IEEE Open Journal of the Computer Society

Received 10 November 2025; revised 15 December 2025; accepted 27 December 2025. Date of publication 12 January 2026; date of current version 23 January 2026. The review of this article was arranged by Associate Editor Huansheng Ning.

Digital Object Identifier 10.1109/OJCS.2026.3651318

Machine Learning for Parkinson's Disease Detection: Analyzing Hybrid Voice Data With Spectral, Topological, and Random Matrix Methods

ANDY DOMÍNGUEZ-MONTERROZA, ALFONSO MATEOS CABALLERO, AND ANTONIO JIMÉNEZ-MARTÍN

Grupo de Análisis de Decisiones y Estadística, Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, 28040 Madrid, Spain

CORRESPONDING AUTHOR: ALFONSO MATEOS CABALLERO (e-mail: alfonso.mateos@upm.es).

This work was supported by MICIU/AEI/10.13039/501100011033 under Grant PID2021-122209OB-C31, Grant PID2024-155179NB-C22, and Grant RED2022-134540-T.

All source code and analysis scripts are available at https://www.parkinson.org/understanding-parkinsons/statistics.

ABSTRACT Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects both motor and speech functions. Advances in machine learning and signal processing have enabled non-invasive PD detection through voice analysis. This study proposes a comprehensive mathematical framework for PD classification that integrates topological, statistical, and spectral representations of speech signals. The framework combines topological descriptors derived from persistent homology, statistical measures based on random matrix theory, and spectral features extracted from frequency-domain analysis to capture complementary information about vocal dynamics. A hybrid training strategy was employed, using synthetic speech data generated from real recordings to train the models, while real samples were reserved exclusively for evaluation. Experimental results demonstrate that spectral features, particularly when fused with statistical descriptors, yield the highest discriminative power, achieving 98.00% accuracy and 97.98% F1-score with a multi-layer perceptron classifier. In contrast, topological descriptors provided limited standalone performance, serving instead as complementary components that enrich the overall representation. The findings highlight the potential of combining diverse mathematical representations to improve speech-based PD detection, especially in scenarios with limited access to clinically annotated data.

INDEX TERMS Parkinson's disease, speech analysis, topological data analysis, random matrix theory, spectral features, machine learning.

I. INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects motor and non-motor functions, including speech production [1]. Speech impairments in PD, collectively referred to as hypokinetic dysarthria, manifest as reduced articulation precision, altered phonation, and variations in speech rate [4], [5].The early detection of PD through non-invasive speech analysis has gained significant attention as a cost-effective and accessible diagnostic tool [6], [36]. In response, machine learning techniques have been widely explored to improve the accuracy and efficency of PD detection based on speech features [9], [10].

Recent studies have investigated various machine learning based approaches for PD classification using speech signals. For instance, [7] proposed a multiclass classification framework leveraging diverse acoustic features extracted from voice recordings to differentiate PD patients at various stages of the disease and distinguish them from healthy controls. The study evaluated various machine learning algorithms, including support vector machines, k-nearest neighbors and random forest, demonstrating that integrating heterogeneous feature sets improves classification performance by effectively capturing the complex vocal alterations associated with PD.

© 2026 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Another study in [8] investigates the classification of PD using acoustic features derived from multilingual speech tasks, emphasizing the importance of utilizing diverse speech data to improve the robustness of machine learning models for PD detection. By incorporating acoustic features extracted from sustained vowels, isolated words, and continuous speech, the study captured a more comprehensive representation of PD-related vocal impairments. Their results demonstrated that integrating features from multiple speech tasks significantly improved classification accuracy.

Similarly, [11] proposed a machine learning-based approach to classify PD stages and differentiate PD patients from healthy controls by using digital biomarkers derived from speech and motor evaluations. The integration of objective digital biomarkers significantly improved in classification performance compared to conventional clinical assessments, highlighting the potential of combining acoustic features with sensor-based measurements for more accurate PD diagnosis and monitoring.

In a multimodal approach, [10] combined acoustic and facial analysis for PD detection, focusing on hypomimia and hypokinetic dysarthria. The study utilized speech and video recordings from PD patients and healthy controls, extracting both acoustic and facial features to train machine learning models, particularly XGBoost. The results demonstrated that combining audio and video features significantly improved diagnostic performance compared to single-modality approaches, underscoring the potential of combining multimodal data for the PD detection.

Furthermore, [12] proposed a novel method for PD detection based on time-frequency representations generated using the empirical wavelet transform and the Hilbert transform. Energy and entropy features were extracted from these representations, followed by feature selection using a genetic algorithm and classification using a support vector machine. The proposed approach outperformed traditional methods, demonstrating the effectiveness of combining time-frequency analysis for the classification of PD.

A. ADVANCED MATHEMATICAL REPRESENTATIONS FOR SPEECH ANALYSIS

Beyond conventional acoustic features, recent research has explored the potential of advanced mathematical frameworks such as topological data analysis (TDA) and random matrix theory (RMT) for speech signal characterization. TDA provides a geometric and nonlinear approach to analyzing complex speech patterns by capturing topological structures within high-dimensional data [13]. Persistent homology, a key TDA technique, extracts topological features such as connected components, loops, and voids that persist across multiple scales, offering novel insights into the geometric complexity of speech signals [37], [19].

In parallel, RMT has been extensively used to analyze covariance structures in complex datasets, offering statistical insights into speech signal variability [26], [30], [31]. By characterizing noise distributions and eigenvalue spectra, RMT has demonstrated effectiveness in differentiating pathological speech patterns from healthy controls [32]. Furthermore, spectral features, particularly *Mel-frequency cepstral coefficients* (MFCCs), have been widely recognized for their discriminative power in PD classifcation [34].

Despite advancements in feature extraction and machine learning models, data scarcitiy remains a critical challenge in PD detection. Synthetic data augmentation has been explored as a strategy to enhance model generalization in biomedical applications [16]. However, studies investigating the impact of synthetic speech data augmentation for PD clasification remain limited. A recent study [15] introduced a novel approach employing generative adversarial networks (GANs) to generate synthetic voice signals, increasing the diversity and robustness of training datasets. The study evaluated deep learning architectures, including *residual network* (ResNet), *long short-term memory - fully convolutional network* (LSTM-FCN), InceptionTime, and *causal dilated inception-like convolutional neural network* (CDIL-CNN), demonstrating that synthetic data augmentation significantly improved classification performance. Nevertheless, even with synthetic augmentation, the highest reported accuracy was approximately 73%, suggesting limitations in the discriminative power of raw waveform features. Furthermore, the study primarily focused on deep learning approaches without integrating heterogeneous feature representations, which could further enhance classification accuracy.

B. CONTRIBUTIONS AND PROPOSED APPROACH

To address these limitations, this study proposes a machine learning framework for PD detection based on speech signals, systematically evaluating the effectiveness of TDA, RMT and spectral features. Unlike previous studies that rely solely on real speech data, we adopt a novel synthetic data-driven approach, where synthetic speech data generated using BigVASAN [15] are employed for training and validation, while real speech samples from the PC-GITA database [14] are used exclusively for final testing. This ensures consistency between training and evaluation conditions while assessing the generalization capabilities of models trained on synthetic speech.

The main contributions of this work are summarized as follows:

- A systematic comparison of topological, statistical, and spectral features for PD classification using SVM and multilayer perceptron models.
- An investigation into the generalization capability of models trained exclusively on synthetic speech data and evaluated on real speech recordings.
- An analysis of feature complementarity, with a specific focus on the performance impact of integrating TDA, RMT, and spectral descriptors.

