STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–23.
Published online in International Academic Press (www.IAPress.org)

Using reduced feature space (2D and 3D) based on entropic measures for detecting Parkinson's disease through voice

Mónica Giuliano [1,2,4] Luis Alberto Fernández [1,3] Walter Legnani [4]

[1]*Institute of Health Sciences, Universidad Nacional del Oeste, San Antonio de Padua, Argentina*
[2]*Experimental Computing Research and Development Laboratory, Universidad Nacional de Hurligham, Villa Tesei, Argentina*
[3]*Department of Engineering and Technological Research, Universidad Nacional de La Matanza, San Justo, Argentina*
[4]*Signal and Image Processing Center, Facultad Regional Buenos Aires, Universidad Tecnológica Nacional, Ciudad Autónoma de Buenos Aires, Medrano 951, Argentina*

Abstract
This research presents a proposal for the integration of different fields, including Parkinson's disease (PD) detection, acoustic voice analysis, and signal processing. The proposal entails the development of two and three dimensional parsimonious models, predicated on feature spaces constituted by variants of Shannon's permutation entropy and autocorrelation measures. These models elucidate the structural and informative nature of vocal signals in individuals with and without the disease (PD-NPD). The reduced-dimensional feature spaces (2D and 3D) are novel and were used for the automatic classification of voices using support vector machines (SVM) with polynomial kernels and cross-validation, achieving average accuracy values between 0.82 and 0.88. Furthermore, the identification of homogeneous subgroups according to the coordinates in the space of 2D characteristics represents significant progress. The variables under consideration are candidates for biomarkers of subtypes of speech disorders for Parkinson's disease. The database used is freely accessible to facilitate reproducibility. The results obtained are compared with a selection of variables proposed by other authors and also with independent databases. The proposed approach is simple and precise and shows promise for the diagnosis and monitoring of PD through the effective use of samples of the vowel /a/ of just one second with a reduced feature space that could improve clinical workflows.

Keywords Parkinson's Disease, Voice, Autocorrelation, Weighted Permutation Entropy, Band Spectral Entropy.

DOI: 10.19139/soic-2310-5070-2638

1. Introduction

Parkinson's disease (PD) is a neuro-degenerative condition that affects millions of people worldwide. Among its most debilitating manifestations are speech disorders, which can have a significant impact on the quality of life of patients. Despite advances in diagnosis and treatment, early detection of these alterations remains a challenge. In this context, voice signal analysis has emerged as a promising tool for the identification of vocal biomarkers associated with the disease.

According to the World Health Organization [1] 2022 report, PD is a neurodegenerative disorder with increasing global incidence. In 1990, there were approximately 2.5 million people diagnosed with PD and 6.1 million in 2016 [2]. Among the possible causes of PD, environmental causes such as exposure to pesticides [3] or genetic factors [4] stand out.

The diagnosis of PD is mainly based on clinical criteria defined by the International Parkinson and Movement Disorder Society (MDS). The characteristic symptoms of PD are motor symptoms such as tremor, rigidity at rest,

*Correspondence to: Mónica Giuliano (Email: mgiuliano@uno.edu.ar). Institute of Health Sciences, Universidad Nacional del Oeste, San Antonio de Padua, Argentina.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

bradykinesia, as well as cognitive and neuropsychiatric signs [5]. PD is a disease with heterogeneous features and considerable differences in its course. More homogeneous subgroups can be identified based on clinical criteria [6].

Regardless of initial symptoms, 60–80% of PD patients will experience voice changes as the disease progresses, ranging from mild hypophonia to severe symptoms that can lead to anarthria [7].

Speech analysis of people with PD has been the subject of recent research, but is not yet included in the recommendations highlighted by the MDS. Biomarkers are currently being sought to identify disease progression in relation to speech disorders or specific speech alterations in patients with PD using different analysis techniques [8, 9, 10].

Speech studies in people with PD use different acoustic measures and phonations of different types (vowels, words, sentences, spontaneous speech). Comparisons between studies are particularly challenging, as discrepancies and difficulties in replicability are observed [11].

In recent years, automatic speech classification of people with and without PD has been studied. Examples include [10, 12, 13, 14, 15, 16]. Due to its simplicity and acoustic richness, many studies use the sustained phonation of /a/ for a few seconds. These studies propose a wide variety of acoustic measures, different criteria for selecting measures, and algorithms for classifying patients with PD and subjects without Parkinson's disease [12].

Recently, machine learning and deep learning techniques have emerged as promising tools for improving PD diagnosis. In [17] a review paper presents a detailed analysis of the current state of the art of this techniques applied to PD diagnosis, focusing on voice, handwriting, and wave spiral datasets. That study also evaluates the effectiveness of various algorithms, including classifiers, on these datasets and highlights their potential in enhancing diagnostic accuracy and aiding clinical decision-making. Several usual measures in speech analysis such as jitter, shimmer and cepstral coefficients appear, however, the measures presented of the present research are not explored.

The possibility of diagnosing PD using permutation entropy measures has not been found in the available literature in the way presented here and is the main objective of the present study. Inspired by the causal plane formed by normalized permutation entropy and statistical complexity, which has been successfully studied in various biomedical and bioengineering research works (for example in [18, 19]), the aim of this study was to define a low-dimensional feature space using entropic measures to differentiate both classes of interest, that is, to allow the separation by machine learning algorithms of voices of people with PD and NPD. To this end, biomarkers have been identified to differentiate phonations of /a/, which explain part of the disturbances observed in the voices. Additionally, although the literature review in this study was not exhaustive, the results were compared with those previously achieved by other authors who used a considerably larger number of variables for the separation between the groups of interest [13, 14, 15].

The work is organized as follows: in Sections 2.1, 2.2 and 2.3, the database used, the autocorrelation and entropy measures studied, and audio preprocessing are described. In Section 2.4, the automatic classification techniques and the selected quality measures are defined. In Section 2.5, the procedure to define the feature spaces for the 2D model (Model 1) and the 3D models (Models 2 to 9) is described. In Section 2.6 analysis of the Sensitivity to Noise. In Section 2.7 comparisons of the 3D model with other authors. In Section 2.8 includes information to enable replication of the work. In Sections 3.1, 3.3 and 3.4, the results obtained with the 2D and 3D models built with polynomial SVM are shown. In Section 3.2 in HUNIPA are analysis the relationship between results and clinical symptoms obtained. In Section 3.5, the classification with other algorithms is performed, and in Section 3.6, 4D models are analyzed. In Section 3.7 analysis of the Sensitivity to Noise. As outlined in Section 3.8, the 3D model is compared to those of other authors in HUNIPA; the features used with other databases are detailed in the same section. In Section 3.9 results with HUNIPA are compared by other independent database. Finally, in Section 4, the findings are presented the discussion and in Section 5 conclusions are summarized.

2. Materials and Methods

The database was constructed using audios of sustained phonation of the vowel /a/. The audios were recorded in a controlled environment with a Focusrite Scarlett 2i2 audio interface and an AT2020 microphone (cardioid condenser microphone). The recordings were made with a depth of 16 bits, at a sampling frequency of 44100 Hz and were encoded in wav format. Autocorrelation and entropy measures were calculated from the audio signal in search of those that best differentiated the automatic classification between PD and NPD.

2.1. Database

The database considered here is part of a speech corpus that includes recorded audios of subjects with and without a diagnosis of PD. The construction was carried out in 2019 with the participation of specialists from two public hospitals, Hospital Nacional Profesor Alejandro Posadas and Hospital General de Agudos Bernardino Rivadavia, and from Universidad Nacional de La Matanza (UNLaM) [20].

The database was named HUNIPA-2019 because it was generated by public hospitals and public universities in Argentina with voices of people with and without Parkinson's disease.The dataset was recorded in compliance with the Helsinki Declaration and approved by the Hospital Rivadavia Ethics Committee.

A corpus was recorded with phonation of vowels, words and phrases from which only the phonations of /a/ were used here and is available for free download in [21].

Patients with PD received two neurological evaluations, using the Hoehn and Yahr (H&Y) scale [22] and the Spanish version of the Unified Parkinson's Disease Rating Scale (UPDRS), sponsored by the MDS [23, 24]. Patients with PD were evaluated and recorded in the "on" state, that is, with the medication taking effect and motor symptoms at their best.

The patients were examined by otorhinolaryngology specialists who performed laryngoscopy and stroboscopic examinations. As exclusion criteria, patients with PD who had more than 16 years of the disease, had undergone basal ganglia surgery, experienced a stroke, had prior voice disorders, laryngeal pathology, or other neurodegenerative diseases were discarded [25].

The non-Parkinson's cases (NPD) were recorded in a controlled environment with the equipment described previously. The whole database was evaluated by several speech therapists who have conducted perceptual studies of voice quality in different instances. In the case of NPD voices, those with severe alterations were discarded.

The database used here consists of 110 cases, 55 PD and 55 NPD. The demographic characteristics of the participants and the clinical status of the PD patients are displayed in Table 1. The PD patients have a mean age of 64 years with a standard deviation of 8 years and a mean disease duration of approximately 6 years. In the NPD cases, the mean age was 60 years, not significantly different from that of the PD cases, with no difference according to the Wilcoxon test (p-value > 0.01).

According to the UPDRS scale, PD patients presented mean values of 30.35, with values ranging from a minimum of 12 to a maximum of 53. On the H&Y scale, mean values of 1.59 were observed. These disease indicators account for patients with PD in early stages.

The construction of the database of patients with PD implied that the different specialist assessments were carried out on the same day. At the time of registration, the patients were medicated with an average amount of 800 mg of levodopa and in a state of mild symptoms.

Table 1. Demographic characteristics of the participants and clinical condition of patients with PD. Mean (deviation) is indicated.

Condition	#	Age (years)	Illness period (years)	UPDRS III	H&Y	*Levodopa* (mg)
PD	55	64.09 (8.24)	6.11 (3.89)	30.35 (11.17)	1.59 (0.59)	880.19 (479.22)
NPD	55	59.95 (11.59)				

2.2. Autocorrelation and entropy measures

Several signal measures were calculated on the database, with the aim being to find the ones that would best help to differentiate the classes of interest, i.e. subjects with PD and NPD.

The autocorrelation (AC) of a signal is a measure of the similarity between the signal and a lagged version of itself and was proposed by [26, 27].

Mathematically, the normalized autocorrelation $r_{xx}(l)$ of a discrete signal subtracting its mean value is:

$$r_{xx}(l) = \frac{\sum_{n=-\infty}^{\infty} (x_n - me)(x_{n-l} - me)}{\sum_{n=-\infty}^{\infty} (x_n - me)^2}, \tag{1}$$

where l is the time lag (or lag) and x_n represents the signal at time t_n, where n is the time index that spans the corresponding range of the signal and me is the mean value.

