Diagnosis of Parkinson's Disease Using Machine Learning Algorithms

*Ilaria Pia Battista, Michele Roccotelli, *Member, IEEE*, Wasim Ali, Maria Pia Fanti, *Fellow, IEEE*

***Abstract*—Parkinson's Disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease, significantly impairing motor functions and quality of life. Early and accurate monitoring of PD progression is essential for improving patient outcomes. Among the innovative approaches, vocal signal analysis has gained traction as a non-invasive tool for assessing disease progression and treatment efficacy. PD patients often experience dysarthria, a neurological speech disorder affecting the pneumo-phono-articulatory system responsible for voice and language production. This study leverages machine learning algorithms to predict the motor and total scores of the Unified Parkinson's Disease Rating Scale (UPDRS), widely used for tracking PD symptoms. Utilizing a dataset of 5,875 samples, various regression models, including Decision Tree, Random Forest, XGBoost, and Extra Tree, were trained and tested. Additionally, an ensemble Stacking Regressor was implemented to enhance prediction accuracy. The analysis of vocal recordings offers an innovative, non-invasive method for monitoring PD progression, reducing reliance on more subjective and invasive traditional approaches. The use of the ensemble model surpassed the performance of individual models, achieving an R^2 of 98.31% for predicting total UPDRS and 98.21% for motor UPDRS. Furthermore, the ensemble approach mitigates the risk of overfitting, ensuring greater robustness and reliability in predictions.**

These findings demonstrate the potential of machine learning in providing reliable and objective tools for PD monitoring, overcoming the subjectivity and limitations of traditional methods.

I. Introduction

Parkinson's Disease (PD) is a progressive neurodegenerative disorder of the central nervous system, first described by James Parkinson in 1817 [1]. It is characterized by the degeneration of dopaminergic neurons in the substantia nigra, leading to reduced dopamine levels. This neurotransmitter deficit impairs the basal ganglia, which are responsible for initiating voluntary movements, suppressing involuntary actions, and coordinating posture changes. PD symptoms generally manifest after a loss of 50% of these neurons, typically appearing around the age of 70 in women and 68 in men, though juvenile-onset cases are also documented.

Motor symptoms [2] include tremors, bradykinesia, rigidity, and speech difficulties, such as hypophonia and dysarthria. Non-motor symptoms [3] encompass cognitive impairments, mood disorders, sleep disturbances, and autonomic dysfunctions. Of particular interest are vocal alterations, including reduced pitch variability, monotony, and vocal tremors, which can precede motor symptoms. These changes stem from hypokinesia and rigidity of laryngeal muscles, affecting phonation and contributing to reduced communication efficacy.

Advanced vocal analysis has emerged as a promising tool for understanding PD progression and early diagnosis, offering insights into its complex interplay of symptoms.

Recent advancements emphasize the collaboration between medicine and engineering to improve the quality of life for Parkinson's Disease (PD) patients. Projects like REMPARK [4] showcase the potential of telemonitoring algorithms and wearable technologies for patient management. REMPARK integrates wearable sensors and smartphone connectivity to monitor motor symptoms in real time and provide feedback to healthcare professionals. This system supports personalized care and treatment optimization, improving patient-doctor interaction while reducing hospitalizations and healthcare costs.

The potential impact of REMPARK spans three critical areas: medical, social, and economic benefits. Medically, it enhances disease management and rehabilitation by providing accurate, real-time insights into patient conditions. Socially, it contributes to a more sustainable European healthcare system by offering high-quality, personalized care. Economically, it reduces hospitalization rates and drives innovation in Personal Health Systems (PHS) and medical technologies. REMPARK also fosters interoperability standards and secure health data communication, improving patient-doctor interaction and promoting a more integrated healthcare ecosystem.