The remainder of this paper is organized as follows. Section II details the dataset, feature extraction techniques, and classification models. Section III reports and discusses. Finally, Section IV provides conclusions and directions for future research.

II. METHODOLOGY

This study introduces a machine learning-based framework for PD detection using speech signals. The proposed methodology consists of three main steps: (i) data acquisition and preprocessing, (ii) feature extraction, and (iii) classification.

A. DATASET AND PREPROCESSING

The dataset employed in this study is the PC-GITA speech corpus [14], which comprises speech recordings from 100 Colombian speakers (50 diagnosed with PD and 50 healthy controls (HC)). Each participant performed a sustained phonation task of the vowel /a/ repeated three times, yielding a total of 300 speech signals (150 PD, 150 HC) [14]. The recordings were acquired under controlled conditions with a sampling rate of 44.1 kHz with 16-bit resolution. For consistency in feature extraction, all signals were subsequently downsampled to 24 kHz.

Additionally, a dataset of 2000 synthetic speech signals was utilized separately for model training and validation. This synthetic dataset consists of 1000 PD samples and 1000 HC samples, generated in a previous study [15]. These synthetic signals were leveraged to assess the impact of data augmentation on classification performance. However, they were not integrated with the real dataset during training. Instead, real speech signals were exclusively used for final validation to ensure an unbiased evaluation of model generalization. All speech signals, both real and synthetic, underwent amplitude normalization and were stored as csv files for further processing.

B. TDA FEATURE EXTRACTION

Topological data analysis (TDA) provides a robust framework for quantifying structural patterns in high-dimensional data. To do this, it employs persistent homology, which captures the evolution of topological features across multiple spatial scales. Given the inherent complexity and dynamic nature of speech signals, TDA is particularly suitable for identifying geometric structures that emerge due to articulatory and phonatory variations.

State-Space Reconstruction via Takens' Embedding Theorem: Takens' embedding theorem [18] provides a method to reconstruct the state space of a dynamical system from time series data. Given a discrete speech signal $x(t)$ of length N, the embedding matrix X is constructed as

$$X = \begin{pmatrix} x_1 & \cdots & x_{1-(d-1)\tau} \\ \vdots & \ddots & \vdots \\ x_{N-(d-1)\tau} & \cdots & x_{N-(d-1)\tau-(d-1)\tau} \end{pmatrix}$$

where, d is the embedding dimension, τ is the time delay, and N is the length of the original signal. This transformation unfolds the time series into a higher-dimensional space, facilitating the identification of meaningful topological structures.

Persistent Homology and Vietoris-Rips Complex Construction: Homology is a mathematical tool used to quantify the shape of data by detecting connected components, loops, voids, and higher-dimensional analogs. In speech signal analysis, homology offers insights into the underlying structure by tracking the emergence and dissolution of topological features as a scale parameter varies. Specifically, 0-dimensional homology classes correspond to connected components, while 1-dimensional classes correspond to loops or cycles. Higher-dimensional homology classes capture voids or cavities, 1-dimensional homology (loops) is typically the most relevant tool in the context of speech data.

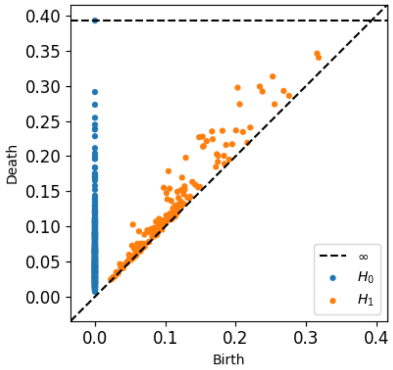

FIGURE 1. Persistence diagram of a pathological PC-GITA speech signal. Blue: H_0 components; orange: H_1 loops. Points above the diagonal represent persistent features.

For the embedded state space X, a Vietoris-Rips complex is constructed by connecting points within a predefined pairwise distance threshold. As this threshold varies, a sequence of nested simplicial complexes emerges, forming a filtration. Persistent homology is then applied to track the birth and death of topological features across this filtration, generating persistence diagrams that encode the longevity of these features [19].

Each persistence diagram consists of birth-death coordinate pairs (b_i, d_i), where b_i and d_i denote the scales on which a topological feature appears and disappears, respectively. Features that persist over a wide range of scales, reflected by points farther from the diagonal, are considered structurally significant, whereas features closer to the diagonal are typically interpreted as noise.

In general terms, a persistence diagram is a graphical representation of the lifespan of topological features across different scales. Each point (b, d) in the diagram corresponds to a feature that appears at b and disappears at d. The difference $d - b$ defines the persistence of the feature, reflecting its relevance to the underlying structure of the data. Features closer to the diagonal are typically interpreted as noise, while those farther away represent significant topological structures. Fig. 1 shows an example of a persistence diagram derived from speech data, illustrating the distribution of topological feature birth and death times.

In this study, we focus on persistent 1-dimensional homology (loops) because it has been shown to capture geometric features relevant to speech signal dynamics, as evidenced in previous studies [25]. The rationale behind this choice lies in

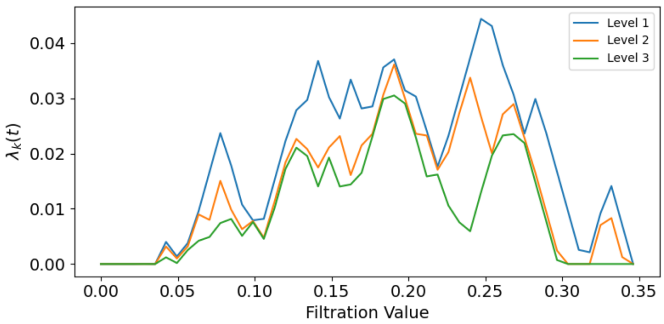

FIGURE 2. Persistence landscape from Fig. 1. Each color denotes a landscape function $\lambda_k(t)$ for $k = 1, 2, 3$.

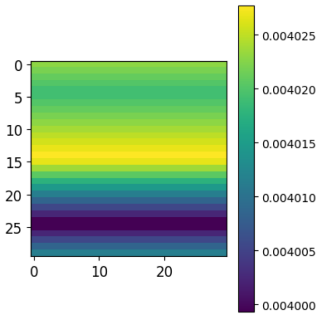

FIGURE 3. Persistence image from 1D homology (H_1) of Fig. 1. Warmer colors reflect longer-lived topological features.

the nature of speech signals, where cyclic patterns and periodicity are critical for capturing articulatory and phonatory characteristics.

Three types of TDA-based features were extracted from persistence diagrams.

1) PERSISTENT LANDSCAPE

A *persistent landscape* (PL) transforms the persistence diagram into a sequence of piecewise linear functions, capturing the topological characteristics across multiple scales [20], [21]. Mathematically, it is defined as $\lambda_k(t) = \max_{(b,d)\in D} \min(t-b, d-t)$, where D represents the persistence diagram, containing topological feature birth-death pairs (b, d), t is a real-valued parameter representing the domain over which the landscape is defined, and k is the landscape level, indexing the sequence of functions that capture different layers of topological persistence.

The function $\lambda_k(t)$ is piecewise linear and encodes the prominence of topological features on different scales (see Fig. 2). The PL is flattened into a vector representation, which serves as an input feature for machine learning models.

2) PERSISTENT IMAGES

Persistent images (PIs) provide a structured representation of persistence diagrams by mapping them onto a discrete grid using a kernel density estimation approach [22]. This method transforms topological features into a fixed-size numerical representation, making them compatible with machine learning models [23].