In information theory, Shannon's entropy [28, 29] measures the uncertainty of a message or information source. It can also be considered as the average amount of information contained in the symbols associated with the information source. If a message source produces N symbols S_i, with probability of occurrence p_i, Shannon's entropy H is calculated by the expression:

$$H = -\sum_{i=1}^{N} p_i \ln(p_i). \tag{2}$$

If the symbols are equiprobable, the entropy reaches its maximum value, $H = \ln(N)$, and if the probability is concentrated in a few symbols it takes small values.

The permutation entropy (PE) of a signal provides a measure of the predictability of the data series comprising it, where the probability is calculated in a specific way. Given a time series $(x_1, x_2, x_3, \ldots, x_N)$ the m-tuples are considered:

$$s_\tau^m(t) = (x_t, x_{t+\tau}, x_{t+2\tau}, \ldots, x_{t+(m-1)\tau}), \quad 1 \le t \le N - (m-1)\tau, \tag{3}$$

where m (embedding dimension) and τ (tau, time embedding delay) are positive integers with $1 < m$ and it is often requested that $m! \ll N$. PE was introduced by Bandt and Pompe in 2002 [30].

If each entry of $s_\tau^m(t)$ is associated with its position and the vector $s_\tau^m(t)$ is ordered in ascending order, a permutation π of $(1, 2, \ldots, m)$ is defined. With this idea, each permutation π defines a pattern in the signal. By selecting the definition of normalized Shannon's entropy and calculating the permutations following the proposal of [30], the normalized Shannon permutation entropy is obtained.

$$H_\tau^m = \frac{-\sum_\pi p_\pi \ln(p_\pi)}{\ln(m!)}, \tag{4}$$

where

$$p_\pi = \frac{\#S_\tau^m(\pi)}{N - (m-1)\tau}, \tag{5a}$$

$$S_\tau^m(\pi) = \left\{ s_\tau^m(t) : \; s_\tau^m(t)_{\pi(i)} < s_\tau^m(t)_{\pi(i+1)} \right\}. \tag{5b}$$

The PE captures the ordinal structure present in a time series, but does not take into account variations in the amplitude of pattern components. In time series, such as those derived from speech signals, the information associated with changes in magnitude is important.

In [31], an alternative way to calculate p_π is proposed, where the contribution of $s_\tau^m(t) \in S_\tau^m(\pi)$ depends on the variation of the amplitude of the components of $s_\tau^m(t)$ by incorporating a weight.

If $s_\tau^m(t)$ contributes in an amount $\omega(t)$,

$$p_\pi = \frac{\Omega(\pi)}{\Omega}, \tag{6a}$$

is defined with

$$\Omega(\pi) = \sum_{s_\tau^m(t)\in S_\tau^m(\pi)} \omega(t), \tag{6b}$$

and

$$\Omega = \sum_{t=1}^{N-(m-1)\tau} \omega(t). \tag{6c}$$

In this work, the following expression was used to calculate the weights

$$\omega(t) = \sum_{k=1}^{m} \left(x_{t+k\tau} - M(t)\right)^2 \tag{7a}$$

and

$$M(t) = \frac{1}{m}\sum_{k=1}^{m} x_{t+k\tau}, \tag{7b}$$

which defines the weighted permutation entropy (WPE).

Slope entropy (SlopePE) is another alternative that takes into account variations in amplitude. Basically, for $\tau = 1$, the procedure consists of calculating the permutation entropy on the slope series quantified from the original time series. The slope is estimated with the previous sample and a quantifier Q with 5 levels is used [32]. In general, the above indicated procedure is repeated to obtain the permutation entropy on the $(m-1)$-tuples $Q_\Delta s_\tau^m(t)$ where:

$$Q_\Delta s_\tau^m(t) = \left(Q(x_{t-2\tau} - x_{t-\tau}), Q(x_{t-3\tau} - x_{t-2\tau}), \ldots, Q(x_{t-m\tau} - x_{t-(m-1)\tau})\right), \tag{8}$$

with $Q(\alpha) \in \{-2, -1, 0, 1, 2\}$.

Spectral entropy (SE) provides a measure of the spectral distribution of the signal energy. If $(f_1, f_2, \ldots, f_N)$ are the frequency components of a given signal $(x_1, x_2, \ldots, x_N)$, then the SE is Shannon's entropy of the set of frequencies $1, 2, \ldots, N$ with which the energy proportion can be calculated to define the "probabilities" given by [33]:

$$p_i = \frac{|f_i|^2}{\sum_{i=1}^{N} |f_i|^2}, \quad 1 \le i \le N. \tag{9}$$

If the set of values in the expression (9) is restricted to those corresponding to a given frequency band, a measure of the spectral distribution of the energy of the signal under analysis in the band considered is obtained.

2.3. Preprocessing

Measures were calculated on a 1 second signal length taken from the central part of the recordings. Autocorrelation and permutation, weighted permutation and slope entropies were calculated for different values of their parameters. In general, formants are better defined in healthy voices than in pathological voices [34]. Additionally, hypernasality disorders may appear in pathological voices, which manifest as a formant below 500 Hz for /a/. Therefore, having measures of the spectral distribution of energy in bands that contain formant information may result in biomarkers of interest for this study. To capture differences in the spectral distribution of energy between PD and NPD, total spectral entropy (SE) and 9-band spectral entropy (BSEi) were calculated [35].

Since the fundamental frequency F0, amplitude and formants [36] are temporally less stable in PD compared to NPD, entropic measures of the F0, amplitude and formants F1 and F2 curves were calculated.

Table 2 shows a summary of the measures calculated with the parameters explored to build the feature space with which the machine learning algorithm was fed to construct the models. The criteria for selecting parameter ranges (such as *m* and *tau*) followed the recommendations of Riedl et al. [37] and Myers and Khasawneh [38].

Table 2. Variable names and parameters of the features explored.

Variables	Name	Parameters explored
WPE	Weighted Permutation Entropy	m: 3 to 6; tau: 1 to 15
AC	Autocorrelation	lag: 1 to 50
PE	Permutation Entropy	m: 3 to 6 tau: 1 to 15
SlopePE	Slope Entropy	m: 3 to 6; tau: 1 to 15
SE	Total Spectral Entropy	-
BSEi	Band Spectral Entropy	Bands between: 0, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000 Hz.
AWPE	Amplitude Weighted Permutation Entropy	m: 3; tau: 1 to 15
F0WPE	Frequency F0 Weighted Permutation Entropy (fundamental frequency)	m: 3; tau: 1 to 15
F1WPE	F1WPE Frequency F1 Weighted Permutation Entropy (lowest frequency formant)	m: 3; tau: 1 to 15
F2WPE	F2WPE Frequency F2 Weighted Permutation Entropy (second frequency formant)	m: 3; tau: 1 to 15

2.4. Classification models and model quality measures

Machine learning (ML) algorithms are reliable even in small samples, as argued by [39]. Here, for the automatic classification between phonations of people with and without PD, a group of three supervised classification techniques were applied: support vector machine (SVM), k-Nearest Neighbors (kNN), and Random Forest (RF). The choice of these techniques was made based on the fact that each of them uses a different principle for the operation of its algorithm for classification.

The selected algorithms (SVM, RF, kNN) are relevant for classification and are among the most widely used in the literature, especially for classification of voices of people with and without Parkinson's disease [17].

In this work, the R language "caret" packages have been used for binary classification algorithms. Random Forest [47], SVM [48] and kNN [49] were used. For a more in-depth and detailed treatment of these classifiers, see references [50, 51].

Linear SVM was initially thought of as a binary classification algorithm in an n-dimensional space that defines a hyperplane separating the classes. This hyperplane is chosen in a way that it maximizes the margin between the classes, that is, the distance between the closest data points of the two different groups. These points that are close to the hyperplane are called support vectors, and they are crucial for the definition of the optimal hyperplane. There are different types of SVMs, mainly differentiated by the type of kernel they use. Most of them are linear, polynomial and radial (method = "svmPoly"; "svmRadial"). SVM is a very stable methodology [40, 41, 42].

kNN classifies an object according to the most common class among its k nearest neighbors. This non-parametric method is simple to implement, robust to noisy training data, and effective. This type of classifier is memory-based, does not require any model for tuning, and has proven successful in a large number of classification problems [43, 44]. The employed methodology is the "kknn" method from the caret package.

The RF method was proposed by [45]. It is a method that uses ensembles of decision trees in splits with randomly generated vectors or random subsets of training data, and calculates a score based on the components of each tree. In the context of classification, the prediction made by RF results from the most frequent class in the set of predictions obtained by each decision tree [46]. In this case, the "rf" method from the caret package is employed.

To assess the quality of the automatic classification, indicators calculated from the confusion table [50] were used. In the classification, a positive label was assigned to the PD condition and a negative label to the NPD condition. The possible predictions can be: true positive (TP), true negative (TN), false positive (FP) or false negative (FN). To measure the quality of the classification, the following indicators were used: precision (Acc) and F1 score (F1). Acc is defined as the proportion of well predicted data, i.e. Acc=(TP + TN)/(TP+FP+TN+FN). The F1 score is the harmonic mean between the precision given by TP/(TP+FP) and the retrieval defined by TP/(TP+FN), that is, F1=2×(precision×retrieval)/(precision+retrieval). Also, Receiver Operating Characteristics (ROC) was generated and the Area Under The Curve (AUC).

To avoid overfitting in the classification, cross-validation (CV) was used in the training process. This involves dividing the data into multiple subsets (partitions), training the model multiple times (iterations) with different combinations of hyperparameter subsets, and evaluating its performance [50]. For each subset of hyperparameters, 10 partitions with 10 randomly selected iterations were used, which allowed the mean Acc to be calculated and then the maximum mean Acc obtained to be used to define the optimal model (method = "repeatedcv", number = 10, repeats = 10).

2.5. Analysis procedure

Considering that the objective of this research is to establish models for classification between the classes of interest (PD and NPD) from feature spaces with the minimum possible number of variables, autocorrelation (AC) and weighted permutation entropy (WPE) were in first place analyzed.

Significant differences were observed in the magnitude of the first coefficients of the signal autocorrelation in several numbers of PD versus NPD cases (Figure 1). Differences were also observed in the weighted permutation entropy (Figure 2). In preliminary tests, AC and together with WPE were found to perform better, with respect to the average Acc using the polynomial SVM algorithm.

To select the parameters of the measurements (*m*, *tau*, *lag*), it was adapted a technique successfully used in other contexts (e.g., [19]). This selection was made seeking its remarkable conceptual simplicity, computational speed, and robustness to noise.

Once AC and WPE were selected as the main variables of the feature space, their distribution, their mean values and their dispersion for different variations of their corresponding parameters (*lag*, *m*, *tau*) were studied separately. This made it possible to determine the range of parameters to define the best model, using a lag from 1 to 50 for AC and, for WPE, an embedding dimension (*m*) from 3 to 6 and an embedding time delay (*tau*) from 1 to 15.

