Check for updates

RESEARCH ARTICLE

Voice Analysis with Machine Learning: One Step Closer to an Objective Diagnosis of Essential Tremor

Antonio Suppa, MD, PhD,[1,2†] Francesco Asci, MD,[1†] Giovanni Saggio, PhD,[3] Pietro Di Leo, BA,[3] Zakarya Zarezadeh, BA,[3] Gina Ferrazzano, MD, PhD,[1] Giovanni Ruoppolo, MD,[4] Alfredo Berardelli, MD,[1,2*] and Giovanni Costantini, PhD[3]

[1]*Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy*
[2]*IRCCS Neuromed Institute, Pozzilli, Italy*
[3]*Department of Electronic Engineering, University of Rome Tor Vergata, Rome, Italy*
[4]*Department of Sense Organs, Otorhinolaryngology Section, Sapienza University of Rome, Rome, Italy*

ABSTRACT: Background: Patients with essential tremor have upper limb postural and action tremor often associated with voice tremor. The objective of this study was to objectively examine voice tremor and its response to symptomatic pharmacological treatment in patients with essential tremor using voice analysis consisting of power spectral analysis and machine learning.
Methods: We investigated 58 patients (24 men; mean age ± SD, 71.7 ± 9.2 years; range, 38–85 years) and 74 age- and sex-matched healthy subjects (20 men; mean age ± SD, 71.0 ± 12.4 years; range, 43–95 years). We recorded voice samples during sustained vowel emission using a high-definition audio recorder. Voice samples underwent sound signal analysis, including power spectral analysis and support vector machine classification. We compared voice recordings in patients with essential tremor who did and did not manifest clinically overt voice tremor and in patients who were and were not under the symptomatic effect of the best medical treatment.
Results: Power spectral analysis demonstrated a prominent oscillatory activity peak at 2–6 Hz in patients who manifested a clinically overt voice tremor. Voice analysis with support vector machine classifier objectively discriminated with high accuracy between controls and patients who did and did not manifest clinically overt voice tremor and between patients who were and were not under the symptomatic effect of the best medical treatment.
Conclusions: In patients with essential tremor, voice tremor is characterized by abnormal oscillatory activity at 2–6 Hz. Voice analysis, including power spectral analysis and support vector machine classification, objectively detected voice tremor and its response to symptomatic pharmacological treatment in patients with essential tremor. © 2021 International Parkinson and Movement Disorder Society

Key Words: voice tremor; essential tremor; spectral analysis; machine learning; beta-blockers

Essential tremor (ET) is the most frequent movement disorder, affecting about 4% of people older than 65 years.[1-5] ET usually manifests with upper limb postural and action tremor without other associated neurologic signs or symptoms.[6-8] In addition to upper limb tremor, ET patients frequently manifest other types of clinically overt tremulous activity, including voice tremor.[2,4,6,7,9] The diagnosis of voice tremor and the evaluation of symptomatic therapies in ET patients are fully dependent on neurologic examination and clinical scales for speech disorders.[10-13] Hence, novel approaches for objective evaluation and treatment assessment of voice tremor in ET patients would provide major clinical advancements in the field.

Voice analysis with spectral analysis allows objective examination of the human voice in healthy subjects (HS) as well as in patients with specific neurologic disorders,

*Correspondence to: Prof. Alfredo Berardelli, Department of Human Neurosciences and Neuromed Institute, Sapienza University of Rome, Viale dell'Università, 30, 00185 Rome, Italy; E-mail: alfredo.berardelli@uniroma1.it

†These authors contributed equally to the article.

Relevant conflicts of interest/financial disclosures: The authors report no financial disclosures or conflicts of interest.

Received: 23 October 2020; **Revised:** 23 December 2020; **Accepted:** 3 January 2021

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.28508

including adductor-type spasmodic dysphonia.[14-17] In ET patients manifesting voice tremor, spectral analysis of voice recordings could be an appropriate methodological tool to assess the main frequency components of voice tremor. Indeed, although the frequency of upper limb tremor in ET typically ranges between 8 and 12 Hz, the frequency of voice tremor in patients manifesting this speech disorder remains unclear.[2,7] No studies have used spectral analysis to examine and compare the main frequency components of voice tremor in a large cohort of ET patients under and not under the symptomatic effect of the best medical treatment.