Given a persistence diagram D, which consists of birth-death pairs (b, d) representing topological features, a Gaussian kernel is applied to smooth the distribution of these features as follows $\mathrm{PI}(x, y) = \sum_{(b,d)\in D} \exp(-\frac{(x-b)^2+(y-d)^2}{2\sigma^2})$, where (x, y) represents the pixel coordinates in the persistence image, and σ is the standard deviation controlling the spread of the Gaussian function.

The resulting continuous function is discretized into a grid of predefined resolution, allowing the persistence diagram to be represented as a matrix of intensity values (see Fig. 3). Unlike PL, which encode topological prominence through piecewise linear functions, PI provide a dense representation by integrating topological features across the entire birth-death space.

The final PI matrix is flattened into a vector representation and used as input for machine learning models.

3) PERSISTENT ENTROPY

Persistent entropy (PE) quantifies the disorder of the persistence diagram, serving as a measure of the topological complexity [24]. It is defined as $PE = -\sum p_i \log p_i, \quad p_i = \frac{a_i}{\sum a_i}$, where $a_i = d_i - b_i$ denotes the persistence of each feature. Higher PE values indicate greater structural complexity within the speech signal.

C. RMT FEATURE EXTRACTION

RMT was employed to examine the eigenvalue-based characteristics of the covariance matrix derived from the Takens' embedding of the speech signal [26]. This methodological framework can capture of statistical properties in high-dimensional systems and has been successfully utilized in signal processing and biological data analysis [27], [28].

Given an embedding matrix X constructed from a speech signal, the covariance matrix C is defined as $C = \frac{1}{N}X^T X$, where N is the number of embedded vectors, X^T is the transpose of X, and C characterizes the estimated correlation structure of the signal.

The eigenvalues λ_i and eigenvectors v_i of C are obtained by solving the characteristic equation $Cv_i = \lambda_i v_i$, where λ_i are the eigenvalues that quantify the variance along the corresponding eigenvector direction v_i.

From the obtained eigenvalues, the following RMT-based features were extracted:

- *Largest Eigenvalue:* The *largest eigenvalue*, $\lambda_{\max}$, reflects the dominant variance component of the covariance matrix. It indicates the direction along which the variance is maximized, providing insight into the most significant variability pattern in the signal.
- *Mean Eigenvalue:* The *mean eigenvalue*, $\bar{\lambda}$, represents the average variance across all dimensions. This metric provides a global assessment of the variance distribution within the feature space offering insights into the overall spread of the data.

- *Eigenvalue Entropy:* The entropy of the eigenvalue distribution quantifies the disorder or randomness within the spectral properties of the covariance matrix. It is defined as $S = -\sum p_i \log p_i, \quad \text{with } p_i = \frac{\lambda_i}{\sum \lambda_j}$.
  Higher entropy indicates a more uniform distribution, whereas lower entropy indicates a concentration around a few dominant eigenvalues, whereas a lower entropy suggests that variance is concentrated in a few dominant eigenvalues, reflecting structural regularities in the speech signal.
- *Inverse participation ratio:* The *inverse participation ratio* (IPR) measures the localization of the components of the eigenvectors and is defined as $\text{IPR}_i = \sum_{j=1}^{N} |v_{ij}|^4$, where v_{ij} is the $j-th$ component of the $i-th$ eigenvector. A higher IPR value indicates that fewer components dominate the eigenvector structure implying stronger localization effects [29]. This metric is particularly useful for detecting the underlying structural constraint in high-dimensional speech representations.

D. SPECTRAL FEATURE EXTRACTION

Spectral features provide critical insights into the frequency content and dynamic variations of a speech signal. In this study, spectral features were computed using the *Librosa* Python library [33], which offers efficient implementations of several frequency-domain descriptors. The extracted spectral features include *Mel-frequency cepstral coefficients* (MFCCs), this first-order derivatives (delta MFCCs), spectral centroid, spectral bandwidth, spectral rolloff, and zero-crossing rate. A detailed description of each feature is provided bellow.

- *Mel-frequency cepstral coefficients (MFCCs)* are widely employed in speech processing as they efficiently represent the spectral envelope of short-term power spectra. Their computation involves applying the discrete cosine transform to the logarithm of the magnitude of the Mel spectogram, which reduces spectral correlation while emphasizing perceptually relevant features. The k-th MFCC coefficient is defined as $\text{MFCC}_k = \sum_{n=1}^{N} S_n \cos(\frac{\pi k(n-0.5)}{N})$, where S_n represent the log-magnitude of the mel spectrogram at index n, N is the number of mel frequency bands, and k is the coefficient index.
  In this study, 13 MFCC coefficients were extracted for each speech signal. In addition, their mean and standard deviation to capture statistical variability.
  To model the temporal evolution of spectral features, first-order time derivates (delta MFCCs) were also computed. These derivates quantify the rate of change in MFCCs over time, thereby capturing dynamic variations in speech production.
- The *spectral centroid* (SC) characterizes the center of mass of the spectral power distribution and is perceptually associated with the brightness of sound. It is mathematically defined as $\text{SC} = \frac{\sum_k f_k S_k}{\sum_k S_k}$, where f_k represents the frequency in the k bin and S_k denotes the spectral magnitude in that frequency bin.
  A higher centroid value indicates in the signal a greater proportion of high-frequency components, which can differ significantly between pathological and healthy speech in the signal. The mean and standard deviation of the spectral centroid were computed as statistical descriptors.
- *Spectral bandwidth (SB)* quantifies the dispersion of frequency components around the spectral centroid by capturing variations of the spread of the spectral energy. It is expressed as $\text{SB} = \sqrt{\frac{\sum_k (f_k - SC)^2 S_k}{\sum_k S_k}}$, where SC is the spectral centroid.
  This provides information on articulation patterns that may exhibit altered frequency distributions in speech affected by PD. The mean and standard deviation values of the spectral bandwidth were extracted as statistical descriptors.
- The *spectral rolloff* (SR) defines the frequency below which a specified percentage (typically 85%) of the total spectral energy is concentrated. It serves as an indicator of the harmonic content of a signal and is computed as $\text{SR} = \min\{f_r : \frac{\sum_{k=1}^{r} S_k}{\sum_{k=1}^{N} S_k} \geq 0.85\}$, where f_r represents the rolloff frequency, and S_k is the spectral magnitude at bin k.
  This feature is particularly useful for identifying variations in spectral energy distribution, which may be characteristic of dysarthric speech in individuals with PD.
- The *zero-crossing rate* (ZCR) quantifies the rate at which the signal waveform changes sign, serving as an indicator of high-frequency content. It is defined as: $\text{ZCR} = \frac{1}{N-1} \sum_{n=1}^{N-1} |\text{sgn}(x_n) - \text{sgn}(x_{n-1})|$, where x_n is the signal amplitude at time n, and $\text{sgn}(\cdot)$ denotes the sign function. The ZCR provides insights into signal variability and articulation differences, which are relevant for distinguishing between pathological and normal speech. As with other spectral features, both the mean and standard deviation of ZCR were calculated.

In total, 60 spectral features were derived, encompassing the mean and standard deviation of 13 MFCC coefficients and their delta derivatives, as well as the mean and standard deviation for the spectral centroid, bandwidth, rolloff, and zero-crossing rate.

E. MODEL TRAINING AND EXPERIMENTAL SETUP

The machine learning models were trained and evaluated using both synthetic and real speech data. The synthetic data were generated using the BigVSAN approach, derived from the PC-GITA database [14]. Specifically, synthetic data were exlusively employed for model training and validation, whereas real speech recording from the PC-GITA database were reserved for final testing to assess model

generalizability. The synthetic dataset consisted of 1,000 signals per class (control and pathological), resulting in a balanced training set of 2,000 samples. All available real speech signals from the PC-GITA dataset were used as the test set.