Artificial intelligence (AI) has emerged as a valuable tool for medical diagnosis [5]. In particular, AI is used for diagnosing PD by analyzing biomarkers, including vocal and handwriting features. Vocal analysis [6] [7] has been used to detect abnormalities such as reduced pitch variability and tremors, while handwriting analysis [8] identifies symptoms like micrographia, which can aid early detection. These technologies aim to create autonomous systems for reliable disease monitoring, enabling earlier interventions and personalized care, ultimately improving patient outcomes.

AI also enhances healthcare systems by optimizing resource allocation, lowering operational costs, and improving diagnostic accuracy. By analyzing complex data, AI systems support better treatment planning and more efficient care, marking a shift toward accessible, patient-centered healthcare.

This study, based on the dataset from Athanasios Tsanas et al. [9], focuses on preprocessing data and evaluating various

Ilaria Pia Battista, Michele Roccotelli, Wasim Ali and Maria Pia Fanti are with the Department of Electrical and Information Engineering of the Polytechnic University of Bari, Italy. email: hillarybattista@gmail.com, (michele.roccotelli,wasim.ali,mariapia.fanti)@poliba.it.

regression algorithms to predict PD severity using the UPDRS scale. Among the models tested, the Extra Trees algorithm achieved the highest prediction accuracy for both motor and total UPDRS scores.

Section II describes the UPDRS limitations and Dataset features for Parkinson's analysis.

Section III explains the regression models used for prediction and data preprocessing.

Section IV shows the evaluation metrics used, the performance of the 4 regression algorithms and finally the Stackiing Regression algorithm.

Section V discusses the results before providing the conclusions

II. UPDRS and Dataset Features for Parkinson's Analysis

A. Unified Parkinson's Disease Rating Scale(UPDRS)

The Unified Parkinson's Disease Rating Scale (UPDRS) is a widely used clinical tool for assessing the severity and progression of Parkinson's Disease (PD) [10]. It evaluates various symptoms through a standardized questionnaire, covering motor function, mood, behavior, and daily activities, providing a comprehensive view of the patient's condition. The scale's reliability and sensitivity make it ideal for tracking subtle changes in PD symptoms over time, a critical factor in both clinical practice and research.

In this study, UPDRS is chosen as the evaluation method for measuring PD progression. Its multidimensional framework ensures that diverse aspects of the disease are systematically quantified, enabling robust correlations between speech impairments and overall disease severity. Additionally, its widespread clinical adoption allows for consistency and comparability across studies. By aligning telemonitoring methods with UPDRS, the study ensures clinically meaningful, standardized, and reproducible results, bridging traditional assessments with innovative remote monitoring solutions.

B. Dataset Features

The dataset used in this study includes several features that provide quantitative and qualitative insights into the vocal characteristics and clinical conditions of Parkinson's Disease (PD) patients. Below is a detailed description of the features:

- **Subject#:** A categorical variable that uniquely identifies each patient in the trial.
- **Age:** A numerical variable indicating the age of the subject at the time of measurement, expressed in years.
- **Sex:** A categorical variable coded as 0 for male and 1 for female.
- **Test_time:** A numerical variable, expressed in days, representing the time elapsed from the start of the trial to the recording session.

Vocal recordings capture a nearly periodic signal caused by the opening and closing of the glottis during phonation. The period of this signal, as illustrated in Fig.1 [11], is called the pitch period, while the vibration frequency of the vocal cords (inverse of the pitch period) is referred to as the fundamental frequency (Fo).

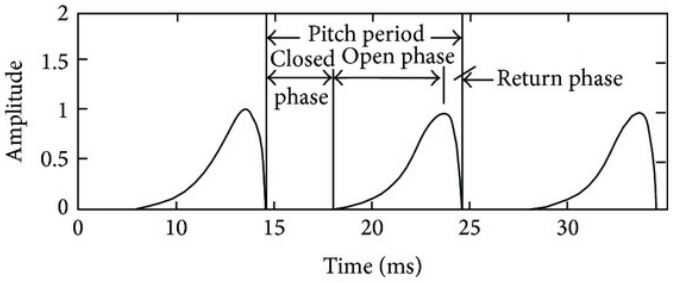

Fig. 1. Tone period representation for the vocal recording [11].