Despite advances made using voice spectral analysis, more advanced techniques are required to objectively classify the human voice under physiologic and pathologic conditions. Previous studies in patients with neurologic as well as nonneurologic disorders have demonstrated that machine-learning algorithms are more useful than spectral analysis in the objective classification of voice samples owing to their ability to dynamically combine and analyze high-dimensional data sets of voice features.[17-23] Machine-learning analysis of the human voice could provide a new advanced tool to objectively classify voice tremor in ET. No studies have previously used voice analysis based on classification algorithms to objectively recognize voice tremor and rigorously examine pharmacologic treatment response in ET patients manifesting this speech disorder. Last, no studies have correlated the clinical features of ET patients manifesting voice tremor with the results obtained through spectral analysis or machine learning,

In this study, we applied spectral analysis to assess the main frequency components of voice tremor in ET patients. Then, to distinguish between HS, ET patients who manifested clinically overt voice tremor, and patients who did not (ET_{VT+} and ET_{VT-}, respectively), we applied a voice analysis based on support vector machine (SVM) classifier that included a large number of features in addition to the main frequency components of voice. A further aim of the study was to evaluate the effect of pharmacological treatment on voice tremor by comparing voice recordings collected in a subgroup of patients who were and were not under the effect of the best medical treatment (ET_{th+} and ET_{th-}, respectively). For these purposes, we assessed in detail the sensitivity, specificity, positive predictive value, and negative predictive value and accuracy of all diagnostic tests. Furthermore, we calculated the area under the receiver operating characteristic (ROC) curves to verify the optimal diagnostic threshold as reflected by the associated criterion (Ass. Crit.) and Youden Index (YI). To assess possible clinical-instrumental correlations, we also used a feed-forward artificial neural network (ANN) analysis to calculate a continuous numerical value (the likelihood ratio [LR]) providing a measure of voice impairment severity for each patient.

Materials and Methods

Subjects

We recruited 58 patients affected by ET (24 men; 71.7 ± 9.2 years; range, 38–85 years) and 74 age- and sex-matched HS (20 men; 71.0 ± 12.4 years; range, 43–95 years). Participants were recruited from the movement disorders clinic at the Department of Human Neurosciences, Sapienza University of Rome, Italy. Participants gave written informed consent, which was approved by the institutional ethics committee according to the Declaration of Helsinki. All subjects were native Italian speakers and nonsmokers. ET clinical diagnosis was made according to current standardized clinical criteria.[6-8] Patients underwent otolaryngologic and phoniatric evaluation to exclude bilateral/unilateral hearing loss, respiratory disorders, and other nonneurologic disorders affecting the vocal cords.

Symptoms and signs associated with ET were scored using the Fahn Tolosa Marin rating scale for essential tremor (FTM).[10-12] In all participants, we assessed cognitive function and mood using the Mini–Mental State Examination (MMSE)[24] and the Hamilton Depression Rating Scale (HAM-D).[25] Participant demographic and clinical features are summarized in Table 1 and reported in detail in Supplementary Materials 1 and 2. According to the presence of clinically overt voice tremor, we classified ET patients into those with voice tremor (ET_{VT+}, n = 34; 9 men; 72.6 ± 8.4 years; range, 55–85 years) or those without voice tremor (ET_{VT-}, n = 24; 15 men; 70.4 ± 10.3 years; range, 38–84 years). Voice impairment in ET patients was evaluated using the Italian version of the Voice Handicap Index (VHI).[13,26]

The entire ET cohort included 41 patients (20 men; mean age ± SD, 72.3 ± 9.6 years; range, 38–85 years) chronically treated with pharmacological compounds to improve tremor (eg, beta-blockers [BBs] or benzodiazepines [BZDs]), whereas 17 patients (4 men; 70.3 ± 8.5 years; range, 55–83 years) were not under drug treatment at the time of the study (Supplementary Material 2). In chronically treated ET patients, voice samples were recorded at baseline (ET_{th+}) and after an appropriate pharmacological withdrawal of at least 1 week (ET_{th-}). To examine the effect of chronic pharmacological treatment on the voice,[27] we compared voice recordings collected in 25 of 41 ET patients (13 patients with ET_{VT+} and 12 patients with ET_{VT-}) while under (ET_{th+}) or not under (ET_{th-}) the effect of the best medical treatment (Supplementary Material 2).