Feature extraction. TDA features: The embedding dimension and time delay for Takens' embedding were selected through a grid search optimization procedure. A predefined range of values was considered for the embedding dimension $d \in \{2, 3, 4, 5\}$ and time delay $\tau \in \{5, 10, 15, 20\}$. For each parameter combination, the topological features were extracted, classification models were trained, and the resulting accuracy was recorded. The optimal configuration $d = 5, \tau = 20$ was selected based on the highest classification accuracy obtained during cross-validation. Using these parameters, persistence images were computed with a pixel resolution of 30×30 and a pixel size of 0.1. Persistence landscapes were generated with 50 evaluation points and three persistence levels.

RMT features: eigenvalues were obtained from the covariance matrix of the embedded signals. The following statistical descriptors were computed: largest eigenvalue ($\lambda_{\max}$), mean eigenvalue ($\bar{\lambda}$), eigenvalue entropy (S), and inverse participation ratio (IPR).

Spectral feature: the spectral features extraction followed a standardized procedure using the *Librosa* library [33], to ensure consistency and reproducibility. The sampling rate of all speech signals was set to 24 KHz. A total of 13 MFCCs were computed using a frame length of 512 samples and a hop length of 256 samples. The extracted spectral features included MFCCs. The MFCCs were derived as the discrete cosine transform of the log-magnitude mel spectrogram. To capture temporal dynamics, first-order derivates (delta MFCCs), representing the rate of change over time. Both MFCCs and delta coefficients were summarized by their mean and standard deviation to capture overall tendencies and variability.

Additional spectral descriptors: To further characterize the frequency distribution and temporal structure of the signals, the following features were computed: the spectral centroid and spectral bandwidth (SB) both estimated using a fast Fourier transform window of 512 samples; the spectral roll-off (SR), determined at a threshold of 85%. The zero-crossing rate (ZCR) was calculated by counting the number of times the signal waveform changed sign within each analysis frame. All spectral descriptors were summarized by computing their mean and standard deviation over the entire signal duration.

Feature normalization: Prior to combining the different representations of characteristics, each feature vector was standardized to ensure balanced contributions from all descriptors. This step is particularly important when merging heterogeneous feature spaces, such as topological, statistical, and spectral features, which may differ significantly in scale and distribution.

Model training and evaluation: Two machine learning models were implemented for classification: support vector machine (SVM) and multi-layer perceptron (MLP). A linear kernel was selected as the SVM as it is effective in high-dimensional feature spaces. A feedforward neural network with one hidden layer and ReLU activation function was employed as the MLP.

Model performance was assessed using stratified 5-fold cross-validation. The following performance metrics were computed: Accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$, Sensitivity $= \frac{TP}{TP+FN}$, Specificity $= \frac{TN}{TN+FP}$, $F1 = \frac{2\times TP}{2\times TP+FP+FN}$, where TP (true positives) denotes correctly classified pathological speech cases, TN (true negatives) corresponds to correctly classified control speech cases, FP (false positives) represents control cases misclassified as pathological, and FN (false negatives) are pathological cases misclassified as controls.

All experiments were conducted on a computer system equipped with an Intel Core Ultra 5 125H processor (3.60 GHz) and 8 GB RAM.

III. RESULTS AND DISCUSSION

This study systematically evaluated the effectiveness of topological, statistical and spectral features in PD classification using machine learning models trained on synthetic data and tested on real data. The experimental design aimed to assess the generalizability of models trained exclusively on artificially generated signals derived from the PC-GITA database, while ensuring that real speech signals were reserved solely for model evaluation. The classification performance across different feature sets and machine learning architectures is summarized in Table 1.

A. CLASSIFICATION PERFORMANCE ACROSS FEATURE SETS

The experimental results indicate that the spectral features consistently achieved the highest classification performance across both SVM and MLP models. Notably, the MLP classifier reached an accuracy of 98.00% with an F1-score of 97.95%, demonstrating the high discriminative capacity of spectral descriptors in distinguishing between control and pathological speech. Similarly, the combination of RMT and spectral features yielded comparable results, with MLP achieving an accuracy of 98.00% and an F1-score of 97.98%, highlighting the robstness of spectral representations when supplemented with statistical descriptors.

The superior performance of spectral-based features aligns with previous findings that underscore their efficacy as biomarkers for neurodegenerative speech impairments [6], [34]. These results can be attributed to the well-documented alterations in articulation and phonation observed in PD patients, which manifest prominently in the spectral domain [5], [36].

Conversely, the classification performance of topological features (TDA) was significantly lower, with the highest accuracy reaching 62.00% when using MLP and 58.67% with SVM. These findings suggest that topological descriptors alone do not effectively capture the vocal alterations associated with PD. Moreover, the combination of TDA and RMT

TABLE 1. Classification Performance With Synthetic Training and Real Testing

Feature Set	Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
TDA	SVM	58.67 ± 5.31	58.00 ± 7.18	59.33 ± 10.41	58.29 ± 5.09
	MLP	62.00 ± 1.94	62.00 ± 7.18	62.00 ± 4.00	61.80 ± 3.79
RMT	SVM	58.67 ± 5.31	48.67 ± 4.99	68.67 ± 12.93	54.11 ± 3.33
	MLP	65.67 ± 3.74	60.67 ± 3.89	70.67 ± 8.27	63.89 ± 3.05
Spectral	SVM	82.00 ± 3.40	80.00 ± 6.32	84.00 ± 2.49	81.52 ± 3.91
	MLP	$\mathbf{98.00 \pm 1.63}$	96.67 ± 2.98	$\mathbf{99.33 \pm 1.33}$	97.95 ± 1.67
RMT+Spectral	SVM	82.00 ± 2.45	82.00 ± 6.86	82.00 ± 8.06	81.95 ± 2.52
	MLP	$\mathbf{98.00 \pm 1.63}$	$\mathbf{97.33 \pm 2.49}$	98.67 ± 1.63	$\mathbf{97.98 \pm 1.65}$
TDA+RMT	SVM	58.00 ± 6.18	58.00 ± 7.77	58.00 ± 8.84	57.90 ± 6.34
	MLP	61.33 ± 1.63	62.67 ± 8.00	60.00 ± 5.96	61.59 ± 3.87
TDA+RMT+Spectral	SVM	70.33 ± 5.21	68.00 ± 8.84	72.67 ± 5.33	69.40 ± 6.37
	MLP	74.67 ± 3.40	70.00 ± 10.95	79.33 ± 4.90	73.02 ± 5.29

did not yield substantial improvements in classification accuracy, indicating a lack of complementarity between these feature sets within this specific classification task.

RMT-based features alone exhibited moderate classification performance, achieving an accuracy of 65.67% with MLP and 58.67% with SVM. Interestingly, the relatively stable performance of RMT-based features across different models suggests that these statistical descriptors capture relevant patterns in speech variability that are less sensitive to distortions introduced by synthetic data.

B. SYNERGISTIC EFFECTS OF FEATURE COMBINATIONS

In addition to the primary accuracy results, a more detailed analysis of the performance metrics reveals important insights into the behavior of the model and feature utility. The integration of RMT and spectral features consistently outperformed individual feature sets, demonstrating the advantages of leveraging both covariance structure analysis and spectral representations. This hybrid approach effectively captures a broader range of speech characteristics affected by PD, contributing to improved classification performance.

The combination of topological, statistical, and spectral features (TDA+RMT+spectral) provided intermediate results, achieving 74.67% accuracy with MLP and 70.33% with SVM. However, this feature fusion did not surpass the performance of spectral or RMT+spectral sets.