Figure 1 shows that PD cases generally tend to have higher autocorrelation values compared to NPD cases. At mean values, the autocorrelation was higher for all lags in PD cases.

Figure 2 shows the WPE mean calculated with an embedding dimension (m) of 6 for different values of time delay (tau), for which values between 1 and 15 were used, showing lower mean values in PD cases.

Figure 3 presents the procedure scheme followed to select the best pair of parameters for AC and WPE. First, the signal is processed considering the lag variation for AC using values from 1 to 50 and WPE parameters with *m* ranging from 3 to 6 and *tau* ranging from 1 to 15. Then, using polynomial SVM with CV, the average Acc is calculated for each partition and the highest value in the iterations was identified. Then, using polynomial SVM with CV, the mean Acc was calculated for each partition and the highest value in the iterations was found to select the model's parameters.

Following the procedure described above, the highest mean Acc was obtained for WPE with tau=13, m=6 and for AC with lag=13. With these parameters, Model 1 was then defined as the WPE x AC plane. It proved stable, varying between 1 and 1.5% in response to variations of 15% in tau and 23% in lag parameters.

By adding a dimension to Model 1, 3-dimensional feature spaces were explored. For this purpose, other entropic measures were considered for the third variable and, by varying their parameters, the model was optimized using a procedure analogous to that used for Model 1. That is to say, the maximum mean Acc was sought using SVM with

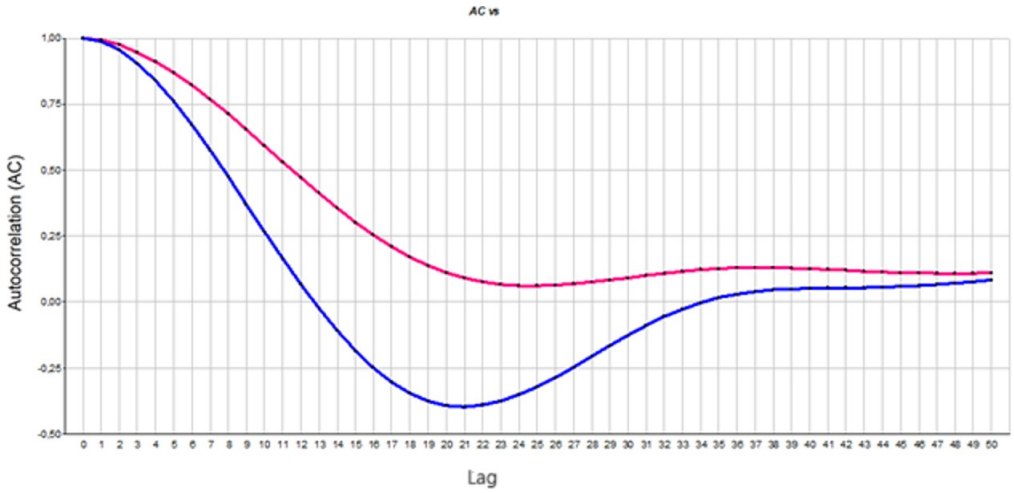

Figure 1. Autocorrelation mean, of the whole databse, for different lags according to disease condition. Red PD (upper curve), Blue NPD (lower curve).

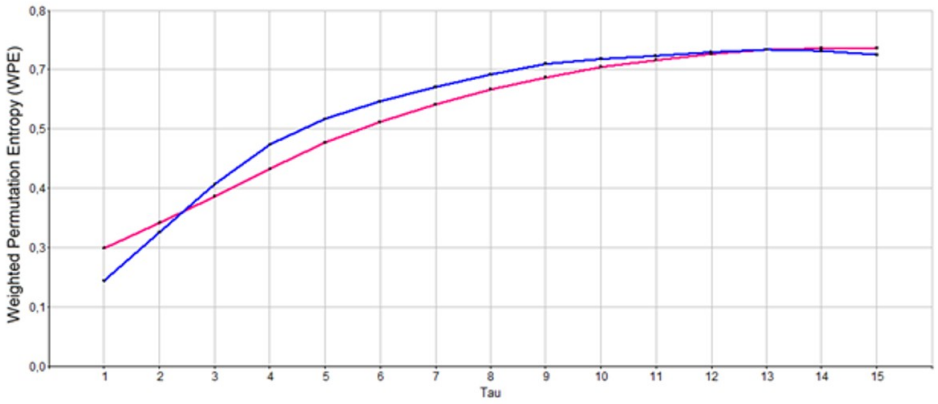

Figure 2. Weighted entropy mean with embedding dimension 6 (WPE), for the whole database, for different embedding times delays (tau) (Red PD, Blue NPD).

CV and the parameters indicated in Table 2. Eight different and parsimonious 3-dimensional models were obtained from the corresponding feature spaces (Models 2 to 9, see for details Table 3).

To analyze the performance of the machine learning algorithm and validate the results, a classification using other techniques (radial SVM, RF, and kNN) was performed on the 9 models presented in Table 6.

Finally, the sample of 110 elements was randomly divided, with 80% used for training and 20% for testing. This partition was used to train and test the 9 proposed models (Table 7) using an SVM with a polynomial kernel, and the hyperparameters of the best model were calculated. For the R caret algorithm, the hyperparameters are: d (degree of the polynomial kernel), γ (scale factor), and C (cost that controls the penalty applied to misclassifications) [48]. The hyperparameters explored were: d: $\{1, 2, 3\}$; γ: $\{0.001, 0.01; 1\}$; C: $\{0.25, 0.5, 1\}$.

Next, 4-dimensional feature spaces were analyzed by adding one dimension to the best 3D model obtained (Model 2). The mean accuracy (Acc) of the 4D spaces was calculated, and none of them performed better than Model 2.

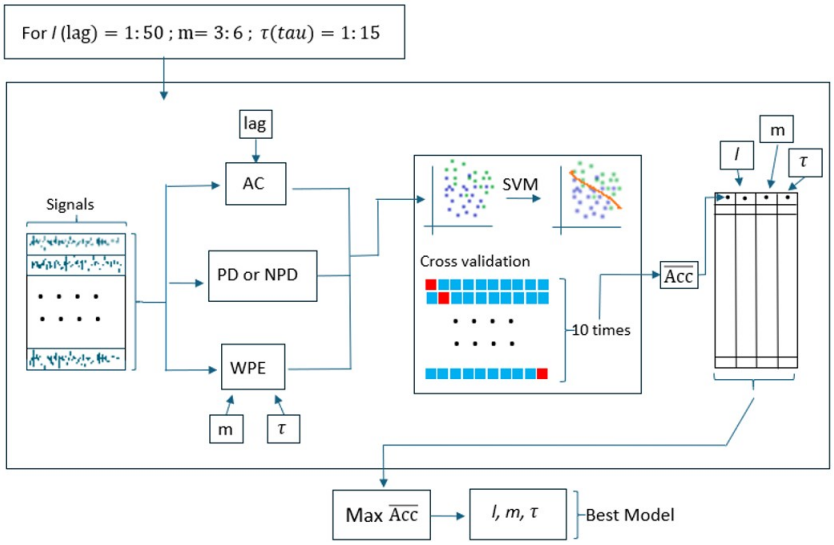

Figure 3. Scheme to explore the autocorrelation (AC) and the weighted permutation entropy (WPE) parameters, which enable the optimization of the 2D space (Model 1), using polynomial SVM. $\overline{\text{Acc}}$ refers to Acc Mean accuracy.

2.6. Sensitivity of the 3D Model to Noise

In order to build large audio sample bases, for remote monitoring or self-tracking, it is important to analyze the performance of classification algorithms with lower quality samples than those used in this work. For example, in PD voices, signs are observed in high frequency bands [52] where different types of noise have an influence. When audios are recorded in uncontrolled environments, the pathological characteristics of the voice are mixed with characteristics of the ambient noise, which in general could negatively affect performance. In addition, spectral distortions produced by the use of low-quality recording devices could also reduce performance.

Since for this research the database does not have low quality recordings, in order to have a reference of the impact of noise on the measurements, computer simulated noise (white noise and pink noise) [60] and noise taken from a real environment were added to the audios used in this work, in different signal-to-noise ratios. The measurements used by the 3D algorithm were calculated from the distorted audios and the validation was repeated with the same audios used in the validation stage without noise.

2.7. Comparisons of the 3D model with other authors and other databases

Section 3.2 analyzes the relationship between the obtained results in section 3.1 and clinical symptoms. It considers the previous study by Giuliano et al. [25], which was conducted at HUNIPA and evaluated perceptual voice quality and estimated the integrated perturbation index (IPI).

To define their classification models, other authors propose higher-dimensional feature spaces than those considered here in HUNIPA-2019 database. In particular, [13] proposes around 300 measures derived from the sustained phonation of the vowel /a/, which can be calculated using the author's own software (Voice Analysis Toolbox). Using these measures as a starting point, [14] and [15] propose a selection of 10 or 11 variables.

In Section 3.8, using the variables proposed in the aforementioned works as a starting point, which represent 10D and 11D models, the classification results are shown using the database employed in this research, applying SVM with a polynomial kernel and using CV. The mean Acc values obtained are compared with the best feature space (3D) previously established.

In the study by [15], the 11 variables group measures are categorized into four distinct types: jitter, shimmer, harmonic noise, and cepstral measures. In the Section 3.6, the HUNIPA database was utilized to examine 4D models, with each of the 11 variables identified serving as the fourth variable in the analysis.

Guatelli et al. [53] presents the analysis of voice recordings from the HUNIPA database through spectrogram analysis and convolutional neural networks (CNN). A multilayer neural network is used to classify convolutional feature vectors and is compared with the use of extreme learning machines (EML). Several experiments have been conducted considering four types of spectrograms with the convolutional neural network models AlexNet, VGG-16, SqueezeNet, Inception V3, and ResNet-50. Section 3.8 shows the best results from Guatelli 2023 in the HUNIPA database.

In Section 3.9, an analysis of the measures proposed in the classification of individuals is conducted from two independent databases.To ensure a comprehensive comparison, two datasets that have been thoroughly analyzed in the literature were considered: GITA [54] and NeuroVoz [55] .

The GITA dataset was recorded by the Noel Clinic in Medellín (Colombia Republic) from 100 native Colombian Spanish speakers, with 50 healthy controls (NPD) and 50 patients with Parkinson's disease (PD). The NeuroVoz dataset was compiled by the Polytechnic University of Madrid in collaboration with the Gregorio Marañón Hospital in Madrid, Spain. This dataset includes material from 107 native adult speakers of Castilian Spanish (56 NPD and 51 PD).

The recordings for both corpora were obtained under controlled environmental conditions, with a sampling rate of 44.1 kHz and 16-bit quantization. Both datasets were recorded in accordance with the Declaration of Helsinki and were approved by their respective ethics committees.

