# Simulated Feature Fusion for Parkinson's Disease Prediction Using Non-aligned Public Handwriting and Voice Datasets

**Rashmi Kariganurun Bheemarao*, Shobana Thippeswamy Sharadha, Madiraju Snigdha, Neichelle Alvina Mendonca, Pranamya Prakash Rayakar, Muskan Gupta & Panchangam Sai Manogna**

B.M.S. College of Engineering
P.O. Box No.: 1908, Bull Temple Road, Bengaluru-560019, KA, India
*E-mail: rashmikb.ise@bmsce.ac.in

**Abstract.** Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects over 10 million people worldwide, with prevalence increasing due to aging populations. Early symptoms such as tremor, micrographia, bradykinesia, and voice changes often remain undetected until substantial neuronal loss has occurred, leading to irreversible damage and increased treatment costs. This paper presents a comparative study of five machine learning models—Logistic Regression, Random Forest, Support Vector Machine (RBF Kernel), K-Nearest Neighbors, and XGBoost—for Parkinson's disease prediction using handwriting and voice-based features. Performance was evaluated using accuracy, precision, recall, F1-score, AUC, 5-fold cross-validation, ROC curve analysis, feature importance analysis, and SHAP-based explainability.XGBoost achieved the highest test-set F1-score (0.7407) and AUC (0.8329), while SVM achieved the highest average cross-validation F1-score. Decision-level fusion further improved recall (0.88) while maintaining competitive overall performance. K-Nearest Neighbors produced the lowest overall performance, while XGBoost achieved the best balance between classification accuracy, F1-score, and AUC.

**Keywords:** *feature-level fusion; handwriting analysis; machine learning; multimodal fusion; non-aligned datasets; Parkinson's disease; simulated alignment; SMOTE; voice analysis; XGBoost.*

## 1 Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder, which is due to the gradual death of dopamine-producing brain cells in the substantia nigra [1-3] It has an impact on over 10 million individuals all over the world today and this number continues to grow. It rises with the increasing age of the population [4, 5]. The initial symptoms–a slight tremor at rest, smaller handwriting (micrographia), slower movements, stiff posture, and softer, hoarse voice—generally do not appear until after 60–80% of these neurons are lost [2, 3, 6 ]. Most of the patients are diagnosed when the damage cannot be reversed and is

Received on December 3rd, 2025, Revised on June 19th, Accepted for publication on July 15th, 2026.
Copyright © 2026 Published by IRCS-ITB, ISSN: 2337-5787, DOI: 10.5614/itbj.ict.res.appl.2026.20.3.2

already too extensive. Treatment is significantly more costly. Research indicates that PD diagnosis even a few years earlier is likely to reduce long-term healthcare expenditure by up to 35% per patient [7].

It is due to this that researchers have been seeking cheap, non-surgical methods of detecting the disease at the earliest stage. Parkinson's disease affects handwriting [1,8], voice characteristics [2], gait patterns [9], and facial movements [10], making these signals useful early indicators for detection [11].

Even with a single modality, accuracies of 80–99% have been reported using both classical machine learning and deep learning methods [1, 2, 12]. Performance often improves further when multiple modalities are collected from the same individuals, with several multimodal studies reporting accuracies above 90% [13–15]. However, most publicly available Parkinson's disease datasets are unimodal and contain different subjects, making multimodal research difficult without dedicated clinical data collection.

The proposed methodology addresses this limitation by investigating whether useful multimodal information can be obtained from independent public handwriting and voice datasets through simulated fusion using simple machine learning techniques.

**Novelty** The proposed work introduces a lightweight and explainable multimodal framework for Parkinson's disease prediction using non-aligned public handwriting and voice datasets. Unlike existing studies that relied on subject-aligned multimodal data and computationally intensive deep learning architectures, the proposed approach employs simulated feature fusion and an interpretable XGBoost classifier. In addition, SHAP-based explainability analysis is incorporated to identify the most influential multimodal biomarkers contributing to prediction. The framework used publicly available datasets to train the model on multiple types of data, while keeping the system transparent, reproducible, and computationally efficient

## 2 Related Work

### 2.1 Unimodal PD Detection

Systems that are handwriting-based acquire frequency, radius deviation, and pen pressure. Pereira *et al*. [1] applied CNNs on spirals image analysis and achieved accuracy equal to 92%. In the case of voice-based techniques, acoustic variations have been considered, such as pitch and amplitude variations. Tsanas *et al*. [2] applied sustained vowel recordings to determine Parkinson's disease using the SVM classifier, achieving an accuracy of 99%.