Focusing on the RMT feature set alone, the SVM model exhibited a sensitivity of only 48.67%, whereas specificity reached 68.67%. This asymmetry suggests that RMT-based representations may be more effective in capturing variability in speech patterns of healthy controls compared to patients with PD. Although the MLP model improved sensitivity to 60.67%, a significant imbalance persisted, indicating that RMT features alone are insufficient to achieve balanced classification performance.

When comparing the performance of SVM and MLP, the results indicate that MLP generally achieves higher accuracy, particularly when multiple feature sets are combined. For example, although the fusion of topological and statistical features (TDA+RMT) resulted in relatively low overall performance, the MLP still outperformed the SVM, achieving an accuracy of 61.33% compared to 58.00%. More notably, in the combined TDA+RMT+spectral configuration, the MLP achieved 74.67% accuracy, surpassing the SVM's 70.33%. This performance gap suggests that neural architectures more effectively integrate heterogeneous feature representations, especially when spectral information is incorporated.

Furthermore, the combination of RMT and spectral features yielded identical sensitivity and specificity values (82.00%) for the SVM, indicating balanced performance between classes. In contrast, the MLP achieved a sensitivity of 97.33% and a specificity of 98.67%, reflecting a slight bias towards the correct classification of healthy controls. This consistently strong performance across both metrics highlights the robustness of this hybrid representation when employed with neural network models.

Another important aspect is the variability in model performance. The MLP consistently exhibited lower standard deviations in most configurations, indicating more stable predictions. For example, when using spectral features, the SVM achieved an accuracy of 82.00% $\pm$ 3.40%, while the MLP reached 98.00% $\pm$ 1.63%. A similar trend was observed for the TDA+RMT+spectral combination, where the MLP demonstrated reduced variability ($\pm$3.40%) compared to the SVM ($\pm$5.21%). These findings support the notion that neural models not only outperform traditional classifiers in terms of average accuracy, but also provide more consistent classification performance across cross-validation folds.

To further investigate the limited synergy observed among the TDA, RMT, and Spectral feature families, we conducted a *redundancy and fusion analysis* aimed at quantifying the relationships, complementarity, and discriminative contribution of each descriptor domain.

A redundancy analysis combining cross-correlation and *principal component analysis* (PCA) was first performed to examine the degree of linear dependency and scaling disparity among feature groups. The correlations between the first principal components of TDA, RMT, and spectral features were consistently low ($r_{\text{TDA,RMT}} = -0.17$, $r_{\text{TDA,SPEC}} = 0.29$,

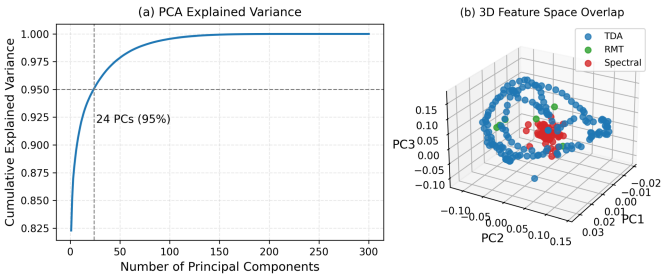

FIGURE 4. PCA-based feature space analysis. (a) The first 24 components explain 95.08% of the total variance in the fused space (TDA, RMT, and spectral descriptors). (b) The 3D loading plot shows spectral features forming a compact, high-variance cluster, whereas RMT and TDA occupy orthogonal regions, providing complementary but unevenly weighted information.

TABLE 2. Feature Breakdown and Dimensionality by Descriptor Family

Family	Component	Params	Count
TDA (1051)	PI	(30×30)	900
	PL	50×3	150
	PE	Scalar	1
RMT (6)	Eig. stats	mean, variance, ratio, spectral entropy, avg. IPR	6
Spectral (60)	MFCCs	13×2	26
	Δ-MFCCs	13×2	26
	Moments	Centroid, bandwidth, roll-off, ZCR (mean and std)	8
Total	All families	–	**1117**

$r_{\mathrm{RMT,SPEC}} = -0.42$), suggesting that each descriptor family captures distinct and complementary aspects of the speech signal rather than redundant information.

A global PCA applied to the fused feature space—comprising 1051 TDA, 6 RMT, and 60 Spectral features for a total of 1117 descriptors—revealed that 24 principal components were required to explain 95.08% of the total variance (Fig. 4(a) and Table 2). This finding indicates a high-dimensional representation that is not dominated by any single feature type, and that most redundancy arises within individual descriptor blocks, rather than from cross-feature overlap.

The three-dimensional PCA projection (Fig. 4(b)) further clarifies the internal organization of the fused feature space. Spectral descriptors (red) form a dense, axis-aligned cluster, suggesting that the largest share of global variance is concentrated within this domain. In contrast, TDA features (blue) exhibit a broader and partially orthogonal dispersion, capturing complementary geometric properties that are not aligned with the dominant spectral directions. RMT features (green) occupy an intermediate region, bridging the statistical and structural representations of the signals.

This spatial organization confirms that the three feature families encode distinct yet partially overlapping information, explaining why hybrid models yield only limited performance gains.

To explore alternative integration strategies, a *weighted feature fusion analysis* was performed, in which each feature block contributed to the composite representation according to a tunable set of weights. The highest accuracy (0.81) was achieved when the Spectral block predominated, corresponding to optimal weights of $[w_{\mathrm{TDA}}, w_{\mathrm{RMT}}, w_{\mathrm{SPEC}}] = [0.0, 0.0, 0.25]$. This configuration indicates that the discriminative variance is largely concentrated in the Spectral domain, while TDA and RMT features contribute secondary but complementary information that does not substantially alter decision boundaries.

Overall, these results demonstrate that the limited synergy among feature families stems primarily from scaling disparities and the dominant variance contribution of spectral descriptors, rather than from intrinsic redundancy. The joint PCA and fusion analyses provide a clear explanation of why hybrid representations yield only marginal improvements, while highlighting the structural complementarity of topological and statistical features that could be further exploited through nonlinear or attention-based fusion strategies in future work.

C. STATISTICAL VALIDATION OF MODEL AND FEATURE DIFFERENCES

To assess whether the observed performance variations across classifiers and feature representations were statistically and practically significant, we conducted a series of parametric tests on cross-validation results, including paired t-tests, one-way ANOVA, and Tukey's HSD post-hoc analyses.

Paired t-tests comparing MLP and SVM classifiers revealed statistically significant improvements in both accuracy and F1-score for the Spectral ($p < 10^{-9}$) and RMT + Spectral ($p < 10^{-3}$) feature sets, confirming the superior discriminative capacity of MLP models under these representations. In contrast, no significant differences were observed for purely topological (TDA) or hybrid topological–statistical combinations ($p > 0.1$), suggesting that the advantages of neural architectures emerge primarily when spectral information is present.

The one-way ANOVA results further indicated substantial performance disparities across feature sets for all models ($p < 10^{-9}$). Subsequent Tukey's HSD post-hoc comparisons showed that both Spectral and RMT + Spectral features achieved significantly higher accuracy and F1-scores than TDA-based or mixed feature sets ($p < 0.001$). These findings demonstrate that spectral descriptors play a dominant role in capturing discriminative information, while topological descriptors alone provide limited predictive power.

