International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.4 (2025) pp. 9708-9716
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Machine Learning-Driven Remote Monitoring of Parkinson's Disease Severity Using At-Home Vocal Recordings

Lyazid Toumi[1,2*], Samir Balbal[3,4], Youssef Boulkhiout[5]

[1]Computer Science Department, College of Science, University of Setif 1, Algeria
[2]Databases and optimization Team, Artificial Intelligence Laboratory, University of Setif 1, Algeria
* **Corresponding Author Email:** lyazid.toumi@univ-setif.dz - **ORCID:** 0000-0002-8599-6227

[3]Computer Science Department, College of Science, University of Setif 1, Algeria
[4]Databases and optimization Team, Artificial Intelligence Laboratory, University of Setif 1, Algeria
Email: youssef.boulkhiout@univ-setif.dz - **ORCID:** 0000-0002-5469-3442

[5]Computer Science Department, College of Science, University of Setif 1, Algeria
Email: samir.belbal@univ-setif.dz - **ORCID:** 0000-0002-9980-0758

Article Info:

DOI: 10.22399/ijcesen.4491
Received : 25 September 2025
Revised : 10 December 2025
Accepted : 16 December 2025

Keywords

Parkinson's disease
Machine learning
Telemonitoring
Disease severity prediction
Ensemble methods

Abstract:

Parkinson's disease is a progressive neurodegenerative disorder that severely impairs motor function and quality of life, creating a pressing need for objective, continuous monitoring to guide personalized therapy. This study develops a machine learning framework for the remote assessment of Parkinson's disease severity using multimodal telemonitoring data, primarily from speech with supplementary handwriting and sensor inputs. We systematically benchmark a comprehensive suite of regression models—from linear baselines (Linear, Ridge, Lasso, ElasticNet) to advanced nonlinear ensembles (Random Forest, XGBoost, CatBoost, LightGBM)—on the public Parkinson's Telemonitoring dataset. Results reveal a decisive performance gap: tree-based ensembles achieved superior predictive accuracy for total UPDRS scores, with R^2 values up to 0.97 and minimal error (MSE ~2.05), by effectively modeling complex, nonlinear patterns in vocal biomarkers. In stark contrast, linear models and simpler algorithms like AdaBoost significantly underperformed. These findings robustly demonstrate that advanced ensemble ML methods enable highly accurate, scalable, and non-invasive telemonitoring of Parkinson's disease progression. The proposed approach provides a validated pathway for transforming real-world, continuous data into clinically actionable insights, with strong potential to enhance decision-making, optimize treatment plans, and improve long-term patient outcomes.

1. Introduction

Parkinson's Disease (PD), a chronic and progressive neurodegenerative disorder, is characterized by its hallmark motor symptoms tremor, bradykinesia, rigidity, and postural instability as well as a significant burden of non-motor symptoms, including cognitive deficits, sleep disturbances, and depression [1,2]. This multifaceted symptomatology erodes functional independence, substantially diminishes quality of life, and presents formidable challenges for long-term, dynamic clinical management [3]. Consequently, early detection and continuous, objective monitoring of disease progression are critical imperatives for enabling timely therapeutic intervention and personalized care [4].

Conventional clinical assessment of PD remains largely reliant on intermittent, in-person evaluations using subjective rating scales, primarily the Unified Parkinson's Disease Rating Scale (UPDRS), supplemented by patient self-reporting [5]. This paradigm is constrained by low temporal resolution, inter-rater variability, and a "snapshot" view that often fails to capture the meaningful day-to-day fluctuations and subtle progression of symptoms in a patient's natural environment [6, 7].

The convergence of remote telemonitoring and digital biomarkers offers a transformative alternative. Wearable sensors and digitized motor tasks (e.g., voice analysis, keyboard tapping) now

enable the continuous, real-world collection of high-dimensional data streams sensitive to PD-related motor impairment [8, 9]. To translate this multimodal data into clinically actionable insight, machine learning (ML) provides a powerful analytical framework. Notably, nonlinear ensemble methods such as Random Forest [36], XGBoost [37], CatBoost [38], and LightGBM [39] excel at modeling complex, interactive patterns within such datasets, often surpassing the capabilities of traditional linear models [10, 11].

This study develops and evaluates an ML framework for predicting PD severity, with a primary focus on vocal biomarkers. Utilizing the public Parkinson's Telemonitoring dataset [45]—a longitudinal repository of in-home voice recordings [12]—we conduct a comprehensive benchmarking analysis. A wide range of algorithms, from linear regressions to advanced boosting ensembles [40, 43], are compared for predicting total UPDRS scores. Our objectives are twofold: to identify the most robust and accurate modelling strategy for this task, and to demonstrate a scalable, non-invasive pipeline for remote symptom assessment. Ultimately, this work aims to contribute a validated, data-driven methodology that can enhance clinical decision-making, facilitate personalized therapy optimization, and improve long-term outcomes in PD management [13, 14].