The databases were used in the bibliography by several authors. In this research were comapared the results presented in Ibarra et al. [56], where a set deep learning (DL) architectures were used for classification.

2.8. Reproducibility of results

For the reproducibility of the work in [21], the audios of the database are freely accessible at the following link:
https://figshare.com/articles/dataset/Audios_vowel_A_PD/21453867/1

The algorithms utilized are available in the open-source toolkit designed for the analysis of entropic data, which can be found at the following link:
https://www.entropyhub.xyz/

The parameterized HUNIPA database and preprocessing algorithms can be found at the following link: https://drive.google.com/drive/folders/19GQjw1i9v7SinjlLs6vhlkAhhdYqgBgu?usp=sharing

The first author can also be contacted directly for further references.

3. Results

With the procedure proposed, it was possible to define one 2D model and 8 3D models. The results showing the best quality measures in the PD and NPD classification were those of Model 1 and Model 2, giving two feature spaces which will be discussed in the conclusions section.

3.1. WPE x AC 2D feature space (Model 1)

Model 1 was defined using the 2D feature space defined by WPE with m=6 and tau=13 and AC taking a lag=13. AC shows differences in the medians between people with and without PD according to Wilcoxon's median difference test for independent samples (p-value < 0.0001). In contrast, the WPE with m=6 and tau=13 does not show any significant differences for the same test (p-value = 0.236).

Figure 4 shows the WPE x AC plane with the identification of disease condition (PD and NPD), in which 4 subgroups can be observed within the cases. A subgroup of PD has a higher AC (SG1-PD) and another one a higher WPE (SG2-PD), showing a clear difference with NPD in both cases. The lower AC and WPE values mostly correspond to NPD people (SG3-NPD). There is also a clear fourth subgroup in which the PD cases do not differ from NPD cases (SG4).

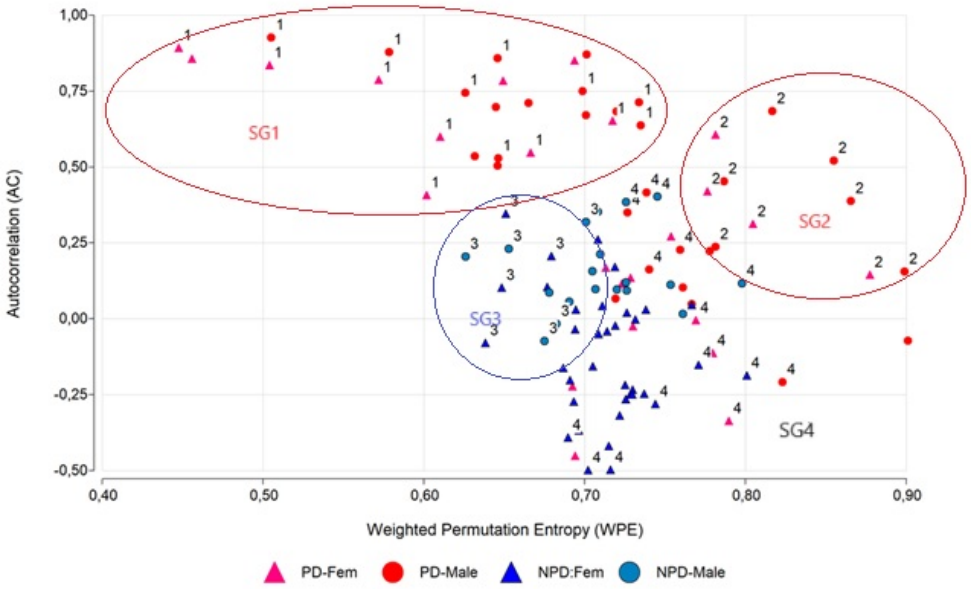

Figure 4. WPE x AC feature space (Model 1): Weighted Permutation Entropy vs Autocorrelation, group number (SG:1;2;3;4), according to disease condition (red PD, blue NPD) and gender (triangle female, circle male).

Table 3. Distribution of the groups according to gender, disease condition and mean in the WPE x AC feature space.

SG	# F	# M	# Total	AC Mean	WPE Mean	Disease condition
SG1	10	15	25	0.717	0.632	PD
SG2	3	8	11	0.332	0.831	PD
SG3	13	11	24	0.084	0.685	NPD
SG4	34	16	50	-0.044	0.738	PD-NPD
Total	60	50	110	0.194	0.711	

Table 3 shows the number of individuals split by gender of each SG and the mean of each group. SG4 has composed by most females and SG2 has a great quantity of males, more than half. SG1 is slightly more composed by males meanwhile SG3 slightly composed by more female.

Table 3 also shows the mean of each group in the characteristics space. The difference of means according to SG is significant by the Lawley-Hotelling statistical test ($p < 0.0001$). There are also differences in the total group (WPE x AC) by gender ($p < 0.01$), but there are no differences by gender combined with SG and neither gender combined with disease condition ($p > 0.01$).

Using the feature space, the best Acc mean value of 0.826 with a deviation of 0.102 was obtained for Model 1 (WPE x AC), using SVM with a polynomial kernel.

3.2. Relationship between results and clinical symptoms

The proposed characteristic measures in Model 1 and the four subgroups were analyzed according to the different variables associated with clinical symptoms in patients with PD in HUNIPA. The nonparametric Kruskal-Wallis test was used to analyze differences in means according to the three SG subgroup with PD (SG1, SG2 and SG4). SG1 and SG2 groups did not show significant differences in disease progression (UPDRS and H&Y) or disease

duration. In turn, the SG4 group stood out with significantly lower differences than SG1 and SG2 in the same disease characteristics (p-value < 0.05).

Variations were identified among the SG groups with respect to age, with a younger demographic observed in SG3 (Mean: 55 years) and no substantial differences observed among the other groups (SG1 and SG2).

Integrated Perturbation Index acoustic analysis (IPI) and perceptual evaluation of voice quality were carried out by a speech therapy specialist [25]. The voice perturbations are defined as any abnormality in the vocal quality, pitch, loudness or vocal effort that disturbs communication or generates a negative effect on the voice-related quality of life.

IPI is a quantitative indicator that allows voice signals to be classified into levels of acoustic alteration: normal, risk zone and severe [57], [58]. The Acoustic Analysis and Graphing of Speech Signals (ANAGRAF) software was used to estimate the IPI values. Table 4 reveals a moderate relationship between IPI and SG, with a Cramér's contingency coefficient of 0.33.

Table 4. Integrated Perturbation Index acoustic analysis (IPI) and SG (N=55).

Condition IPI	SG1	SG2	SG4	Total
Normal	1	1	5	7
Risk	14	1	12	27
Alteration	10	9	2	21
Total	25	11	19	55

The perceptual analysis was carried out by a speech therapy specialist, who rated the stimuli in GRBAS (Dysphonia Grade, Roughness, Breathiness, Asthenia and Strain) scale with the help of the Audio-Perceptual Evaluation System (EVAPER) [59]. The value of alteration was classified as low, mild or severe. As shown in Table 5 there was a low relationship between perceptual evaluation and SG, with a Cramér's contingency coefficient of 0.24.

Table 5. Perceptual evaluation of voice quality an SG (N=55).

Alteration perceptual	SG1	SG2	SG4	Total
Low	20	5	17	42
Mild	2	4	2	8
Severe	3	2	0	5
Total	25	11	18	55

3.3. 3D feature spaces (Models 2 to 9)

Table 6 shows the results obtained with the 3-dimensional models. The 3D models were nominalized according to the descending order of the mean Acc from Models 2 to 9. For each model, the optimized features or parameters can be observed in each component of the feature space. The mean Acc ranges from 0.879 to 0.821, with similar deviations, in all cases, of approximately 0.1.

Summarizing, the best 3D model is Model 2 (WPE $\times$ AC $\times$ BSE), which is the space formed by the AC (*lag*=13), WPE (*m*=6 and *tau*=13) and BSE4, the spectral entropy of the band ranging from 2000 to 3000 Hz. Figure 5 shows the differences in the values of the variables in Model 2 according to disease condition (PD-NPD).

Table 6. Parameters selected for the 3rd variable added to the 2D space in order to generate a 3D space, with the entire sample and with maximum mean Acc and deviation using SVM with a polynomial kernel.

Model	3rd variable	Variable name	Selected parameters	Mean Acc (Max obtained)	Deviation
Model 2	Spectral entropy by bands	BSE4	Spectral Band 4 between 2000 and 3000 Hz	0.879	0.094
Model 3	Entropy on amplitude curve	AWPE	m=3, tau=2	0.867	0.102
Model 4	F2WPE Entropy on F2 frequency curve	F2WPE	m=3, tau=9	0.843	0.103
Model 5	Entropy on F1 frequency curve	F1WPE	m=3, tau=15	0.841	0.094
Model 6	Permutation Entropy	PE	m=6, tau=13	0.835	0.111
Model 7	Spectral entropy	SE	-	0.824	0.097
Model 8	SlopePE – Slope entropy	SlopePE	m=6, tau=7	0.830	0.119
Model 9	Entropy on F0 frequency curve	F0WPE	m=3, tau=13	0.821	0.107

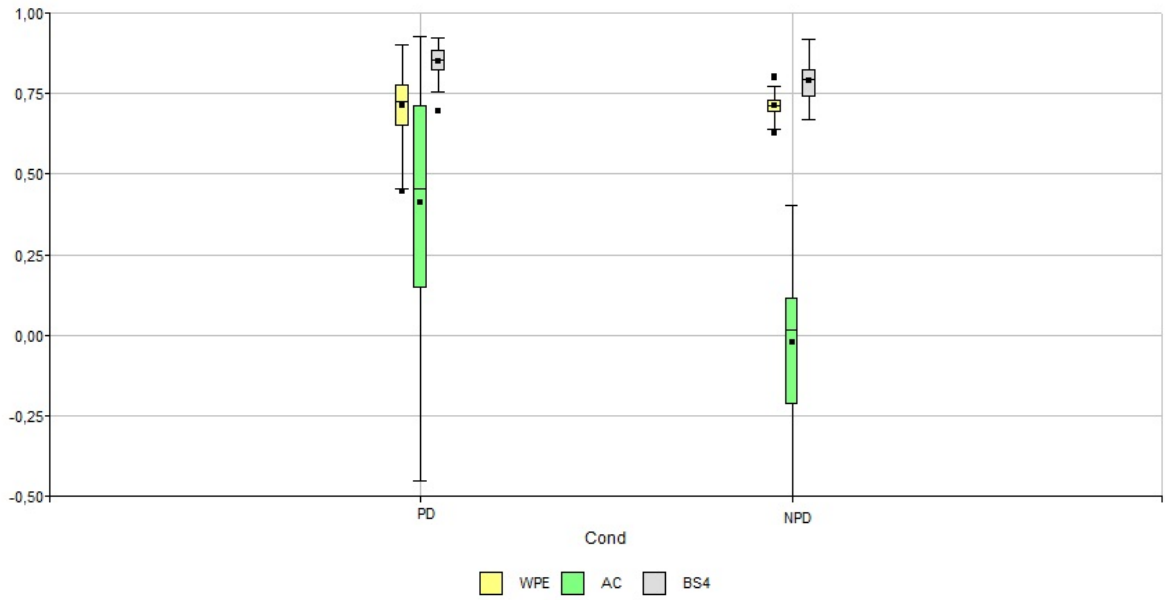

Figure 5. Boxplot WPE (Weighted Permutation Entropy) and AC (Autocorrelation) and BS4 (Band Spectral Entropy 2000-3000 Hz), according to disease condition (PD, NPD).

