Enhanced Multimodal approach for Parkinson's Disease Detection : fusing deep handwriting and Voice Features with Optimized Classification

Moez Mathlouthi* Imen Hamrouni Trimech † Najoua Essoukri Ben Amara ‡
* University of Charthage, The Higher Institute of Information Technologies and Communication, 1164, Ben Arous, Tunisie;
† University of Charthage, The Higher Institute of Information Technologies and Communication
LATIS- Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisie;
‡ University of Sousse, Ecole Nationale d'Ingénieurs de Sousse,
LATIS- Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisie;

***Abstract*—Parkinson's disease (PD) is a progressive neurodegenerative disorder that leads to motor impairments. Early detection is crucial for effective treatment; however, conventional diagnostic methods are often costly, time-consuming, and inaccessible, limiting their widespread clinical adoption. To address these challenges, we propose a novel hybrid approach that merges Deep Learning (DL) features extrated from handwriting images with voice characteristics using an optimized machine learning (ML) classification technique. The integration of multimodal data enhances robustness by reducing dependency on a single biomarker, making PD diagnosis more reliable. We begin by augmenting both of the datasets size to expand samples diversity. Then, we combine DenseNet201's detailed features with ResNet50's robust spatial features to enhance analytical precision and capture both fine-grained and high-level patterns of handwriting images. The obtained DL features are fused with voice characteristics leveraging complementary information. Afterwards, feature selection is performed using Fisher's score to retain the most relevant attributes, further boosting classification accuracy. We achieve an accuracy of 92.31%, demonstrating superior performance compared to state-of-the-art methods.**

***Index Terms*—Parkinson disease detection, Deep features, Feature fusion, Fisher score, Extratrees.**

I. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder that primarily affects the central nervous system, specifically targeting the dopaminergic neurons in the midbrain [1]. As the disease progresses, the loss of dopamine-producing neurons disrupts normal brain function, leading to a range of motor and non-motor symptoms.

Non-motor symptoms often appear in the early stages of PD, preceding the onset of motor impairments. These symptoms, which include loss of smell, memory decline and sleep disturbances [2], can emerge years before motor symptoms. However, their use for diagnostic purposes is challenging due to their overlap with other medical conditions and variability among individuals. Motor symptoms, on the other hand, typically develop later and are more characteristic of PD. The primary motor impairments include bradykinesia (slowed movement), muscle rigidity, and resting tremors.These symptoms are critical for clinical diagnosis, as neurologists rely heavily on them to evaluate disease progression.

Early diagnosis is crucial for effective intervention, as prompt treatment can slow disease progression and significantly enhance patients' quality of life. However, traditional diagnostic methods, such as clinical evaluations and neuroimaging, are often time-consuming, and not universally accessible.

Various Machine Learning (ML) and Deep Learning (DL) techniques have been consequently developed to facilitate the detection of PD by analyzing different modalities [3]. DL has revolutionized PD analysis by enabling automated feature extraction and classification. Unlike traditional ML methods that depend on handcrafted features [4], DL models autonomously learn hierarchical and intricate representations from raw data, capturing subtle patterns.

Most state of the art methods often exploit unimodal analysis for PD detection which is insufficient for accurate diagnosis. PD presents diverse motor and non-motor symptoms, which require multiple data sources, such as handwriting dynamics, and speech signals to provide a comprehensive assessment. Combining these modalities improves diagnostic accuracy by leveraging complementary information, mitigating the limitations of unimodal analysis.

To tackle these challenges, we propose a hybrid approach for PD detection that integrates extracted DL features from drawing images with voice features, using ExtraTrees for optimized classification. By integrating multimodal features, our method significantly enhances classification accuracy and improves the model's ability to differentiate PD patients from healthy individuals. This approach not only boosts overall performance but also ensures better generalization, making PD detection more reliable, adaptable, and clinically applicable.

To summarize, our key contributions are:

- Multimodal data exploitation: We integrate both handwriting images and voice data to leverage complementary information from multiple modalities. Handwriting images capture spatial and fine-grained motor impairments, while voice data provide dynamic insights into speech patterns and motor control. By combining these modalities, our approach enhances feature representation and strengthens the robustness of PD detection.