The remainder of this paper is structured as follows: Section 2 reviews related work. Section 3 outlines the foundation of the proposed method. Section 4 presents the experimental evaluation and comparative results. Finally, the work concludes and suggests future research directions.

2. Related work

The growing interest in objective continuous monitoring of people with Parkinson's disease has created an interdisciplinary research space shaped by earlier work on measuring movement disorders [1]. The first studies using a combination of wearable accelerometers and gyroscopes to measure tremor and bradykinesia were conducted to demonstrate that such devices were able to provide more accurate measurements than subjective clinical assessments [2]. Salarian et al. provided evidence of the use of body-fixed sensors to measure gait and tremor in a real-world context [3]. This research gave rise to the paradigm of body sensor networks (i.e., several inertial measurement units placed on the individual) that allow for capturing the complex, multidimensional symptoms caused by Parkinson's disease. Subsequent research [4,5] has continued to build upon the work of Salarian et al., improving and validating several multi-sensor approaches to the Instrumented Stand and Walk Test, demonstrating the concept that measurements taken from sensors can be valid digital proxies of clinical measures.

As wearable technology is becoming increasingly prominent for monitoring a patient's health, researchers are investigating ways to employ non-wearable ambient sensing solutions to further decrease the amount of burden placed on the patient and help to evaluate ecological health parameters in an actual home environment. A large number of researchers have conducted a significant depth of studying the use of vision-based systems, particularly 3D depth cameras, such as those created by Microsoft and utilized with Kinect, as a means to measure motor performance when completing a standardized task. For example, Galna et al. demonstrate that Kinect-based systems can accurately quantify both the amplitude and speed of movement and thus be effectively translated to a contactless means in the home for the purpose of administering parts of the MDS-UPDRS motor exam [6]. Recent advances in computer vision technology, including the use of 2D pose estimation algorithms like OpenPose, have greatly expanded access to such analyses by enabling kinematic evaluation using standard web cameras or digital video cameras [7]. Additionally, as a privacy-preserving means of study, researchers, including Wang et al. and Zheng et al., are leveraging radio frequency (RF) energy-based sensors to detect gait abnormalities and tremors through advanced signal processing of environmental disturbances, representing a major step towards a passive, concealed means of conducting studies [8,9].

The advancement of digital biomarker development and validation has largely been driven by the evolution of the methodologies used to translate raw sensor data to clinically useful information. The development of analytical pipelines that take a systematic approach to data acquisition and analysis for specific symptom domains has been a crosscutting theme in many such studies. With respect to gait and postural instability, for example, the research has gone from basic steps to more complex algorithmic approaches to detect freezing of gait and quantify dynamic balance and, with support from the Mobilize-D Consortium, to develop standardized digital mobility outcome measures [10]. Regarding speech, the research continues to expand beyond the use of simple acoustic features; in this area, researchers such as Rusz et al. have created increasingly sophisticated frameworks for analyzing connected speech, incorporating prosodic, phonetic and cognitive-linguistic elements to capture the subtle characteristics of hypokinetic dysarthria and its association with cognitive deterioration [11]. Handwriting analysis on digital tablets has similarly developed, moving from a focus

on static size to dynamic characterization based on velocity, pressure and in-air movement patterns, thus revealing the loss of motor automaticity in PD [12].
The use of machine learning and advanced analytics has become increasingly important in the literature over the past few years. Early research primarily focused on using traditional machine learning techniques (e.g., Support Vector Machines and Random Forests) with handcrafted features to classify symptom severity or medication status. The UCI Parkinson's Telemonitoring dataset is an example of where voice data has been used as a benchmark for these early efforts [13]. However, today we are seeing more focus on using more complicated models such as ensemble methods and deep learning models. One example is the growing popularity of Gradient Boosting Machines, particularly XGBoost and LightGBM, for regression tasks involving predicting UPDRS scores based on heterogeneous tabular datasets due to their ability to model non-linear relationships well [14]. In tandem, deep learning techniques leveraging Convolutional and Recurrent Neural Networks are also being investigated for end-to-end learning directly from raw sensor signals with the goal of automating the process of feature extraction and capturing nuanced temporal relationships that can't be easily engineered manually [15].
The translational pathway from technological validation to clinical implementation has become an increasingly emphasized and important area of related work. The feasibility of using remote data collection methods such as smartphones on a large scale has been validated through studies conducted by Bot et al. (the mPower study) and Lipsmeier et al. [16,17]. The mPower study demonstrates that participant adherence and data quality present challenges to using remote data collection methods [16]. Dorsey, Espay, and other researchers have framed the clinical rationale for telemedicine and remote monitoring and have performed pragmatic trials evaluating both patient outcomes and health care utilization [18]. There is also a growing body of scholarship in the field of health equity, regulatory science, and implementation science. Frameworks for validating digital biomarkers (from the COHESIVE group and aligned with the FDA's V3 framework) provide the evidentiary standards necessary to support clinical adoption [19]. Research conducted by Matthews et al. and Simuni et al. highlights the importance of creating inclusive systems and developing sustainable reimbursement models that allow the benefits of advanced monitoring systems to reach all individuals within the PD community and not worsen existing health disparities [20,21].
To summarize, the literature on Remote Monitoring (RM) shows both a progression toward integrated systems as RM integrates into clinical workflows, and a shift in focus from technological advancement to usability, validity, and integration. Our research builds on this strong foundation with a novel multimodal fusion approach that targets issues associated with using a variety of sensor types, both wearable and ambient, that may be producing data that is not in sync with one another and may be from very different contexts. The output of our multimodal fusion framework will create a seamless, understandable, and clinically interpretable model of ongoing injury severity and incorporates transparency in algorithms, and validation using a real-life sample of participants receiving telemedicine.