3.4. Hyperparameters of the defined Models and quality measures

Table 7 presents the hyperparameters (d, γ, C), adjusted with CV for SVM with a polynomial kernel, in each of the defined feature spaces models. The Acc values are within the expected order but slightly lower than the mean value shown in Table 6.

Table 7. Description of the 2D and 3D models. Hyperparameters for the best polynomial SVM model trained with 80% of the sample and CV: d (degree of the polynomial kernel), γ (scale factor), and C (cost that controls the penalty applied to misclassifications). Quality measures with the 20% test: Acc, F1 and AUC.

Models		Polynomial SVM hyperparameters with 80% training (N=88) (d - γ - C)	Quality measures with the 20% test (N=22)		
			Acc	F1score	AUC
Model 1	WPE × AC	3 - 0.1 - 0.5	0.773	0.706	0.818
Model 2	WPE × AC × BS4	1 - 0.1 - 0.5	0.864	0.842	0.909
Model 3	WPE × AC × AWPE	2 - 0.1 - 1	0.818	0.778	0.892
Model 4	WPE × AC × F2WPE	3 - 0.1 - 0.5	0.818	0.778	0.818
Model 5	WPE × AC × F1WPE	2 - 0.1 - 1	0.773	0.737	0.810
Model 6	WPE × AC × PE	2 - 0.1 - 1	0.727	0.625	0.810
Model 7	WPE × AC × SE	2 - 0.1 - 1	0.818	0.778	0.818
Model 8	WPE × AC × SlopePE	2 - 0.1 - 1	0.773	0.706	0.818
Model 9	WPE × AC × F0WPE	2 - 0.1 - 0.25	0.682	0.667	0.818

3.5. Model comparison using other classification techniques: kNN, radial SVM and RF

In order to compare the results obtained with models 1 to 9 using SVM with a polynomial kernel, the procedures are repeated with other classification models and CV: SVM with a radial kernel, RF and kNN. Table 8 presents the results obtained.

Table 8. Mean accuracy of the Classification with SVM, RF and kNN for the techniques used for the comparison of the performance of the classification of the 2D and 3D Models.

Model	Space	Acc Polynomial SVM	Acc Radial SVM	Acc RF	Acc kNN
Model 1	WPE × AC	0.826	0.830	0.782	0.798
Model 2	WPE × AC × BS4	0.879	0.857	0.806	0.830
Model 3	WPE × AC × AWPE	0.867	0.822	0.791	0.802
Model 4	WPE × AC × F2WPE	0.843	0.816	0.795	0.789
Model 5	WPE × AC × F1WPE	0.841	0.801	0.777	0.792
Model 6	WPE × AC × PE	0.835	0.828	0.791	0.799
Model 7	WPE × AC × SE	0.824	0.843	0.788	0.840
Model 8	WPE × AC × SlopePE	0.830	0.850	0.821	0.862
Model 9	WPE × AC × F0WPE	0.821	0.805	0.831	0.799

The results are consistent and, in general, a better mean Acc is observed using SVM with a polynomial kernel and results a worse one with RF. Only for the spectral entropy (SE) and the slope entropy (SlopePE) variables have better mean Acc results been observed with kNN and radial SVM.

3.6. 4D Model evaluation

With the addition of the third feature from Models 3 to 9 to Model 2 (WPE × AC × BSE), 4-dimensional feature space models were studied. Table 9 shows the mean Acc values obtained. It is observed that none of the 4D models outperforms Model 2. In some cases, the result is similar, with the disadvantage of increased dimensionality and the resulting computational and overall procedural cost.

Table 9. Four-dimensional models, with different variables (Y) added to the space of Model 2 (WPE × AC × BSE × Y). Mean accuracy with the indicated fourth variable.

Y	4th variable	Parameters	Mean Acc	Deviation
PE	PE Permutation Entropy	m=6 tau=13	0.874	0.097
AWPE	AWPE Entropy on Amplitude Curve	m=3 tau=2	0.861	0.114
FOWPE	FOWPE Entropy on F0 Frequency Curve	m=3 tau=13	0.858	0.105
F1WPE	F1WPE Entropy on F1 Frequency Curve	m=3 tau=15	0.871	0.093
F2WPE	F2WPE Entropy on F2 Frequency Curve	m=3 tau=15	0.866	0.088
SlopePE	SlopePE	m=6 tau=7	0.873	0.102
SE	Spectral entropy (SE)	–	0.850	0.103

Table 10. The 4D models were evaluated by incorporating a variable Y into the three-dimensional models from Giuliano et al. [15] (**WPE × AC × BS4 x Y**). The acoustic measurements were evaluated according to the classification groups, with accuracy and deviation values obtained from the HUNIPA-2019 database.

Group[15]	Type of measure	Y	Feature Name according to Tsanas [13]	Mean Acc	Deviation
G1	Jitter	V1	Jitter → $F0\_abs\_dif$	0.866	0.095
		V51	GQ → std_cycle_closed	0.856	0.111
G2	Shimmer	V34	Shimmer → $F0\_abs0th\_perturb$	0.849	0.105
G3	HNR and NHR	V338	DFA	0.832	0.118
	(Harmonic Noise)	V59	VFER → std	0.872	0.097
		V70	IMF → $NSR\_entropy$	0.842	0.110
G4	MFCC	V137	std_9th delta	0.831	0.115
	(Mel-frequency	V141	std_delta delta log energy	0.828	0.116
	cepstral coefficients)	V152	std_10th delta-delta	0.831	0.111
		V71	mean_Log energy	0.869	0.081
		V75	mean_MFCC_3rd coef	0.877	0.102

As stated in the work of [15], 339 dysphonia measures proposed by Tsanas [13] were analyzed from four groups that characterize the most common problems in the dysfunctional voices of patients with PD. Problems in periodicity are quantified in the variations of the signal cycle by cycle. These variations are measured in relation to two groups: group 1 (G1), which focuses on fundamental frequency (F0) jitter, and group 2 (G2), which focuses on amplitude shimmer. Group 3 (G3) problems are associated with incomplete closure of the vocal folds; this condition can be linked to noise, specifically to harmonic noise (HNR-NHR). Finally, the variables in Group 4 (G4) are associated with articulatory problems in the vocal tract and are cepstral measures of the MFCC type (mel-frequency cepstral coefficients). From the set of 339 features, 11 were selected using different statistical techniques: principal

components (PCA), Analysis of Variance (ANOVA), logistic regression with penalty (RL). A final model was achieved with an accuracy between 0.80 and 0.82 in classifying people with and without Parkinson's, identifying as relevant variables.

As outlined in Table 10, each of the 11 variables is considered separately as the 4th variable in Model 2, and the mean classification accuracy is calculated using a polynomial SVM. This table indicates the original name proposed in Tsanas et al. [13] for variables. The mean Acc is high, but it does not exceed the level achieved with only the three variables in Model 2.

3.7. Analysis Sensitivity of the 3D models to the noise

To analyze the performance of the 3D models in the presence of noise, white, pink and real ambient noise were added to the original database records. The ambient noise was taken from an environment with electronic equipment, air conditioning units, and the presence of distant murmurs of voices. The audios was evaluated using the trained Model 2 and the performance in the 20% validation set was measured. The results are shown in Table 11, the trained Model 2 (Table 7) was applied and the quality de la classification was observed for different signal-to-noise ratios (SNR).

Table 11. Evaluating the performance of Model 2 (WPE × AC × BSE) with the addition of pink noise, white noise, gaussian noise, and ambient noise at the validation stage (N=22).

SNR	Pink noise			White noise			Ambient noise		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
0 dB	0.500	0.667	0.760	0.500	0.667	0.760	0.500	0.667	0.760
5 dB	0.500	0.667	0.760	0.500	0.667	0.760	0.591	0.710	0.893
10 dB	0.681	0.764	0.876	0.636	0.714	0.884	0.773	0.783	0.917
15 dB	0.767	0.777	0.893	0.818	0.833	0.893	0.864	0.842	0.896
20 dB	0.864	0.842	0.901	0.818	0.800	0.876	0.864	0.842	0.876
25 dB	0.864	0.842	0.893	0.864	0.842	0.900	0.864	0.842	0.896
30 dB	0.864	0.842	0.909	0.864	0.842	0.909	0.864	0.842	0.909
Sin ruido	0.864	0.842	0.909	0.864	0.842	0.909	0.864	0.842	0.909

3.8. Comparison of the 3D model with 10D, 11D and CNN models by other authors

Using the Hunipa database, 10D and 11D variable models were considered as a starting point for binary classification (PD-NPD), using SVM with a polynomial kernel and cross-validation (CV). The feature spaces proposed in Arora et al. [14] (10D models) and Giuliano et al. [15] (11D model) were taken, and the mean Acc value was calculated. The results are shown in Table 12. The selection of feature selection (FS) proposed by the authors was based on the 339 measures proposed by Tsanas [13].

In Arora et al. [14] six techniques are considered for the selection of 10 variables (10D models) and results are compared (see Table III in [14]). The FS techniques proposed are: (1) minimum redundancy maximum relevance (mRMR); (2) Gram-Schmidt orthogonalization (GSO), (3) Relief, (4) least absolute shrinkage and selection operator (LASSO), (5) variable ensemble proposed [14], and (6) the method proposed in [13]. The authors highlight the 6 models of 10 variables (10D) as robust and parsimonious.

In HUNIPA dataset with the 10D models, from (1) to (6), mean Acc values between 0.602 and 0.712 were obtained, which are much lower than the models presented here. With the 11D model identified as (7), a mean Acc of 0.850 was achieved, which is higher than the one previously obtained, but lower than that obtained here with Models 2 and 3, which are developed from a feature space with only 3 variables. (Table 12).

Table 12. For the HUNIPA dataset. Variable groups (FS) and mean Acc and standard deviation for the classification using SVM with a polynomialkernel (1) to (7). CNN classification (8) and (9).

