Check for updates

Effects of deep brain stimulation of the subthalamic nucleus on patients with Parkinson's disease: a machine-learning voice analysis

Antonio Suppa[1,2]*[†], Francesco Asci[1,2†], Giovanni Costantini[3†], Francesco Bove[4†], Carla Piano[4], Francesca Pistoia[5,6], Rocco Cerroni[7], Livia Brusa[8], Valerio Cesarini[3], Sara Pietracupa[1,2], Nicola Modugno[2], Alessandro Zampogna[1], Patrizia Sucapane[6], Mariangela Pierantozzi[7], Tommaso Tufo[9,10], Antonio Pisani[11,12], Antonella Peppe[13], Alessandro Stefani[7], Paolo Calabresi[4], Anna Rita Bentivoglio[4], Giovanni Saggio[3] and Lazio DBS Study Group

[1]Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy, [2]IRCCS Neuromed Institute, Pozzilli, IS, Italy, [3]Department of Electronic Engineering, University of Rome Tor Vergata, Rome, Italy, [4]Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy, [5]Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, Coppito, AQ, Italy, [6]Neurology Unit, San Salvatore Hospital, Coppito, AQ, Italy, [7]Department of System Medicine, University of Rome Tor Vergata, Rome, Italy, [8]Neurology Unit, S. Eugenio Hospital, Rome, Italy, [9]Neurosurgery Unit, Policlinico A. Gemelli University Hospital Foundation IRCSS, Rome, Italy, [10]Neurosurgery Department, Fakeeh University Hospital, Dubai, United Arab Emirates, [11]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy, [12]IRCCS Mondino Foundation, Pavia, Italy, [13]IRCSS Fondazione Santa Lucia, Rome, Italy

Introduction: Deep brain stimulation of the subthalamic nucleus (STN-DBS) can exert relevant effects on the voice of patients with Parkinson's disease (PD). In this study, we used artificial intelligence to objectively analyze the voices of PD patients with STN-DBS.

Materials and methods: In a cross-sectional study, we enrolled 108 controls and 101 patients with PD. The cohort of PD was divided into two groups: the first group included 50 patients with STN-DBS, and the second group included 51 patients receiving the best medical treatment. The voices were clinically evaluated using the Unified Parkinson's Disease Rating Scale part-III subitem for voice (UPDRS-III-v). We recorded and then analyzed voices using specific machine-learning algorithms. The likelihood ratio (LR) was also calculated as an objective measure for clinical-instrumental correlations.

Results: Clinically, voice impairment was greater in STN-DBS patients than in those who received oral treatment. Using machine learning, we objectively and accurately distinguished between the voices of STN-DBS patients and those under oral treatments. We also found significant clinical-instrumental correlations since the greater the LRs, the higher the UPDRS-III-v scores.

Discussion: STN-DBS deteriorates speech in patients with PD, as objectively demonstrated by machine-learning voice analysis.

KEYWORDS

Parkinson's disease, voice analysis, machine-learning, deep brain stimulation, subthalamic nucleus

1. Introduction

Patients with Parkinson's disease (PD) manifest variable degrees of voice abnormalities characterized by hypophonia, mono-pitch, and mono-loudness speech, and hypophonic and hypokinetic articulation. These specific voice impairments have been collectively identified as *hypokinetic dysarthria* (1, 2). PD patients may experience voice disorders from the prodromal phase of the disease, with speech deteriorating as the disease progresses (2–6). Accordingly, it is important to investigate voice changes in PD under pharmacological as well as advanced treatments such as deep brain stimulation of the subthalamic nucleus (STN-DBS).

DBS is a well-established therapeutic option for advanced-stage patients with PD (7), as demonstrated by short- and long-term follow-up studies (7–10). Besides the well-known beneficial effects of STN-DBS on the cardinal motor symptoms in PD (i.e., bradykinesia, rigidity, and tremor), the effect of this surgical procedure on specific axial functions such as voice remains elusive (7, 11–14). Following STN-DBS procedures, the estimated prevalence of speech disorders, as a post-surgical side-effect, has been reported in PD to vary between 1% after 6 months and 70% at 3 years of follow-up (12, 15–17). Hence, STN-DBS may lead to a significant worsening of parkinsonian *hypokinetic dysarthria*, resulting in a rather different voice abnormality characterized by a hypophonic voice with a strained and spastic speech mainly associated with stuttering, as suggested by previous studies in the field (18). Therefore, DBS-related voice impairments in PD patients have been identified as DBS-related dysarthria (19).

The complexity of voice as a biological phenomenon, the heterogeneity of *dysarthria* in PD, and, finally, the variable effect of STN-DBS on the voice would therefore require more advanced techniques, including artificial intelligence that allows the analysis and dynamic combination of high-dimensional datasets of voice features (20–22). Machine learning offers a potentially useful methodology to investigate voice abnormalities, especially in complex and multifactorial neurologic disorders, including PD (2, 20, 21, 23, 24).

To date, no study has assessed voice abnormalities in a large cohort of STN-DBS patients with PD compared to chronically treated L-Dopa patients with PD through objective procedures based on machine-learning analysis. Moreover, no study has correlated the clinical and instrumental assessments of voice in patients with PD by using machine-learning output measures. Filling these knowledge gaps would be relevant for the objective recognition of voice abnormalities in STN-DBS patients with PD.

In the present cross-sectional study, we examined voice performances in a large cohort of STN-DBS and chronically treated L-Dopa patients with PD using machine-learning analysis for automatic classification purposes. Therefore, we compared voice samples recorded from STN-DBS and chronically treated L-Dopa patients as well as from healthy subjects (HS), using standardized perceptual analysis as well as advanced analysis based on machine-learning procedures. We assessed the sensitivity, specificity, positive and negative predictive values, and accuracy of all diagnostic tests and calculated the area under the receiver operating characteristic (ROC) curves. Finally, by providing an objective instrumental measure of voice impairment, the likelihood

ratio (LR), for each patient based on machine-learning analysis, we also assessed possible clinical-instrumental correlations.

2. Materials and methods

2.1. Subjects

This cross-sectional study enrolled 101 patients with PD (61.9 ± 7.5 years, range 41–81 years) and 108 patients with HS (60.3 ± 10.3 years, range 42–76 years). Participants were progressively recruited during regular follow-up clinical evaluations in the outpatient clinic for movement disorders at IRCCS Neuromed and University Departments and Public Hospitals on behalf of the "*Lazio DBS Study Group.*" All participants were native Italian speakers and non-smokers. None of the participants reported bilateral/unilateral hearing loss, respiratory disorders, or other non-neurologic disorders affecting the vocal cords. The participants provided written informed consent, which was approved by the institutional ethics committee of the IRCCS Neuromed Institute (NCT04846413), according to the Declaration of Helsinki.

The clinical diagnosis of PD was made according to the current standardized clinical criteria of the International Parkinson and Movement Disorder Society (25). Symptoms and signs of PD were scored using the Hoehn and Yahr (H&Y) scale and the Unified Parkinson's Disease Rating Scale Part III (UPDRS-III) (26). The clinical (i.e., perceptual) evaluation of speech abnormalities in PD was achieved by an independent rater using the specific item (item 3.1) for speech evaluation included in the UPDRS-III scale (UPDRS-III-v) during the overall motor assessment (26). In all participants, we excluded cognitive and mood impairments potentially affecting speech production through the Mini-Mental State Examination (MMSE) (27) corrected for years of education, the Beck's Depression Inventory (BDI) (28), and the Frontal Assessment Battery (FAB) (29).

The cohort of PD included patients in the mid-to-advanced phase (H&Y scores > 2) (30) and those who were chronically treated with L-Dopa. The PD cohort included two separate subgroups of patients: the first subgroup included 50 STN-DBS patients (61.6 ± 6.6 years, range 45–75 years), whereas the second subgroup included 51 patients (62.1 ± 8.3 years, range 41–81 years) chronically treated with the best medical treatment (i.e., L-Dopa). To specifically recognize the effect of STN-DBS on voice in PD, patients were enrolled and assigned to each of the two subgroups according to the inclusion criteria, attempting to statistically match the age, gender, H&Y, UPDRS, disease duration, and the L-Dopa equivalent daily doses (LEDDs) (all measures were calculated for each patient before the enrollment in the study). All patients were evaluated clinically and instrumentally 1–2 h after the administration of their chronic dopaminergic therapy (i.e., in the ON state). All implanted patients received chronic bilateral non-directional and non-interleaving STN-DBS with stable treatment and stimulation parameters for longer than 3 months. Most of the STN-DBS received bilateral monopolar stimulation ($n = 43$), the remaining being treated with bilateral bipolar stimulation ($n = 7$). Moreover, most of the patients received DBS at a frequency higher than 100 Hz ($n = 35$), whereas the remaining

patients received DBS at a frequency lower than 100 Hz ($n = 15$). Stimulation parameters were set to optimize motor symptoms and fluctuations (31, 32). DBS pulse width was set at 60 μs for all STN-DBS patients. All STN-DBS patients were evaluated, clinically and instrumentally, on stimulation (i.e., when ON DBS) and on medication. Participant demographic and clinical features (including the STN-DBS parameters) are reported in Table 1.