Voice Recordings

Voice recordings were performed by asking participants to produce a specific speech task with their usual

TABLE 1. Demographic and clinical features of ET patients and HS

	Age (years)	Weight (kg)	Height (cm)	DD (years)	MMSE	HAM-D	FTM	FTM-v	VHI	LRth−	LRth+
ET	71.7 ± 9.2	71.1 ± 13.6	166.1 ± 7.9	11.3 ± 10.9	29.6 ± 0.9	3.7 ± 2.0	22.8 ± 15.2	1.7 ± 1.5	21.3 ± 26.7	0.21 ± 0.18	0.47 ± 0.25
HS	71.0 ± 12.4	69.5 ± 12.0	163.7 ± 9.5	—	29.9 ± 0.4	3.2 ± 1.4	—	—	—	—	—

DD, disease duration; MMSE, Mini–Mental State Examination; HAM-D, Hamilton Depression Rating Scale; FTM, Fahn Tolosa Marin rating scale for essential tremor; FTM-v, Fahn Tolosa Marin rating scale for essential tremor, voice impairment subitem; VHI, Voice Handicap Index; LR, likelihood ratio.
LR scores were calculated from the sustained emission of a vowel in patients with ET not under (th−) and under (th+) the best medical treatment. Results are expressed as average ± standard deviation.

voice intensity, pitch, and quality. The speech task consisted of a sustained emission of a close mid-front unrounded vowel /e/ for at least 5 seconds. We selected this specific speech task because voice tremor is usually best detected during sustained vocalization.[28] Voice recordings were collected by using a high-definition audio recorder H4n Zoom (Zoom Corporation, Tokyo, Japan), connected with a Shure WH20 Dynamic Headset Microphone (Shure Incorporated, Niles, IL), which was placed at a distance of 5 cm from the mouth. Voice samples were recorded in linear PCM format (.wav) at a sampling rate of 44.1 kHz, with 24-bit sample size. The signal was subsequently converted to 44.1-kHz 16-bit linear PCM format. Last, we applied a segmentation procedure through Audacity, dedicated software for audio editing, to acquire only the sustained emission of the vowel /e/ from all voice recordings.

Spectral Analysis

To investigate the main frequency components of voice tremor in ET, we first calculated and compared the spectral analysis of voice samples recorded in HS and in patients with ET_{VT+} under and not under the best medical treatment and finally in ET_{VT-} by using *Praat* software, a dedicated computer program for analyzing, synthesizing, and manipulating speech audio recordings.[29] More in detail, for each participant, we calculated the frequency tremor frequency as a measure of the dominant frequency band and the frequency tremor intensity index as a measure of the intensity, according to standardized procedures.[29-31]

Machine-Learning Analysis

We performed voice analysis by using dedicated machine-learning algorithms.[32-35] Each voice sample underwent feature extraction preprocess using OpenSMILE (audEERING GmbH, Germany), dedicated software.[36] For each voice sample, we extracted 6139 voice features using the configuration file that was prescribed for the INTERSPEECH2016 Computational Paralinguistics Challenge (IS ComParE 2016) feature data set.[34]

To identify a subset of the most relevant features for the objective analysis of voice tremor,[37] the extracted voice features underwent feature selection preprocess using the Correlation Features Selection.[38] Selected features were ranked for class relevance, using gain ratio concerning the class through Gain Ratio Attribute Evaluation (GRAE).[39,40]

To further improve the accuracy of results, we applied Fayyad and Irani's multi-interval discretization method, which is a binary recursive method based on the information entropy minimization and adopts a decision criterion based on the minimum description length principle.[41] This discretization method allows