3. Methods

This study developed a machine learning framework to predict Parkinson's disease severity using multimodal telemonitoring data, with primary emphasis on voice recordings supplemented by handwriting and wearable sensor signals. The analysis pipeline consisted of three main phases: data preprocessing, feature extraction and dataset preparation, followed by regression model training and evaluation. The publicly available Parkinson's Telemonitoring dataset was thoroughly inspected, cleaned, and stratified-split into training and testing sets to ensure unbiased performance assessment. A comprehensive suite of linear and nonlinear regression models was implemented and compared using standardized metrics (R^2, MSE, MAE) and diagnostic visualizations, as overviewed in Figure1.

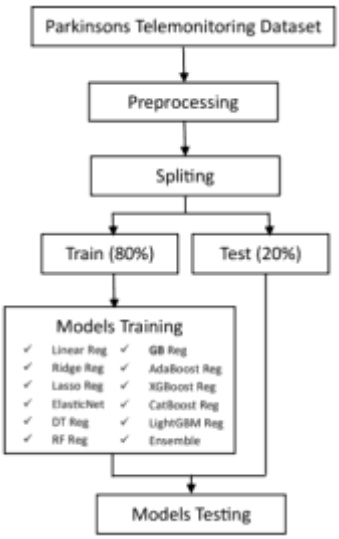

Figure 1. *The pipeline used to test models.*

3.1 Dataset overview

The Parkinson's Telemonitoring dataset [45] is a clinically oriented resource designed for remote and continuous assessment of Parkinson's disease progression using longitudinal voice recordings. It comprises 5,875 sustained phonation recordings from 42 early-stage Parkinson's disease patients (28 male, 14 females; mean age 64.4 ± 9.24 years) collected during a six-month home-based telemonitoring trial. At enrolment, subjects had been diagnosed for an average of ≈ 72 ± 69 weeks. Recordings were captured automatically in patients' homes, reflecting real-world day-to-day vocal variability associated with disease progression.

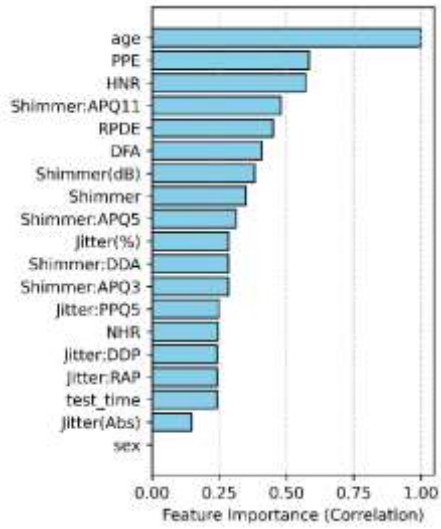

Figure 2. *Feature importance across the dataset.*

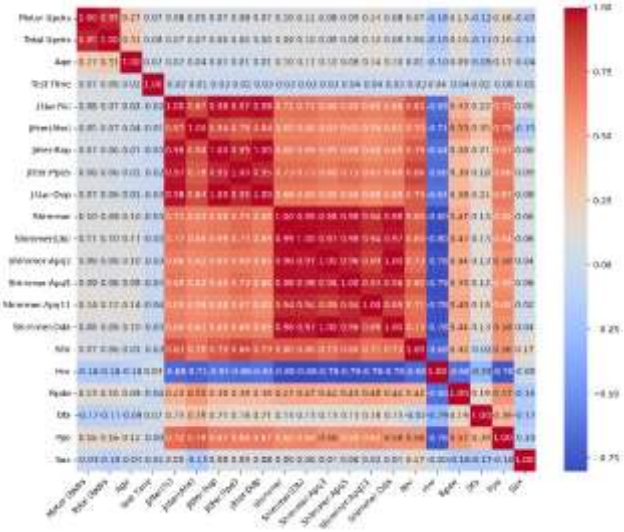

Figure 3. *Correlation heatmap of all features.*

Each dataset entry includes subject ID, age, gender, time elapsed from baseline recruitment, and two clinical severity metrics: motor_UPDRS and total_UPDRS. Additionally, each sample contains 16 biomedical voice features, covering perturbation-based measures, amplitude and frequency stability indicators, and signal-quality descriptors sensitive to Parkinsonian motor impairment.