- Feature fusion: We employ transfer learning with pre-trained models, specifically DenseNet201 and ResNet50, to efficiently extract high-quality features from handwriting images. DenseNet201 captures intricate feature representations through dense connectivity, while ResNet50 uses residual connections to preserve critical spatial information. By fusing both extracted handwriting Deep features with voice characteristics, our approach enhances feature diversity and affords a more informative and discriminative representation.
- Improved efficiency : We use Fisher's score to reduce dimensionality by selecting the most relevant features, and employ the ExtraTrees algorithm combined with Grid Search Cross-Validation to optimize hyperparameters and ensure robust classification.

The remainder of this article is structured as follows: Section II provides a review of previous research on PD detection. Section III details the proposed methodology. Section IV describes the experimental setup, presents the results, and discusses key findings. Finally, Section V concludes the study with final remarks and future research directions.

II. Related Work

This section offers an in depth overview of recent research on PD detection, focusing on the application of DL techniques to handwriting-based data, voice recordings, besides multi-modal approaches.

A. PD detection using handwriting images

Handwriting plays a crucial role in PD detection, as it directly reflects fine motor control.For instance, Arasavali et al. [6] explored the use of a Deep Neural Network Multi-Layer Perceptron (DNN-MLP) model to classify individuals as either Parkinson's patients or healthy controls by analyzing handwriting dynamics. The dataset comprised features such as root mean square values, mean relative tremor, and radial deviations between reference templates and handwritten samples. The trained DNN-MLP achieved a classification accuracy of 83%. Similarly, Wang et al. [7] propose a method for Parkinson's disease classification that captures fine-grained temporal dynamics within handwriting signals. The approach involves segmenting the handwriting trajectory into fixed-length sequences and employing a compact one-dimensional hybrid neural network to extract meaningful temporal features from each segment. This strategy effectively captures localized signal variations, achieving a high diagnostic accuracy of 90.7% on the widely used PaHaW dataset. In an other work [8], the authors investigated PD detection by extracting structural features using the Histogram of Oriented Gradients technique. To enhance dataset diversity and model robustness, data augmentation was applied. A Convolutional Neural Network was then trained on the augmented feature set, achieving an accuracy of 86.67%.

B. PD detection using voice data

Voice data is essential in Parkinson's disease detection, as speech abnormalities frequently emerge early in the disease, offering a non-invasive, easily accessible, and sensitive means of identifying motor and neurological impairments. As an example, Malekroodi et al. [9] applied the VGG16 deep learning model to classify different stages of PD based on vocal acoustic features, achieving an accuracy of 91.8%. The model, known for its capacity to extract complex patterns through convolutional layers, was trained on voice recordings. Key vocal features, such as fundamental frequency, jitter, and shimmer, proved essential for accurately distinguishing between PD progression levels. This method highlights the power of DL in utilizing subtle speech variations for non-invasive and reliable PD progression assessment. Additionally, [10] introduces a hybrid system for PD detection using speech signals, employing two datasets and extracting 17 acoustic features. The top eight features are selected through a genetic algorithm, and four classifiers are tested, achieving 90% accuracy on the King's College London dataset. In an other work, [11] evaluates various ML classifiers for PD detection using speech features. The results show that the Gradient Boost model outperforms other classifiers, achieving the highest test accuracy of 91.53%.

C. PD detection using multimodal approaches

Multimodal approaches in PD detection are valuable as they combine different data types, offering a comprehensive understanding of the disease and capturing its complex symptoms. In [12], the authors introduced a robust framework that integrates DL models with interpretability techniques to enhance PD diagnosis. The framework leverages Convolutional Neural Networks for analyzing imaging data, Recurrent Neural Networks for sequential sensor data, and fully connected networks for structured clinical data, achieving a classification accuracy of 92%. Similarly, [13], the authors present a preliminary study that utilizes a Vision Transformer for analyzing handwriting data and an Audio Spectrogram Transformer for vocal data to detect PD. The study focuses on spiral and meander drawings, as well as sustained phonation of vowels /a/ and /o/, achieving a classification accuracy of 90%. Additionally, [14], investigates various speech and language representations for the automatic detection of PD patients. Initially, each modality is analyzed separately. General representations, such as Wav2Vec and BETO, are combined with disease-specific representations, such as phonemic identifiability in speech and grammatical unit analysis in language. The most effective speech and language representations are then fused using a Gated Multimodal Units strategy, achieving an accuracy of 87.3%.