Variations between successive vocal cycles can be analyzed through the following features:

- **Jitter (%):** The percentage variation in the fundamental frequency (Fo).
- **Jitter (Abs):** The absolute variation in the fundamental frequency.
- **Jitter: RAP:** Relative Average Perturbation, measuring pitch variation averaged over three consecutive periods.
- **Jitter: PPQ5:** Five-point Period Perturbation Quotient, averaging pitch variation over five consecutive periods.
- **Shimmer (dB):** The variation in amplitude expressed in decibels.
- **Shimmer: APQ3:** Three-point Amplitude Perturbation Quotient, representing amplitude variation averaged over three consecutive periods.
- **Shimmer: APQ5:** Five-point Amplitude Perturbation Quotient, averaged over five consecutive periods.
- **Shimmer: APQ11:** Eleven-point Amplitude Perturbation Quotient, averaged over eleven consecutive periods.
- **Shimmer: DDA:** Difference of Differences of Amplitude, calculated as the mean absolute difference in amplitude among three consecutive vocal cycles.
- **HNR (Harmonics-to-Noise Ratio):** Indicates the proportion of harmonic components relative to noise in the voice.
- **NHR (Noise-to-Harmonics Ratio):** Measures the noise present in the voice relative to harmonic components.
- **RPDE (Recurrence Period Density Entropy):** Assesses the ability of vocal folds to sustain simple vibrations and quantifies deviations from exact periodicity.
- **DFA (Detrended Fluctuation Analysis):** Quantifies the stochastic similarity of turbulent noise in the vocal signal.
- **PPE (Pitch Period Entropy):** Measures the unpredictability of pitch periods, reflecting compromised pitch control during sustained phonation.
- **Motor_UPDRS:** The score for the motor section of the original UPDRS scale (maximum 108).
- **Total_UPDRS:** The total UPDRS score, summing all sections of the original scale (maximum 199).

These features, derived from voice recordings and clinical observations, are essential for training machine learning models to predict *motor_UPDRS* and *total_UPDRS* scores. Using this dataset, models can accurately estimate these clinical scores, bypassing the issue of neurologist misinterpretation and patient misperception. Each feature correlates with key symptoms of PD, such as changes in pitch, amplitude, and voice quality, providing a quantitative representation of the patient's condition.

III. "The Machine Learning-based Methodology"

The procedure adopted to process the data and to implement the machine learning models for PD prediction is represented in Fig. 2. The procedure begins with data preprocessing, which includes handling missing values, data visualization, and normalization or scaling. The optimized dataset is then used to implement various machine learning algorithms, including Random Forest, Decision Tree, XGBoost, and Extra Trees. Finally, the stacking regressor combines these models to produce the final results.

The dataset used in this study consists of 5,875 vocal recordings from 42 early-stage PD patients, collected during a six-month clinical trial. Recordings were performed weekly at patients' homes using the At-Home Testing Device (AHTD), which enables remote telemonitoring. Each session included sustained phonation of the vowel "ahhh," recorded via a head-mounted microphone. The dataset comprises 4,008 male and 1,867 female samples, highlighting PD's higher prevalence in men. The patients' ages range is from 36 to 85 years, with a mean age of approximately 65 years. The *motor_UPDRS* and *total_UPDRS* scores, key clinical indicators of PD severity, show considerable variability, with means of 21.3 and 29.0, respectively. Variations in vocal features such as *Jitter(%)* and *Jitter(Abs)* are also observed, reflecting subtle changes in pitch stability.

A. Data Preprocessing

The development environment was configured using Anaconda, the most widely used platform for Data Science and Machine Learning with Python. Anaconda simplifies the process of setting up an integrated development environment by providing tools such as Spyder and Jupyter Notebook, along with a Python interpreter and a package manager named Conda. Additionally, it includes over 300 pre-installed libraries, enabling an immediate and efficient workflow.