To complement statistical significance with a measure of practical relevance, we computed effect size indices, including Cohen's d for t-tests and η^2 for ANOVA. Both measures indicated large effect magnitudes (Cohen's $d > 0.8$; $\eta^2 > 0.14$),

TABLE 3. Statistical Significance and Effect Size Across Classifiers and Feature Sets

Test	Comparison	Feature Set	t/F	p-value	Effect Size	Significant ($p<0.05$)	Interpretation
Paired t-test	MLP vs. SVM	Spectral	51.00	8.8×10^{-7}	$d = 3.85$ (Large)	Yes	MLP > SVM
Paired t-test	MLP vs. SVM	RMT+Spectral	11.16	3.7×10^{-4}	$d = 8.28$ (Large)	Yes	MLP > SVM
Paired t-test	MLP vs. SVM	TDA	0.14	8.9×10^{-1}	$d = 0.10$ (Small)	No	n.s.
Paired t-test	MLP vs. SVM	TDA+RMT	1.56	1.9×10^{-1}	$d = 0.05$ (Small)	No	n.s.
ANOVA	Across feature sets	MLP	361.18	9.2×10^{-22}	$\eta^2 = 0.97$ (Large)	Yes	Feature sets differ
ANOVA	Across feature sets	SVM	34.47	3.4×10^{-10}	$\eta^2 = 0.89$ (Large)	Yes	Feature sets differ
Tukey HSD	MLP (Accuracy)	Spectral vs. others	–	$p < 0.001$	–	Yes	Spectral best
Tukey HSD	MLP (F1-score)	RMT+Spectral vs. others	–	$p < 0.001$	–	Yes	RMT+Spectral best

TABLE 4. Confusion Matrix for the Best-Performing Configuration (MLP Trained on RMT + Spectral Features)

True / Predicted	Control	PD
Control	148	2
PD	4	146

confirming that the observed improvements are not only statistically significant but also practically meaningful. Overall, these findings confirm the dominant role of spectral representations and the limited standalone contribution of TDA descriptors when not supported by spectral information. A summary of the statistical analyses is provided in Table 3, highlighting how feature selection, model architecture, and metric stability jointly shape the performance of synthetic-data-based PD detection systems.

D. ERROR ANALYSIS

To complement the main performance evaluation, we conducted an error analysis aimed at characterizing the distribution of misclassifications and assessing the robustness and fairness of the proposed model. This analysis was designed to identify any systematic bias or asymmetric error patterns between Parkinson's disease (PD) and control samples, which could comprise the model's reability in practial applications.

The best-performing configuration—an MLP trained on the RMT+Spectral feature set—was used for this assessment. Table 4 presents the aggregated confusion matrix, obtained by summing predictions across all cross-validation folds. This configuration achieved an overall accuracy of 98.0%, with a sensitivity of 97.3% and a specificity of 98.7%. The macro-averaged precision, recall, and F1-score across folds were consistently balanced between the two classes (PD vs. control), yielding a macro F1-score of 0.98.

The small number of off-diagonal entries in the confusion matrix confirms that misclassifiations are rare and symmetrically distributed, indicating that the model generalizes effectively without systematic bias towards either class.

Furthermore, according to Pah et al. [38], participants in the PC-GITA dataset are clinically homogeneous across both demographic and disease-related variables. No statistically significant differences were reported between male and female subjects in age ($p = 0.966$), *unified Parkinson's disease rating scale* (UPDRS) scores ($p = 0.760$), or Hoehn & Yahr (H&Y) stages ($p = 0.927$), with mean H&Y values of 2.30 ± 0.94 for males and 2.28 ± 0.54 for females. Likewise, the disease duration did not differ significantly between groups ($p = 0.157$).

This balanced demographic and clinical composition ensures that the model's performance reflects genuine pathological voice differences rather than confounding effects associated with age, gender, or disease severity. Consequently, the observed accuracy of 98% can be interpreted as robust and unbiased within a clinically controlled and statistically comparable cohort, reinforcing the reliability and geneeralizability of the proposed approach.

E. DOMAIN SHIFT ANALYSIS BETWEEN SYNTHETIC AND REAL SPEECH DATA

To quantitatively and visually assess the degree of alignment between synthetic and real speech data, a *domain shift analysis* was conducted. This evaluation aimed to determine whether the distribution of features derived from synthetic signals faithfully reproduces the statistical structure of real recordings, ensuring that models trained on synthetic data can generalize effectively to real-world conditions.

A *maximum mean discrepancy* (MMD) test was first applied to measure the statistical divergence between the tow domains. The analysis employed a *radial basis function* (RBF) kernel with a width parameter of $\gamma = 0.5$. The resulting MMD statistic was 0.00649, with a permutation-based p-value of 0.0010, indicating a small yet statistically significant difference between the distribution of synthetic an real samples. This outcome suggests that, although synthetic speech is not identical to natural recordings, their overall statistical structure remains highly consistent.

To complement the quantitative assessment, a *t-distributed stochastic neighbor embedding* (t-SNE) visualization was generated using the RMT + spectral feature embeddings (Fig. 5). The resulting projection revealed a substantial overlap between synthetic (blue) and real (red) samples, forming a continuous manifold without distinct cluster separation. This overlap indicates that the generative process successfully preserved the acoustic and structural characteristics of the original PC-GITA dataset, producing synthetic samples that are statistically and geometrically aligned with real speech data.

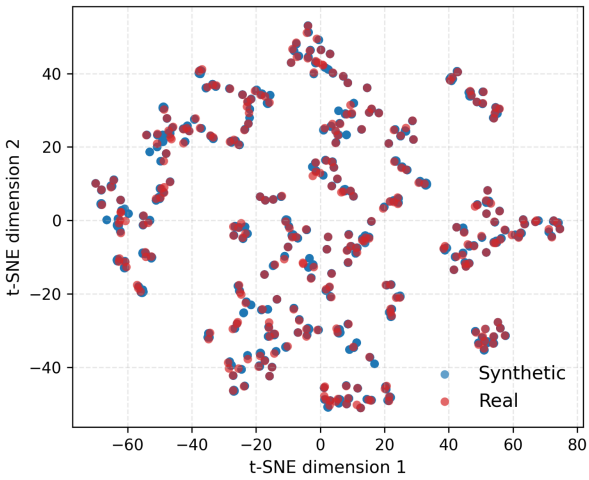

FIGURE 5. t-SNE projection of the RMT+Spectral feature embeddings comparing synthetic (blue) and real (red) speech samples. The substantial overlap between the two distributions indicates that synthetic data effectively capture the statistical and structural characteristics of real speech, suggesting a minimal domain shift.

TABLE 5. Comparative Accuracy of Methodologies Applied to the PC-GITA Dataset Reported in Recent Studies

Methodology	Reference	Accuracy
Spectral & cepstral features + SVM	[39]	89
IEDCC + SVM	[40]	91
TF features + SVM	[41]	92
Glottal signal + CNN + LSTM	[42]	68.56
HCC + MLP	[43]	91
Autoencoder + CNN	[44]	84
Semantic + grammatical + SVM	[45]	66
Articulatory + prosodic + phonemic +SVM	[46]	84
Time-distributed 2D-CNN	[47]	85.33
EWT + HT based TF + SVM	[12]	82.3
TDA + RMT + Spectral + MLP (synthetic)	Our work	**98.0**

The combination of a low MMD value and overlapping t-SNE embeddings demonstrates that the proposed generative approach effectively captures the essential variability of the target domain, resulting in minimal domain shift and strong generalization capacity for classifiers trained exclusively on synthetic data.

Nevertheless, to further mitigate residual discrepancies, future work could explore adaptive learning strategies such as adversarial domain adaptation, transfer learning, or instance weighting. These techniques could enhance robustness to linguistic variability, recording conditions, and other environmental factors, thereby improving the scalability of the proposed framework to more diverse real-world scenarios.