**Table 1** Comparison with other multimodal studies.

| Study | Modalities | Subjects | Acc (%) |
|---|---|---|---|
| Meghashree [13] | Voice + gait + hand | 1000 | 96.5 |
| Benredjem [14] | Hand + clinical | 150 | 96.0 |
| Babui [15] | Voice + hand | 804(samples) | 90.0 |
| **Proposed methodology** | Voice + hand (sim.) | 264 (sim.) | 73.58 |

## 2.2 Multimodal PD Detection

Recent research on multimodal detection of PD has indicated that the fusion of the various types of data can be utilized to create a more reliable detector of the disease. Classification performance typically improves when modalities come from the same subjects. As mentioned in Table 1, Meghashree et al. (2025) obtained 96.5% accuracy by combining voice, gait, and handwriting using a Transformer-SSL framework on 1,000 participants [13]. Benredjem et al. (2024) achieved 96% accuracy by fusing handwriting features with clinical scores using attention-based models [14]. Babui et al. used handwriting and voice data for Parkinson's disease prediction and achieved 90% accuracy using an ensemble classifier [15]. Unlike references [13] and [14], the present study used non-aligned public datasets and evaluated both feature-level and decision-level fusion approaches.

# 3 Method

## 3.1 Datasets and Features

Two publicly available datasets representing handwriting and voice modalities were used in this study [16, 17]. The HandPD New Spiral dataset [17] (https://wwwp.fc.unesp.br/˜papa/pub/datasets/Handpd/) contains 264 spiral drawings of 63 subjects (30 PD, 33 healthy), from which 9 features were derived, namely, RMS error, Mean Relative Tremor (MRT), standard deviation of height (STDHT), radius, velocity and acceleration.

The UCI Parkinson's Disease Voice Dataset [16] (https://archive.ics.uci.edu/dataset/174/parkinsons) contains 195 voice samples of 31 participants (23 PD, 8 healthy) and 19 acoustic features such as jitter percent, jitter absolute, RAP, PPQ5, shimmer. The following parameters are included: shimmer (dB), APQ3, APQ5, HNR, RPDE, DFA, and PPE.

## 3.2 Preprocessing

Data preprocessing was used to guarantee data quality and compatibility. Noisy data were removed and noise was reduced by applying smoothing. All features were standardized using Eq. (1), to normalize modalities. Stratified sampling was

then used to split the dataset into a training and a testing set (80% and 20%, respectively), preserving the class distribution. Dataset splitting was performed before feature standardization. Scaling parameters were learned exclusively from the training data and subsequently applied to the testing data to prevent data leakage.

$$Z = \frac{x-\mu}{\sigma} \tag{1}$$

Single-modality baselines were created to measure unimodal performance. XGBoost was used for both the voice and the handwriting datasets [18] with 200 estimators, a maximum depth of 3, a learning rate of 0.1, and a subsample ratio of 0.8. Early stopping was applied after approximately 20 rounds to reduce overfitting. These models served as reference baselines for evaluating the effectiveness of feature fusion.

## 4 Methodology

The proposed system combines voice and handwriting features to recognize the presence of Parkinson's disease. The workflow is illustrated in Figure 1.