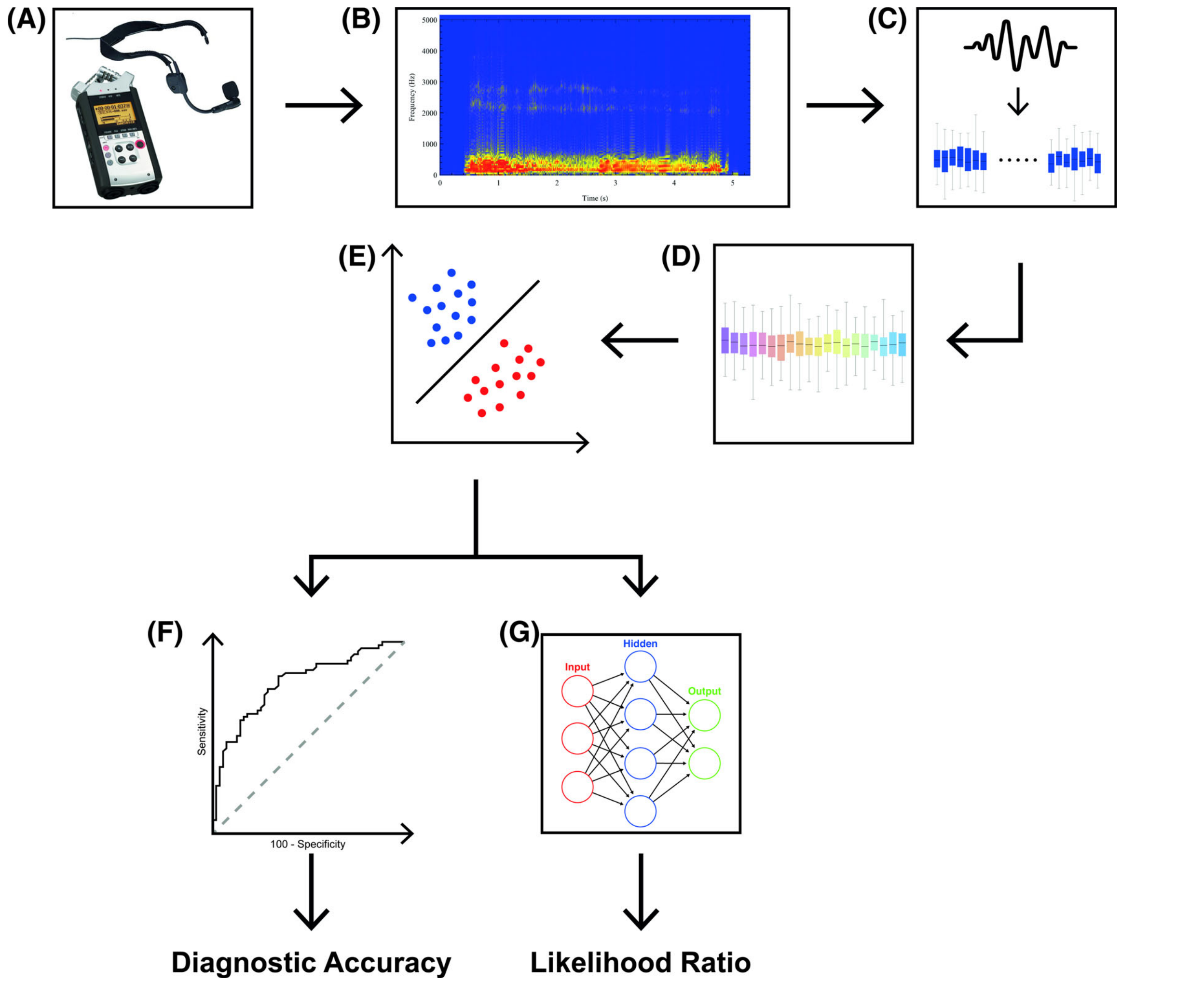

FIG 1. .Experimental paradigm. (**A**) Recording of voice samples by a high-definition audio recorder; (**B**) narrow-band spectrogram of the acoustic voice signal; (**C**) feature extraction; (**D**) feature selection; (**E**) feature classification; (**F**) ROC curve analysis; (**G**) LR values calculated by means of ANN. [Color figure can be viewed at wileyonlinelibrary.com]

improving performances of SVM classifier, also reducing the classification's time.[42-45] Moreover, this method, in combination with classifiers, has already been successfully applied in speech analysis.[45]

Then we trained a linear kernel SVM classifier through the first 20 most relevant discretized features, ranked GRAE.[46,47] A complete list of the first 20 features — functionals applied to low-level descriptors — during the sustained emission of a vowel for the 6 comparisons are reported in Supplementary Material 3. Specifically, the SVM was trained using the sequential minimal optimization method, which is considered a fast and efficient machine-learning algorithm to implement an SVM classifier.[48] Both the feature selection and the classification were performed by Weka (Waikato Environment for Knowledge Analysis, University of Waikato, New Zealand), software containing a collection of algorithms for data analysis and predictive modeling.[38,49]

Given that SVM represents a classifier providing binary output to verify possible clinical-instrumental correlations, we used a feed-forward ANN to calculate a continuous numerical value (LR) reflecting the degree of voice impairment in each patient. ANN consisted of a 20-neurons input layer, a 10-neurons hidden layer, and a 1-neuron output layer. ANN was trained using the same selected features used to train the SVM. Also, the ANN output was normalized between 0 and 1 to obtain continuous numerical values (LRs) allowing objective measures of the severity of voice impairment in each patient. LR values are shown as averages in Table 1 as well as individual values in Supplementary Materials 2. The experimental procedures are summarized in Figure 1.