2.2. Voice recordings

Voice recordings were performed by asking healthy subjects and patients to produce a specific vocal task that consisted of the sustained emission of a close-mid front unrounded vowel /e/ for at least 5 s (2). All audio signals were collected in a quiet and echo-free room. Voice recordings were recorded by expert neurologists. All voice samples collected in this study from controls and patients were recorded using a specific smartphone available on the market, equipped with a high-definition microphone and a dedicated application allowing for recording in linear pulse-code modulation (PCM) format (.wav) at a sampling rate of 44.1 kHz, 16-bit depth, without compressions or filtering. Participants were asked to hold the smartphone in front of their face, at ∼30 cm from the mouth, and then to speak with their usual voice intensity, pitch, and quality (33) (Figure 1).
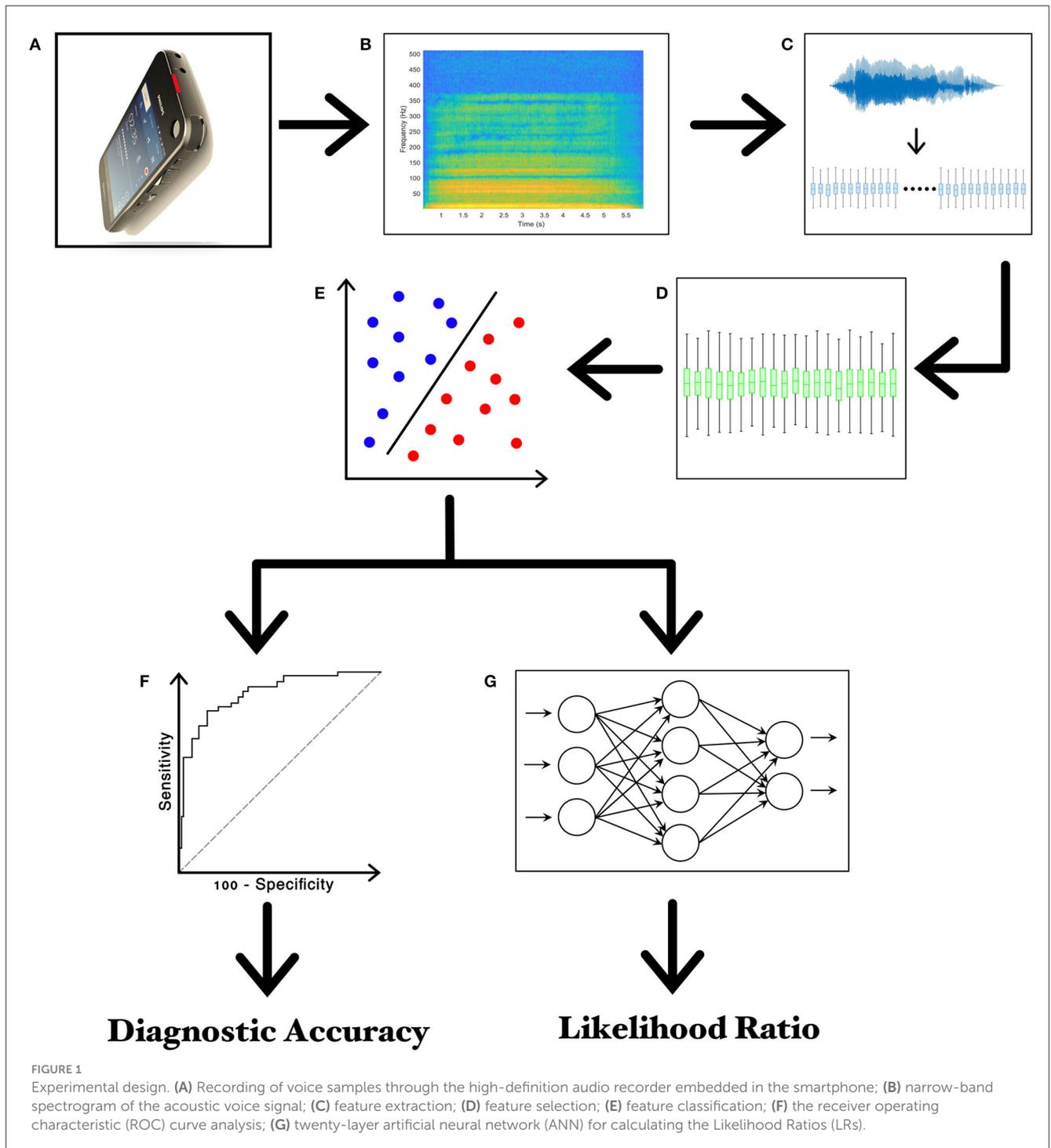
2.3. Machine-learning analysis

Specific spectral subtraction techniques, such as multi-band spectral subtraction, were initially used to remove background noise and other artifacts from each audio track of the voice sample. Spectral subtraction is a powerful noise reduction method based on a "learned" noise profile estimated during speech pauses and subtracted from the noisy spectrum to enhance speech. Specifically, we detected the frequency spectrum of the background noise by selecting specific sections of the audio tracks, including noise and other artifacts without biological signals (i.e., voice). The toolbox specifically employed in our analysis was the Izotope RX7 (iZotope, n.d.), which offers fine-tuning capabilities, enabling the algorithm to prioritize gating-like effects over the "musical noise" that exacerbates distortion. This procedure allowed us to reduce file corruption, possibly affecting the following analysis (34, 35).

Then, voice features underwent feature extraction and pre-processing through the Data Analytics Research and Technology in Healthcare group's Voice Analysis Toolbox (DARTH-VAT) (36). The DARTH-VAT Toolbox is open-source software provided by MATLAB (MathWorks, USA) that allows the extraction of a grand total of 345 acoustic features specific to the analysis of pathological voices. The main domains of extracted features are jitter, shimmer, HNR, glottal model-based features, empirical mode decomposition (EMD), entropy, Teager-Kaiser energy operator (TKEO), pitch period entropy (PPE), recurrence period density entropy (RPDE), and detrended fluctuation analysis (DFA) (36). Each domain entails several single-value descriptors, such as mean or standard deviation, computed as the result of a moving average on the original signal evolving in time. In

TABLE 1 Demographic and clinical features of healthy subjects (HS), STN-DBS and L-Dopa patients.

	Age (years)	Weight (kg)	Height (cm)	BMI	DD (years)	STN-DBS (years)	LEDDs	MMSE	BDI	FAB	H&Y	UPDRS-III	UPDRS-III-v	DBS freq. (Hz)	DBS pulse width (μs)	DBS intens. (mA) dx	DBS intens. (mA) sn
PD STN-DBS	61.6 ± 6.6	74.8 ± 13.7	168.7 ± 9.6	26.3 ± 4.4	15.7 ± 6.2	2.9 ± 5.2	747.3 ± 386.4	28.0 ± 2.1	9.6 ± 6.9	14.5 ± 2.8	2.7 ± 0.5	20.6 ± 8.1	2.5 ± 0.5	134.5 ± 41.1	58.1 ± 9.7	2.8 ± 1.5	3.0 ± 1.3
PD L-Dopa	62.1 ± 8.3	72.5 ± 12.0	172.1 ± 8.3	24.4 ± 3.1	14.0 ± 4.5	–	758.3 ± 317.6	28.0 ± 2.7	9.5 ± 4.0	14.8 ± 2.5	2.8 ± 0.4	24.5 ± 14.7	2.0 ± 0.6	–	–	–	–
HS	60.3 ± 10.3	68.5 ± 10.6	169.0 ± 10.1	23.9 ± 3.0	–	–	–	29.0 ± 0.8	3.3 ± 1.7	17.9 ± 1.1	–	–	–	–	–	–	–

DD, disease duration; LEDDs, L-Dopa equivalent daily doses; MMSE, Mini-Mental State Evaluation; BDI, Beck Inventory Scale for Depression; FAB, Frontal Assessment Battery; H&Y, Hoehn and Yahr Scale for assessment stage of PD; HS, healthy subjects; PD, patients with Parkinson's disease; UPDRS-III, Unified Parkinson's Disease Rating Scale part III; UPDRS-III-v, Unified Parkinson's Disease Rating Scale part III, voice impairment subitem; results are expressed as average ± standard deviation.

FIGURE 1
Experimental design. **(A)** Recording of voice samples through the high-definition audio recorder embedded in the smartphone; **(B)** narrow-band spectrogram of the acoustic voice signal; **(C)** feature extraction; **(D)** feature selection; **(E)** feature classification; **(F)** the receiver operating characteristic (ROC) curve analysis; **(G)** twenty-layer artificial neural network (ANN) for calculating the Likelihood Ratios (LRs).

addition, the DARTH Toolbox also provides additional algorithms, including SWIPE, that can extract vectors of values relative to the variation of fundamental frequency (f0) over time, such as mean, median, standard deviation, minimum, maximum, and 70% trim mean, which is the mean computed, excluding the 15% top and bottom values. Moreover, jitter, shimmer, HNR, and F0 are selected since they are common feature domains in the analysis of voice abnormalities in PD patients. Therefore, DARTH-VAT has been specifically implemented for detecting

voice abnormalities in PD patients, as shown by previous research (37–39).

Moreover, extracted features underwent feature selection pre-processing using the correlation-based feature selector (CFS) (40, 41) available as an open-source toolkit in Weka (42, 43). The optimal subset was chosen with the help of a (non-greedy) Best First Search method, which involves the selection of the optimal subset and path via progressive enlargement of the cardinality while evaluating the factor of merit. The most relevant features selected

by the CSF were ranked by relevance using the Information Gain Attribute Evaluation (IGAE) algorithm (44) available as an open-source toolkit in Weka (40, 41). The IGAE algorithm measures the information gained concerning the class.