III. Methodology

In this study, we propose a novel framework for PD detection that leverages multimodal data, primarily handwriting images and voice measurements, by integrating feature fusion, dimensionality reduction, and optimized classification techniques into a unified pipeline. Handwriting images capture fine motor impairments such as tremors and micrographia, while voice data reflects subtle speech abnormalities and vocal impairments associated with PD, making their combination highly complementary. Feature fusion integrates these distinct yet related modalities into a richer and more informative representation. This enhanced feature space is subsequently refined through dimensionality reduction, improving computational efficiency and reducing the risk of overfitting. Finally, an optimized classification technique is applied to the compacted feature set, enhancing decision-making accuracy and ensuring more robust PD detection. Figure 1 provides an overview of the proposed approach.

A. Data Augmentation

Data augmentation (DA) is crucial for enhancing model generalization and increasing feature diversity, particularly in scenarios involving limited or imbalanced datasets.

1) DA for handwriting images: The limited size of the exploited handwriting dataset [21], which is described in the experiments section, poses a significant challenge to the performance and generalization ability of the model. To mitigate this issue, classic DA techniques were employed. DA involves applying label-preserving transformations such as random translations, scaling, flipping, and rotations, which help increase dataset diversity and reduce overfitting. In this study, augmentation was carried out by rotating each image by 90°, 180°, and 270°, as well as by applying a vertical flip. As a result of these transformations, the original dataset of 204 images, comprising samples from both healthy individuals and patients with PD, was expanded to a total of 3264 images. This augmented dataset was subsequently used to evaluate the performance of the proposed approach.

2) DA for voice data: Since the utilized voice dataset [22] exhibits a class imbalance comprising 31 individuals, of whom 23 are diagnosed with PD, we address this issue using the Synthetic Minority Over-sampling Technique (SMOTE) [23]. Rather than simply replicating samples from the minority class, SMOTE intelligently generates new synthetic instances by interpolating between existing samples within the minority class feature space. This approach enhances the diversity of the training data, mitigates overfitting, and promotes more balanced and robust model learning, especially when distinguishing between healthy individuals and those affected by PD. In the SMOTE algorithm, each synthetic sample s is produced by linearly interpolating between a minority class instance x and one of its 5 nearest neighbors x_R, which is randomly selected. The interpolation is expressed as:

$$s = x + u \cdot (x_R - x), \tag{1}$$

where u is a random scalar such that $0 \leq u \leq 1$ and x_R is randomly selected from among the five closest neighbors of x within the minority class. .

B. Deep feature Extraction

We utilize transfer learning with two pretrained DL models, DenseNet201 [15] and ResNet50 [16], to efficiently extract high-level features from handwriting images. These models are particularly useful in the context of PD detection, as they are capable of identifying subtle patterns in handwriting that may indicate motor impairments associated with the disease.

1) DensNet201: is chosen for its distinctive architecture, where each layer is connected to every other layer in a dense manner. This approach allows the network to learn highly detailed features by promoting feature reuse and improving gradient flow. These characteristics are particularly beneficial when working with limited datasets like those in PD handwriting analysis, as DenseNet201 can capture intricate variations in handwriting, such as reduced stroke size, irregular pen pressure, and changes in writing fluidity, which are often seen in Parkinsonian patients.

2) ResNet50: employs residual connections, which enable layers to skip over others, making it easier for the network to learn deeper representations without losing essential information. This architecture helps preserve spatial features, such as the shape and continuity of letters and lines, which are crucial for detecting subtle motor symptoms in handwriting. In PD detection, these spatial features can reveal issues like tremors or difficulty in maintaining consistent pressure or stroke alignment.

C. Feature fusion

Feature fusion [17] [18] across two levels is a sophisticated approach for integrating diverse feature sets to improve the performance of ML models. In the case of PD detection, combining handwriting and voice features allows the model to leverage complementary information from different data sources, leading to more precise and reliable predictions. This two-stage feature fusion process can be described as follows:

1) Level 1: Fusion of Deep Features from Handwriting Data: The first stage of fusion focuses on combining the deep features extracted from the handwriting dataset. We utilize two powerful pre-trained models, DenseNet201 and ResNet50, to extract meaningful features from handwriting images. Afterwards, we concatenate the obtained feature vectors from the final fully connected layers of DenseNet201 and ResNet50. This fusion operation is represented as:

$$\mathbf{F}_{\mathbf{handwriting}} = \text{concat}(\mathbf{F}_{\mathbf{DenseNet}}, \mathbf{F}_{\mathbf{ResNet}})$$

Here:

$\mathbf{F}_{\mathbf{DenseNet}}$ denotes the feature vector from DenseNet201.