Statistical Analysis

The normality of the demographic and anthropometric parameters of HS and ET patients (age, sex, height, and weight) was assessed using the Kolmogorov–Smirnov test. The Mann–Whitney U test was used to

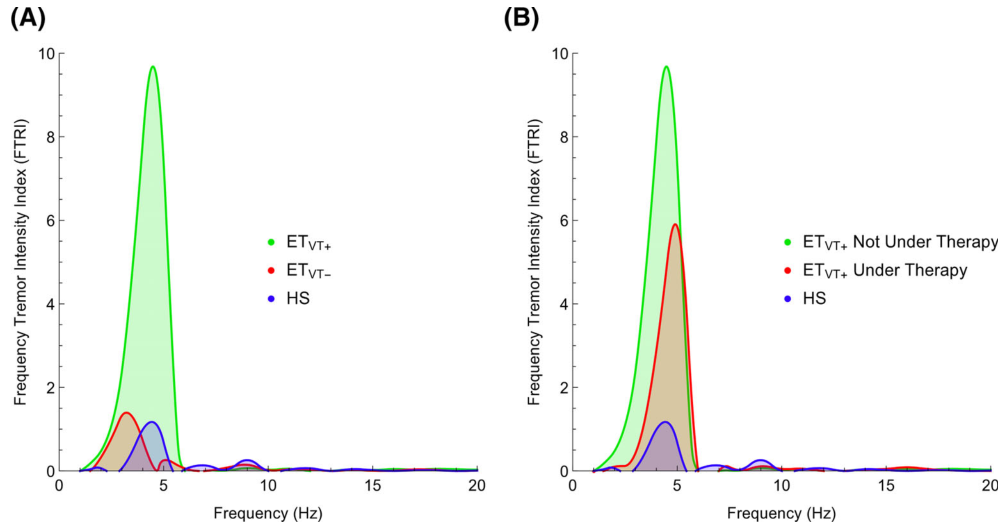

FIG. 2. Spectral analysis of voice samples. (**A**) frequency tremor intensity index (FTRI) in representative HS and ET_{VT+} and ET_{VT-} patients. Note activity peaks in the 2- to 6-Hz frequency band in the 3 groups and the increased FTRI in the ET_{VT+} patient compared with the HS and ET_{VT-} patient. (**B**) FTRI in a representative HS and an ET_{VT+} patient not under and under the best medical therapy. Note the lower FTRI values in the ET_{VT+} patient under therapy than in the ET_{VT+} patient not under therapy. [Color figure can be viewed at wileyonlinelibrary.com]

compare demographic and anthropometric parameters in HS and ET patients, as well as clinical scores (eg, FTM, FTM voice impairment subitem [FTM-v], VHI, MMSE, and HAM-D scores) in the various ET subgroups (eg, ET_{VT+} and ET_{VT-}, ET_{th+} and ET_{th-}). The chi-square test was used to compare the frequency of familial cases in ET_{VT+} and ET_{VT-} patients. The Wilcoxon signed rank test was used to compare FTM, FTM-v, and VHI scores in ET_{th+} and ET_{th-} patients. The unpaired Student *t* test was used to compare possible peaks in the spectral analysis of voice samples recorded in ET_{VT+} and ET_{VT-} patients, whereas the paired Student *t* test was used to compare the same measures in ET_{VT+} patients under and not under therapy.

ROC analyses were performed to identify the optimal diagnostic cutoff values to discriminate between HS and ET_{VT+}, ET_{VT-}, ET_{th-}, and ET_{th+} patients according to standardized procedures.[17,19]

Spearman's rank correlation coefficient was used to assess correlations between clinical scores (FTM, FTM-v, VHI, MMSE, and HAM-D scores) and output measures of the spectral and neural network analyses (LR values).

A $P < 0.05$ was considered statistically significant.

Results

Demographic and anthropometric parameters were normally distributed in HS and ET patients ($P > 0.05$) and were comparable in the 2 groups ($P > 0.05$). In our cohort, 58% of ET patients manifested clinically overt voice tremor. The chi-square test showed a higher number of familial cases in the ET_{VT-} group (62.5%) than in the ET_{VT+} group (29.4%); $P < 0.05$. Clinical scores (MMSE and HAM-D) were also comparable in the various ET subgroups ($P > 0.05$ for all comparisons). ET_{VT+} patients showed higher scores on the FTM, FTM-v, and VHI scales than ET_{VT-} patients (Table 1, Supplementary Materials 1 and 2).