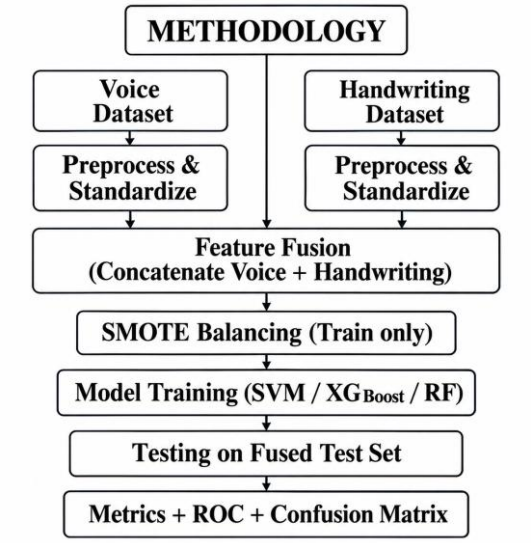



**Figure 1** Proposed flow of methodology in the Parkinson disease detection.

### 4.1 Simulated Fusion Pipeline

1. **Bootstrapping:** A random sample of 264 voice records is selected, with replacement, resizing the UCI dataset to the same size as HandPD.

2. **Concatenation:** Create 28 voice dimension vectors from 19 voice and 9 handwriting features.
3. **Balancing:** Oversample the PD class with SMOTE 1:1 with healthy controls.
4. **Training:** Train the XGBoost classifier using the same hyperparameter as the unimodal baselines. SMOTE is a minority class oversampling technique that creates synthetic minority samples by finding a point between each minority sample and its closest neighbors [19].

Table 2 compares the performances of the evaluated machine learning models on the fused dataset. Figure 2 compares the test-set F1-scores obtained by all evaluated models. XGBoost achieved the highest test-set F1-score. Figure 3 presents the distribution of 5-fold cross-validation F1-scores across all models. SVM achieved the highest average cross-validation performance, while XGBoost showed the lowest variance.

**Table 2** Comparison of machine learning models on fused handwritten–voice data based on test-set F1-score, AUC, 5-fold cross-validation F1-score (mean ± SD), and model stability.

| Model | Test F1 | AUC | 5-fold CV F1 | Stability |
|---|---|---|---|---|
| XGBoost | 0.741 | 0.833 | 0.670 ± 0.044 | High |
| Random Forest | 0.731 | 0.824 | 0.646 ± 0.063 | Moderate |
| SVM (RBF Kernel) | 0.714 | 0.799 | 0.675 ± 0.064 | Moderate |
| Logistic Regression | 0.679 | 0.797 | 0.670 ± 0.095 | Moderate |
| K-Nearest Neighbors | 0.698 | 0.716 | 0.604 ± 0.085 | Low |