Variable selection technique	Model dimension	Mean Acc	Deviation
(1) mRMR [14]	10D	0.703	0.123
(2) GSO [14]	10D	0.657	0.140
(3) Relief [14]	10D	0.702	0.103
(4) LASSO [14]	10D	0.602	0.114
(5) Ensemble ranking [14]	10D	0.673	0.091
(6) Tsanas et al. [13]	10D	0.712	0.123
(7) Giuliano et al. [15]	11D	0.850	0.112
(8) Guatelli et al. [53]	CNN	0.839	–
(9) Guatelli et al. [53]	EML	0.800	–
This research in Model 2	3D (WPE $\times$ AC $\times$ BSE)	0.879	0.094

In addition, the maximum average accuracy obtained by Guatelli et al. [53] with the HUNIPA database was included. The authors used spectrogram images and classified them using convolutional neural networks (CNN) and extreme learning machines (EML), obtaining Acc 0.83 and 0.80, respectively, as shown in (8) and (9) of the Table 12.

3.9. A comparison of the 3D model with independent databases

Three independent databases (HUNIPA, GITA, and NeuroVoz) were analyzed, both separately and in combination, and the results are shown in Table 13. Better results were achieved by changing BS4 (2000 - 3000 Hz) to BS7 (6000 - 8000 Hz) for 3D models, from Model 2 (WPE x AC x BD4) to Model 10 (WPE x AC x BD7). The mean Acc values were comparable to those reported by Ibarra et al. [56] using CNN (Table 14). Figure 6 illustrates the differences in spectral band entropy (BS4 and BS7) according to the databases and disease conditions.

In Ibarra et al. [56] used several deep learning (DL) architectures: multilayer perceptrons (MLP), a combination of convolutional neural networks (CNN) and MLP, recurrent neural networks (RNN), a combination of CNN and long short-term memory (LSTM) networks, and combinations of time-distributed 2D CNN and 1D CNN. Table 14 shows the best classification results obtained with DL architectures in the NeuroVoz and GITA databases by the authors.

4. DISCUSSION OF RESULTS AND FUTURES WORKS

This work is framed within the field of acoustics and signal processing, specifically applied to the biomedical field, with special attention to the identification of speech deficiencies in individuals with Parkinson's disease. The study emphasizes signal measures such as autocorrelation, weighted permutation entropy, and band spectral entropy, which are well established in signal processing and show strong discriminative power in this context. However, these measures are novel because they are not commonly used in the classification of voices with and without PD.

As is well established, AC accounts for differences and similarities in the temporal structure of the signal, and here it allows the recognition of signal properties to characterize groups of similar cases and differentiate the classes of interest.

The incorporation of WPE not only characterizes its temporal sequential structure but also reflects the variations in signal magnitude between the classes of interest, accounting for short-term patterns and instantaneous amplitude variations. Permutation entropy measures are not found in the literature in the style presented here for voices with

Table 13. The mean and standard deviation of accuracy (Acc.) were calculated for the three features with the same parameters Model 2 (WPE x AC x BD4) and Model 10 (WPE x AC x BD7). These metrics were evaluated separately for each corpus (HUNIPA, GITA and Neurovoz) for the vowel /a/ (in 1 second of phonation) separately and all together. .

Database	Model	Mean Acc.	Deviation	Name	Selected parameters
HUNIPA	Model 2	0.879	0.094	AC WPE BSE4	lag = 13 m = 6, τ = 13 Band 4: 2000–3000 Hz
HUNIPA	Model 10	0.846	0.099	AC WPE BSE4	lag = 13 m = 6, τ = 13 Band 7: 6000–8000 Hz
GITA	Model 10	0.711	0.136	AC WPE BSE7	lag = 13 m = 6, τ = 13 Band 7: 6000–8000 Hz
NeuroVoz	Model 10	0.708	0.124	AC WPE BSE7	lag = 13 m = 6, τ = 13 Band 7: 6000–8000 Hz
Mixed datasets	Model 10	0.724	0.068	AC WPE BSE7	lag = 13 m = 6, τ = 13 Band 7: 6000–8000 Hz

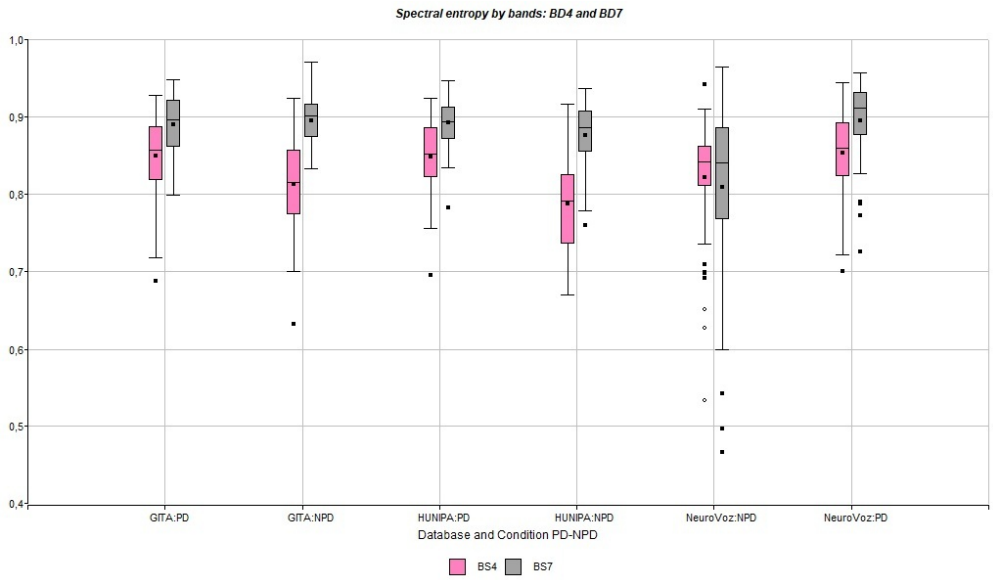

Figure 6. Spectral entropy by bands (BS) according to disease condition and database (BS4: 2000-3000hz; BS7: 62000-8000hz).

and without PD. [61] propose analyzing permutation entropy for phonations of PD and NPD individuals, but only the difference in means was analyzed.

The feature space defined by AC and WPE was found to provide a good representation of the cases analyzed and was identified as the privileged feature plane (Model 1: WPE $\times$ AC) and shows strong discriminative power. It

Table 14. Performance of different neural networks on the vowel /a/ in GITA and NeuroVoz corpora, retaining only the best configurations. Deep Learning architectures: convolutional neural networks (CNN), long short-term memory (LSTM) networks, and combinations of time-distributed 2D-CNN and 1D-CNN, Time-CNN-LSTM. Values reported as mean and standard deviation in Ibarra et al. [56].

Corpus	Network	Acc.	Deviation
GITA	2D-CNN	70.5	9.8
	Time-CNN-LSTM	72.4	10.6
	1D-CNN	72.6	10.2
NeuroVoz	2D-CNN	74.9	10.9
	Time-CNN-LSTM	71.8	17.0
	1D-CNN	71.5	6.6

allowed the identification of four subgroups of individuals with similar characteristics in both PD and NPD cases. Future research could investigate the relationship between these groups and voice disorders, in order to differentiate the types of alterations between cases.

On the other hand, the inclusion of the BSE provides information about the energy distribution by bands, which were selected based on phono-audiological criteria [34]. This proved to be another feature that emerged from the research, confirming its discriminatory nature between PD and NPD subjects, which gives the model using the 3D feature space a better performance as a discriminator of the classes. Considering that this entropy contains information in the frequency domain about how the energy of the signal is distributed and is a direct characteristic of how sounds are produced, there is a difference found between the two groups analyzed.

Different entropy options were evaluated as possible third dimensions to improve the 2D model proposed, and in all cases, the mean accuracy was improved. The best results were obtained with the 3D model incorporating BSE, which highlighted an average correct classification rate of almost 88% (Model 2: WPE $\times$ AC $\times$ BSE). In addition, the four-dimensional models (4D) showed a worse mean accuracy, further confirming the applicability of the three-dimensional models.

The proposed models offer advantages such as low differential dimensionality, explanatory power, computational efficiency, and result interpretability. However, they also have disadvantages, including dependencies on signal length, sampling theorem conditions, signal reconstruction, and selection of some parameters like the embedding time dimension, embedding time delay, time lag and spectral band.

A large number of dysphonia measures available for voice study can potentially lead to a degradation in analysis performance during the statistical analysis phase. Although modern algorithms aim to mitigate this issue, even powerful classifiers can be degraded due to computational complexity and the presence of high-dimensional datasets, especially with few cases. Therefore, it is particularly useful to reduce the dimensionality to a minimal set of features, as this facilitates inference and understanding of the problem by analyzing the most relevant features [50].

The performance of the best 3D model was compared to several 10D models proposed by [14] and the 11D model introduced by [15]. While these authors highlighted the strengths of their models, the 3D model demonstrated superior accuracy when applied to the used database with their feature selection criteria.

This article obtains accuracy results comparable to the performance of CNN models proposed by other authors using the same databases studied. In the HUNIPA database, better performance is obtained than that presented in Guatelli et al.[53], 0.88 compared to 0.84.

The results obtained with the two independent databases considered (GITA and NeuroVoz) in the models with the same parameters are close to those obtained by Ibarra et al. However, this author uses DL techniques that are more computationally demanding and less explainable. On the other hand, when proposing models with a mix of all databases, the result improves contrary to what is reported in Ibarra et al. [56] The dependence on the data domain does not seem to be as strong with the proposed models.

As stated in Ibarra et al.[56], it is necessary to explore techniques for analyzing the voices of Parkinson's patients that are independent of the origin of the voices. The 3D models that include weighted permutation entropy are competitive. However, the optimal characteristics depend on the database used. Future work will seek to analyze the generalizable characteristics that are not dependent on the domain of the data.

In this research was proposed a working hypothesis that low-dimensional spaces are preferable when they can elucidate the structural and informational nature of the problem, as was the case in this study. It has been considered that a reduction in dimensions reduces noise or overfitting, thereby allowing 3D models to obtain superior classification results compared to higher-dimensional models.

The study employed supervised learning techniques for binary classification, a common approach in machine learning [50]. This involved identifying a function the feature set to the binary outcome, distinguishing between individuals with PD and healthy controls (NPD). The results obtained show the consistent of the methodology, despite the sample size, as the models were compared their performance using different classification techniques, including polynomial and radial SVM, kNN and RF. To assess model performance and prevent overfitting 10-fold cross-validation was employed. The results were similar for all methods, with polynomial SVM performing best in terms of mean accuracy. This suggests that the models are well configured and effective in capturing the relevant patterns by using cross-validation.