Spectral Analysis

Spectral analysis demonstrated that ET_{VT+} and ET_{VT-} patients manifested a clear oscillatory activity peak at 2–6 Hz that mostly overlapped with that observed in HS. However, oscillatory activity at 2–6 Hz was higher in ET_{VT+} patients than in HS ($t = 6.49$, $P < 0.01$) and ET_{VT-} patients ($t = 4.74$, $P < 0.01$), whereas it was comparable in HS and ET_{VT-} patients ($t = 0.17$, $P = 0.45$); see Figure 2A. Furthermore, spectral analysis showed that the best medical treatment decreased the power of oscillatory activity at 2–6 Hz in ET_{VT+} patients ($t = -2.15$, $P < 0.05$). However, pharmacological treatment was not able to restore oscillatory activity at 2–6 Hz ($t = 1.75$, $P < 0.05$; Fig. 2B).

Voice Analysis With SVM

We first compared HS and ET_{VT+} patients and achieved a significant diagnostic performance of our artificial classifier. ROC analyses calculated using SVM identified an optimal diagnostic threshold value of 0.88 (Ass. Crit.) when applying discretization and 10-fold cross-validation procedures, with a YI of 0.94 (Fig. 3A, Table 2).

Given the high statistical significance of the above analysis, we assessed whether our artificial classifier based on the SVM algorithm was also able to

