Parkinson's Disease Progression Prediction using Machine Learning Algorithms

Moiz Hassan
moiz1.hassan@torontomu.ca
Department of Electrical and Computer Engineering
Toronto Metropolitan University

November 2023

Abstract

Parkinson's disease (PD) progression prediction is a crucial aspect of early diagnosis and management. This study employs machine learning algorithms to enhance predictive models, focusing on the Unified Parkinson's Disease Rating Score (UPDRS) using the Parkinson's Telemonitoring Dataset. The research explores various methodologies, including ensemble learning and MLP, drawing insights from related works. Feature engineering strategies, such as multicollinearity removal, contribute to model stability. Mean Absolute Error (MAE) is adopted as the performance metric for consistency with prior studies. Experimental findings spotlight the Decision Tree Regressor, boasting an MAE of 0.7131, as a notable performer.

1 Literatire Review

1.1 Background

Parkinson's disease is a gradually advancing neurological disorder that impacts the nervous system and various bodily functions regulated by nerves. While it manifests more frequently in males than females[1], the key symptoms encompass tremors or shaking, typically commencing in the hands or fingers, muscle stiffness, alterations in speech, and a decline in automatic movements[2]. This condition significantly disrupts daily activities, diminishing the overall quality of life for both patients and their families. Notably, it stands as the second most common neurodegenerative disorder affecting individuals aged 60 and above, following Alzheimer's disease[3].

Diagnosing Parkinson's disease in its advanced stages poses fewer challenges, but the early stages present a formidable task due to the nonspecific and variable nature of symptoms across individuals[4]. This complexity has spurred the

application of machine learning algorithms by researchers to predict the severity of Parkinson's Disease (PD) and monitor symptom progression.

A prominent approach involves leveraging voice datasets collected from patients, with the Parkinson's Telemonitoring Dataset being a well-established and widely used resource[5]. Comprising a diverse array of biomedical voice measurements from 42 individuals in the early stages of PD, the dataset spans six months and offers a non-invasive technique. Recorded by patients in their homes, the dataset contains 5,875 voice recordings, with each row representing an individual. The primary objective of utilizing this data is to predict both the motor and total Unified Parkinson's Disease Rating Score (UPDRS) by extracting features from the recordings.

Typically, observations from physical tests are translated into a metric expressly crafted to track disease progression, providing insight into the Unified Parkinson's Disease Rating Score (UPDRS) of a patient[5]. The UPDRS serves as a comprehensive gauge, reflecting both the existence and intensity of symptoms. It spans a scale from 0 to 176m, where a score of 0 denotes a healthy state and 176 indicates total disability.

1.2 Related Work

Max et al. [5] developed the Oxford Parkinson's Disease Telemonitoring Dataset, selecting 42 subjects diagnosed within the previous five years, unmedicated for six months. All subjects remained unmedicated for the six-month duration of the study. UPDRS was assessed at baseline, and after three and six months as shown in table1.

Dysphonia measures were derived from speech signals using various techniques, resulting in a 5923×16 feature matrix. Linear and nonlinear regression methods were applied to map dysphonia measures to interpolated UPDRS values. Linear techniques included classical least squares (LS), iteratively reweighted LS (IRLS), and least absolute shrinkage and selection operator (LASSO). The study also explored nonlinear regression using classification and regression trees (CART). The dataset was split into training and testing subsets, repeated 1000 times, with mean absolute error (MAE) recorded in each repetition.

Linear methods had training and testing errors around 8.5, while CART performed better with a training MAE of 6.0 and testing MAE of 7.5. By tuning LASSO's regularization parameter, the study identified a subset of six dysphonia measures, achieving testing errors around 8.47 for total UPDRS using IRLS and 5.95 using CART.

	Mean	**SD**	**Min**	**Max**	**Median**
Baseline	26.39	10.80	8	54	25.5
After 3-Month	29.36	11.82	7	55	28
After 6-Month	29.57	11.92	7	54	26

Table 1: UPDRS Assessment of subjects in [5]

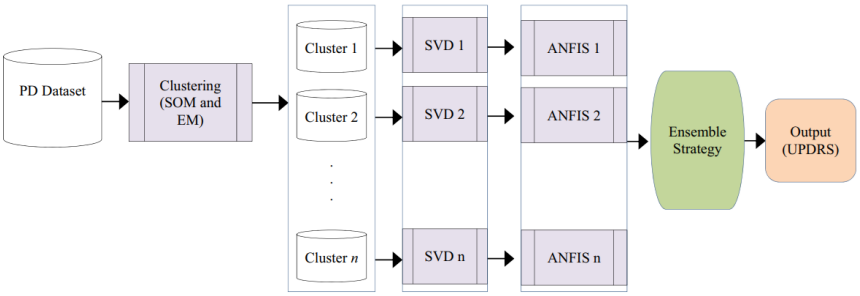

Figure 1: Proposed Method of Nilashi et al. in [6]