$\mathbf{F}_{\mathbf{ResNet}}$ is the feature vector from ResNet50.

$\text{concat}(\cdot)$ refers to the operation of concatenating the two feature vectors.

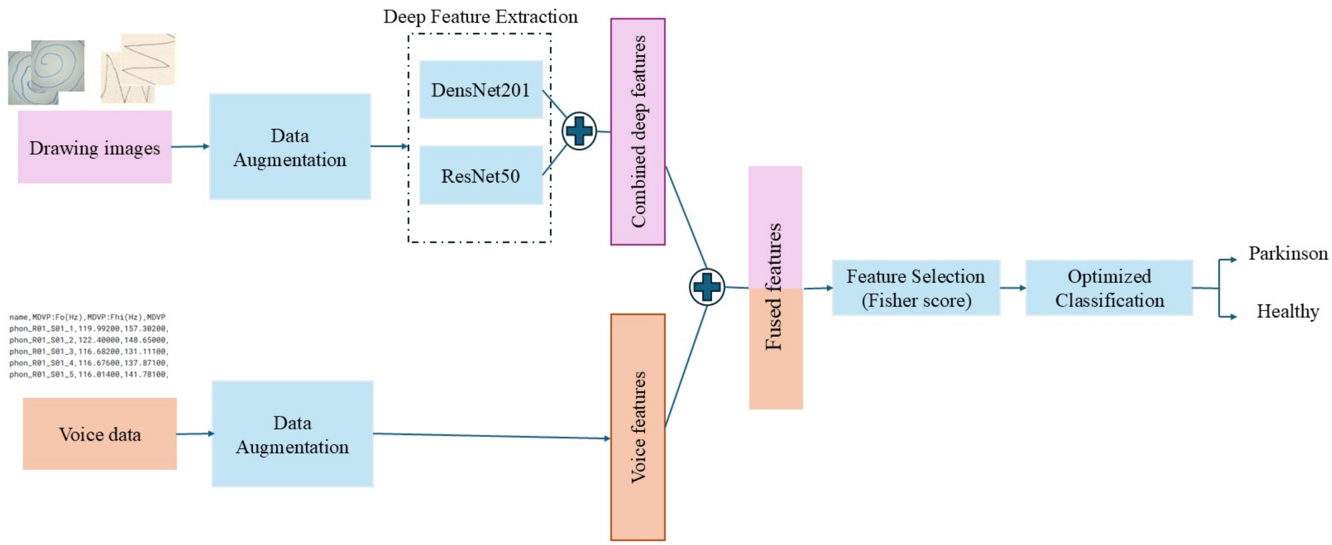

Fig. 1. Overview of the proposed approach.

This combined feature vector captures the unique capabilities of both models, making it more informative for subsequent processing.

2) Level 2: Fusion of Handwriting and Voice Features: The second stage involves merging the handwriting features $\mathbf{F}_{\text{handwriting}}$ with the voice features, which provide valuable insights into Parkinson's disease detection. Voice features such as fundamental frequency (Fo), jitter, shimmer, and nonlinear dynamical measures (e.g., RPDE and D2) offer important clues about speech abnormalities related to Parkinson's disease.

Let $\mathbf{F}_{\text{voice}}$ represent the feature vector derived from the voice dataset. This vector typically includes various characteristics of speech signals, such as pitch and spectral information.

At this level, we combine the handwriting features $\mathbf{F}_{\text{handwriting}}$ and the voice features $\mathbf{F}_{\text{voice}}$ to form a unified feature vector. The fusion process can be performed by concatenating these two feature vectors:

$$\mathbf{F}_{\text{combined}} = \text{concat}(\mathbf{F}_{\text{handwriting}}, \mathbf{F}_{\text{voice}})$$

D. Feature Selection using Fisher Score

Feature selection is a critical step when combining handwriting and voice data for PD detection. It ensures that the most informative features are selected, reducing computational complexity while maintaining high classification performance. To achieve this, we utilize the Fisher Score [19] method to measure the relevance of each feature.

The Fisher Score for a feature f_j in the combined feature vector F_{combined} is calculated using the ratio of between-class variance (S_B) to within-class variance (S_W):

$$F_j = \frac{S_B}{S_W}$$

Where:

- S_B represents the between-class scatter, which captures the separation between the means of the two classes (PD and healthy).
- S_W denotes the within-class scatter, which measures the variance within each class.