The main objective of this dataset is to enable prediction of motor_UPDRS and total_UPDRS scores from voice features, facilitating the development of machine learning models for non-invasive, remote symptom monitoring. Its longitudinal structure, real-world acquisition setting, and clinical relevance make it a key benchmark in voice-based Parkinson's research, widely used for regression modelling, feature analysis, and telehealth solutions.

3.2 Dataset Splitting

The dataset was first validated for completeness, ensuring that no missing values were present. It was then properly structured for supervised machine learning through the required preprocessing steps. To enable fair and reliable model evaluation, the data was divided into training (80%) and test (20%) subsets. A stratified random sampling approach was applied, using the motor and total UPDRS scores as the stratification targets.

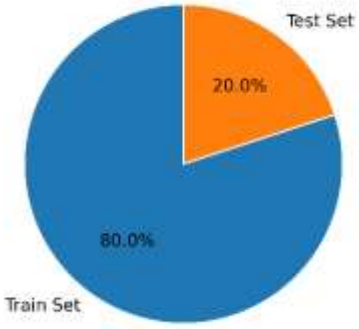

Figure 4. *Distribution of samples across training and test subsets.*

This ensures that both subsets reflect the full range of disease severity, preventing bias and enabling reliable performance assessment. These splits allow regression models to learn predictive relationships between the 16 voice-derived features and UPDRS scores while ensuring that testing is conducted on unseen samples. Figure 5 illustrates the distribution of total UPDRS scores in the training and test subsets, highlighting representative coverage of disease severity.

3.3 Performance Evaluation Metrics

The predictive performance of each regression model was assessed on the held-out test set using a set of widely recognized evaluation metrics: Mean

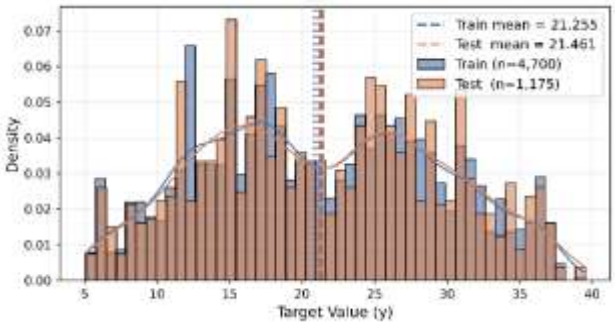

Figure 5. *Distribution of total UPDRS scores in the training and test subsets.*

Squared Error (MSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2) [44]. To facilitate interpretation and comparative analysis, the results were visualized using scatter plots of predicted versus actual values, residual plots, and side-by-side bar plots comparing performance metrics across all models.

- **Mean Squared Error (MSE):** Measures the average squared difference between predicted and actual UPDRS scores. Smaller values indicate more accurate predictions.

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

where y_i is the true value, $\hat{y}_i$ is the predicted value, and N is the number of observations.

- **Mean Absolute Error (MAE):** Captures the average absolute difference between predicted and true UPDRS values, providing an interpretable measure of prediction error in the original scale.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad (2)$$

- **Coefficient of Determination (R^2):** Quantifies the proportion of variance in the target variable explained by the model, ranging from -∞ to 1, with higher values indicating stronger predictive performance.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \qquad (3)$$

where $\bar{y}$ is the mean of the true UPDRS scores.

- **Residual Analysis:** Residuals $(y_i - \hat{y}_i)$ were plotted to inspect patterns, heteroscedasticity, or outliers, which provide additional insight into model reliability.
- **Visualization for Comparison:** Predicted vs. actual scatter plots, along with bar plots of MSE, MAE, and R^2 across models, were generated to facilitate model comparison and interpretability.

Notably, lower MSE and MAE values for ensemble models indicate that predicted UPDRS scores closely match true patient severity, suggesting these models could support day-to-day clinical monitoring and inform individualized treatment decisions in a home-based telemonitoring context.

4. Results and Discussion

In this section, the performance of the machine learning models developed for predicting Parkinson's disease severity is presented and analysed. A comprehensive evaluation of each regression algorithm is conducted, with attention to error trends, model generalization capability, and suitability for real-time telemonitoring applications. All models were assessed using the same standardized performance metrics, namely MSE, MAE, and R^2, computed on both the training and testing sets. The complete numerical results for all algorithms including linear regressions, tree-based methods, boosting models, and ensemble techniques, are summarized in Table 1.

Collectively, these metrics provide a rigorous and balanced basis for comparing how effectively each model captures disease-related patterns and predicts patient severity scores.