After the pre-processing, the audio features underwent classification procedures. The classification focused on the 20 most relevant features, as ranked by the IGAE (23), and streamlined the data needed for machine-learning purposes (2, 20) (see Table 2). Given the relatively small dataset considered here, an SVM with a linear kernel and soft margins was used as a classifier. The SVM classifier is suitable for small datasets and noisy data since it allows for reducing the likelihood of "overfitting". Then, we applied Platt's Sequential Minimal Optimization method to perform the supervised training of voice features (45, 46). Platt's method is an algorithm used to train SVMs and solve their quadratic programming problem. Platt's method is a fast methodology based on iteratively solving analytically small sub-problems of minimization, which only involve two Lagrange multipliers (22). The SVM was also calibrated using a logistic regressor to convert its score-like output into probabilistic values suitable for producing ROC curves. Calibration essentially works by fitting a probabilistic model to various sub-versions of the main classifier to cast the observed likelihood of their outputs into probabilities (47). However, a hyperparameter optimization was also performed to find the best-performing setup for the SVM. The main hyperparameters of the SVM are complexity (or C), which quantifies the amount of penalization for a classification error within the training set, allowing for softer or harder margins, and the ridge of the calibrator. The optimization was performed automatically owing to a look-up table of discrete values for each parameter, effectively training various versions of SVM and then posteriorly choosing the best combination of hyperparameters.

Finally, in order to improve the biological interpretation of our results by providing automatic binary discriminations among the three classes of participants (i.e., HS, STN-DBS, and L-Dopa), we identified the smallest subset of features, which were then included in further analysis. As reported in the next section of results, among the most relevant and representative extracted features, we identified Jitter.F0_TKEO_mean, Shimmer.F0_TKEO_mean, and HNR_mean. The jitter and shimmer indicate the frequency and amplitude of micro-instability in vocal fold vibrations, respectively, and both contribute to rough speech. Conversely, HNR represents the amount of noise in voice signals. In the case of our analysis, the Jitter.F0_TKEO_mean and the Shimmer.F0_TKEO_mean were both calculated as the average of the jitter and the shimmer, respectively, as computed with the aid of a Teager-Kaiser energy operator, whereas the HNR_mean was calculated as the average of the HNR.

Finally, we performed a further machine-learning analysis for clinical-instrumental correlation purposes after achieving feature extraction and selection in parallel with the SVM classification procedures. We used a feed-forward artificial neural network (ANN) consisting of a 20-neuron input layer, a 10-neuron hidden layer, and a 1-neuron output layer. Input for ANN consisted of the first 20 most relevant selected features, which thus matched the 20-neuron input layer. Then, the ANN was trained to calculate a continuous numerical value (the likelihood ratio, or LR), ranging from 0 to 1 and reflecting the degree of voice impairment in each patient with PD (i.e., the closer the LRs are to 1, the higher the degree of voice impairment). ANN was trained by using the same selected features used to train the SVM. The experimental paradigm is also summarized in Figure 1 (22).

2.4. Statistical analysis

The normality of all parameters was assessed using the Kolmogorov-Smirnov test. The Mann-Whitney U-test was used to compare demographic and anthropometric parameters in HS, STN-DBS, and L-Dopa patients. The Mann-Whitney U-test was also used to compare the UPDRS-III and UPDRS-III-v scores between STN-DBS and L-Dopa patients. Finally, the Mann–Whitney U-test was used to compare UPDRS-III, UPDRS-III-v, and LRs values in STN-DBS patients who received monopolar or bipolar stimulation as well as in patients who received low (<100 Hz) and high STN-DBS frequencies (>100 Hz).

ROC analyses were performed to identify the optimal diagnostic cutoff values to discriminate between HS vs. L-Dopa patients, HS vs. STN-DBS patients, and STN-DBS vs. L-Dopa patients. We provided detailed values for sensibility, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and area under the curve (AUC). Moreover, we showed the output of the ROC analysis by calculating the Youden index and its optimal criterion value, the associated criterion.

Spearman's rank correlation coefficient was used to assess correlations between clinical scores (including the STN-DBS parameters) and LR values. A p-value of <0.05 was considered statistically significant.

3. Results

Demographic and anthropometric parameters were normally distributed and comparable in HS, STN-DBS, and L-Dopa patients (all $p > 0.05$). MMSE scores were comparable among groups (all $p > 0.05$). BDI was higher, and FAB was lower in PD patients than in controls (all $p < 0.05$). Disease duration, LEDDs, MMSE, BDI, FAB, H&Y, and UPDRS-III were all similar between STN-DBS and L-Dopa patients (all $p > 0.05$) (Table 1).

3.1. Voice impairment in STN-DBS and L-Dopa patients

According to our results, all PD patients included in our cohort manifested a variable degree of clinically overt voice impairment (UPDRS-III subitem voice, UPDRS-III-v ≥ 1). STN-DBS patients scored higher at UPDRS-III-v than L-Dopa patients ($p < 0.01$), suggesting greater voice impairment in the first study group. When considering only STN-DBS patients, UPDRS-III and UPDRS-III-v scores were comparable in the patients who received monopolar or bipolar stimulation ($p > 0.05$). UPDRS-III was also similar in the patients who received low (<100 Hz) and high STN-DBS

TABLE 2 Most relevant voice features selected by correlation-based feature selector (CFS) algorithm during the recording of the sustained emission of vowel/e/in healthy subjects (HS) vs. L-Dopa patients; HS vs. STN-DBS patients; and STN-DBS vs. L-Dopa patients.

Ranking position	HS vs. L-Dopa	HS vs. STN-DBS	STN-DBS vs. L-Dopa
1	DFA (Detrended Fluctuation Analysis)	F0s_median	HNR_mean
2	Jitter->F0_FM	Jitter->F0_FM	Jitter->F0_abs0th_perturb
3	mean_0th delta	Jitter->F0_TKEO_mean	Jitter->F0_TKEO_prc5
4	mean_10th delta	mean_MFCC_10th coef	Jitter->F0range_5_95_perc
5	mean_11th delta-delta	mean_MFCC_11th coef	mean_1st delta delta
6	mean_12th delta	mean_MFCC_1st coef	mean_delta delta 0th
7	mean_12th delta-delta	mean_MFCC_4th coef	mean_MFCC_4th coef
8	mean_1st delta delta	mean_MFCC_6th coef	NHR_mean
9	mean_2nd delta-delta	mean_MFCC_7th coef	Shimmer->F0_abs_dif
10	mean_4th delta-delta	PPE (Pitch Period Entropy)	Shimmer->F0_abs0th_perturb
11	mean_5th delta-delta	Shimmer->F0_abs0th_perturb	Shimmer->F0_FM
12	mean_MFCC_3rd coef	Shimmer->F0_FM	Shimmer->F0_PQ11_classical_Baken
13	mean_MFCC_6th coef	Shimmer->F0_TKEO_mean	Shimmer->F0_TKEO_mean
14	mean_MFCC_7th coef	Shimmer->F0_TKEO_prc75	Shimmer->F0_TKEO_prc5
15	NHR_mean	Shimmer->F0_TKEO_prc95	Shimmer->F0_TKEO_std
16	NHR_std	std_MFCC_10th coef	std_MFCC_10th coef
17	Shimmer->F0_FM	std_MFCC_11th coef	std_MFCC_12th coef
18	Shimmer->F0_PQ11_classical_Baken	std_MFCC_12th coef	std_MFCC_4th coef
19	Shimmer->F0_PQ5_classical_Baken	–	std_MFCC_5th coef
20	Shimmer->F0_TKEO_prc25	–	std_MFCC_6th coef