By computing the Fisher Score for all features, we can identify those that provide the most discriminatory power for PD classification, allowing for a more efficient and effective model.

E. Optimized Classification using Extratrees

In PD detection, the ExtraTrees algorithm [20] is utilized as an ensemble method to improve classification accuracy by aggregating predictions from multiple decision trees. ExtraTrees builds trees with random splits, allowing the model to explore diverse decision boundaries, making it effective for complex feature sets like fused handwriting and voice data.

To optimize the ExtraTrees model, we apply Grid Search Cross-Validation (GridSearchCV), which systematically searches for the best combination of hyperparameters, ensuring robust performance. The optimization can be represented as:

$$\hat{h} = \arg\max_{\theta} \left(\text{GridSearchCV} \left(\text{ExtraTrees}, X, y, \theta \right) \right)$$

where:

- $\hat{h}$ is the optimal hyperparameter configuration.
- ExtraTrees is the classifier.
- X is the feature matrix (fused handwriting and voice features).
- y is the label vector (Parkinson's or healthy).
- θ represents the hyperparameters, such as the number of trees and tree depth.

This process ensures optimal tuning of the model for accurate PD detection.

IV. Experiments

A. Dataset

1) Handwriting dataset: The dataset used in this study [21] was collected from participants who were asked to draw spirals and waves on paper, which were then digitized. Individuals with PD often produce irregular, shaky patterns in their drawings, while healthy individuals tend to create smoother, more consistent curves. The dataset contains a total of 204 images, equally divided into two categories: 102 spiral drawings and 102 wave drawings. Each image is labeled to indicate whether it corresponds to a PD patient or a healthy individual, enabling the analysis of motor control differences between the two groups. Fig 2 illustrates an example of the exploited handwriting samples.

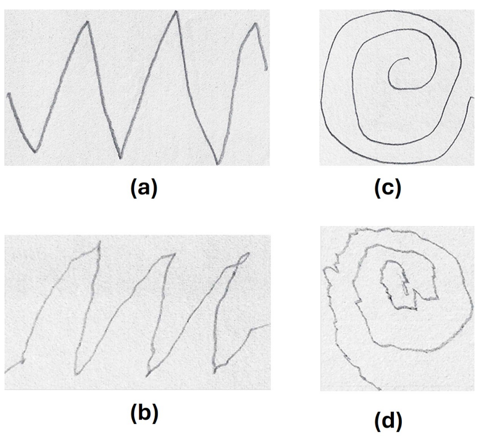

Fig. 2. Example of wave handwriting samples for: (a) healthy and (b) PD Patient and example of spiral handwriting samples for : (c) healthy and (d) PD patient.

2) Voice dataset: This exploited UCI dataset [22] comprises biomedical voice measurements obtained from 31 individuals, including 23 diagnosed with PD. The data is structured in a tabular format, with each column representing a distinct voice feature (such as jitter, shimmer, and other vocal characteristics), while each row corresponds to one of the 195 individual voice recordings from these participants.

B. Evaluation metrics

Evaluation metrics assess a model's performance in classification tasks using the confusion matrix, which consists of True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN). Key metrics include:

- **Accuracy (Acc)**: The ratio of correct predictions to total predictions:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision**: The proportion of true positives among positive predictions:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall**: The proportion of true positives identified by the model:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1 Score**: The harmonic mean of Precision and Recall:

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics provide a balanced assessment of model performance, focusing on classification accuracy, precision, sensitivity, and balance between positive predictions.

C. Results and discussion

This section focuses on evaluating the performance of the proposed method by utilizing three key metrics: accuracy curves, confusion matrices, and loss curves. These metrics provide a comprehensive view of the model's effectiveness.The confusion matrix presented in Fig. 3, reveals that the classification model performs effectively in differentiating between Healthy and PD cases. It accurately classifies 31 PD cases (True Positives) and 5 Healthy cases (True Negatives). However, it incorrectly labels 2 Healthy individuals as PD (False Positives) and 1 PD case as Healthy (False Negative). This indicates that the model excels in detecting PD, demonstrating a high recall for PD cases. Fig. 4 illustrates a strong learning