Table 1. *Performance of Machine Learning Models for PD Severity Prediction.*

Model	Train			Test		
	MSE	MAE	R^2	MSE	MAE	R^2
Linear Regression	55.76	6.31	0.15	55.72	6.37	0.16
Ridge Regression	55.76	6.31	0.15	55.72	6.37	0.16
Lasso Regression	61.88	6.76	0.06	61.65	6.80	0.07
ElasticNet	60.39	6.67	0.08	60.41	6.73	0.09
Decision Tree	0.00	0.00	1.00	3.63	0.54	0.95
Random Forest	0.27	0.25	1.00	2.05	0.64	0.97
Gradient Boosting	13.89	2.95	0.79	15.90	3.16	0.76
AdaBoost	39.84	5.66	0.40	41.31	5.77	0.38
XGBoost	0.20	0.31	1.00	2.37	1.04	0.96
CatBoost	0.89	0.69	0.99	2.42	1.06	0.96
LightGBM	1.19	0.77	0.98	2.53	1.05	0.96
VotingEnsemble	13.53	3.17	0.79	15.03	3.35	0.77

Table 1 summarizes the performance of eleven machine-learning models for Parkinson's disease (PD) severity prediction, evaluated using mean squared error (MSE), mean absolute error (MAE), and the coefficient of determination (R^2) on both the training and test sets. The results highlight substantial differences in predictive capability across algorithms.

Linear, Ridge, and Lasso Regression models show limited predictive power, with high errors and low

R^2 scores, failing to capture nonlinear relationships. ElasticNet behaves similarly, confirming linear models are insufficient. Tree-based ensembles, including Random Forest, XGBoost, CatBoost, and LightGBM, achieve very low training errors and R^2 near 1.0, modeling complex interactions effectively. These models generalize well to the test set, with R^2 between 0.95 and 0.97. Random Forest attains the lowest test MSE (2.05), while XGBoost provides consistently strong performance. Overall, ensembles outperform linear models in both accuracy and stability.

Gradient Boosting and the Voting Ensemble achieve moderate performance, with R^2 values around 0.76–0.79, outperforming linear models but remaining below advanced boosted tree methods. In contrast, Decision Tree and AdaBoost show clear overfitting. The Decision Tree attains perfect training performance but drops to an R^2 of 0.95 on the test set, while AdaBoost exhibits similarly large train–test discrepancies, indicating limited generalization. Overall, the results confirm that nonlinear ensemble learning methods particularly Random Forest, XGBoost, CatBoost, and LightGBM, are best suited for PD severity prediction, offering superior predictive accuracy and robust generalization across evaluation metrics.

Six visualizations were generated, with the first being a scatter plot comparing true and predicted values across all models. Points close to the diagonal indicate higher accuracy, where ensemble models (Random Forest, XGBoost, CatBoost, LightGBM) clearly perform best. Linear models (Linear, Ridge, Lasso, ElasticNet) display wider dispersion (especially at higher values) highlighting their difficulty in capturing nonlinear patterns. The Decision Tree shows the most scatter, reflecting typical overfitting behavior.

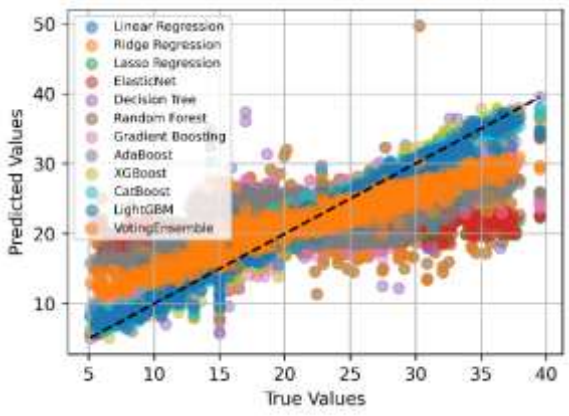

Figure 6. *True vs Predicted values of all models.*

The error distribution plot reveals how prediction errors are spread for each model. Here, the ensemble methods like XGBoost, LightGBM, and CatBoost display narrow, centralized distributions centered around zero, indicating consistent and accurate predictions. Conversely, linear regressions and AdaBoost show wider error distributions, reflecting less reliability. The Decision Tree model again stands out for the wrong reasons, displaying long tails in its error distribution that further corroborate its overfitting.

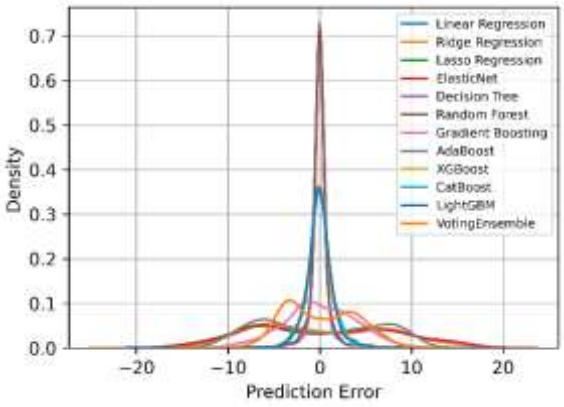

Figure 7. *Distribution of prediction errors.*

When comparing Mean Squared Error (MSE), a key metric for regression performance, the results confirm the trends observed. Random Forest, XGBoost, CatBoost, and LightGBM achieve the lowest MSE values, solidifying their status as the top performers. Linear regressions and AdaBoost, however, record high MSEs, which reflects their poor fit to the dataset's nonlinear patterns. The Decision Tree model also suffers from a higher MSE, a consequence of its high variance.