frequencies ($>100\,Hz$) ($p = 0.53$). Conversely, the patients who received STN-DBS at a frequency of $>100\,Hz$ manifested higher UPDRS-III-v scores than patients treated with a frequency of $<100\,Hz$ ($p < 0.05$).
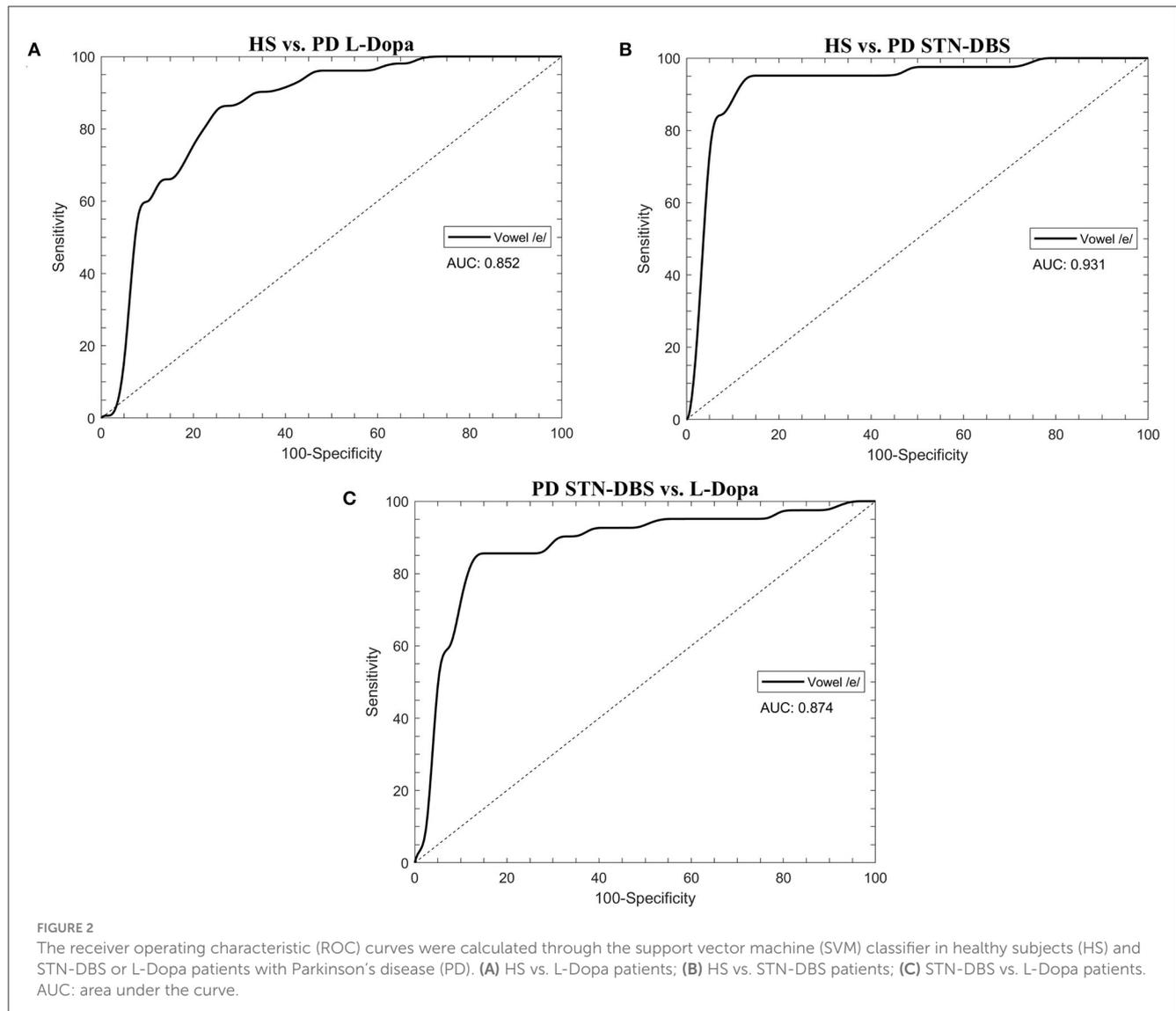
Concerning machine-learning analysis, voice samples collected from eight patients with PD (3 STN-DBS and 5 L-Dopa patients) were excluded from further analysis due to unexpected file corruption. When discriminating between HS and L-Dopa patients, the artificial classifier based on SVM allowed us to achieve a significant performance on our test. Specifically, when comparing the 20 most relevant selected features extracted from the sustained emission of the vowel, the ROC curve analyses identified an optimal diagnostic threshold value of 0.39 (associated criterion) when applying discretization and 10-fold cross-validation (Youden index = 0.63). Using this cutoff value, the performance of our diagnostic test was as follows: sensitivity = 80.0%, specificity = 78.9%, PPV = 78.4%, NPV = 80.4%, accuracy = 79.4%, and AUC = 0.852 (Figure 2A; Table 3).

When comparing HS and STN-DBS patients using the SVM, we achieved a significant diagnostic performance of our test, identifying an optimal diagnostic threshold value of 0.82 (associated criterion) when applying discretization and 10-fold cross-validation (Youden index = 0.83) to the 20 most relevant voice features. Using this cutoff value, we obtained the following: sensitivity = 88.6%, specificity = 95.8%, PPV = 95.1%, NPV = 90.2%, accuracy = 92.4%, and AUC = 0.874 (Figure 2B; Table 3).

When classifying STN-DBS and L-Dopa patients, the SVM applied to the 20 most relevant selected features extracted from the sustained emission of the vowel identified an optimal diagnostic threshold value of 0.51 (associated criterion) when applying discretization and 10-fold cross-validation (Youden index = 0.74). Using this cutoff value, the performance of our diagnostic test was consistent, as suggested by the following values: sensitivity = 85.4%, specificity = 88.2%, PPV = 85.4%, NPV = 88.2%, accuracy = 87.0%, and AUC = 0.874 (Figure 2C; Table 3).

Concerning the analysis of binary discriminations between the three groups of classes (i.e., HS, STN-DBS, and L-Dopa), we identified Jitter.F0_TKEO_mean, Shimmer.F0_TKEO_mean, and HNR_mean, among the most relevant and representative extracted features. We achieved the following results for the L-Dopa vs. HS classification: sensitivity = 67.3%, specificity = 57.1%, PPV = 81.5%, NPV = 38.5%, and accuracy = 64.7%. Moreover, concerning the DBS vs. L-Dopa classification, we obtained the following results: sensitivity = 78%, specificity = 72.9%, PPV = 66.7%, NPV = 82.7%, and accuracy = 75%. Finally, for the DBS vs. HS classification, we reported the following significant statistic output: sensitivity = 78.3%, specificity = 75.7%, PPV = 88.9%, NPV = 58.3%, and accuracy = 77.5%. The output of this further analysis is visually displayed in a 3D scatter plot (Figure 3).

The Mann-Whitney U-test showed comparable LR scores in STN-DBS patients receiving bilateral monopolar or bipolar

FIGURE 2
The receiver operating characteristic (ROC) curves were calculated through the support vector machine (SVM) classifier in healthy subjects (HS) and STN-DBS or L-Dopa patients with Parkinson's disease (PD). **(A)** HS vs. L-Dopa patients; **(B)** HS vs. STN-DBS patients; **(C)** STN-DBS vs. L-Dopa patients. AUC: area under the curve.

stimulation ($p > 0.05$), as well as in the patients who received low (<100 Hz) and high (>100 Hz) frequencies of STN-DBS ($p > 0.05$).

3.2. Correlation analysis

In L-Dopa patients, we found a positive correlation between UPDRS-III and UPDRS-III-v scores ($r = 0.40$, $p < 0.01$). A similar positive correlation between UPDRS-III and UPDRS-III-v ($r = 0.48$, $p < 0.01$) was also found when considering the cohort of STN-DBS patients. These findings demonstrate that the greater the disease severity, the higher the impairment of voice in L-Dopa patients as well as in STN-DBS. We also found that UPDRS-III and UPDRS-III-v scores did not correlate with years from the STN-DBS implant ($r = 0.09$, $p = 0.58$; $r = 0.08$, $p = 0.54$, respectively), the frequency ($r = 0.02$, $p = 0.92$; $r = 0.14$, $p = 0.37$, respectively),
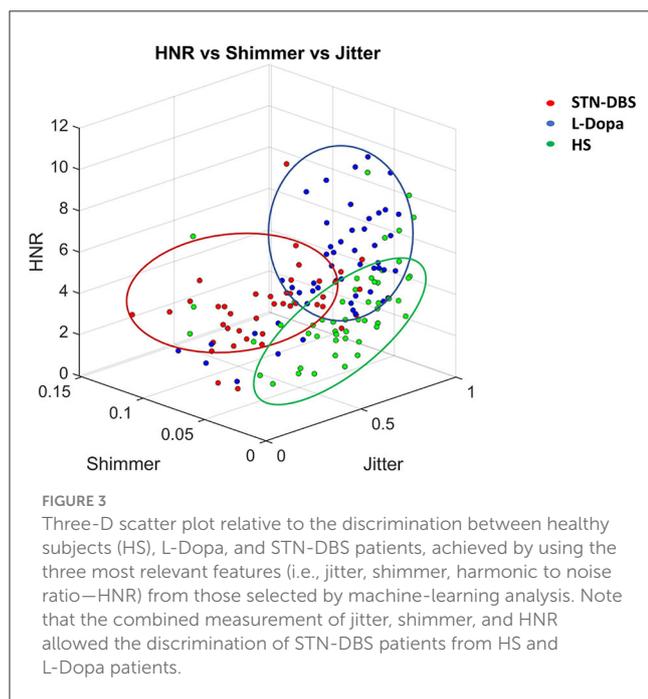
and the intensity of STN-DBS (mean value between the right and left STN-DBS electrodes) ($r = 0.12$, $p = 0.45$; $r = -0.18$, $p = 0.25$, respectively).

Concerning machine-learning analysis, we found that LR scores collected in L-Dopa patients positively correlated with UPDRS-III ($r = 0.31$, $p < 0.05$) and UPDRS-III-v ($r = 0.41$, $p < 0.01$) values. Moreover, when considering STN-DBS patients, we found a correlation between LR scores and UPDRS-III ($r = 0.51$, $p < 0.01$) as well as UPDRS-III-v values ($r = 0.33$, $p < 0.05$). Accordingly, our analysis showed that the higher the LR values calculated by machine learning, the greater the severity of motor (UPDRS-III) as well as voice (UPDRS-III-v) symptoms in both groups of PD patients (i.e., L-Dopa and STN-DBS). Finally, LR scores also correlated with the intensity (mean value between the right and left STN-DBS electrodes) ($r = 0.33$, $p < 0.05$) but not with the years from the STN-DBS implant ($r = 0.06$, $p > 0.05$) or the frequency of STN-DBS ($r = 0.08$, $p > 0.05$) (Figure 4).

TABLE 3 Performance of the machine learning algorithm.

Comparisons	Instances	Cross validation	Associated criterion	Youden index	Se (%)	Sp (%)	PPV (%)	NPV (%)	Acc (%)	AUC
HS vs. L-Dopa	154	10 folds	0.39	0.63	80.0	78.9	78.4	80.4	79.4	0.852
HS vs. STN-DBS	155	10 folds	0.82	0.83	88.6	95.8	95.1	90.2	92.4	0.931
STN-DBS vs. L-Dopa	93	10 folds	0.51	0.74	85.4	88.2	85.4	88.2	87.0	0.874

Performance of Support Vector Machine (SVM) linear classifier elaborating the 20 most relevant selected features during the sustained emission of the vowel/e/for three independent conditions: (1) Healthy subjects (HS) vs. L-Dopa patients; (2) HS vs. STN-DBS patients; (3) STN-DBS patients vs. L-Dopa patients. Selected features refer to the number of features able to obtain the best results; instances refer to the number of subjects considered in each comparison; cross validation refers to standardized validation procedures (see Methods for details). Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; Acc, accuracy; AUC, area under the curve.