F. COMPARATIVE ANALYSIS WITH PREVIOUS STUDIES

To contextualize the results of this study, Table 5 provides a comparative summary of recent works that have employed the PC-GITA dataset for the classification of PD. The proposed approach, which utilizes synthetic data for training and real data for evaluation, achieved an accuracy of 98.00%, demonstrating its competitiveness with existing methods that rely solely on real speech recordings.

This outcome supports the feasibility of using synthetic data for training effective PD classifiers, particularly when combined with spectral and statistical features. The observed performance exceeds that of several state-of-the-art approaches based on traditional spectral and cepstral features [39], empirical mode decomposition [40], and deep learning architectures [44], [47]. These results highlight the potential of synthetic data, when properly generated and paired with informative feature representations, to support the development of robust PD detection models. In contrast to previous studies, the present work demonstrates that synthetic speech data can generalize effectively to real-world scenarios, particularly when enriched with statistically and spectrally informative features. Moreover, the integration of RMT with spectral features appears to enhance robustness by capturing complementary aspects of speech variability.

The findings of this study suggest that synthetic data augmentation, coupled with informed feature selection, may serve as a viable alternative in scenarios where real data availability is constrained. This is particularly relevant in clinical applications, where the collection of large-scale annotated speech datasets is often limited by practical constraints, including time, cost, and privacy considerations. Overall, the results highlight the potential of synthetic training in biomedical voice analysis and the value of combining diverse signal representations to enhance diagnostic performance. Future work should integrate synthetic and real data within deep learning frameworks to improve the generalizability and robustness of PD classification.

IV. CONCLUSION

This work introduced a machine learning framework for the detection of Parkinson's disease from speech signals, emphasizing the integration of hybrid mathematical features and the role of synthetic data in model development. The results demonstrated that the combination of spectral and statistical descriptors provides the most discriminative representation of PD-related speech alterations, achieving 98.0% accuracy and an F1-score of 97.98% with a multi-layer perceptron classifier. In contrast, topological descriptors exhibited limited sensitivity to pathological variation, confirming their usefulness as auxiliary rather than primary features.

Beyond achieving state-of-the-art performance on the PC-GITA dataset, the analysis also showed a strong alignment between synthetic and real speech domains, validating the effectiveness of the generative process in preserving essential acoustic and structural properties. These findings reinforce the potential of synthetic data generation as a viable strategy to alleviate data scarcity in biomedical speech analysis.

Nevertheless, the current study is limited to a single dataset and linguistic domain, and the generative approach relied on a single synthesis technique. Future work will extend

the framework to multilingual and cross-corpus evaluations, incorporate multiple data augmentation pipelines, and integrate interpretability-driven analyses to better understand the interplay between topological, statistical, and spectral representations within the learned feature space. Such developments will enhance the scalability, transparency, and clinical reliability of speech-based diagnostic systems for Parkinson's disease.

ACKNOWLEDGMENT

The authors thank Prof. Orozco Arroyave and the GITA group at the University of Antioquia for access to the PC-GITA database.

REFERENCES

[1] C. Tanner, Y. Ben-Shlomo, and M. San Luciano, "Epidemiology of Parkinson's disease," *Lancet*, vol. 403, no. 10423, pp. 283–292, 2024.
[2] Parkinson's Foundation, "Parkinson's disease statistics," 2024. [Online]. Available: https://www.parkinson.org/understanding-parkinsons/statistics
[3] S. Hauser and S. A. Josephson, *Harrison's Neurology in Clinical Medicine*, 4th Ed. New York, NY, USA: McGraw-Hill, 2017.
[4] L. Brabenec, J. Mekyska, Z. Galaz, and I. Rektorova, "Speech disorders in Parkinson's disease: Early diagnostics and effects of medication and brain stimulation," *J. Neural Transmiss.*, vol. 124, no. 3, pp. 303–334, 2017.
[5] S. Skodda, W. Grönheit, and U. Schlegel, "Intonation and speech rate in Parkinson's disease: General and dynamic aspects and responsiveness to levodopa admission," *J. Voice*, vol. 25, no. 4, pp. e199–e205, Jul. 2011, doi: 10.1016/j.jvoice.2010.04.007.
[6] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.
[7] S. Srinivasan et al., "Detection of Parkinson disease using multiclass machine learning approach," *Sci. Rep.*, vol. 14, 2024, Art. no. 13813, doi: 10.1038/s41598-024-64004-9.
[8] S.-M. Jeong, Y.-D. Song, C.-L. Seok, J.-Y. Lee, E. C. Lee, and H.-J. Kim, "Machine learning-based classification of Parkinson's disease using acoustic features: Insights from multilingual speech tasks," *Comput. Biol. Med.*, vol. 182, 2024, Art. no. 109078, doi: 10.1016/j.compbiomed.2024.109078.
[9] F. Amato, G. Saggio, V. Cesarini, G. Olmo, and G. Costantini, "Machine learning- and statistical-based voice analysis of Parkinson's disease patients: A survey," *Expert Syst. Appl.*, vol. 219, 2023, Art. no. 119651, doi: 10.1016/j.eswa.2023.119651.
[10] J. Skibińska and J. Hosek, "Computerized analysis of hypokinetic dysarthria for improved diagnosis of Parkinson's disease," *Heliyon*, vol. 9, no. 11, 2023, Art. no. e21175.
[11] J. M. Templeton, C. Poellabauer, and S. Schneider, "Classification of Parkinson's disease and its stages using machine learning," *Sci. Rep.*, vol. 12, 2022, Art. no. 14036, doi: 10.1038/s41598-022-18015-z.
[12] S. Anap, S. Jondhale, B. Agarkar, P. Deshmukh, and R. Kshirsagar, "Parkinson's disease detection from speech using combination of empirical wavelet transform and Hilbert transform," *Int. J. Speech Technol.*, vol. 28, pp. 185–194, 2025, doi: 10.1007/s10772-025-10172-6.
[13] E. Tulchinskii et al., "Topological data analysis for speech processing," in *Proc. Interspeech*, 2023, pp. 311–315, doi: 10.21437/Interspeech.2023-1861.
[14] C. Gonzalez-Rátiva and E. Nöth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proc. 9th Int. Conf. Lang. Resour. Eval.*, 2014, pp. 342–347.
[15] M. Rey-Paredes, C. J. Pérez, and A. Mateos-Caballero, "Time series classification of raw voice waveforms for Parkinson's disease detection using generative adversarial network-driven data augmentation," *IEEE Open J. Comput. Soc.*, vol. 6, pp. 72–84, 2025, doi: 10.1109/OJCS.2024.3504864.
[16] N. P. Narendra, B. Schuller, and P. Alku, "Detection of Parkinson's disease from speech using GAN-generated data," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 1925–1936, 2021.
[17] H. Edelsbrunner and J. Harer, "Persistent homology: A survey," *Contemporary Math.*, vol. 453, pp. 257–282, 2008.
[18] F. Takens, "Detecting strange attractors in turbulence," in *Proc. Dynamical Syst. Turbulence*, 1981, vol. 898, pp. 366–381.
[19] R. Ghrist, *Elementary Applied Topology*. Providence, RI, USA: Amer. Math. Soc., 2014.
[20] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 77–102, 2015.
[21] P. Bubenik, "The persistence landscape and some of its properties," in *Proc. Topological Data Anal.*, N. G. Baas, G. Carlsson, M. Quick Szymik, and M. Thaule, Eds., Cham, Switzerland: Springer, 2020, vol. 15, pp. 97–117, doi: 10.1007/978-3-030-43408-3_4.
[22] H. Adams et al., "Persistence images: A stable vector representation of persistent homology," *J. Mach. Learn. Res.*, vol. 18, no. 8, pp. 1–35, 2017.
[23] D. Barnes, L. Polanco, and J. A. Perea, "A comparative study of machine learning methods for persistence diagrams," *Front. Artif. Intell.*, vol. 4, 2021, Art. no. 681174, doi: 10.3389/frai.2021.681174.
[24] N. Atienza, R. Gonzalez-Díáz, and M. Soriano-Trigueros, "On the stability of persistent entropy and new summary functions for topological data analysis," *Pattern Recognit.*, vol. 107, 2020, Art. no. 107509, doi: 10.1016/j.patcog.2020.107509.
[25] T. Fireaizen, S. Ron, and O. Bobrowski, "Alarm sound detection using topological signal processing," in *Proc. 2022 IEEE Int. Conf. Acoust., Speech Signal Process.*, Singapore, 2022, pp. 3069–3073, doi: 10.1109/ICASSP43922.2022.9747228.
[26] M. L. Mehta, *Random Matrices*, 3rd ed. Amsterdam, The Netherlands: Elsevier, 2004.
[27] J. Bun, J.-P. Bouchaud, and M. Potters, "Cleaning large correlation matrices: Tools from random matrix theory," *Phys. Rep.*, vol. 666, pp. 1–109, 2017.
[28] G. Livan, M. Novaes, and P. Vivo, *Introduction to Random Matrices: Theory and Practice*. New York, NY, USA: Springer, 2018.
[29] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, "Random-matrix theories in quantum physics: Common concepts," *Phys. Rep.*, vol. 299, no. 4-6, pp. 189–425, 1998.
[30] P. Šeba, "Random matrix analysis of human EEG data," *Phys. Rev. Lett.*, vol. 91, no. 19, Nov. 2003, Art. no. 198104, doi: 10.1103/PhysRevLett.91.198104.
[31] H. Liu et al., "A data driven approach for resting-state EEG signal classification of schizophrenia with control participants using random matrix theory," 2017, *arXiv:1712.05289*.
[32] M. Pajovic, "The development and application of random matrix theory in adaptive signal processing in the sample deficient regime," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2024.
[33] B. McFee et al., "Librosa: Audio and music signal analysis in Python," in *Proc. 14th Python Sci. Conf.*, 2015, pp. 18–25.
[34] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 884–893, Apr. 2010.
[35] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," in *Proc. Int. Conf. Learn. Representations*, 2018, *arXiv:1711.04340*.
[36] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, May 2012.
[37] G. Carlsson, "Topology and data," *Bull. Amer. Math. Soc.*, vol. 46, no. 2, pp. 255–308, 2009.
[38] N. D. Pah, M. A. Motin, and D. K. Kumar, "Phonemes based detection of Parkinson's disease for telehealth applications," *Sci. Rep.*, vol. 12, 2022, Art. no. 9687, doi: 10.1038/s41598-022-13865-z.
[39] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Spectral and cepstral analyses for Parkinson's disease detection in Spanish vowels and words," *Expert Syst.*, vol. 32, no. 6, pp. 688–697, 2015.
[40] B. Karan, S. Sahu, and K. Mahto, "Parkinson disease prediction using intrinsic mode function based features from speech signal," *Biocybernetics Biomed. Eng.*, vol. 40, no. 1, pp. 249–264, 2020.

