along with important biomarkers, is presented. The authors model a classifier using several machine learning models that have not been used before. Specifically, Multilayer Perceptron, BayesNet, Random Forest, and Boosted Logistic Regression models were developed and evaluated. The results showed that Boosted Logistic Regression provided the best performance, with an accuracy of 97.159% and an area under the ROC curve of 98.9%. Based on these findings, the authors concluded that the developed models have the potential for early prediction of Parkinson's disease [4].

Another research paper was also published in IEEE titled Predicting the Progression of Parkinson's Disease MDS-UPDRS-III Motor Severity Score from Gait Data using Deep Learning. The paper proposes a deep learning-based approach for predicting the progression of Parkinson's disease (PD) using gait data. The authors collected gait data from PD patients over a period of one year and used the Movement Disorder Society-Unified Parkinson's Disease Rating Scale Part III (MDS-UPDRS-III) motor severity score as the gold standard for disease progression. They preprocessed the gait data and used a convolutional neural network (CNN) to predict the MDS-UPDRS-III scores. The results indicated that the proposed method outperformed traditional machine learning approaches in predicting the MDS-UPDRS-III scores with a mean absolute error of 1.45. The authors concluded that their approach could potentially be used as a tool for monitoring PD progression and evaluating treatment effectiveness [5].

Assessment Methodology

The performance of our method will be assessed using several metrics such as mean squared error (MSE), root mean squared error (RMSE), and R-square. The MSE determines the average squared difference between the predicted and actual values, while the RMSE provides an error measure in the same units as the target variable. On the other hand, the R-square evaluates the fraction of the target variable's variance that can be accounted for by the model.

Cross-validation will be employed to ensure the model performs well on testing data and to avoid overfitting. Additionally, the project will involve feature selection as mention above and an ablation analysis to determine the impact of each variable on the model's performance. In this analysis, variables will be systematically removed from the model, and the resulting performance will be evaluated to assess their importance in predicting Parkinson's disease progression. Through this assessment methodology, the optimal model approach and feature selection will be determined, leading to the best possible results.

Reference

[1] "Parkinson disease," World Health Organization, 13-Jun-2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/parkinson-disease#:~:text=Current%20estimates%20suggest%20that%2C%20in,of%20over%20100%25%20since%202000. [Accessed: 03-Mar-2023].

[2] S. K. Holden, T. Finseth, S. H. Sillau, and B. D. Berman, "Progression of MDS-UPDRS scores over five years in de novo parkinson ...," Progression of MDS-UPDRS Scores Over Five Years in De Novo Parkinson Disease from the Parkinson's Progression Markers Initiative Cohort, 22-Sep-2017. [Online]. Available: https://movementdisorders.onlinelibrary.wiley.com/doi/10.1002/mdc3.12553. [Accessed: 03-Mar-2023].

[3] "AMP®-Parkinson's Disease Progression Prediction," *Kaggle*. [Online]. Available: https://www.kaggle.com/competitions/amp-parkinsons-disease-progression-prediction/data?select=supplemental_clinical_data.csv. [Accessed: 03-Mar-2023].

[4] K. N. Challa, V. S. Pagolu, G. Panda, and B. Majhi, "An improved approach for prediction of parkinson's disease using Machine Learning Techniques," 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Jun. 2017.

[5] R. Zia Ur Rehman, L. Rochester, A. J. Yarnall, and S. Del Din, "Predicting the progression of parkinson's disease MDS-UPDRS-III Motor

Group 17 Project Proposal

Leaksmy Heng

Khoury College of Computer Science, Northeastern University

March 3, 2023

Severity Score from gait data using Deep Learning," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Dec. 2021.