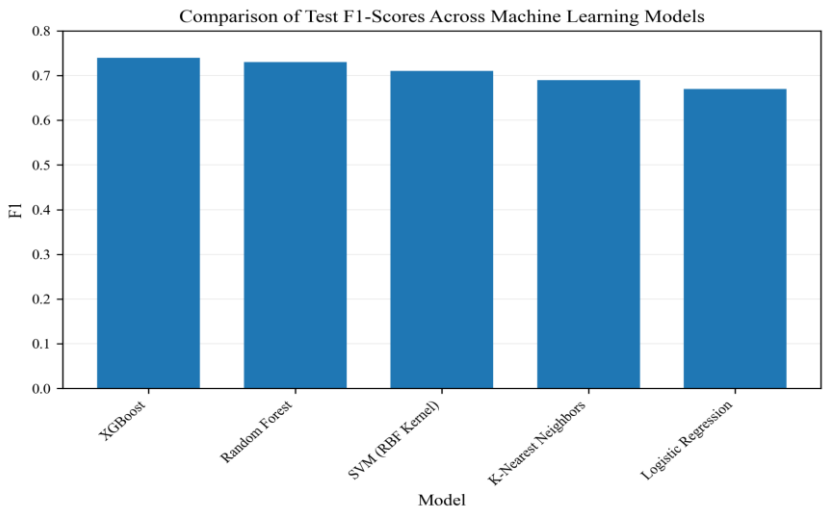



**Figure 2** Comparison of test-set F1-scores across machine learning models.

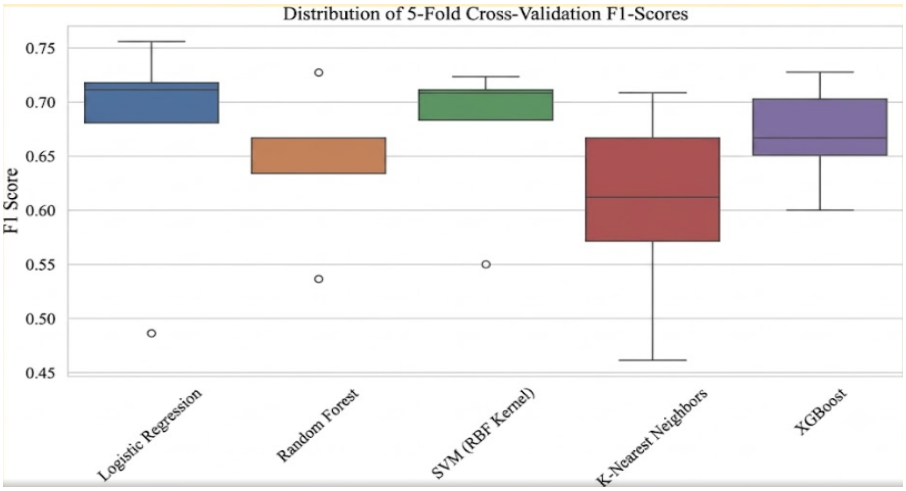



**Figure 3** Comparison of 5-fold cross-validation F1-scores across models.

# 5 Results

## 5.1 Model Performance

Table 3 presents the 5-fold cross-validation and test-set results of all evaluated models. SVM achieved the highest average cross-validation F1-score, while XGBoost showed the lowest variance across folds. On the testing set, XGBoost achieved the highest F1-score and AUC, indicating the best overall performance on the fused dataset. Figure 4 summarizes the ROC results. XGBoost and Random Forest achieved the highest AUC values, indicating better discrimination between Parkinson's disease and healthy subjects.

**Table 3** Performance metrics for unimodal and fused models.

| Model | Acc (%) | Prec | Rec | F1 |
|---|---|---|---|---|
| Voice only | 92.31 | 0.93 | 0.965 | 0.95 |
| Handwriting | 73.58 | 0.70 | 0.76 | 0.73 |
| **Fused** | 73.58 | 0.69 | 0.80 | 0.74 |

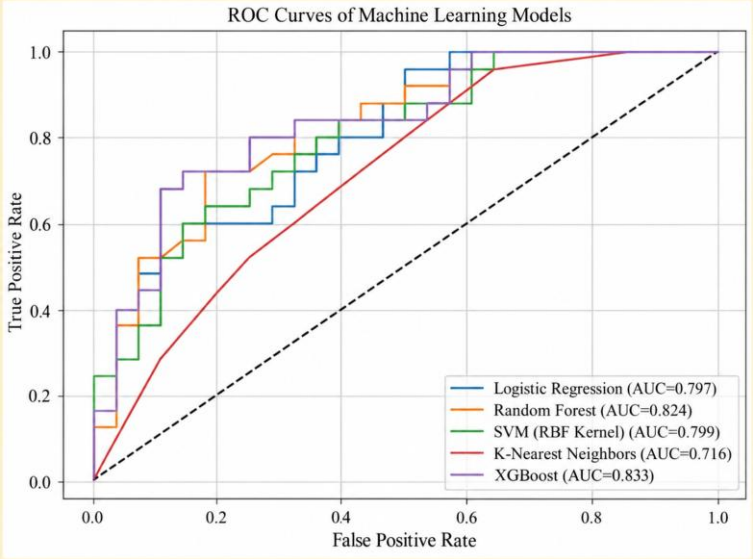



**Figure 4** ROC curves for all models. Random Forest and XGBoost achieve the highest AUC values.

## 5.2 Feature Importance Analysis

Figure 5 shows the top-15 features ranked by XGBoost importance. Handwriting-derived features such as STD HT, MRT, MAX HT and ET–HT were among the most influential predictors, while several voice features also contributed to classification.