FIGURE 3
Three-D scatter plot relative to the discrimination between healthy subjects (HS), L-Dopa, and STN-DBS patients, achieved by using the three most relevant features (i.e., jitter, shimmer, harmonic to noise ratio—HNR) from those selected by machine-learning analysis. Note that the combined measurement of jitter, shimmer, and HNR allowed the discrimination of STN-DBS patients from HS and L-Dopa patients.
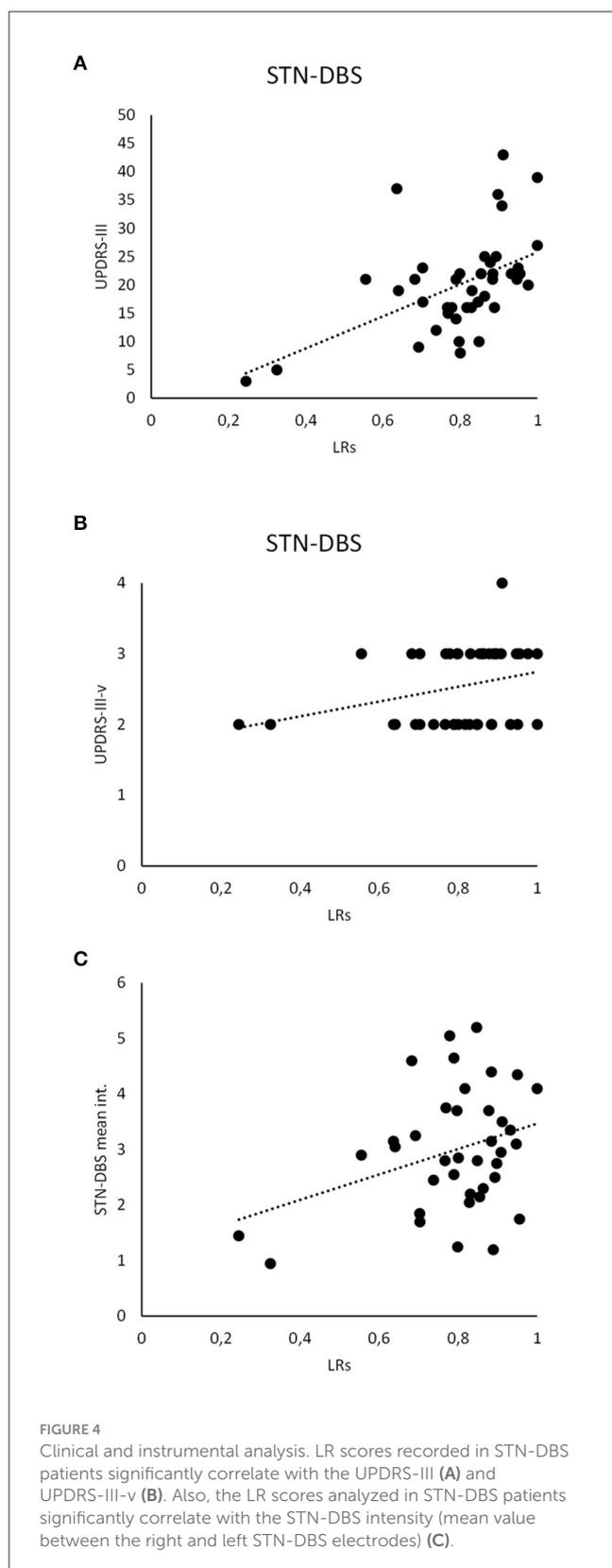
4. Discussion

The present study provided convergent data from perceptive (i.e., clinical) as well as instrumental analysis (i.e., machine-learning), showing the effect of STN-DBS on voice in PD patients. Indeed, STN-DBS significantly worsened *dysarthria* in patients with PD, leading to DBS-related dysarthria. Supporting this conclusion, we found significant clinical-instrumental correlations between machine-learning output measures (LRs) and the clinical assessment of voice impairment (UPDRS-III-v). Our study, therefore, indicates that machine-learning analysis is a reliable tool to assess voice abnormalities objectively in STN-DBS patients with PD.

The strengths of the study include the large sample of patients and their rigorous selection based on comparable demographic, anthropometric, and clinical parameters among groups. All patients were assessed clinically and instrumentally when ON L-Dopa. STN-DBS patients were clinically and instrumentally assessed when ON DBS and ON L-Dopa, with their chronic stimulation parameters (i.e., polarity, frequency, and intensity)

based on efficacy and safety on motor and non-motor symptoms, according to the best clinical practice (48, 49). The comparable LEDDs in STN-DBS and L-Dopa patients allowed us to exclude confounding factors due to dopaminergic stimulation when comparing implanted and not-implanted PD patients. The specific vocal task (i.e., sustained emission of the vowel/e/) was selected since it represents a language- and culture-free vocal task, according to previous reports (20, 21, 33, 50). All corrupted vocal samples were excluded from the analysis to avoid confounding factors due to non-biologic audio signals. Finally, our machine-learning analysis included the RASTA filtering technique, which allowed us to reduce the irrelevant and potentially misleading information added to the signal by the background noise or electromagnetic interference of the implantable pulse generator (33).

4.1. Clinical assessment of voice

The clinical observation that all patients manifested a certain degree of voice impairment (UPDRS-III-v $\geq$ 1) is consistent with the estimated prevalence of *hypokinetic dysarthria,* reaching 90% of the global PD population in the advanced stages of the disease (1, 4). Since our patients manifested higher BDI and lower FAB scores than controls, it might be argued that *hypokinetic dysarthria* also reflected a mild decline in mood and/or frontal functions. However, STN-DBS and L-Dopa patients were characterized by comparable overall disability (H&Y scores) and disease severity (UPDRS-III values), as well as BDI and FAB scores. The clinical observation that STN-DBS patients showed a higher degree of voice impairment (i.e., UPDRS-III-v) than L-Dopa patients indicates a more severe *dysarthria* in STN-DBS patients, in line with previous reports (7, 15–17, 51–54). Previous studies indeed reported prominent voice impairments characterized by a harsh, breathy, strained voice, hypernasality, imprecise consonant emission, speech rhythm disturbances, stammering, and stuttering in STN-DBS patients (13, 14, 53, 55). We also found a significant correlation between voice impairment (UPDRS-III-v scores) and overall disease severity (UPDRS-III scores), both in patients treated with STN-DBS and in those under L-Dopa, in line with previous observations (2, 13). Finally, concerning the specific STN-DBS parameters, we found that voice prominently deteriorated in patients receiving a higher ($>$100 Hz) rather than a lower frequency ($<$100 Hz) of STN-DBS. This finding t fully agrees

FIGURE 4
Clinical and instrumental analysis. LR scores recorded in STN-DBS patients significantly correlate with the UPDRS-III **(A)** and UPDRS-III-v **(B)**. Also, the LR scores analyzed in STN-DBS patients significantly correlate with the STN-DBS intensity (mean value between the right and left STN-DBS electrodes) **(C)**.

It is posited that high-frequency (>100 Hz) STN-DBS severely affects laryngeal coordination due to the current spreading to contiguous brain structures (57). Overall, our clinical assessment showed that STN-DBS patients manifest a significant worsening of *dysarthria* compared with those who received only L-Dopa therapy (17, 58).

4.2. Machine-learning analysis of voice

The accuracy achieved in discriminating L-Dopa patients from HS confirmed and expanded a recent observation from our group (2), showing that voice is altered in advanced-stage patients with PD under chronic L-Dopa treatment. This observation receives further support from the significant correlation we found between the instrumental scores (i.e., LRs) and the clinical impairment of voice (UPDRS-III-v scores) as well as motor symptoms (UPDRS-III scores) (2).

Machine learning achieved robust accuracy (92.4%) in the comparison between STN-DBS patients and controls, and the performance of the algorithm was significantly higher than that observed in the discrimination between controls and L-Dopa patients (79.4%). Moreover, machine learning achieved consistent accuracy in the comparison between STN-DBS and L-Dopa patients (87%). Again, the severity of motor (UPDRS-III) and voice (UPDRS-III-v) impairment significantly correlated with the instrumental scores (i.e., LRs) provided by the algorithm. Overall, these findings objectively demonstrate a significant worsening of voice in STN-DBS patients. Concerning the specific output of our machine-learning analysis in PD, it is worth noting that, when discriminating between STN-DBS and L-Dopa patients and, finally, healthy controls, the 20 most relevant features selected by our classifier included those reported in previous reports on spectral analysis, such as jitter, shimmer, HNR, and fundamental frequency (F0) (16, 53, 58). Further relevant biological information came from our final machine learning analysis concerning the most relevant voice features allowing discrimination among STN-DBS and L-Dopa patients and healthy controls. We demonstrated that the combination of only three independent voice features (Jitter.F0_TKEO_mean, Shimmer.F0_TKEO_mean, and the HNR_mean) allowed discrimination among the three groups of participants. Indeed, we found that jitter (i.e., the Jitter.F0_TKEO_mean) and shimmer (i.e., the Shimmer.F0_TKEO_mean) were both lower in HS than in L-Dopa and STN-DBS patients, whereas HNR (i.e., the HNR_mean) was higher in HS than in L-Dopa and STN-DBS patients. Jitter and shimmer indicate the frequency and amplitude of micro-instability in vocal fold vibrations, respectively, and both contribute to rough speech. Conversely, HNR represents the amount of noise in voice signals. Hence, we conclude that L-Dopa and STN-DBS patients are mostly characterized by abnormally rough and noisy speech compared with healthy controls. Overall, we confirm that jitter, shimmer, and HNR are very common domains in voice analysis in PD, allowing us to objectively recognize *dysarthria* in STN-DBS and L-Dopa patients.

with previous observations (13, 14, 31, 56), which outlined the well-known detrimental effect of high-frequency (>100 Hz) STN-DBS on phonatory and articulatory aspects of speech production.