The following libraries were utilized in this study, each serving a specific purpose:

- **Pandas**: Used for data manipulation and analysis, enabling efficient handling of numerical tables and time-series data.
- **NumPy**: Facilitated operations with large vectors and matrices, improving computational efficiency for numerical data processing.
- **Seaborn**: Employed to generate statistical plots, providing insights that could not be easily derived from tabular data.

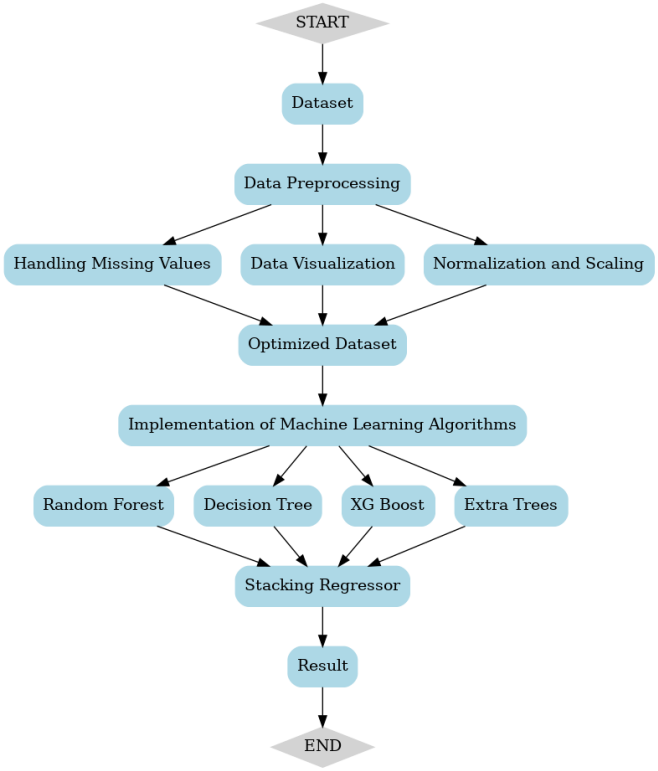

Fig. 2. Workflow of data preprocessing and model implementation.

- **Matplotlib**: Used for creating static, animated, and interactive plots. Scatter plots in Chapter 2 were created using this library.
- **Scikit-learn**: An open-source Machine Learning library, utilized for implementing algorithms such as decision trees and evaluating model performance using metrics like accuracy and Mean Squared Error (MSE).

During the data preprocessing phase, the 'subject#' feature was removed as it introduced unnecessary variability without contributing meaningful information to the regression models. Missing values were addressed by verifying the completeness of the dataset, which was confirmed to contain no null values. Data visualization shows *Jitter(%)* values with positive bias, indicating reduced vocal stability in some patients due to bradykinesia and rigidity. *Motor_UPDRS* scores range from 5 to 39, clustering between 15 and 28, reflecting mild motor symptoms. Similarly, *total_UPDRS* scores range from 7 to 54, with a mean of 29.0 and moderate variability, suggesting that patients present with early Parkinson's disease with varying severity of symptoms. Before training the regression models, additional preprocessing steps were performed to prepare the data. The target variables *motor_UPDRS* and *total_UPDRS* were separated from the dataset to create the dependent variables *y_motor* and *y_total*, while the independent features were stored in a variable *X* with a shape of (5875, 19), representing all observations and 19 independent features.

A normalization step was then applied to ensure that all independent features contribute equally during model training.

This technique standardizes the features to have a mean of 0 and a standard deviation of 1, calculated using the formula:

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma}$$

where:

- X is the original feature value,
- μ is the mean of the feature,
- σ is the standard deviation of the feature.