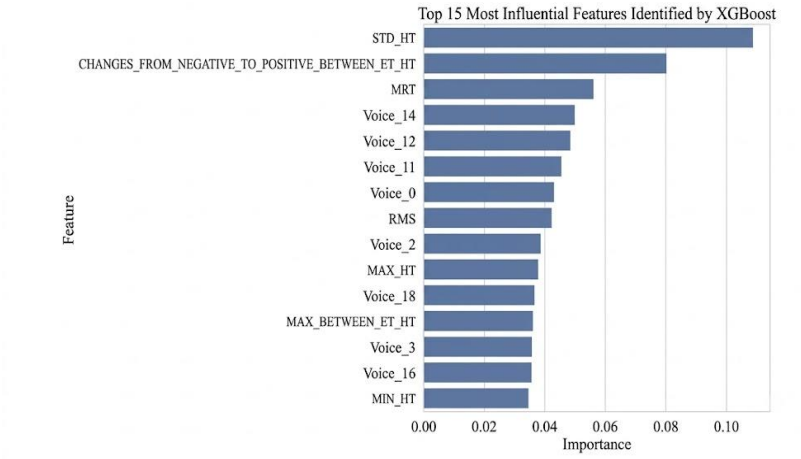



**Figure 5** Top 15 features ranked by XGBoost importance.

### 5.3 SHAP-based Explainability Analysis

SHAP analysis was performed on the best-performing model, XGBoost [20]. Figure 6 shows that handwriting-related features had the strongest influence on predictions, while voice features also contributed to the final decision. Features such as STD HT, MRT, MAX HT, and ET–HT exhibited the highest impact on model output, indicating that handwriting-derived motor characteristics were important indicators of Parkinson's disease. The results also confirm that both handwriting and voice modalities contributed to classification, supporting the effectiveness of the proposed multimodal approach.

### 5.4 Decision-level Fusion Analysis

Besides feature-level fusion, another fusion method, namely, decision-level fusion was employed in our study. Decision-level fusion combines predictions generated independently from multiple modalities and is widely used in multimodal machine learning systems to improve robustness and complementarity between data sources [21, 22]. For the handwriting and voice datasets, two separate XGBoost classifiers were trained individually. The prediction probabilities obtained from these two classifiers were used for score averaging to get the final predictions. Interpretable learning-based models have also been applied to Parkinson's disease diagnosis from voice-related EEG signals, showing that explainability can be combined with multimodal biomedical data [23].

With the implementation of decision-level fusion, accuracy, precision, recall, F1-score, and AUC values of 69.8%, 62.8%, 88.0%, 0.733, and 0.848, respectively, were produced, as shown in Table 4. Decision-level fusion gave better recall scores than feature-level fusion and had a relatively higher AUC, but feature-level fusion had higher accuracy and F1-score values than decision-level fusion. As can be seen in Figure 7 decision-level fusion achieved higher recall and AUC, while feature-level fusion achieved higher accuracy and F1 scores.

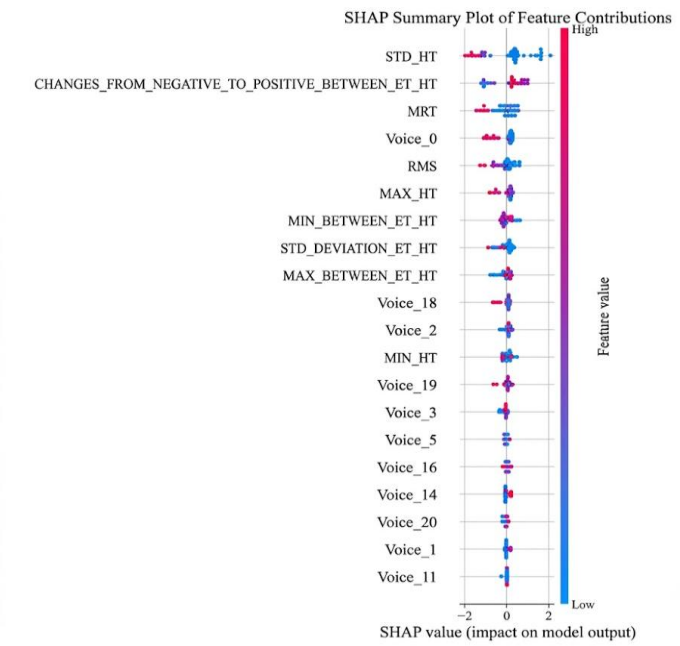



**Figure 6** SHAP summary plot illustrating the contribution of the most influential multimodal features toward Parkinson's disease prediction. Features positioned farther from zero have a stronger impact on model decisions.