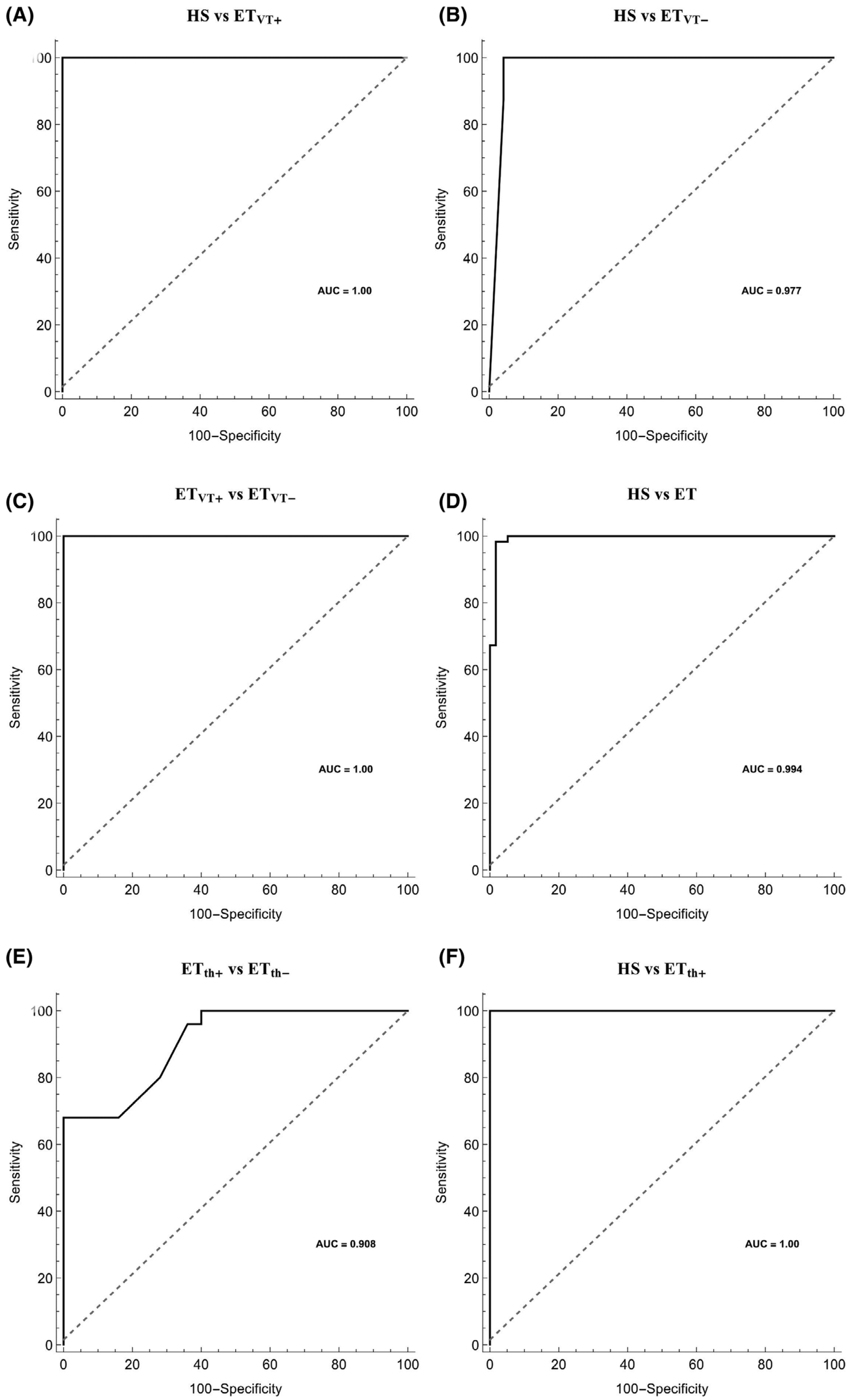

FIG. 3. Support vector machine analysis of voice samples. Receiver operating characteristic curves calculated with a support vector machine to differentiate HS and ET_{VT+} patients (**A**); HS and ET_{VT-} patients (**B**); ET_{VT+} and ET_{VT-} patients (**C**); HS and ET patients (**D**); ET_{th+} and ET_{th-} patients (**E**); and HS and ET_{th+} patients (**F**).

TABLE 2. Performance of the support vector machine classifier

Comparisons	Selected features	Instances	Optimization	Cross-validation	Associated criterion	Youden Index	Se (%)	Sp (%)	PPV (%)	NPV (%)	Acc (%)	AUC
HS vs ET_{VT+}	20	68	Discretize	10-fold	0.88	0.94	94.4	100	100	94.1	97.1	1.000
HS vs ET_{VT-}	20	48	Discretize	10-fold	0.87	0.96	96.0	100	100	95.8	97.9	0.977
ET_{VT+} vs ET_{VT-}	20	58	Discretize	10-fold	0.75	0.97	100	97.1	95.8	100	98.3	1.000
HS vs ET	20	116	Discretize	10 fold	0.99	0.97	98.3	98.3	98.3	98.3	98.3	0.994
ET_{th+} vs ET_{th-}	20	50	Discretize	10 fold	0.53	0.64	89.5	74.2	68.0	92.0	80.0	0.908
HS vs ET_{th+}	20	50	Discretize	10-fold	0.52	0.96	96.2	100	100	96.0	98.0	1.000

HS, healthy subjects; ET: essential tremor; ET_{VT+}, ET patients with clinically overt voice tremor; ET_{VT-}, ET patients without clinically overt voice tremor; ET_{th+}, ET patients under therapy; ET_{th-}, ET patients not under therapy; Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; Acc, accuracy; AUC, area under the curve.
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; optimization and cross-validation refer to standardized machine learning algorithm procedures (see the text for details).
Support vector machine performance elaborating 20 selected features during the sustained emission of a vowel for the 6 conditions: (1) HS vs ET_{VT+} patients; (2) HS vs ET_{VT-} patients; (3) ET_{VT+} vs ET_{VT-} patients; (4) HS vs ET patients; (5) ET_{th+} vs ET_{th-} patients; (6) HS vs ET_{th+} patients.

objectively identify subclinical voice impairment in ET_{VT-} patients. For this purpose, we compared voice recordings collected in HS and ET_{VT-} patients and again obtained a highly significant diagnostic performance (Fig. 3B, Table 2).

To objectively discriminate patients with ET based on clinically overt voice tremor, we applied artificial classifiers comparing ET_{VT+} and ET_{VT-} patients. We achieved highly significant diagnostic performance using the SVM classifier (Fig. 3C, Table 2).

Finally, we used an artificial classifier based on the SVM algorithm in the entire patient group to objectively classify ET and again achieved a highly significant diagnostic performance (Fig. 3D, Table 2).

Effect of Pharmacological Treatment

We found that the best medical treatment induced a significant clinical improvement in overall tremor in ET, as demonstrated by reduced FTM scores (ET_{th+}, 21.9 ± 12.6; ET_{th-}, 31.2 ± 19.6; $z = -4.3$; $W = 0$; $P < 0.05$). Similarly, pharmacological treatment also improved voice tremor in ET as shown by reduced scores on the FTM-v (ET_{th+}, 1.8 ± 0.7; ET_{th-}, 2.8 ± 0.8; $z = -3.2$; $W = 0$; $P < 0.05$) and VHI (ET_{th+}, 26.7 ± 29.0; ET_{th-}, 38.2 ± 35.1; $z = -2.4$; $W = 5.5$; $P < 0.05$) scales.