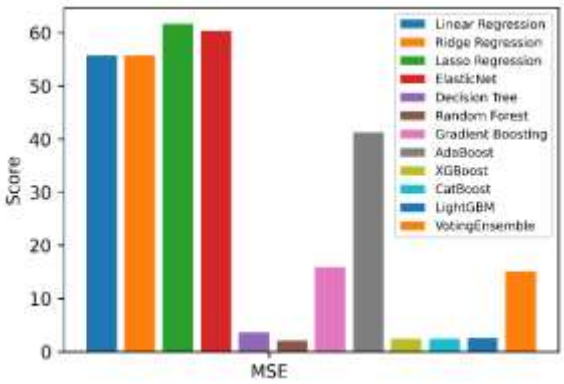

Figure 8. *Comparison of MAE values across different models.*

The analysis of Mean Absolute Error (MAE) provides another perspective on prediction accuracy and tells a similar story. The ensemble models maintain the lowest MAE, meaning their predictions have the smallest average deviation from the actual values. The linear models and AdaBoost are again

less accurate, with consistently larger average errors in their forecasts.

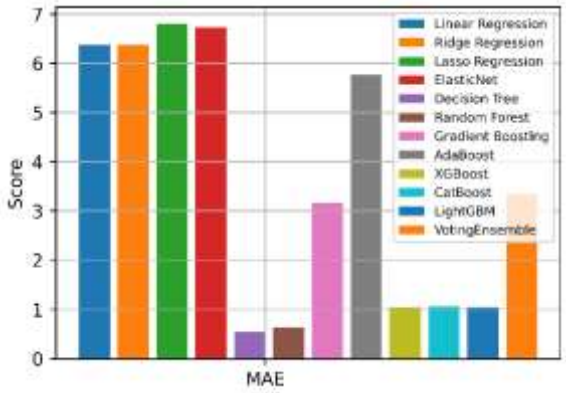

Figure 9. *Comparison of MAE values across different models.*

The R^2 comparison confirms the superiority of ensemble models, which achieve scores around 0.9 and capture most of the variance in PD severity. Linear regressions perform poorly, with R^2 below 0.2, failing to model the complex nonlinear relationships. The Decision Tree shows moderate explanatory power but remains unstable.

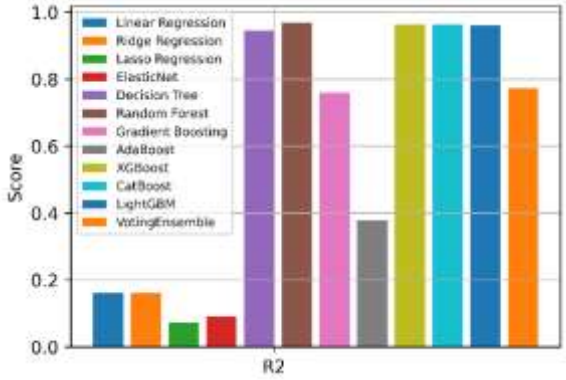

Figure 10. *Comparison of R^2 values across different models.*

Finally, the residual plot shows that ensemble methods produce residuals randomly scattered around zero, indicating unbiased predictions. Linear regressions and some boosting models show structured patterns, while the Decision Tree displays widely spread residuals consistent with overfitting.

Overall, ensemble models (Random Forest, XGBoost, CatBoost, and LightGBM) clearly outperform all others, achieving high R^2, low MSE and MAE values, and well-behaved residuals. Linear models and AdaBoost remain the weakest performers, unable to capture the complex nonlinear patterns present in the Parkinson's disease telemonitoring data.

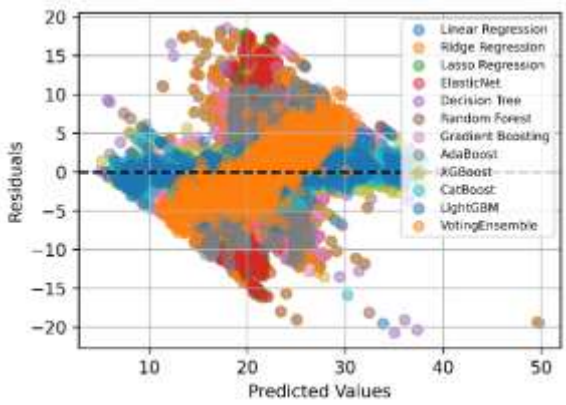

Figure 11. *Residuals of the model predictions.*

5. Conclusion

The study confirms that machine learning's ability to analyze multimodal data (speech, handwriting, and wearables) offers an innovative avenue for precise and continuous remote tele-monitoring of patients suffering from Parkinson's Disease (PD), while also providing useful insights into how these evolving PD pathologies can be modeled. The nonlinear ensemble models (Random Forest, XGBoost, CatBoost, and LightGBM) exhibited superior results compared to their linear model counterparts (in this case, predictors) and demonstrated the importance of utilizing more complex and nonlinear approaches to track the overall progression of PD due to the complexity and non-linear nature of the progression of its pathophysiological conditions.