In another study by Nilashi et al.[6], a hybrid method for PD diagnosis was developed, combining ensemble learning and dimensionality reduction techniques. Singular Value Decomposition (SVD) and an ensemble of Adaptive Neuro-Fuzzy Inference System (ANFIS) were used to improve time complexity and accuracy in predicting Total- and Motor-UPDRS. In their approach, they employed Self-Organizing Map (SOM) and Expectation Maximization (EM) for clustering data, followed by SVD for dimensionality reduction and ANFIS ensemble for UPDRS prediction. The integration by average approach was used to obtain final results from ensemble learning, Figure 1 shows the steps followed for UPDRS prediction. The authors utilized the Oxford Parkinson's Disease Telemonitoring Dataset[5] for evaluation.

In a series of experiments, the authors compared their method with others, including ANFIS, Neural Network (NN), Support Vector Regression (SVR), and Multiple Linear Regression (MLR). For SVR, the RBF kernel was selected, and a feedforward back-propagation with three layers was applied for the neural network. The results, presented in Table 2, indicated that methods employing clustering (EM and SOM), SVD, and ANFIS ensemble (SOM-SVD-ANFIS ensemble and EM-SVD-ANFIS ensemble) exhibited better predictive accuracy for Total UPDRS in the PD dataset.

Method	**MAE**
MLR	0.987
NN	0.951
ANFIS	0.743
SVR	0.689
SOM-SVD-ANFIS ensemble	0.489
EM-SVD-ANFIS ensemble	0.480

Table 2: The results of MAE of the methods for UPDRS prediction in [6]

In [7], the authors implemented various configurations of Multi-Layer Perceptron (MLP) models to distinguish between severe and non-severe cases and

assess the progression of disease in individual patients. The proposed architecture incorporates four dense intermediate layers, with 100, 200, 300, and 100 neurons respectively, along with two Dropout layers to prevent overfitting.

Before conducting classification and/or regression tasks, the authors employed an unsupervised autoencoder for feature reduction, eliminating non-essential information. The autoencoder architecture consists of an encoder, featuring a dense layer of 200 neurons between the input layer and the latent space, and a symmetric decoder.

A distinctive aspect of their methodology is the joint integration of the MLP classifier or regressor with the autoencoder, as opposed to the conventional approach of reducing features first and then classifying the data. Experiments were conducted using the latent space for classification or regression, and the architecture is noteworthy for the encoder simultaneously learning to reduce data dimensionality and accurately classify or predict the data.

The reported results indicate the effectiveness of this integrated approach. Applying the MLP network with the autoencoder yielded a Mean Absolute Error (MAE) of 0.2857, a notable improvement compared to the 1.2739 MAE obtained when using the MLP without the autoencoder. This underscores the potential of combining MLP with autoencoder-based feature reduction for enhanced accuracy and efficiency in disease classification and progression prediction, as observed in the referenced study.

2 Feature Engineering

We omitted the subject number feature as it was deemed irrelevant to the predictive task.

To enhance the stability of our model, we employed the *remove multicollinearity* parameter within the pycaret setup function. Multicollinearity, also known as collinearity, is a phenomenon where one feature variable in a dataset exhibits high linear correlation with another feature variable in the same dataset. This correlation can lead to increased coefficient variance, rendering them unstable and noisy, particularly in linear models. The Pycaret documentation highlights the utility of addressing multicollinearity by eliminating one of the highly correlated features through the *remove multicollinearity* parameter[8].

Post multicollinearity removal, the feature count dwindled from 19 to a concise 5 features: age, sex, test time, HNR, and PPE. Restricting our model to these essential features resulted in more robust models with improved Mean Absolute Error (MAE) values.

The deployed model is configured to predict the Total UPDRS based on the values of these 5 features, ensuring a streamlined and effective approach to the prediction task.

3 Performance Metric

The metric chosen for evaluating model performance is the Mean Absolute Error (MAE). MAE holds prominence due to its unique feature—its units align with those of the target value being predicted. This characteristic simplifies the interpretability of the error score.

In contrast to Root Mean Squared Error (RMSE) and Mean Squared Error (MSE), MAE exhibits linear changes, providing an intuitive understanding of error progression. Unlike MSE and RMSE, which accentuate larger errors by squaring them, thereby inflating the mean error score, MAE treats all errors equally. It assigns a consistent weight to errors of varying magnitudes, ensuring a linear increase in scores with escalating errors.

A pivotal reason for adopting MAE as the performance metric is its alignment with previous studies, specifically works cited as [5], [6], and [7]. The consistent use of MAE in these works serves as a benchmark, fostering comparability between our models and those established in prior research. This strategic choice not only upholds methodological consistency but also facilitates a more comprehensive assessment of our models within the broader context of existing literature.