## 6 Discussion

The fused model performed with an accuracy of 73.58%, which was similar to the handwriting-only model. The fusion approach did not improve the overall accuracy, but it did achieve a better trade-off between precision, recall and F1 score by combining complementary information from hand-writing and voice modality. However, the performance was still below that reported in studies employing subject-aligned multimodal data [13-15]. The difficulties of bringing together non-aligned public data sets are high-lighted. The performance of the unimodal and fused models is summarized in Figure 9 and Table 3. Figure 8 shows the confusion matrices of the voice-only, handwriting-only, and fused models. The voice-only model achieved the highest accuracy, while the fused model improved recall and F1-score over the handwriting-only model.

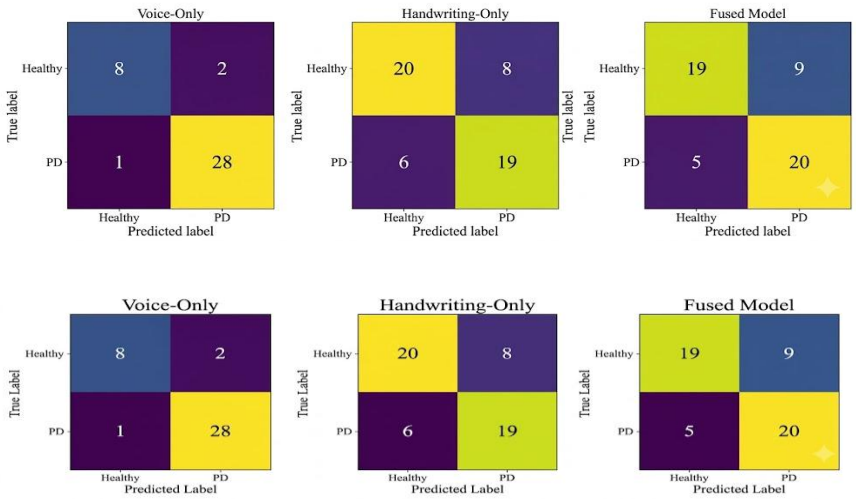



**Figure 7** Confusion matrices for voice-only, handwriting-only, and fused models.

**Main limitation** The handwriting and voice datasets used in this study were collected from different individuals, with no subject overlap between modalities. As a result, voice and handwriting samples were paired randomly for fusion. This may have weakened the natural relationship between the two signals and affected performance. Therefore, the proposed approach is limited primarily by the availability of non-aligned public datasets. Future studies could also draw on low-cost, passively collected signals, such as computer keyboard typing dynamics, which have been shown to separate early Parkinson's disease patients from controls [24], to build larger aligned multimodal datasets.