4.3. Effect of STN-DBS on voice in PD: putative mechanisms

The prominent voice abnormalities observed in STN-DBS patients may reflect several mechanisms. We have recently reported that L-Dopa may improve, even though it does not restore *dysarthria* in PD (2). Following STN-DBS procedures, patients experience a significant reduction of LEDDs by ∼50% (59), as a result of relevant improvements in motor and non-motor symptoms (60, 61). If not given the STN-DBS procedure, patients would have probably required at least twice the dose of L-dopa. Accordingly, following STN-DBS, our patients would be characterized by prominent voice changes simply because of decreased LEDDs. However, we did not examine voice in STN-DBS patients after a further increase of LEDDs; both implanted and non-implanted subgroups received the best medical treatment and had comparable disease stages, severity, and duration, thus making the hypothesis of suboptimal LEDDs rather unlikely. Alternatively, a mechanism for explaining the STN-DBS-related worsening of *dysarthria* in PD would imply a specific pathophysiological effect of electric stimulation on target neuronal populations. The DBS implanted in the STN may activate antidromically axons of the hyperdirect pathway (i.e., cortico-subthalamic fibers), which in turn may lead to abnormal activation of cortical areas involved in voice production, thus leading to stuttering and spastic speech (16, 19, 53, 62–64). Another reasonable mechanism would imply the spread of current from STN to contiguous brain structures owing to horizontal propagation of the electric field and the related volume of tissue activation (VTA) (65). Accordingly, STN-DBS would deteriorate voice in PD owing to the spread of the VTA to the descending corticobulbar and corticospinal tracts (53, 54, 58, 66). Moreover, an additional mechanism would imply the propagation of VTA to ascending fibers traveling in the cerebellothalamic and pallid-thalamic radiation, including those in the adjacent medial Zona Incerta, Hassler's pre-lemniscal radiation, and Forel's prerubral field or H-field (16, 17). Finally, an alternative hypothesis of DBS-related dysarthria would imply the lead location of DBS within the STN in our PD patients, as suggested by previous reports showing differential motor outcomes following the stimulation of the posterolateral/dorsomedial portion of the STN (67). Although our study lacks the neuroimaging reconstruction of electrode position and VTA for each patient, the correlation we found between the intensity used for STN-DBS and LRs values (i.e., the higher the STN-DBS intensity, the greater the voice impairment) provides support to the hypothesis of STN-DBS deleterious effect on voice in PD as a result of VTA propagation to contiguous brain structures (12, 15, 17). Accordingly, we speculate that STN-DBS deteriorates voice in PD through an abnormal engagement of specific brain structures included in the human phonological loop (68). The phonological loop is a complex cortico-subcortical network that mediates speech planning, programming, and articulation and includes regions such as the inferior frontal gyrus, supplementary motor area, primary somatosensory cortex, superior temporal gyrus, and inferior parietal lobule (69, 70). The phonological loop also includes subcortical regions, such as the striatum (i.e., the putamen) and interconnected basal ganglia nuclei (69). The cortical output of the phonological loop is the laryngeal primary motor cortex and its descending projections directed to alpha-motoneurons in the brainstem structure responsible for speech articulation, such as the nucleus ambiguous (69). In patients with PD and *hypokinetic dysarthria*, previous neuroimaging studies indeed reported abnormal activation of cortical and subcortical areas included in the phonological loop, such as the supplementary motor area, inferior lateral premotor cortex, and putamen. We, therefore, conjecture that STN-DBS may deteriorate *dysarthria* in patients with PD by degrading the activity of the phonological loop, a hypothesis that requires further investigation in future studies.

4.4. Limitations

When interpreting our results, several limitations should be considered. We did not record vocal samples before and after surgery or examine patients in a pharmacological OFF state or with the stimulator turned off (OFF DBS). Hence, our results do not fully explain the specific interaction of STN-DBS with dopaminergic stimulation and their combined effect on the voice in PD. This will be the topic of a future study. Moreover, the variable timing of observation after surgery (2.9 ± 5.2 years) would not affect the overall interpretation of our findings since we found no correlation between UPDRS-III-v scores as well as LRs and years from the STN-DBS implant. Our artificial intelligence could not discriminate between various components of DBS-related dysarthria (i.e., spastic and hypokinetic) in patients with PD during the analysis. Thus, this will be the topic of a future study. Furthermore, in the absence of neuroimaging data allowing the reconstruction of electrode position within the STN and the resulting VTA for each patient, our new pathophysiological interpretation based on STN-DBS interference on the human phonological loop remains rather speculative. Also, we recognize that a speech task based on the sustained emission of a vowel would be judged as not sufficient for analyzing speech production thoroughly and that short language-specific sentences based on various phonological features would provide additional results. However, as demonstrated in our previous studies, vowel emission can provide diagnostic accuracies similar to those achieved by more detailed speech tasks, including the reading of sentences (2, 71, 72). Moreover, the vowel emission gives the advantage of a language- and culture-free speech task that is useful for cohorts of advanced-stage PD patients (20, 33). Hence, we believe that sustained vowel emission represents a useful task for interpreting speech-related abnormalities in STN-DBS patients.

5. Conclusions

We here report the first machine learning study of voice in a homogeneous and clinically well-characterized cohort of PD patients and provide instrumental evidence of significant worsening of *dysarthria* in STN-DBS patients, thus leading to DBS-related dysarthria. Owing to an accurate methodology based on a cross-sectional design, our findings demonstrate that STN-DBS exerts a relevant impact on *dysarthria*, particularly when given at high frequency and intensity of stimulation.

Our observations in PD can pave the way for new approaches based on machine-learning analysis of voice associated with current steering technology or adaptive stimulation to optimize the overall management of motor symptoms and fluctuations without worsening *dysarthria* in STN-DBS patients (65, 73, 74). Future studies based on a comparative analysis between vowel emission and short language-specific sentences would also be of help in clarifying the pathophysiologic underpinnings of DBS-related dysarthria.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Institutional Ethics Committee of IRCCS Neuromed Institute (NCT04846413). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

ASu: Conceptualization, Data curation, Investigation, Supervision, Validation, Writing—review and editing, Writing—original draft. FA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing—original draft, Writing—review and editing. GC: Data curation, Formal analysis, Methodology, Project administration, Software, Supervision, Validation, Writing—original draft. FB: Data curation, Investigation, Project administration, Writing—original draft. CP: Investigation, Methodology, Writing—original draft. FP: Investigation, Methodology, Writing—original draft. RC: Investigation, Methodology, Writing—original draft. LB: Investigation, Methodology, Writing—original draft. VC: Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Writing—original draft. SP: Investigation, Methodology, Writing—original draft. NM: Investigation, Methodology, Writing—original draft. AZ: Investigation, Methodology, Writing—original draft. PS: Investigation, Methodology, Writing—original draft. MP: Investigation, Methodology, Writing—original draft. TT: Investigation, Methodology, Writing—original draft. APi: Writing—review and editing. APe: Investigation, Methodology, Writing—original draft. ASt: Writing—review and editing. PC: Writing—review and editing. AB: Writing—review and editing. GS: Writing—review and editing.

Lazio DBS study group

Maria Concetta Altavista (Neurology Unit, San Filippo Neri Hospital ASL Roma 1, Roma, Italy); Alessandra Calciulli

(Department of Brain and Behavioral Sciences, University of Pavia and IRCCS Mondino Foundation, Pavia, Italy); Marco Ciavarro (IRCCS Neuromed, Pozzilli, Italy); Francesca Cortese (Neurology Unit, San Filippo Neri Hospital ASL Roma 1, Roma, Italy); Antonio Daniele (Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS and Department of Neuroscience, Università Cattolica del Sacro Cuore, Rome, Italy); Alessandro De Biase (Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Manuela D'Ercole (Neurosurgery, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Lazzaro Di Biase (Unit of Neurology, Neurophysiology and Neurobiology, Department of Medicine, Campus Bio-Medico of Rome University, Rome, Italy); Daniela Di Giuda (Nuclear Medicine Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Pietro Di Leo (Department of Electronic Engineering, University of Rome Tor Vergata, Rome, Italy); Danilo Genovese (Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Isabella Imbimbo (Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Alessandro Izzo (Neurosurgery, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Rosa Liperoti (Geriatrics, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Giuseppe Marano (Institute of Psychiatry and Psychology, Department of Geriatrics, Neuroscience and Orthopedics, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy); Massimo Marano (Unit of Neurology, Neurophysiology and Neurobiology, Department of Medicine, Campus Bio-Medico of Rome University, Rome, Italy); Marianna Mazza (Institute of Psychiatry and Psychology, Department of Geriatrics, Neuroscience and Orthopedics, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy); Alessandra Monge (Department of Neurology, St John the Baptist Hospital, ACISMOM, Rome, Italy); Nicola Montano (Neurosurgery, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Michela Orsini (Clinical Psychology, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Leonardo Rigon (Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Marina Rizz (Neurology Unit, Azienda Ospedaliera Ospedali Riuniti Villa Sofia e Cervello, Palermo, Italy); Camilla Rocchi (Department of Systems Medicine, University of Rome Tor Vergata, Rome, Italy); Gennaro Saporito (Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, L'Aquila, Italy Department of Neurology, San Salvatore Hospital, L'Aquila, Italy); Laura Vacca (University and Institute for Research and Medical Care IRCCS San Raffaele, Rome, Italy); Fabio Viselli, MD (Department of Neurology, St John the Baptist Hospital, ACISMOM, Rome, Italy).