To evaluate the performance of the models on unseen data, the dataset was split into a training set and a test set. The split ratio was set to 80% for training and 20% for testing. The split was performed twice, once for each target variable (*motor_UPDRS* and *total_UPDRS*), while maintaining the same partitioning for the independent features *X*.

The resulting dimensions of the datasets are as follows:

- *X_train*: (4700, 19) *X_test*: (1175, 19)
- *y_train_total*: (4700, 1) *y_test_total*: (1175, 1)
- *y_train_motor*: (4700, 1) *y_test_motor*: (1175, 1)

This dual splitting approach allowed for the independent handling of the two output variables, *motor_UPDRS* and *total_UPDRS*, while ensuring that the same set of independent features was used for both tasks.

B. Regression Model

- **Decision Tree:** The Decision Tree algorithm is a simple and efficient model for both classification and regression tasks [12]. It splits data hierarchically using binary decisions at each node, making the decision process interpretable. However, it is prone to overfitting, especially when no regularization is applied.
- **Random Forest:** Random Forest, introduced by Leo Breiman in 2001, builds an ensemble of decision trees using random subsets of data and features at each split. This randomness reduces overfitting and enhances stability. It aggregates predictions from all trees by averaging for regression tasks, and is robust against noise with minimal data preparation required.
- **XGBoost:** Based on Gradient Boosting, XGBoost builds regression trees iteratively to minimize errors from previous models. It employs L1 and L2 regularization to penalize model complexity, reducing overfitting. XGBoost is known for its computational efficiency and high accuracy, requiring fewer resources than other methods.
- **Extra Trees:** Extra Trees (Extremely Randomized Trees) enhances model diversity by using random cut-points for splits and the entire dataset instead of bootstrap samples. This reduces variance while maintaining low bias, leading to highly accurate model [13]. It does not require optimal split-point selection, which significantly speeds up computations.

The implementation of the regression algorithms followed a structured approach aimed at predicting the output variables *motor_UPDRS* and *total_UPDRS*. For each algorithm, two separate models were created, one for each target variable. The implementation consisted of the following steps:

1) **Library Import**: Required libraries were imported, including `mean_square_error` and `r2_score` from the `sklearn.metrics` module.
2) **Model Creation**: Two models were created for each algorithm, one to predict `motor_UPDRS` and the other for `total_UPDRS`. Both models were trained using the training dataset (`X_train`) with the `fit()` function.
3) **Prediction Generation**: Using the trained models, predictions were made on the test dataset (`X_test`) with the `predict()` function.
4) **Model Evaluation**: The performance of each model was assessed using three key metrics: the coefficient of determination (R^2), the Mean Square Error (MSE), and the Root Mean Square Error (RMSE).

For each model, basic hyperparameter values were adopted to ensure comparability and reproducibility: `random_state=0` for the Decision Tree, and `n_estimators=100` with `random_state=42` for Random Forest, XGBoost, and Extra Trees. No extensive hyperparameter optimization or cross-validation was performed, as the focus was on baseline model behavior and relative performance under uniform conditions.

IV. Evaluation, Performance Analysis, and Stacking Regression Results

A. Evaluation

To assess the performance of the implemented regression models, three key evaluation metrics were used: the coefficient of determination (R^2), the Mean Square Error (MSE), and the Root Mean Square Error (RMSE). These metrics provide a comprehensive evaluation of the predictive accuracy and error magnitude of each model.

Coefficient of Determination (R^2)

The R^2 score measures how well the regression model predicts the target variable compared to the mean of the data. It is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where y_i represents the true values, $\hat{y}_i$ are the predicted values, and $\bar{y}$ is the mean of the true values. A R^2 value close to 1 indicates a good fit.

Mean Square Error (MSE)

The MSE measures the average squared difference between the true and predicted values, penalizing large errors. It is defined as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

A lower MSE indicates better predictive performance.