4 Experimental Results

Various models were trained and fine-tuned on the dataset, yielding different performance metrics. The Decision Tree Regressor achieved an MAE score of 0.7131. In comparison, an ensemble model with boosting outperformed it, attaining an MAE score of 0.6793. A blend model, combining the top three regression models based on the MAE metric, achieved a score of 0.7931, while a stack model, similarly constructed, achieved a score of 0.8762. These results are summarized in Table 3.

Method	**MAE**
Decision Tree	0.7131
DT ensemble model with boosting	0.6793
Blend model	0.7931
Stack Model	0.8762
MLP	5.7754

Table 3: The results of different models on the Parkinson's Disease Telemonitoing Dataset

Although the ensemble model with boosting surpassed the Decision Tree model in MAE, it's worth noting that the latter is computationally lighter. The difference in MAE between the two models is deemed insignificant.

An MLP (Multi-Layer Perceptron Model) was crafted with six dense layers, each containing 200, 400, 600, 400, 200, and 100 neurons, respectively, deter-

mined through experimentation. However, the obtained results were unsatisfactory, leading us to believe that the MLP's performance could be enhanced with more feature engineering. Specifically, implementing an autoencoder for feature extraction, as demonstrated in [7], is expected to boost the MLP model's performance.

The model deployed on Clouds is a Decision Tree model, which we consider to be delivering consistently good and stable results.

5 Conclusions

In concluding this project on predicting Parkinson's disease progression, our exploration has illuminated the potential of machine learning in addressing the complexities of early-stage diagnosis and progression monitoring. Leveraging the Parkinson's Telemonitoring Dataset, our study integrated various machine learning models and methodologies to predict the Unified Parkinson's Disease Rating Score (UPDRS), emphasizing the significance of feature engineering and model evaluation metrics.

Our approach utilized the powerful Pycaret library for streamlined model setup, providing a comprehensive suite of tools and functionalities. Additionally, we employed scikit-learn for the implementation of the Multi-Layer Perceptron (MLP) model. Sklearn proves to be a versatile choice, offering a plethora of functions and tools that greatly facilitate the development and fine-tuning of MLPs, aligning well with the intricacies of our predictive modeling.

However, in the dynamic landscape of machine learning, considering alternative libraries becomes crucial for expanding our toolkit. TensorFlow, an open-source library developed by Google, stands out for its prowess in building and training neural networks. Its flexibility and scalability could potentially provide an advantage, particularly when dealing with complex neural network architectures.

Another library worth exploring is XGBoost, known for its efficiency in gradient boosting. Its capabilities in handling diverse data types and conducting feature importance analysis could complement our existing ensemble models, offering additional insights into Parkinson's disease progression.

As we navigate the future of this research, incorporating these libraries alongside the current toolkit may unlock further potential and improve the overall performance of our predictive models. Continuous exploration and adaptation to emerging technologies and methodologies, coupled with the robust functionalities of Pycaret and scikit-learn, are key to advancing our understanding of neurodegenerative diseases and enhancing the accuracy of predictive models.

References

[1] S. K. Van Den Eeden, C. M. Tanner, A. L. Bernstein, R. D. Fross, A. Leimpeter, D. A. Bloch, and L. M. Nelson, "Incidence of Parkinson's Disease: Variation by Age, Gender, and Race/Ethnicity," *American Journal of Epidemiology*, vol. 157, pp. 1015–1022, 06 2003.

[2] S. Grover, S. Bhartia, Akshama, A. Yadav, and S. K.R., "Predicting severity of parkinson's disease using deep learning," *Procedia Computer Science*, vol. 132, pp. 1788–1794, 2018. International Conference on Computational Intelligence and Data Science.

[3] C. Shi, Z. Zheng, Q. Wang, C. Wang, D. Zhang, M. Zhang, P. Chan, and X. Wang, "Exploring the effects of genetic variants on clinical profiles of parkinson's disease assessed by the unified parkinson's disease rating scale and the hoehn–yahr stage," *PLOS ONE*, vol. 11, pp. 1–11, 06 2016.

[4] R. Prashanth and S. Dutta Roy, "Novel and improved stage estimation in parkinson's disease using clinical scales and machine learning," *Neurocomputing*, vol. 305, pp. 78–103, 2018.

[5] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, 2010.

[6] M. Nilashi, O. Ibrahim, S. Samad, H. Ahmadi, L. Shahmoradi, and E. Akbari, "An analytical method for measuring the parkinson's disease progression: A case on a parkinson's telemonitoring dataset," *Measurement*, vol. 136, pp. 545–557, 2019.

[7] M. García-Ordás, J. Benítez-Andrades, J. Aveleira, J. Alija-Perez, and C. Benavides, "Determining the severity of parkinson's disease in patients using a multi task neural network," *Multimedia Tools and Applications*, 06 2023.

[8] pycaret, "Pycaret docs."