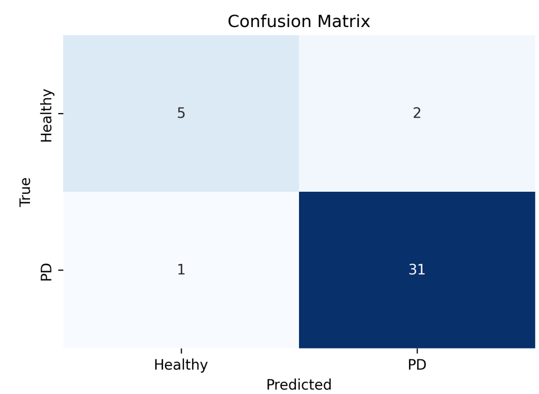

Fig. 3. Confusion matrix of the proposed approach.

performance with a high accuracy and a low loss. The close alignment between the training and validation curves suppose a good generalization, with no evident overfitting. Table I provides a comparison of various classification methods utilizing handwriting and voice features, emphasizing their respective accuracy rates. The proposed approach, which integrates feature fusion, Fisher score, and ExtraTrees, achieves the highest accuracy of 92.31%, surpassing all other methods.

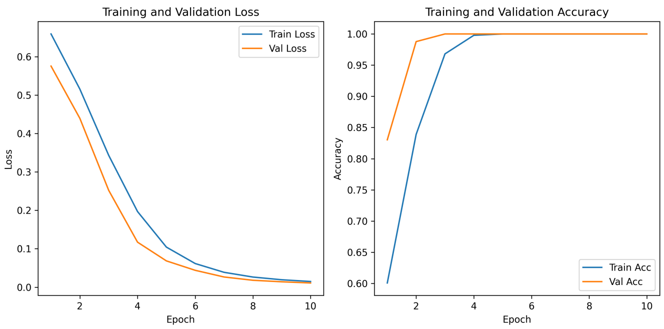

Fig. 4. Training and loss curves of the proposed approach.

TABLE I
EVALUATION OF THE PROPOSED APPROACH PERFORMANCE IN COMPARISON WITH STATE OF THE ART METHODS

Reference	Method	Fetures type	Accuracy
[6]	DNN + MLP	Handwriting	83%
[8]	HOG + CNN	Handwriting	86.67%
[9]	VGG16	Voice	91.8%
[11]	Gradient Boost model	Voice	91.53%
[13]	Vision Transformer	Handwriting + voice	90%
Our proposed approach	**Feature fusion + Fisher score + Extratrees**	Handwriting + voice	**92.31%**

V. CONCLUSIONS

Integrating advanced ML techniques with multimodal data, such as handwriting and voice features, holds significant potential in revolutionizing the early detection and diagnosis of PD. Thus, the proposed approach demonstrates superior performance compared to traditional methods that rely on single data sources, reaching an accuracy of 92.31%. This performance outperforms existing state-of-the-art methods, highlighting the efficacy of multimodal data in capturing the complex nature of PD symptoms. As research progresses, incorporating additional data types will further enhance the precision and robustness of PD detection, paving the way for earlier and more effective interventions.

REFERENCES

[1] J. D. Gallo-Aristizabal, D. Escobar-Grisales, C. D. Ríos-Urrego, J. F. Vargas-Bonilla, A. M. García, and J. R. Orozco-Arroyave, "Towards Parkinson's Disease Detection Through Analysis of Everyday Handwriting," *Diagnostics*, vol. 15, no. 3, article 381, 2025. doi: 10.3390/diagnostics15030381.

[2] R. F. Pfeiffer, "Non-motor symptoms in Parkinson's disease," *Parkinsonism & Related Disorders*, vol. 22, pp. S119–S122, 2016. doi: 10.1016/j.parkreldis.2015.09.004.

[3] M. Thomas, A. Lenka, and P. K. Pal, "Handwriting analysis in Parkinson's disease: current status and future directions," *Movement Disorders Clinical Practice*, vol. 4, no. 6, pp. 806–818, 2017. doi: 10.1002/mdc3.12526.

[4] H. Walhazi, L. R. Haddada, A. Maalej, and N. E. B. Amara, "Preprocessing Latent-Fingerprint Images For Improving Segmentation Using Morphological Snakes," in *Proc. 2020 5th Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, Sousse, Tunisia, 2020, pp. 1–6, doi: 10.1109/ATSIP49331.2020.9231908.

[5] J. Joy and M. P. Selvan, "Efficient Multi-modal Image Fusion Deep Learning Techniques for Classifying Neurodegenerative Disease Type," *Traitement du Signal*, vol. 42, no. 1, 2025.