This research goes beyond validating the technology's usefulness; it shows that the future of neurology will be defined by a data-driven and personalized approach. By converting diverse electronic data into accurate and comprehensive estimates of disease severity, clinicians will be able to use the real-time tracking capabilities of such devices to assess when and where patient behaviors have positive or negative impacts on PD, and that the current constraints of temporary in-office assessments will be replaced with continuous and non-invasive methods, specifically, utilizing patients' homes for their assessments. As a result, digital health technologies have the potential to help define more effective personalized treatment protocols, create more effective therapies, and improve long-term outcomes for patients by enabling clinicians to make better, more informed decisions about each individual patient.

To translate this successful proof-of-concept into practical and clinically available products, there are several important areas for future research. First, the best-performing model was the ensemble model; however, various other models using deep learning could potentially outperform this model in terms of

robustness and precision, including using convolutional neural networks (CNNs) for spatial feature extraction from handwriting or sensor data and using recurrent neural networks (RNNs) for temporal analysis of speech and gait.

Finally, significant implementation science research must address the practical challenges of implementing these tools, including safeguarding data privacy and security, creating simple, user-friendly interfaces for both patients and clinicians, and validating these types of systems in different populations to promote equitable access to these technologies and creating specific regulatory and reimbursement modalities for their use. By advancing on these fronts, next-generation telehealth solutions can evolve from research prototypes into integral components of a modern, proactive, and patient-centric Parkinson's disease care ecosystem.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Poewe, W., Seppi, K., Tanner, C. M., Halliday, G. M., Brundin, P., Volkmann, J., ... & Lang, A. E. (2017). Parkinson disease. *Nature Reviews Disease Primers*, 3(1), 1-21.

[2] Chaudhuri, K. R., Healy, D. G., & Schapira, A. H. (2006). Non-motor symptoms of Parkinson's disease: diagnosis and management. *The Lancet Neurology*, 5(3), 235-245.

[3] Schrag, A., Jahanshahi, M., & Quinn, N. (2000). What contributes to quality of life in patients with Parkinson's disease?. *Journal of Neurology, Neurosurgery & Psychiatry*, 69(3), 308-312.

[4] Espay, A. J., Bonato, P., Nahab, F. B., Maetzler, W., Dean, J. M., Klucken, J., ... & Merola, A. (2016). Technology in Parkinson's disease: challenges and opportunities. *Movement Disorders*, 31(9), 1272-1282.

[5] Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez- Martin, P., ... & Dodel, R. (2007). Movement Disorder Society- sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS- UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders*, 23(15), 2129-2170.

[6] Heldman, D. A., Giuffrida, J. P., Chen, R., Payne, M., Mazzella, F., Duker, A. P., ... & Revilla, F. J. (2011). The modified Bradykinesia Rating Scale for Parkinson's disease: reliability and comparison with kinematic measures. *Movement Disorders*, 26(10), 1859-1863.

[7] Papapetropoulos, S., Mitsi, G., & Espay, A. J. (2017). Digital health and Parkinson's disease: a new paradigm for continuous monitoring and management. *Journal of Parkinson's Disease*, 7(s1), S49-S53.

[8] Arora, S., Baig, F., Lo, C., Barber, T. R., Lawton, M. A., Zhan, A., ... & Foltynie, T. (2019). Smartphone motor testing to distinguish idiopathic REM sleep behavior disorder, controls, and PD. *Neurology*, 93(16), e1526-e1536.

[9] Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2010). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Transactions on Biomedical Engineering*, 57(4), 884-893.

[10] Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).

[11] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.

[12] Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 59(5), 1264-1271.

[13] Adams, J. L., Dinesh, K., Snyder, C. W., Xiong, M., Tarolli, C. G., Sharma, S., ... & Dorsey, E. R. (2020). A real-world study of wearable sensors in Parkinson's disease. *NPJ Parkinson's Disease*, 6(1), 1-8.

[14] Motin, M. A., Karmakar, C. K., & Palaniswami, M. (2022). Speech-based estimation of Parkinson's disease severity using hierarchical neural networks. *Computer Methods and Programs in Biomedicine*, 220, 106810.

[15] Espay AJ, et al. Technology in Parkinson's disease: challenges and opportunities. Mov Disord. 2016.

[16] Hoff JI, et al. Accelerometric assessment of levodopa-induced dyskinesias in Parkinson's disease. Mov Disord. 2001.

[17] Salarian A, et al. Gait assessment in Parkinson's disease: toward an ambulatory system for long-term monitoring. IEEE Trans Biomed Eng. 2004.