Root Mean Square Error (RMSE)

The RMSE is the square root of the MSE, providing an error metric in the same units as the target variable. It is expressed as:

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

The RMSE is useful for interpreting model error in practical terms, as it reflects the magnitude of prediction errors.

These metrics were systematically applied to evaluate the models for both target variables: *motor_UPDRS* and *total_UPDRS*. This approach ensured a consistent and reliable comparison of the regression algorithms, including Decision Tree, Random Forest, Extra Trees, and XGBoost.

B. Performance Analysis

The comparison of R², MSE, and RMSE scores across the regression algorithms is illustrated in Fig. 3 and Fig. 4, which shows histograms generated using Python's Matplotlib library, for the total_UPDRS and motor_UPDRS respectively.

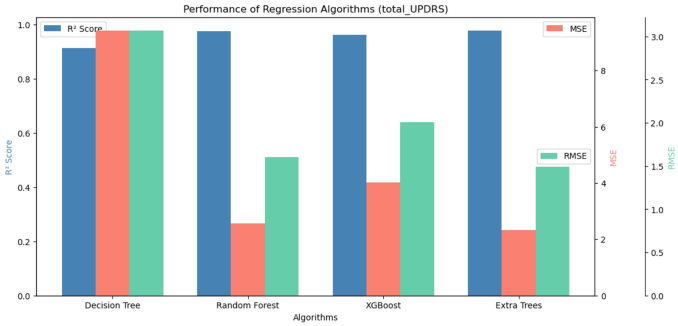

Fig. 3. Evaluation metrics for *total_UPDRS* prediction

For the total_UPDRS, the Decision Tree algorithm achieved an R^2 of 91.49%, an MSE of 9.42, and an RMSE of 3.07. Despite a good fit ($R^2 > 90\%$), it has the highest MSE among the models, indicating greater prediction error variability.

The Random Forest model demonstrated excellent performance with an R^2 of 97.68%, an MSE of 2.57, and an RMSE of 1.60, benefiting from ensemble learning to reduce overfitting and improve robustness.

XGBoost, with an R^2 of 96.37%, an MSE of 4.02, and an RMSE of 2.00, performed well, though slightly less effectively than Random Forest and Extra Trees.

Extra Trees provided the best overall results with an R^2 of 97.90%, an MSE of 2.32, and an RMSE of 1.49, showcasing the model's accuracy and low error variability.

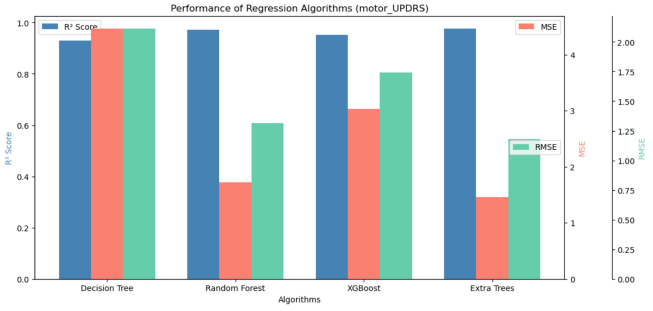

Fig. 4. Evaluation metrics for *motor_UPDRS* prediction

For the motor_UPDRS, the Extra Trees model achieved the best performance with an R² of 97.72%, an MSE of 1.39, and an RMSE of 1.18, demonstrating exceptional accuracy and low prediction error variability.

The Random Forest model followed closely with an R² of 97.29%, an MSE of 1.72, and an RMSE of 1.31, benefiting from ensemble learning to enhance robustness and reduce overfitting.

XGBoost delivered solid results with an R² of 95.25%, an MSE of 3.02, and an RMSE of 1.74, though slightly less effective than the Random Forest and Extra Trees models.

The Decision Tree model, with an R² of 92.99%, an MSE of 4.46, and an RMSE of 2.11, showed relatively lower performance, highlighting its limitations compared to ensemble approaches. These results underline the superiority of ensemble models like Random Forest and Extra Trees, which leverage multiple decision trees to improve prediction accuracy and robustness. Moreover, XGBoost, through its boosting mechanism, achieves a balanced trade-off between bias and variance, significantly outperforming the standalone Decision Tree model.