[6] N. Arasavali, S. Jaffir, B. Lavanya, D. Ashok, and P. T. Kumar, "Handwriting Dynamic Assessment using Deep Neural Network Multi Layer Perceptron for Parkinson's Disease Classification," in *Proceedings of the 2024 International Conference on Communication, Control, and Intelligent Systems (CCIS)*, pp. 1–5, 2024. doi: 10.1109/CCIS63231.2024.10932041.

[7] X. Wang, J. Huang, M. Chatzakou, K. Medijainen, A. Toomela, S. Nõmm, and M. Ruzhansky, "LSTM-CNN: An efficient diagnostic network for Parkinson's disease utilizing dynamic handwriting analysis," *Computer Methods and Programs in Biomedicine*, vol. 247, p. 108066, 2024. doi: 10.1016/j.cmpb.2024.108066.

[8] K. A. Shastry, "Deep Learning-Based Diagnostic Model for Parkinson's Disease Using Handwritten Spiral and Wave Images," *Current Medical Science*, 2025. doi: 10.1007/s11596-025-00017-3.

[9] H. S. Malekroodi, N. Madusanka, B. I. Lee, and M. Yi, "Leveraging Deep Learning for Fine-Grained Categorization of Parkinson's Disease Progression Levels through Analysis of Vocal Acoustic Patterns," *Bioengineering (Basel)*, vol. 11, no. 3, article 295, 2024. doi: 10.3390/bioengineering11030295.

[10] R. Lamba, T. Gulati, A. Jain, and P. Rani, "A Speech-Based Hybrid Decision Support System for Early Detection of Parkinson's Disease," *Arabian Journal for Science and Engineering*, vol. 48, no. 2, pp. 2247–2260, Feb. 2023. doi: 10.1007/s13369-022-07249-8.

[11] D. A. Vigneswari and J. Aravinth, "Parkinson's disease Diagnosis using Voice Signals by Machine Learning Approach," in *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, Bangalore, India, 2021, pp. 869-872. doi: 10.1109/RTEICT52294.2021.9573689.

[12] N. Bibi, J. Courtney, and K. McGuinness, "Enhancing Brain Disease Diagnosis with XAI: A Review of Recent Studies," *ACM Transactions on Computing for Healthcare*, vol. 6, no. 2, pp. 1–35, 2025.

[13] M. Mohaghegh and J. Gascon, "Identifying Parkinson's Disease using Multimodal Approach and Deep Learning," in *Proceedings of the 2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA)*, pp. 1–6, 2021. doi: 10.1109/CITISIA53721.2021.9719945.

[14] D. Escobar-Grisales, T. Arias-Vergara, C. D. Rios-Urrego, E. Nöth, A. M. García, and J. R. Orozco-Arroyave, "An automatic multimodal approach to analyze linguistic and acoustic cues on Parkinson's disease patients," in *Proceedings of INTERSPEECH*, 2023.

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[17] L. R. Haddada, B. Dorizzi, and N. Essoukri Ben Amara, "Watermarking signal fusion in multimodal biometrics," in *Proc. Int. Image Process., Appl. and Syst. Conf. (IPAS)*, Sfax, Tunisia, 2014, pp. 1–6, doi: 10.1109/IPAS.2014.7043280.

[18] I. H. Trimech, N. Messaoudi, and N. E. Ben Amara, "A Novel Feature Combination Approach For Driver Fatigue Detection," in *Proc. 2022 19th Int. Multi-Conf. Syst., Signals Devices (SSD)*, Sétif, Algeria, 2022, pp. 475–479, doi: 10.1109/SSD54932.2022.9955654.

[19] Q. Gu, Z. Li, and J. Han, "Generalized Fisher score for feature selection," *arXiv preprint arXiv:1202.3725*, 2012.

[20] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, pp. 3–42, 2006. Springer.

[21] P. Zham, D. K. Kumar, P. Dabnichki, S. Poosapadi Arjunan, and S. Raghav, "Distinguishing different stages of Parkinson's disease using composite index of speed and pen-pressure of sketching a spiral," *Frontiers in Neurology*, vol. 8, p. 435, 2017. doi: 10.3389/fneur.2017.00435.

[22] M. A. Little, P. E. McSharry, E. J. Hunter, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 4, pp. 1079–1086, 2008. doi: 10.1109/TBME.2007.907919.

[23] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.