The high classification ability of the selected features with the SVM polynomial model supports the suitability of this scheme, in coincidence with the findings reported in the review by [62]. SVM are a powerful and versatile class of supervised learning algorithms, particularly well-suited for classification tasks, including binary classification problems. While not strictly a nonparametric test in the traditional statistical sense, SVM exhibit robustness and flexibility, making them suitable for complex, non-linearly separable data often encountered in real-world scenarios. In study of [39] of commonly used machine learning techniques in small sample settings revealed that SVM is among the most frequently employed methods.

A review of machine learning and deep learning algorithms for PD detection is presented in [17] and it shows a frequently election of the SVM technique among researchers.

The selection of these classification techniques was predicated on their widespread utilization within the extant literature and their efficacy in this particular instance. Their effectiveness was demonstrated in the context of binary classification of PD and NPD, as well as the selection of optimal parameters for the measures employed in 2D and 3D spaces. One of the objectives of the research was the interpretability of the parameters and their possible relationship with the clinical symptoms of the patients and with the results obtained it was accomplished.

In [17] are analyzed various databases and machine learning and deep learning algorithms to classify individuals with and without PD. The studies identify distinct acoustic measures utilized for classification, with results demonstrating an accuracy ranging upper of 60%.

The present study aligns with the machine learning systems described by [17], following usual methodological guidelines. Additionally, it explores heterogeneous subgroups within Parkinson's patients, a topic that has received little attention in the existing literature.

From the authors' point of view, the work presented here represents a contribution regarding previous studies [16, 15], where lower accuracy was obtained with a greater number of computational resources, a higher number of variables, and longer audio lengths. Furthermore, this study contributes to the reproducibility of the present research and facilitates further investigations by making the audio from the constructed database freely accessible to the scientific community [21].

The signal characteristics associated with Parkinson's disease, as found in the literature [7, 10] and in review studies [63, 12, 17] are largely related to variations in measures such as jitter, shimmer, cepstral coefficients, among others.

PD is a heterogeneous disease with considerable differences in progression, and more homogeneous subgroups can be identified based on clinical observations [6, 10]. The defined 2D and 3D characteristic spaces and subgroups were partially analyzed in terms of gender, age, stage of disease progression, among other factors, and dysphonia measures used by speech therapists were considered. It was observed that the SG3 group, composed solely of NPD, was the youngest on average and had the fewest voice alterations, while the SG2 group, composed solely of PD, had the most alterations. In the future, these analyses will be further explored with health specialists.

A particularly significant finding is the identification of subgroups of individuals with homogeneous characteristics within the feature space. The subgroups could be associated with subtypes of PD. Furthermore, it is possible to correlate cases based on clinical symptoms previously studied in the same database, such as alterations identified in laryngoscopy with a laryngeal stroboscope in patients with PD [25].

The UPDRS and H&Y scales are indicators of disease progression and are widely used by clinicians. In the particular case of the database used, the patients correspond to PD in the early stages of the disease. This makes the results more promising, with patients in more severe condition could be expected to show greater alterations in their voice.

Future work will focus on evaluating the robustness and generalizability of the method using larger and more diverse datasets. Gender-specific considerations in voice analysis will also be explored, taking into account potential differences in vocal characteristics between males and females. The parameters that define the variables for male or female gender could be optimized and defining models for each gender. It is also interesting to consider patients with different degrees of disease progression using scales such as H&Y or UPDRS, which are commonly used by physicians.

The measurements proposed here are accessible to low-cost equipment and in controlled low-noise environments, such as a medical center. So far, a quality microphone and a signal-to-noise ratio greater than or equal to 20 dB are required to record one second of sustained phonation of the vowel /a/. It is also recommended a computer system that links the patient's medical history to the analysis of the phonation signal, which is currently under development [64].

A collaboration with the Posadas National Hospital in Argentina is developed to provide tools for integrating voice analysis into clinical assessments. The methodology proposed for voice analysis has the potential to simplify PD screening processes, reduce computational requirements, and improve accessibility. In the near future, we will seek to work together with Parkinson's treating physicians in the workflow for the acquisition, preprocessing, feature extraction and classification of signals such as voice in patients with neurological disorders.

The results obtained in this research agree with [17] that advances in the analysis of multimodal information about PD (audio, writing, gait, depression, etc.) will have a significant impact on the identification, monitoring and treatment of PD in the near future. These innovations may lead to the development of PD diagnostic tools that are easy to use, non-invasive and easily integrated into clinical settings. They also offer the possibility of identifying people at risk of developing PD, allowing early detection and intervention. They also have the potential to improve patients' quality of life and inform the development of new treatment strategies.

The results reinforce the interest in the line of research related to acoustic parameters and entropic measures. A new data corpus is also being developed, which will allow the replication of the analyses and testing of the models proposed here. One issue to consider, which is a question to be answered in the near future, is how to record with a mobile phone instead of a higher-quality microphone to facilitate the work of medical professionals and minimize the impact that recording methods have on the performance of analysis techniques.

5. CONCLUSIONS

The manuscript presents an innovative methodology for PD detection through speech analysis, yielding promising results in low-dimensional (2D and 3D) feature spaces. The simplicity and accuracy of the proposed approach offer great potential for advancing the diagnosis and monitoring of PD. In particular, the methodology proves effective with phonation samples of the vowel /a/ as short as one second, which could help streamline clinical workflows.

The work presented here identifies just three signal features that their use is not widespread as are the AC, WPE, and BSE measures, holistically model and explain signal characteristics, directly derived from the temporal and spectral properties of speech signals. Also, it allowed the identification of four subgroups with similar characteristics in both classes (PD - NPD).

The proposed two-dimensional and three-dimensional feature spaces provide parsimonious models that achieve a classification accuracy of between 80% and 90%. In the studies carried out, the classification performance remains practically stable for SNR values above 20 dB. As the database is small, the study carried out can be considered

a preliminary stage to identify informational measures and subgroups oriented towards the search for biomarker proxies that allow for the knowledge of subtypes of Parkinson's disease.

6. ACKNOWLEDGMENT

The author M. Giuliano wishes to express her gratitude for the research internship at the Signal and Image Processing Center of the Regional School of Universidad Tecnológica Nacional during part of the development of her doctoral thesis.

M. Giuliano also thanks her affiliation institutions: the Institute of Health Sciences of Universidad Nacional del Oeste and Computing Research and Development Laboratory of Universidad Nacional de Hurlingham.

7. AUTHOR DECLARATIONS

Conflict of Interest

The authors declare that there are no conflicts of interest.

Ethics Approval

The present work was performed under a protocol approved by the Ethics Committee of the Hospital de Agudos Bernardino Rivadavia of the Autonomous City of Buenos Aires, Argentina.

REFERENCES

1. World Health Organization, *Parkinson disease: a public health approach*, Technical Brief, Geneva, 2022. https://iris.who.int/bitstream/handle/10665/355973/9789240050983-eng.pdf?sequence=1
2. GBD 2016 Neurology Collaborators, *Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016*, Lancet Neurol., vol. 18, no. 5, pp. 459–480, 2019. doi:10.1016/S1474-4422(18)30499-X
3. A. Ascherio and M. A. Schwarzschild, *The epidemiology of Parkinson's disease: risk factors and prevention*, The Lancet Neurology, vol. 15, no. 12, pp. 1257–1272, 2016. doi:10.1016/s1474-4422(16)30230-7
4. H. Deng, P. Wang, and J. Jankovic, *The genetics of Parkinson disease*, Ageing Research Reviews, vol. 42, pp. 72–85, 2018. doi:10.1016/j.arr.2017.12.007
5. R. B. Postuma, D. Berg, M. Stern, W. Poewe, C. W. Olanow, W. Oertel et al., *MDS clinical diagnostic criteria for Parkinson's disease: MDS-PD Clinical Diagnostic Criteria*, Movement Disorders, vol. 30, no. 12, pp. 1591–1601, 2015. doi:10.1002/mds.26424
6. S.-M. Fereshtehnejad, S. R. Romenets, J. B. M. Anang, V. Latreille, J.-F. Gagnon, and R. B. Postuma, *New Clinical Subtypes of Parkinson Disease and Their Longitudinal Progression*, JAMA Neurology, vol. 72, pp. 863–873, 2015. doi:10.1001/jamaneurol.2015.0703
7. M. De Letter, P. Santens, and J. V. Borsel, *The effects of levodopa on word intelligibility in Parkinson's disease*, Journal of Communication Disorders, vol. 38, no. 3, pp. 187–196, 2005. doi:10.1016/j.jcomdis.2004.09.001
8. I. Karabayir, S. M. Goldman, S. Pappu, and O. Akbilgic, *Gradient boosting for Parkinson's disease diagnosis from voice recordings*, BMC Medical Informatics and Decision Making, vol. 20, no. 1, pp. 228, 2020. doi:10.1186/s12911-020-01250-7
9. S. Yang, F. Wang, L. Yang, F. Xu, M. Luo, X. Chen, X. Feng et al., *The physical significance of acoustic parameters and its clinical significance of dysarthria in Parkinson's disease*, Scientific Reports, vol. 10, no. 1, 2020. doi:10.1038/s41598-020-68754-0
10. A. Suppa, G. Costantini, F. Asci, P. Di Leo, M. S. Al-Wardat, G. Di Lazzaro, S. Scalise et al., *Voice in Parkinson's Disease: A Machine Learning Study*, Frontiers in Neurology, vol. 13, 2022. doi:10.3389/fneur.2022.831428
11. A. Tsanas and S. Arora, *Assessing Parkinson's Disease Speech Signal Generalization of Clustering Results across Three Countries: Findings in the Parkinson's Voice Initiative Study*, Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, 2021. doi:10.5220/0010383001240131
12. P. Hecker, N. Steckhan, F. Eyben, B. W. Schuller, and B. Arnrich, *Voice Analysis for Neurological Disorder Recognition–A Systematic Review and Perspective on Emerging Trends*, Frontiers in Digital Health, vol. 4, 2022. doi:10.3389/fdgth.2022.842301
13. A. Tsanas, *Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning*, PhD thesis, Oxford University, UK, 2012. https://ora.ox.ac.uk/objects/uuid:2a43b92a-9cd5-4646-8f0f-81dbe2ba9d74
14. S. Arora, L. Baghai-Ravary, and A. Tsanas, *Developing a large scale population screening tool for the assessment of Parkinson's disease using telephone-quality voice*, The Journal of the Acoustical Society of America, vol. 145, no. 5, pp. 2871–2884, 2019. doi:10.1121/1.5100272