C. Stacking Regression Results

To further enhance the performance achieved by the four regression algorithms—Random Forest, Decision Tree, Extra Trees, and XGBoost—a Stacking Regressor ensemble method was implemented. Ensemble methods [14] are advanced machine learning techniques that combine the predictions of multiple base models, often referred to as "weak learners" or "base learners," to produce a final prediction that is more robust and accurate. The primary goal of such methods is to reduce generalization errors by compensating for the weaknesses of individual models through their combination.

Stacking is an ensemble technique that combines the predictions of multiple base models through a "meta-model." While the base models are trained on the full training dataset, the meta-model is trained on their predictions. This setup allows the meta-model to correct for the limitations of each base model, often leading to improved overall performance. The goal of using this method is to achieve more accurate and reliable predictions by addressing the inherent variability of individual models.

The implementation consisted of the following steps:

1) **Importing Libraries:** Necessary libraries, including `StackingRegressor` and `LinearRegression` from `sklearn`, were imported.
2) **Defining Base Models:** The previously trained regression models—Decision Tree, Random Forest, Extra Trees, and XGBoost—were defined as base models.
3) **Defining the Meta-Model:** A Linear Regression model was chosen as the meta-model to aggregate predictions from the base models.
4) **Creating, Training, and Predicting:** The `StackingRegressor` was configured with the defined base models and meta-model, trained on the training dataset (`X_train` and `y_train_total`), and used to make predictions on the test dataset (`X_test`).

5) **Performance Evaluation:** The model's performance was assessed using using the same metrics through which the 4 regression algorithms were evaluated.

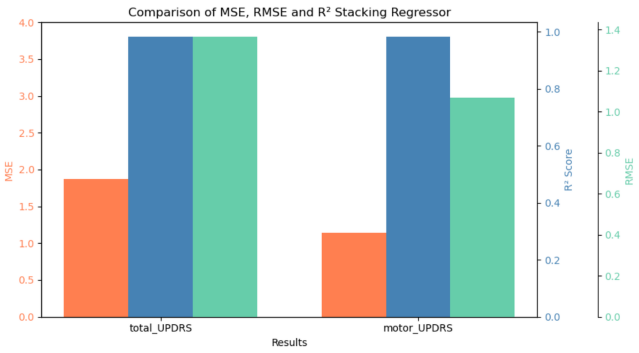

Fig. 5. Performance metrics of the Stacking Regressor algorithm

The implementation of the Stacking Regressor demonstrated significant improvements in predictive performance for both `motor_UPDRS` and `total_UPDRS`. For `motor_UPDRS`, the stacking model achieved an MSE of 1.13, an RMSE of 1.06, and an R^2 of 0.98. Similarly, for `total_UPDRS`, the model obtained an MSE of 1.86, an RMSE of 1.36, and an R^2 of 0.98. These results, shown in Fig. 5, confirm substantial enhancements in predictive accuracy and generalization capabilities. The reduction in MSE and RMSE, coupled with a high R^2, underscores the effectiveness of combining the predictive strengths of individual base models through the stacking approach.

V. Discussion

Our methodology addresses several critical points identified in a recent systematic review that examined articles recently published [15]. Notably, only 38.9% of studies reviewed explicitly described how hyperparameters were optimized. In contrast, this paper provides a detailed description of the preprocessing steps, including handling missing data, normalization, and the division of data into training and test sets. This transparency enhances the reproducibility of the study.