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Ma A, Lau KK, Thyagarajan D. Voice changes in Parkinson's disease: what are they telling us? *J Clin Neurosci.* (2020) 72:1–7. doi: 10.1016/j.jocn.2019.12.029

2. Suppa A, Costantini G, Asci F, Di Leo P, Al-Wardat MS, Di Lazzaro G, et al. Voice in Parkinson's disease: a machine learning study. *Front Neurol.* (2022) 13:831428. doi: 10.3389/fneur.2022.831428

3. Miglis MG, Adler CH, Antelmi E, Arnaldi D, Baldelli L, Boeve BF, et al. Biomarkers of conversion to α-synucleinopathy in isolated rapid-eye-movement sleep behaviour disorder. *Lancet Neurol.* (2021) 20:671–84. doi: 10.1016/S1474-4422(21)00176-9

4. Rusz J, Tykalová T, Novotný M, Růžička E, Dušek P. Distinct patterns of speech disorder in early-onset and late-onset de-novo Parkinson's disease. *NPJ Parkinsons Dis.* (2021) 7:98. doi: 10.1038/s41531-021-00243-1

5. Rusz J, Tykalova T, Novotny M, Zogala D, Sonka K, Ruzicka E, et al. Defining speech subtypes in *de novo* Parkinson disease: response to long-term levodopa therapy. *Neurology.* (2021) 97:e2124–35. doi: 10.1212/WNL.0000000000012878

6. Rusz J, Hlavnička J, Novotný M, Tykalová T, Pelletier A, Montplaisir J, et al. Speech biomarkers in rapid eye movement sleep behavior disorder and Parkinson disease. *Ann Neurol.* (2021) 90:62–75. doi: 10.1002/ana.26085

7. Limousin P, Foltynie T. Long-term outcomes of deep brain stimulation in Parkinson disease. *Nat Rev Neurol.* (2019) 15:234–42. doi: 10.1038/s41582-019-0145-9

8. Bove F, Mulas D, Cavallieri F, Castrioto A, Chabardès S, Meoni S, et al. Long-term outcomes (15 years) after subthalamic nucleus deep brain stimulation in patients with Parkinson disease. *Neurology.* (2021) 97:e254–62. doi: 10.1212/WNL.0000000000012246

9. Follett KA, Weaver FM, Stern M, Hur K, Harris CL, Luo P, et al. Pallidal versus subthalamic deep-brain stimulation for Parkinson's disease. *N Engl J Med.* (2010) 362:2077–91. doi: 10.1056/NEJMoa0907083

10. Zampogna A, Cavallieri F, Bove F, Suppa A, Castrioto A, Meoni S, et al. Axial impairment and falls in Parkinson's disease: 15 years of subthalamic deep brain stimulation. *NPJ Parkinsons Dis.* (2022) 8:121. doi: 10.1038/s41531-022-00383-y

11. Baudouin R, Lechien JR, Carpentier L, Gurruchaga J-M, Lisan Q, Hans S. deep brain stimulation impact on voice and speech quality in Parkinson's disease: a systematic review. *Otolaryngol Head Neck Sur.g.* (2022) 168:307–18. doi: 10.1177/01945998221120189

12. Fabbri M, Natale F, Artusi CA, Romagnolo A, Bozzali M, Giulietti G, et al. Deep brain stimulation fine-tuning in Parkinson's disease: short pulse width effect on speech. *Parkinson Relat Disord.* (2021) 87:130–4. doi: 10.1016/j.parkreldis.2021.05.007

13. Tripoliti E, Strong L, Hickey F, Foltynie T, Zrinzo L, Candelario J, et al. Treatment of dysarthria following subthalamic nucleus deep brain stimulation for Parkinson's disease. *Mov Disord.* (2011) 26:2434–6. doi: 10.1002/mds.23887

14. Tripoliti E, Zrinzo L, Martinez-Torres I, Frost E, Pinto S, Foltynie T, et al. Effects of subthalamic stimulation on speech of consecutive patients with Parkinson disease. *Neurology.* (2011) 76:80–6. doi: 10.1212/WNL.0b013e318203e7d0

15. Kleiner-Fisman G, Herzog J, Fisman DN, Tamma F, Lyons KE, Pahwa R, et al. Subthalamic nucleus deep brain stimulation: summary and meta-analysis of outcomes. *Mov Disord.* (2006) 21(Suppl. 14):S290–304. doi: 10.1002/mds.20962

16. Tanaka Y, Tsuboi T, Watanabe H, Nakatsubo D, Maesawa S, Kato S, et al. longitudinal speech change after subthalamic nucleus deep brain stimulation in Parkinson's disease patients: a 2-year prospective study. *J Parkinsons Dis.* (2020) 10:131–40. doi: 10.3233/JPD-191798

17. Tripoliti E, Limousin P, Foltynie T, Candelario J, Aviles-Olmos I, Hariz MI, et al. Predictive factors of speech intelligibility following subthalamic nucleus stimulation in consecutive patients with Parkinson's disease. *Mov Disord.* (2014) 29:532–8. doi: 10.1002/mds.25816

18. Koeglsperger T, Palleis C, Hell F, Mehrkens JH, Bötzel K. Deep brain stimulation programming for movement disorders: current concepts and evidence-based strategies. *Front Neurol.* (2019) 10:410. doi: 10.3389/fneur.2019.00410

19. Phokaewvarangkul O, Boonpang K, Bhidayasiri R. Subthalamic deep brain stimulation aggravates speech problems in Parkinson's disease: objective and subjective analysis of the influence of stimulation frequency and electrode contact location. *Parkinson Relat Disord.* (2019) 66:110–6. doi: 10.1016/j.parkreldis.2019.07.020

20. Asci F, Costantini G, Saggio G, Suppa A. Fostering voice objective analysis in patients with movement disorders. *Mov Disord.* (2021) 36:1041. doi: 10.1002/mds.28537

21. Asci F, Costantini G, Di Leo P, Saggio G, Suppa A. Reply to: "reproducibility of voice analysis with machine learning". *Mov Disord.* (2021) 36:1283–4. doi: 10.1002/mds.28601

22. Saggio G, Costantini G. Worldwide healthy adult voice baseline parameters: a comprehensive review. *J Voice.* (2022) 36:637–49. doi: 10.1016/j.jvoice.2020.08.028

23. Suppa A, Asci F, Saggio G, Marsili L, Casali D, Zarezadeh Z, et al. Voice analysis in adductor spasmodic dysphonia: objective diagnosis and response to botulinum toxin. *Parkinson Relat Disord.* (2020) 73:23–30. doi: 10.1016/j.parkreldis.2020.03.012

24. Suppa A, Asci F, Saggio G, Di Leo P, Zarezadeh Z, Ferrazzano G, et al. Voice analysis with machine learning: one step closer to an objective diagnosis of essential tremor. *Mov Disord.* (2021) 36:1401–10. doi: 10.1002/mds.28508

25. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord.* (2015) 30:1591–601. doi: 10.1002/mds.26424

26. Antonini A, Abbruzzese G, Ferini-Strambi L, Tilley B, Huang J, Stebbins GT, et al. Validation of the Italian version of the Movement Disorder Society–Unified Parkinson's Disease Rating Scale. *Neurol Sci.* (2013) 34:683–7. doi: 10.1007/s10072-012-1112-z

27. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* (1975) 12:189–98. doi: 10.1016/0022-3956(75)90026-6

28. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry.* (1961) 4:561–71. doi: 10.1001/archpsyc.1961.01710120031004

29. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. *Neurology.* (2000) 55:1621–6. doi: 10.1212/WNL.55.11.1621

30. Hacker ML, Turchan M, Heusinkveld LE, Currie AD, Millan SH, Molinari AL, et al. Deep brain stimulation in early-stage Parkinson disease: five-year outcomes. *Neurology.* (2020) 95:e393–401. doi: 10.1212/WNL.0000000000009946

31. Grover T, Georgiev D, Kalliola R, Mahlknecht P, Zacharia A, Candelario J, et al. Effect of Low versus High Frequency Subthalamic Deep Brain Stimulation on Speech Intelligibility and Verbal Fluency in Parkinson's Disease: a double-blind study. *JPD.* (2019) 9:141–51. doi: 10.3233/JPD-181368

32. Zibetti M, Moro E, Krishna V, Sammartino F, Picillo M, Munhoz RP, et al. Low-frequency subthalamic stimulation in Parkinson's disease: long-term outcome and predictors. *Brain Stimul.* (2016) 9:774–9. doi: 10.1016/j.brs.2016.04.017

33. Asci F, Costantini G, Di Leo P, Zampogna A, Ruoppolo G, Berardelli A, et al. Machine-learning analysis of voice samples recorded through smartphones: the combined effect of ageing and gender. *Sensors.* (2020) 20. doi: 10.3390/s20185022

34. Er MB, Isik E, Isik I. Parkinson's detection based on combined CNN and LSTM using enhanced speech signals with variational mode decomposition. *Biomed Signal Process Control.* (2021) 70:103006. doi: 10.1016/j.bspc.2021.103006

35. Upadhyay N, Karmakar A. Speech enhancement using spectral subtraction-type algorithms: a comparison and simulation study. *Proc Comput Sci.* (2015) 54:574–84. doi: 10.1016/j.procs.2015.06.066

36. Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng.* (2010) 57:884–93. doi: 10.1109/TBME.2009.2036000

37. Ozbolt AS, Moro-Velazquez L, Lina I, Butala AA, Dehak N. Things to consider when automatically detecting Parkinson's disease using the phonation of sustained vowels: analysis of methodological issues. *Appl Sci.* (2022) 12:991. doi: 10.3390/app12030991

38. Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *J R Soc Interface.* (2011) 8:842–55. doi: 10.1098/rsif.2010.0456

39. Scimeca S, Amato F, Olmo G, Asci F, Suppa A, Costantini G, et al. Robust and language-independent acoustic features in Parkinson's disease. *Front Neurol.* (2023) 14:1198058. doi: 10.3389/fneur.2023.1198058

40. Hall MA, Smith LA. *Practical Feature Subset Selection for Machine Learning.* Springer (1998). p. 181–191.

41. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explor Newsl.* (2009) 11:10. doi: 10.1145/1656274.1656278

42. Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-source audio feature extractor. In: *Proceedings of the International Conference on Multimedia - MM '10.* Firenze: ACM Press (2010). 1459 p.