[41] B. Karan, S. Sahu, J. R. Orozco-Arroyave, and K. Mahto, "Non-negative matrix factorization-based time–frequency feature extraction of voice signal for Parkinson's disease prediction," *Comput. Speech Lang.*, vol. 69, 2021, Art. no. 101216.
[42] N. Narendra, B. Schuller, and P. Alku, "The detection of Parkinson's disease from speech using voice source information," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 1925–1936, 2021.
[43] B. Karan and S. Sahu, "An improved framework for Parkinson's disease prediction using variational mode decomposition-hilbert spectrum of speech signal," *Biocybernetics Biomed. Eng.*, vol. 41, no. 2, pp. 717–732, 2021.
[44] J. C. Vasquez-Correa, T. Arias-Vergara, M. Schuster, J. R. Orozco-Arroyave, and E. Nöth, "Parallel representation learning for the classification of pathological speech: Studies on Parkinson's disease and cleft lip and palate," *Speech Commun.*, vol. 122, pp. 56–67, 2020.
[45] A. M. Garcıá et al., "How language flows when movements don't: An automated analysis of spontaneous discourse in Parkinson's disease," *Brain Lang.*, vol. 162, pp. 19–28, 2016.
[46] A. M. Garcıá et al., "Cognitive determinants of dysarthria in Parkinson's disease: An automated machine learning approach," *Movement Disord.*, vol. 36, no. 12, pp. 2862–2873, 2021.
[47] C. Quan et al., "End-to-end deep learning approach for Parkinson's disease detection from speech signals," *Biocybernetics Biomed. Eng.*, vol. 42, no. 2, pp. 556–574, 2022.
[48] S. Anap, S. Jondhale, B. Agarkar, and S. Chaudhari, "Parkinson's disease detection from speech using combination of empirical wavelet transform and Hilbert transform," *Int. J. Speech Technol.*, vol. 28, pp. 185–194, 2025, doi: 10.1007/s10772-025-10172-6.

ANDY DOMÍNGUEZ-MONTERROZA received the B.Sc. degree in mechanical engineering with thesis laureate from the Universidad Nacional de Colombia, Bogotá, Colombia, and the M.Sc. degree in applied mathematics and advanced physics from Universitat de les Illes Balears, Palma de Mallorca, Spain, supported by the Ibero-American Santander Bank Scholarship. He is currently working toward the Ph.D. degree in mathematical engineering, statistics, and operations research with Universidad Complutense de Madrid, Madrid, Spain. His research interests include integrating modern mathematical and statistical methods with data science and artificial intelligence for the analysis of complex systems.

ALFONSO MATEOS CABALLERO received the M.Sc. degree in mathematical science from the University of Extremadura, Badajoz, Spain, in 1991, and the Ph.D. degree in informatics from Universidad Politécnica de Madrid, Madrid, Spain. He is currently a Full Professor with Artificial Intelligence Department, Polytechnic University of Madrid, Madrid. He has written more than 100 papers in Spanish and international journals 44 of which are JCR listed, such as EJOR, *Computational Optimization and Applications, Annals of Operations Research,* JORS, *Reliability Engineering and System Safety,* DSS, GDN, and *Computers and Operations Research*. He has participated in more than 64 projects four of which are European projects. He has coauthored five books and was reader of a paper more than 180 times in conferences.

ANTONIO JIMÉNEZ-MARTÍN received the Ph.D. degree in computer science from Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2002. He is also a Principal investigator with Decision Analysis and Statistics Research Group. He was a Coordinator of the Spanish Group on Multi-Criteria Decision Making from 2018 to 2025. He is currently a Full Professor with the Department of Artificial Intelligence, UPM. He has authored or coauthored 40 journal articles indexed in JCR, and authored 24 book chapters with international publishers. He is co-author of five books on operations research and simulation methods, and has supervised five Ph.D. dissertations. His research interests include decision support systems, multi-criteria decision making, group decision making and negotiation, metaheuristics-based optimization, machine learning, and simulation methods. He was awarded the Spanish Royal Academy of Doctors Research Prize in 2003 in the category of experimental and technological sciences for his doctoral thesis. He has participated in seven European projects funded under various framework programs, nine national projects, and three regional initiatives. He has also led to 24 industrial research projects.