15. M. Giuliano, L. Fernandez, and S. Perez, *Selección de Medidas de Disfonía para la Identificación de Enfermos de Parkinson*, In Proceedings IEEE Congreso Bienal de Argentina (ARGENCON), pp. 1–8, 2020. doi:10.1109/ARGENCON49523.2020.9505554
16. R. Guatelli, M. Giuliano, V. Aubin, L. Fernández, S. Pérez, and L. Pepe, *Análisis comparativo entre CNN y Modelos Logísticos para detección de la Enfermedad de Parkinson utilizando la voz*, Memorias del 10mo Congreso Nacional de Ingeniería Informática / Sistemas de Información (CONAIISI), 2022. doi:10.33414/ajea.1146
17. M. A. Islam, M. Z. H. Majumder, M. A. Hussein, K. M. Hossain, and M. S. Miah, *A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets*, Heliyon, vol. 10, no. 3, pp. e25469, 2024.
18. F. Rosso, L. Olivares, L. Zunino, L. De Micco, A. L. Aquino, A. Plastino, and H. A. Larrondo, *Characterization of chaotic maps using the permutation Bandt-Pompe probability distribution*, European Physical Journal B, vol. 86, pp. 1–13, 2013. doi:10.1140/epjb/e2013-30764-5
19. P. Martínez Coq, A. Rey, O. A. Rosso, R. Armentano, and W. Legnani, *Detection of cardiac arrhythmia patterns in ECG through H×C plane*, Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 32, no. 12, 2022. doi:10.1063/5.0118717
20. M. Giuliano, D. Adamec, and M. I. Debas, *Construcción de una base de voz de personas con y sin enfermedad de Parkinson*, Revista Digital Departamento de Ingeniería e Investigaciones Tecnológicas (REDDI), pp. 6–10, 2021. https://reddi.unlam.edu.ar/index.php/ReDDi/article/view/141
21. S. N. Pérez and M. Giuliano, *Audios_vowel_A_PD*, Figshare, 2022. doi:10.6084/M9.FIGSHARE.21453867.V1
22. M. M. Hoehn and M. D. Yahr, *Parkinsonism: onset, progression, and mortality*, Neurology, vol. 50, no. 2, pp. 318–335, 1998. doi:10.1212/wnl.50.2.318
23. C. G. Goetz, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, G. T. Stebbins et al., *Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan*, Movement Disorders, vol. 22, pp. 41–47, 2007.
24. S. Fahn, R. Elton, and Members of the UPDRS Development Committee, *Unified Parkinson's Disease Rating Scale (UPDRS)*, In: Fahn S, Marsden C.D., Calne D.B., Goldstein M., editores. Recent developments in Parkinson's disease, vol. 2, pp. 153–163, 293–304, 1987.
25. M. Giuliano, M. I. Debas, H. Torres, S. Pérez, L. De Leon, D. Adamec, and A. Berduc, *Clinical Correlate of Dysphonia Measures in Patients with Parkinson's Disease*, Medical Research Archives, vol. 11, no. 12, 2023. doi:10.18103/mra.v11i12.4827
26. G. U. Yule, *On a method of investigating periodicities in disturbed series with special reference to Wolfers sunspot number*, Philosophical Transactions of the Royal Society A, vol. 226, pp. 267, 1927. https://www.jstor.org/stable/91170
27. G. T. Walker, *On periodicity in series of related terms*, Proceedings of the Royal Society of London A, vol. 131, pp. 518–532, 1931. doi:10.1098/rspa.1931.0069
28. C. E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal, vol. 27, no. 3, pp. 379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x
29. C. E. Shannon and W. Weaver, *The mathematical theory of communication*, University of Illinois Press, Urbana, IL, 1949.
30. C. Bandt and B. Pompe, *Permutation Entropy: A Natural Complexity Measure for Time Series*, Physical Review Letters, vol. 88, no. 17, pp. 174102, 2002. doi:10.1103/physrevlett.88.174102
31. B. Fadlallah, B. Chen, A. Keil, and J. Príncipe, *Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information*, Physical Review E, vol. 87, no. 2, pp. 022911, 2013. doi:10.1103/physreve.87.022911
32. D. Cuesta Frau, *Slope Entropy: A New Time Series Complexity Estimator Based on Both Symbolic Patterns and Amplitude Information*, Entropy, vol. 21, no. 12, pp. 1–22, 2019. doi:10.3390/e21121167
33. G. E. Powell and I. C. Percival, *A spectral entropy method for distinguishing regular and irregular motion of Hamiltonian systems*, Journal of Physics A: Mathematical and General, vol. 12, no. 11, pp. 2053, 1979. doi:10.1088/0305-4470/12/11/017
34. L. A. Cecconello, *Aplicación del análisis acústico en la clínica vocal. Trabajando con Anagraf*, Editorial Akadia, Buenos Aires, 2012.
35. J. Shen, J. Hung, and L. Lee, *Robust Entropy-Based Endpoint Detection for Speech Recognition in Noisy Environments*, In Proceedings of ICSLP, vol. 98, pp. 232–235, 1998.
36. I. R. Titze, *Principles of voice production*, National Center for Voice and Speech, 2000.
37. M. Riedl, A. Müller, and N. Wessel, *Practical considerations of permutation entropy: A tutorial review*, *The European Physical Journal Special Topics*, vol. 222, no. 2, pp. 249–262, 2013.
38. A. Myers and F. A. Khasawneh, *On the automatic parameter selection for permutation entropy*, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 3, p. 033123, 2020.
39. P. Kokol, M. Kokol, and S. Zagoranski, *Machine learning on small size samples: A synthetic knowledge synthesis*, Science Progress, vol. 105, 2022. doi:10.1177/00368504211029777
40. B. E. Boser, I. M. Guyon, and V. N. Vapnik, *A training algorithm for optimal margin classifiers*, Proceedings of the fifth annual workshop on Computational learning theory, pp. 144–152, 1992. doi:10.1145/130385.130401
41. J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, *A comprehensive survey on support vector machine classification: Applications, challenges and trends*, Neurocomputing, vol. 408, pp. 189–215, 2020. doi:10.1016/j.neucom.2019.10.118
42. C. W. Hsu, C. C. Chang, and C. J. Lin, *A practical guide to support vector classification*, Technical Report, Department of Computer Science and Information Engineering, University of National Taiwan, Taipei, pp. 1–12, 2003.
43. E. Fix and J. L. Hodges, *Discriminatory analysis. Nonparametric discrimination: Consistency properties*, International Statistical Review, vol. 57, pp. 238–247, 1989. doi:10.2307/1403797
44. M. Sharma and S. K. Sharma, *Generalized K-nearest neighbour algorithm-a predicting tool*, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 11, pp. 1–4, 2013.
45. L. Breiman, *Random forests*, Machine Learning, vol. 45, pp. 5–32, 2001. doi:10.1023/A:1010933404324
46. M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, *Ensemble deep learning: A review*, Engineering Applications of Artificial Intelligence, vol. 115, 2022. doi:10.1016/j.engappai.2022.105151
47. A. Liaw and M. Wiener, *Classification and regression by random Forest*, R News, vol. 2, pp. 18–22, 2002. http://CRAN.R-project.org/doc/Rnews/

48. M. Kuhn, *Building Predictive Models in R Using the caret Package*, Journal of Statistical Software, vol. 28, no. 5, pp. 1–26, 2008. doi:10.18637/jss.v028.i05
49. K. Schliep and K. Hechenbichler, *kknn: Weighted k-Nearest Neighbors*, R package version 1.3.1, 2016.
50. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, New York, 2009.
51. C. Aggarwal, *Data Classification*, Data Mining, Springer, Cham, pp. 285–344, 2015. doi:10.1007/978-3-319-14142-8_10
52. R. Das, *A comparison of multiple classification methods for diagnosis of Parkinson disease*, Expert Systems with Applications, vol. 37, no. 2, pp. 1568–1572, 2010.
53. R. Guatelli, V. Aubin, M. Mora, J. Naranjo-Torres, and A. Mora-Olivari, *Detection of Parkinson's disease based on spectrograms of voice recordings and Extreme Learning Machine random weight neural networks*, Engineering Applications of Artificial Intelligence, vol. 125, p. 106700, 2023.
54. J. Orozco-Arroyave and J. Arias-Londoño, *New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease*, in Proceedings of the International Conference on Language Resources and Evaluation (LREC), 2014.
55. J. Mendes-Laureano, J. A. Gómez-García, A. Guerrero-López, E. Luque-Buzo, J. D. Arias-Londoño, F. J. Grandas-Pérez, and J. I. Godino-Llorente, *NeuroVoz: a Castilian Spanish corpus of parkinsonian speech*, *Scientific Data*, vol. 11, no. 1, p. 1367, 2024.
56. E. J. Ibarra, J. D. Arias-Londoño, M. Zañartu, and J. I. Godino-Llorente, *Towards a corpus (and language)-independent screening of Parkinson's disease from voice and speech through domain adaptation*, *Bioengineering*, vol. 10, no. 11, p. 1316, 2023.
57. J. A. Gurlekian and N. Molina, *Índice de perturbación, de precisión vocal y de grado de aprovechamiento de energía para la evaluación del riesgo vocal*, *Rev. Logop. Foniatr. Audiol.*, vol. 32, no. 4, pp. 156–163, 2012. Available at: http://dx.doi.org/10.1016/j.rlfa.2012.03.007
58. L. Sigal and J. A. Gurlekian, *Aplicación de los índices de perturbación integrado y de precisión articulatoria en pacientes con disfonía espasmódica*, *Rev. Investig. Logop.*, vol. 4, no. 2, pp. 132–150, 2014. Available at: http://dx.doi.org/10.5209/rlog.58665
59. J. A. Gurlekian, H. M. Torres, and M. R. Cediel, *A perceptual method to rate dysphonic voices*, *J. Voice*, vol. 33, no. 4, pp. 453–464, 2019. Available at: http://dx.doi.org/10.1016/j.jvoice.2018.01.007
60. H. Zhivomirov, *A Method for Colored Noise Generation*, Romanian Journal of Acoustics and Vibration, vol. 15, no. 1, 2018.
61. H. Zhang, N. Yan, L. Wang, and M. L. Ng, *Energy distribution analysis and nonlinear dynamical analysis of phonation in patients with Parkinson's disease*, 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 630–635, 2017. doi:10.1109/apsipa.2017.8282102
62. E. Boateng, J. Otoo, and D. Abaye, *Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review*, Journal of Data Analysis and Information Processing, vol. 8, pp. 341–357, 2020. doi:10.4236/jdaip.2020.84020
63. R. Chiaramonte and M. Bonfiglio, *Análisis acústico de la voz en la enfermedad de Parkinson: revisión sistemática de la discapacidad vocal y metaanálisis de estudios*, Revista de Neurología, vol. 70, no. 11, pp. 393–405, 2020. doi:10.33588/rn.7011.2019414
64. P. Occhipinti, M. Giuliano, S. N. Pérez, and A. Berduc, *SYNAPSIS: Plataforma de datos Multimodales y Gestión Clínica para la Enfermedad de Parkinson*, Libro de actas: 30º Congreso Argentino de Ciencias de la Computación (CACIC 2024), pp. 1079–1084, 2024. https://cacic2024.info.unlp.edu.ar/libro-de-actas-cacic-2024-ebook/