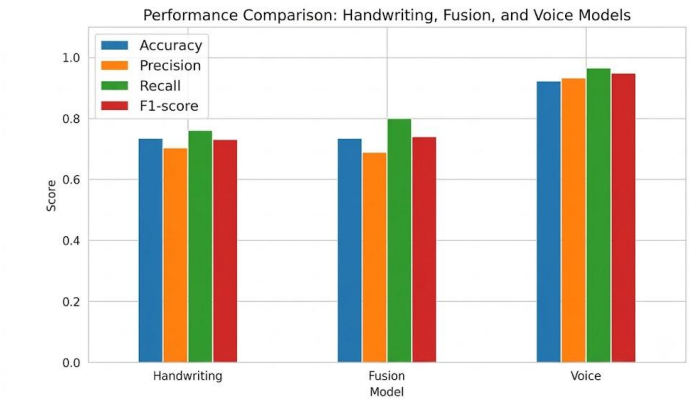



**Figure 8** Performance comparison of unimodal and fused models.

## 7 Conclusion

This paper presented a simulated multimodal framework for Parkinson's disease prediction using non-aligned public handwriting and voice datasets. The proposed feature-level fusion approach achieved an accuracy of 73.58%, an F1-score of 0.7407, and an AUC of 0.8329, demonstrating that useful multimodal representations can be obtained even when subject-level alignment is unavailable. SHAP analysis showed that both handwriting and voice features contributed to classification decisions, while the comparison of feature-level and decision-level fusion highlighted the trade-off between overall performance and sensitivity. The proposed framework provides a reproducible and cost-effective approach for multimodal Parkinson's disease prediction using publicly available datasets. The results support growing evidence that multimodal machine learning can effectively integrate complementary information from different data sources for disease diagnosis and prognosis [18, 20-22]. Beyond motor symptoms, machine learning has also been reviewed for detecting cognitive impairment in Parkinson's disease, indicating further scope for extending multimodal frameworks [25]. Future work will focus on collecting matched handwriting and voice data from the same subjects and incorporating additional modalities to improve multimodal prediction performance.

## Acknowledgement

The authors acknowledge the open data provided by the UCI Parkinson's dataset team and the HandPD NewSpiral dataset contributors.

## References

[1] Pereira, C.R., Pereira, D.R., Rosa, G.H., Albuquerque, V.H.C., Weber, S.A.T., Hook, C. & Papa, J.P., *Handwritten Dynamics Assessment Through Convolutional Neural Networks: An Application to Parkinson's Disease Identification*, Artif. Intell. Med., **87**, pp. 67-77, 2018.

[2] Tsanas, A., Little, M.A., McSharry, P. E., Spielman, J. & Ramig, L.O., *Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease*, IEEE Trans. Biomed. Eng., **59**(5), pp. 1264-1271, 2012 .

[3] Zhou, Z.D., Yi, L.X., Wang, D.Q., Lim, T.M. & Tan., E.K., *Role l Dopamine in the Pathophysiology of Parkinson's Disease*, Transl. Neurodegener., **12**(1), 44, 2023. doi: 10.1186/s40035-023-00378-6.

[4] Dorsey, E., Elbaz, A. & Nichols, E., *Global, Regional, and National Burden of Parkinson's Disease, 1990–2016*, Lancet Neurol., **17**(11), pp. 939-953, 2018.

[5] Bloem, B.R., Okun, M.S. & Klein, C., Parkinson's Disease, Lancet, **397**(10291), pp. 2284-2303, 2021. doi: 10.1016/S0140-6736(21)00218-X.

[6] Schapira, A.H.V., *Recent Developments in Biomarkers in Parkinson Disease,* Curr. Opin. Neurol, **28**(4), pp. 374-379, 2015.

[7] Findley, L.J., *The Economic Impact of Parkinson's Disease. Parkinsonism* Relat Disord., **13**, S8-S12, 2007. doi: 10.1016/j.parkreldis.2007.06.003.

[8] Thomas, M., Lenka, A. & Pal, P.K., *Handwriting Analysis in Parkinson's Disease: Current Status and Future Directions,* Mov Disord Clin Pract, **4**, pp. 806-818, 2017. doi: 10.1002/mdc3.12552

[9] Pham, T.D., *Texture Classification and Visualization of Time Series of Gait Dynamics*, Phys. Rev. E, **97**(5), 053301, 2018.

[10] Bandini, A., Orlandi, S., Escalante, H.J., Giovannelli, F., Cincotta, M., Reyes-Garcia, C.A., Vanni, P., Zaccara, G. & Manfredi, C.. *Analysis of Facial Expressions in Parkinson's Disease Through Video-Based Automatic Methods*, J Neurosci Methods, **281**, pp. 7-20, 2017. doi: 10.1016/j.jneumeth.2017.02.006.