43. Schuller B, Vlasenko B, Eyben F, Wollmer M, Stuhlsatz A, Wendemuth A, et al. Cross-corpus acoustic emotion recognition: variances and strategies. *IEEE Transact Affect Comp.* (2010) 1:119–131. doi: 10.1109/T-AFFC.2010.8

44. Yücelbaş C. A new approach: information gain algorithm-based k-nearest neighbors hybrid diagnostic system for Parkinson's disease. *Phys Eng Sci Med.* (2021) 44:511–24. doi: 10.1007/s13246-021-01001-6

45. Kullback S, Leibler RA. On Information and sufficiency. *Ann Math Statist.* (1951) 22:79–86. doi: 10.1214/aoms/1177729694

46. Platt J. *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines.* Microsoft (1998).

47. Morrison GS. Tutorial on logistic-regression calibration and fusion:converting a score to a likelihood ratio. *Aust J For Sci.* (2013) 45:173–97. doi: 10.1080/00450618.2012.733025

48. Bratsos S, Karponis D, Saleh SN. Efficacy and safety of deep brain stimulation in the treatment of parkinson's disease: a systematic review and meta-analysis of randomized controlled trials. *Cureus.* (2018) 10:e3474. doi: 10.7759/cureus.3474

49. Mansouri A, Taslimi S, Badhiwala JH, Witiw CD, Nassiri F, Odekerken VJJ, et al. Deep brain stimulation for Parkinson's disease: meta-analysis of results of randomized trials at varying lengths of follow-up. *J Neurosurg.* (2018) 128:1199–213. doi: 10.3171/2016.11.JNS16715

50. Rusz J, Tykalova T, Ramig LO, Tripoliti E. Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders. *Mov Disord.* (2021) 36:803–14. doi: 10.1002/mds.28465

51. Fenoy AJ, McHenry MA, Schiess MC. Speech changes induced by deep brain stimulation of the subthalamic nucleus in Parkinson disease: involvement of the dentatorubrothalamic tract. *J Neurosurg.* (2017) 126:2017–27. doi: 10.3171/2016.5.JNS16243

52. Tsuboi T, Watanabe H, Tanaka Y, Ohdake R, Hattori M, Kawabata K, et al. Early detection of speech and voice disorders in Parkinson's disease patients treated with subthalamic nucleus deep brain stimulation: a 1-year follow-up study. *J Neural Transm.* (2017) 124:1547–56. doi: 10.1007/s00702-017-1804-x

53. Tsuboi T, Watanabe H, Tanaka Y, Ohdake R, Yoneyama N, Hara K, et al. Distinct phenotypes of speech and voice disorders in Parkinson's disease after subthalamic nucleus deep brain stimulation. *J Neurol Neurosurg Psychiatry.* (2015) 86:856–64. doi: 10.1136/jnnp-2014-308043

54. Tsuboi T, Watanabe H, Tanaka Y, Ohdake R, Yoneyama N, Hara K, et al. Characteristic laryngoscopic findings in Parkinson's disease patients after subthalamic nucleus deep brain stimulation and its correlation with voice disorder. *J Neural Transm.* (2015) 122:1663–72. doi: 10.1007/s00702-015-1436-y

55. Tanaka Y, Nishio M, Niimi S. Vocal acoustic characteristics of patients with Parkinson's disease. *Folia Phoniatr Logop.* (2011) 63:223–30. doi: 10.1159/000322059

56. Moreau C, Pennel-Ployart O, Pinto S, Plachez A, Annic A, Viallet F, et al. Modulation of dysarthropneumophonia by low-frequency STN DBS in advanced Parkinson's disease: modulation of dysarthropneumophonia. *Mov Disord.* (2011) 26:659–63. doi: 10.1002/mds.23538

57. Morello ANDC, Beber BC, Fagundes VC, Cielo CA, Rieder CRM. Dysphonia and dysarthria in people with Parkinson's disease after subthalamic nucleus deep brain stimulation: effect of frequency modulation. *Journal of Voice.* (2020) 34:477–84. doi: 10.1016/j.jvoice.2018.10.012

58. Klostermann F, Ehlen F, Vesper J, Nubel K, Gross M, Marzinzik F, et al. Effects of subthalamic deep brain stimulation on dysarthrophonia in Parkinson's disease. *J Neurol Neurosurg Psychiatry.* (2008) 79:522–9. doi: 10.1136/jnnp.2007.123323

59. Lachenmayer ML, Mürset M, Antih N, Debove I, Muellner J, Bompart M, et al. Subthalamic and pallidal deep brain stimulation for Parkinson's disease-meta-analysis of outcomes. *NPJ Parkinsons Dis.* (2021) 7:77. doi: 10.1038/s41531-021-00223-5

60. Kurtis MM, Rajah T, Delgado LF, Dafsari HS. The effect of deep brain stimulation on the non-motor symptoms of Parkinson's disease: a critical review of the current evidence. *NPJ Parkinsd Dis.* (2017) 3:16024. doi: 10.1038/npjparkd.2016.24

61. Muthuraman M, Koirala N, Ciolac D, Pintea B, Glaser M, Groppa S, et al. Deep brain stimulation and L-DOPA therapy: concepts of action and clinical applications in Parkinson's disease. *Front Neurol.* (2018) 9:711. doi: 10.3389/fneur.2018.00711

62. Karlsson F, Blomstedt P, Olofsson K, Linder J, Nordh E, van Doorn J. Control of phonatory onset and offset in Parkinson patients following deep brain stimulation of the subthalamic nucleus and caudal zona incerta. *Parkinson Relat Disord.* (2012) 18:824–7. doi: 10.1016/j.parkreldis.2012.03.025

63. Lange F, Eldebakey H, Hilgenberg A, Weigl B, Eckert M, DeSunda A, et al. Distinct phenotypes of stimulation-induced dysarthria represent different cortical networks in STN-DBS. *Parkinson Relat Disord.* (2023) 109:105347. doi: 10.1016/j.parkreldis.2023.105347

64. Pinto S, Nebel A, Rau J, Espesser R, Maillochon P, Niebuhr O, et al. Results of a randomized clinical trial of speech after early neurostimulation in Parkinson's disease. *Mov Disord.* (2023) 38:212–22. doi: 10.1002/mds.29282

65. Frey J, Cagle J, Johnson KA, Wong JK, Hilliard JD, Butson CR, et al. Past, present, and future of deep brain stimulation: hardware, software, imaging, physiology and novel approaches. *Front Neurol.* (2022) 13:825178. doi: 10.3389/fneur.2022.825178

66. Dayal V, Grover T, Tripoliti E, Milabo C, Salazar M, Candelario-McKeown J, et al. Short versus conventional pulse-width deep brain stimulation in Parkinson's disease: a randomized crossover comparison. *Mov Disord.* (2020) 35:101–8. doi: 10.1002/mds.27863

67. Vitek JL, Patriat R, Ingham L, Reich MM, Volkmann J, Harel N. Lead location as a determinant of motor benefit in subthalamic nucleus deep brain stimulation for Parkinson's disease. *Front Neurosci.* (2022) 16:1010253. doi: 10.3389/fnins.2022.1010253

68. Baddeley AD, Hitch GJ. The phonological loop as a buffer store: an update. *Cortex.* (2019) 112:91–106. doi: 10.1016/j.cortex.2018.05.015

69. Fuertinger S, Horwitz B, Simonyan K. The functional connectome of speech control. *PLoS Biol.* (2015) 13:e1002209. doi: 10.1371/journal.pbio.1002209

70. Valeriani D, Simonyan K. The dynamic connectome of speech control. *Philos Trans R Soc Lond B Biol Sci.* (2021) 376:20200256. doi: 10.1098/rstb.2020.0256

71. Lee VS, Zhou XP, Rahn DA, Wang EQ, Jiang JJ. Perturbation and nonlinear dynamic analysis of acoustic phonatory signal in Parkinsonian patients receiving deep brain stimulation. *J Commun Disord.* (2008) 41:485–500. doi: 10.1016/j.jcomdis.2008.02.001

72. Skodda S. Effect of deep brain stimulation on speech performance in Parkinson's disease. *Parkinson's Disease.* (2012) 2012:1–10. doi: 10.1155/2012/850596

73. Krauss JK, Lipsman N, Aziz T, Boutet A, Brown P, Chang JW, et al. Technology of deep brain stimulation: current status and future directions. *Nat Rev Neurol.* (2021) 17:75–87. doi: 10.1038/s41582-020-00426-z

74. Little S, Tripoliti E, Beudel M, Pogosyan A, Cagnan H, Herz D, et al. Adaptive deep brain stimulation for Parkinson's disease demonstrates reduced speech side effects compared to conventional stimulation in the acute setting. *J Neurol Neurosurg Psychiatry.* (2016) 87:1388–9. doi: 10.1136/jnnp-2016-313518