[18] El-Gohary M, Pearson S, McNames J, Mancini M, Horak F, Mellone S, Chiari L. Continuous monitoring of turning in patients with movement disability. Sensors (Basel). 2013 Dec 27;14(1):356-69. doi: 10.3390/s140100356. PMID: 24379043; PMCID: PMC3926561.

[19] MA, Maetzler W, Aminian K, Chiari L. Technical and clinical view on ambulatory assessment in Parkinson's disease. Acta Neurol Scand. 2014 Sep;130(3):139-47. doi: 10.1111/ane.12248. Epub 2014 Apr 1. PMID: 24689772.

[20] Galna B, et al. Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. Gait Posture. 2014.

[21] Kidziński Ł, et al. Deep neural networks enable quantitative movement analysis using single-camera videos. Nat Commun. 2020.

[22] Xu, Zhaoyi et al. "Simultaneous Monitoring of Multiple People's Vital Sign Leveraging a Single Phased-MIMO Radar." IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology 6 (2021): 311-320.

[23] C. R. Dhivyaa, K. Nithya and S. Anbukkarasi, "Enhancing Parkinson's Disease Detection and Diagnosis: A Survey of Integrative Approaches Across Diverse Modalities," in *IEEE Access*, vol. 12, pp. 158999-159024, 2024, doi: 10.1109/ACCESS.2024.3487001

[24] Rochester L, et al. A roadmap to inform development, validation and approval of digital mobility outcomes: the Mobilise-D approach. Digit Biomark. 2020.

[25] Brabenec L, Mekyska J, Galaz Z, Rektorova I. Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. J Neural Transm (Vienna). 2017 Mar;124(3):303-334. doi: 10.1007/s00702-017-1676-0. Epub 2017 Jan 18. PMID: 28101650.

[26] Drotár P, et al. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. Artif Intell Med. 2016.

[27] Tsanas A, et al. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. IEEE Trans Biomed Eng. 2010.

[28] Alshammri R, Alharbi G, Alharbi E, Almubark I. Machine learning approaches to identify Parkinson's disease using voice signal features. Front Artif Intell. 2023 Mar 28;6:1084001. doi: 10.3389/frai.2023.1084001. PMID: 37056913; PMCID: PMC10086231.

[29] Hammerla NY, et al. Deep, convolutional, and recurrent models for human activity recognition using wearables. IJCAI. 2016.

[30] Bot BM, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data. 2016.

[31] Lipsmeier F, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord. 2018.

[32] Faheem MSB, Haroon-Ul-Rasheed M, Aftab R, Faheem Q, Feroze F, Ain HQU, Samadi S. The Next Frontier in Parkinson's Disease Treatment: A Narrative Review of Innovative Neurosurgical and Gene Therapy Approaches. Health Sci Rep. 2025 Aug 27;8(9):e71197. doi: 10.1002/hsr2.71197. PMID: 40895248; PMCID: PMC12391019.

[33] Coravos A, et al. Digital Medicine: A Primer on Measurement. Digit Biomark. 2019.

[34] Hall AM, Allgar V, Carroll CB, Meinert E. Digital health technologies and self-efficacy in Parkinson's: a scoping review. BMJ Open. 2025 Jan 21;15(1):e088616. doi: 10.1136/bmjopen-2024-088616. PMID: 39842912; PMCID: PMC11784209.

[35] Dobkin RD, Amondikar N, Kopil C, Caspell-Garcia C, Brown E, Chahine LM, Marras C, Dahodwala N, Mantri S, Standaert DG, Dean M, Shoulson I, Marek K, Katz A, Korell M, Riley L, Tanner CM; Fox Insight Study. Innovative Recruitment Strategies to Increase Diversity of Participation in Parkinson's Disease Research: The Fox Insight Cohort Experience. J Parkinsons Dis. 2020;10(2):665-675. doi: 10.3233/JPD-191901. PMID: 32250321; PMCID: PMC7242847.

[36] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324

[37] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794. doi: 10.1145/2939672.2939785

[38] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in Advances in Neural Information Processing Systems (NeurIPS), 2018, vol. 31. [Online]. Available: https://github.com/catboost/catboost

[39] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems (NeurIPS), 2017, vol. 30. Available: https://github.com/Microsoft/LightGBM

[40] J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, no. 1, pp. 81–106, 1986. doi: 10.1007/BF00116251

[41] T. G. Dietterich, "Ensemble methods in machine learning," in Proc. Int. Workshop Multiple Classifier Systems, 2000, pp. 1–15, Springer. doi: 10.1007/3-540-45014-9_1

[42] R. Polikar, "Ensemble based systems in decision making," IEEE Circuits and Systems Magazine, vol. 6, no. 3, pp. 21–45, 2006. doi: 10.1109/MCAS.2006.1688199

[43] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 2002. doi: 10.1109/5.726791

[44] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, no. 6088, pp. 533–536, 1986. doi: 10.1038/323533a0

[45] A.Tsanas, & M.Little, Parkinsons Telemonitoring Dataset. UCI Machine Learning Repository, 2009. doi: 10.24432/C5ZS3N