[11] Islam, M.A., Majumder, M.Z.H., Hussein, M.A., Hossain, K.M., Miah, M.S., *A Review of Machine Learning and Deep Learning Algorithms for Parkinson's Disease Detection Using Handwriting and Voice Datasets*, Heliyon, **10**(3), e25469, 2024. doi: 10.1016/j.heliyon.2024.e25469

[12] Alshammri, R., Alharbi, G., Alharbi, E. & Almubark, I., *Machine Learning Approaches to Identify Parkinson's Disease Using Voice Signal Features*, Front. Artif. Intell., **6**,1084001, 2023. doi: 10.3389/frai.2023.1084001

[13] Meghashree, M.B., Kumar, K.D., Vinutha, N. & Akash, D., *The Multimodal Fusion of Voice, Gait, and Handwriting Detection of Parkinson's Disease Using Machine Learning*, Int. Res. J. Multidiscip. Scope, **6**(4), pp. 1062-1074, 2025. doi: 10.47857/irjms.2025.v06i04.06552.

[14] Benredjem, S., Mekhaznia, T., Rawad, A., Turaev, S., Bennour, A., Sofiane, B., Aborujilah, A. & Al Sarem, M., *Parkinson's Disease Prediction: An Attention-Based Multimodal Fusion Framework Using Handwriting and Clinical Data*, Diagnostics, **15**(1), 4, 2025. doi: 10.3390/diagnostics15010004 .

[15] Babui, Y.A., Yamini, Ch., Narendra, G. & Ali, Sk.B., *Parkinson's Disease Prediction using Handwritten and Voice Dataset*, in Proceedings of the International Conference on Computer Science and Communication Engineering (ICCSCE 2025), Adv. Comput. Sci. Res., **124**, pp. 1047-1061, Atlantis Press, 2025. doi: 10.2991/978-94-6463-858-5_87.

[16] Little, M., *Parkinsons,* UCI Machine Learning Repository, 2007.

[17] Pereira, C.R., Pereira, D.R., Silva, F.A., Masieiro, J.P., Weber, S.A.T., Hook, C. & Papa, J.P, *HandPD NewSpiral Dataset*, 2018. https: //wwwp.fc.unesp.br/˜papa/pub/datasets/Handpd/

[18] Chen, T. & Guestrin, C., *XGBoost: A Scalable Tree Boosting System,* in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, USA, pp. 785-794, Aug, 2016.

[19] Chawla, N.V., Bowyer, K W., Hall, L.O. & Kegelmeyer, W.P., *SMOTE: Synthetic Minority Over-Sampling Technique,* J. Artif. Intell. Res., **16**, pp. 321-357, 2002.

[20] Lundberg S.M. & Lee, S.-I., *A Unified Approach to Interpreting Model Predictions,* Advances in Neural Information Processing Systems, **30**, pp. 4765-4774, 2017.

[21] Cui, C., Yang, H., Wang, Y., Zhao, S., Asad, Z., Coburn, L.A. & Wilson, K.T., *Deep Multimodal Fusion of Image and Non-Image Data in Disease Diagnosis And Prognosis: A Review*, npj Digit. Med., **6**(1), 92, 2023.

[22] Baltrusaitis, T., Ahuja, C. & Morency, L.-P., *Multimodal Machine Learning: A Survey and Taxonomy,* IEEE Trans. Pattern Anal. Mach. Intell., **41**(2), pp. 423-443, Feb. 2019.

[23] Zhao, S., Dai, G., Li, J., Zhu, X., Huang, X., Li, Y., Tan, M., Wang, L., Fang, P., Chen, X., Yan, N. & Liu, H., *An Interpretable Model Based on Graph Learning for Diagnosis of Parkinson's Disease with Voice-related EEG*, npj Digit. Med., **7**(1), 3, 2024. doi: 10.1038/s41746-023-00983-9.

[24] Giancardo, L., Sánchez-Ferro, A., Arroyo-Gallego, T., Butterworth, I., Mendoza, C.S., Montero, P., Matarazzo, M., Obeso, J.A., Gray, M.L. & San José Estépar, R., *Computer Keyboard Interaction as an Indicator of Early Parkinson's Disease*, Sci. Rep., **6**, 34468, 2016. doi: 10.1038/srep34468.

[25] Altham, C., Zhang, H. & Pereira, E., *Machine Learning for the Detection and Diagnosis of Cognitive Impairment in Parkinson's Disease: A Systematic Review*, PLOS ONE, **19**(5), e0303644, 2024. doi: 10.1371/journal.pone.0303644.