Another significant issue identified in the review pertains to demographic imbalances in datasets: 68.1% of studies exhibited imbalanced data, with few addressing this limitation. While our dataset also presents a demographic imbalance, this is justified by the epidemiological distribution of PD, which predominantly affects older populations. The histogram of the 'age' feature reveals a skew towards patients aged 50–80 years, consistent with PD's prevalence in older individuals. This demographic representation ensures that the model's results are highly relevant for clinical applications. However, we acknowledge that the model's reliability for younger age groups may be limited, an aspect that warrants further investigation in future studies.

By addressing key limitations in the existing literature and providing a transparent workflow, this study contributes to advancing the application of machine learning in neurodegenerative disease research.

Conclusions

This study leveraged machine learning techniques to analyze vocal recordings from Parkinson's Disease (PD) patients, demonstrating the potential of these non-invasive tools for early diagnosis and monitoring disease progression. By extracting features from the vocal recordings, we were able to predict UPDRS scores, specifically targeting the motor abilities and total scores, which are critical for assessing the severity of PD symptoms. The Extra Trees model proved to be particularly effective, with high R^2 values of 97.71% for `motor_UPDRS` and 97.90% for `total_UPDRS`, showcasing its ability to handle complex regression tasks. The implementation of the Stacking Regressor further improved prediction accuracy, achieving R^2 values of 98.21% for `motor_UPDRS` and 98.31% for `total_UPDRS`, highlighting the strength of ensemble methods in enhancing model performance. The dataset's demographic imbalance, skewed toward older age groups, mirrors the epidemiological characteristics of PD, which predominantly affects the elderly. While this makes the model highly relevant for the clinical context, future research should explore its reliability for younger populations.

References

[1] James Parkinson, *An Essay on the Shaking Palsy*, 1817.

[2] Ahmed A. Moustafaa, et al., "Motor symptoms in Parkinson's disease: A unified framework", *Neuroscience and Biobehavioral Reviews*, vol. 68, pp. 727–740, Sep. 2016.

[3] W. Poewe, "Non-motor symptoms in Parkinson's disease", *European Journal of Neurology*, vol. 15, pp. 14–20, 2008.

[4] A. Samà, et al., "A double closed loop to enhance the quality of life of Parkinson's Disease patients: REMPARK system", *Studies in Health Technology and Informatics*, vol. 207, pp. 115–124, 2014.

[5] V. Dambra, M. Roccotelli and M. P. Fanti, "Diabetic Disease Detection using Machine Learning Techniques," *10th International Conference on Control, Decision and Information Technologies (CoDIT)*, Vallette, Malta, 2024, pp. 1436-1441.

[6] M. Shahbakhi, et al., "Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine", *Journal of Biomedical Science and Engineering*, vol. 7, no. 4, pp. 147–156, 2014.

[7] E. Vaiciukynas, A. Verikas, A. Gelzinis, M. Bacauskiene, "Detecting Parkinson's disease from sustained phonation and speech signals", *PLoS ONE*, vol. 12, no. 10, 2017.

[8] A. L. C. S. Afonso, G. H. Rosa, C. R. Pereira, S. A. T. Weber, C. Hook, V. H. C. Albuquerque, J. P. Papa, "A recurrence plot-based approach for Parkinson's disease identification", *Future Generation Computer Systems*, vol. 94, pp. 282–292, 2019.

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

[10] C. G. Goetz, et al., "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results", *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, Nov. 2008.

[11] T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*, Prentice Hall PTR, 2001.

[12] M. Xu, P. Siqueira, S. Hensley, B. Chapman, "Decision tree regression for land cover classification in remote sensing", *Remote Sensing of Environment*, vol. 97, no. 3, pp. 322–336, 2005.

[13] P. Geurts, D. Ernst, L. Wehenkel, "Extremely randomized trees", *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.

[14] D. Opitz, R. Maclin, "Popular Ensemble Methods: An Empirical Study", *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, 1999.

[15] T. Tabashum, R. C. Snyder, M. K. O'Brien, M. V. Albert, "Machine Learning Models for Parkinson Disease: Systematic Review", *JMIR Medical Informatics*, vol. 12, 2024.