To assess whether our classifier based on SVM could identify voice tremor improvement in ET, we applied our voice analysis in ET_{th+} and ET_{th-} patients. Our classifier and ROC analyses identified an optimal diagnostic threshold of 0.53 (Ass. Crit.; YI = 0.64; Fig. 3E, Table 2).

To verify whether the best medical treatment could improve voice in ET patients, our algorithm was tested in classifying HS versus ET_{th+} patients. SVM achieved a highly significant diagnostic performance, and ROC analyses identified a diagnostic threshold value of 0.52 (Ass. Crit.; YI = 0.96; Fig. 3F, Table 2).

Correlation Analysis

A positive correlation was found between FTM-v and VHI scores in the entire ET group ($r = 0.76$, $P = 0.001$). Regarding spectral analysis, we found a positive correlation between power at 2–6 Hz and FTM-v score in ET_{VT+} patients who were not under the best medical treatment ($r = 0.35$, $P < 0.05$). Furthermore, we also found a positive correlation between power at 2–6 Hz and FTM-v score in ET_{VT+} patients under the best medical treatment ($r = 0.59$, $P < 0.05$). Voice analysis with ANN classifier comparing HS and ET_{VT+} patients allowed us to demonstrate a negative correlation of LR with FTM-v ($r = -0.37$, $P = 0.03$) and with VHI ($r = -0.53$, $P = 0.001$). When considering the correlation between HS and the entire ET group, we also found a negative correlation of LR with FTM-v ($r = -0.42$, $P = 0.001$) and with VHI ($r = -0.39$, $P = 0.003$). Finally, when considering the correlation between HS and ET_{th+} patients, we found a negative correlation of LR with FTM-v ($r = -0.44$, $P = 0.03$) and with VHI ($r = -0.39$, $P = 0.05$).

Discussion

In this study, we applied voice analysis based on spectral analysis and SVM classifier in ET patients with and without clinically overt voice tremor. Spectral analysis showed a prominent oscillatory activity peak at 2–6 Hz in ET_{VT+} patients. Voice analysis with machine learning disclosed highly significant results in discriminating HS from ET_{VT+} patients. We also discriminated between HS and ET_{VT-} patients, as well as between HS and the entire ET patient group. Last, our algorithm distinguished between ET_{th+} and ET_{th-} patients. Our study, therefore, provides the first evidence of objective voice tremor recognition in ET through voice analysis.

In this study, all voice samples were recorded in a dedicated sound-attenuated room using a high-definition audio recorder. Participants were all nonsmokers and native Italian speakers. Furthermore, because ET patients may complain of progressive voice worsening during the day, all voice samples were recorded in the morning. Patients and HS had similar demographic and anthropometric characteristics. We used the narrow-band spectrogram to control for and exclude patients with isolated voice tremor or those showing evidence of adductor/abductor spasmodic dysphonia.[50] Concerning the speech task, participants were asked to produce the sustained emission of a vowel according to standardized procedures.[16,17] As a speech task, we selected the sustained emission of a vowel because voice tremor is usually best detected during nonlinguistic tasks, including sustained phonation.[28]

Spectral Analysis

In HS, spectral analysis demonstrated a clear peak at the 2- to 6-Hz frequency. Physiological oscillations at 2–6 Hz recorded in HS were consistent with the recent report by Brückl et al,[30] possibly reflecting the integrated activity of diaphragmatic, laryngeal, and vocal fold vibration combined with resonant structures, including articulatory jaw and tongue movements.

In ET_{VT+} patients, we also found an oscillatory activity at 2–6 Hz, in agreement with a recent observation from Hlavincka et al.[15] Moreover, the spectral analysis also demonstrated increased power of the oscillatory activity observed at 2–6 Hz in ET_{VT+} patients compared with HS. We suggest that in ET_{VT+} patients, voice tremor would reflect the abnormal activation of a physiologic neuronal oscillator.

When comparing voice samples recorded in ET_{th+} and ET_{th-} patients, we found that pharmacological treatment improved voice tremor in ET patients, as demonstrated by lower scores on the FTM, FTM-v, and VHI scales, consistent with current clinical consensus.[27] In ET_{VT+} patients, pharmacological treatment decreased but did not restore the power of oscillatory activity at 2–6 Hz. This finding points to the pharmacological effect of BBs and BZDs on the abnormal 2- to 6-Hz oscillations responsible for voice tremor in ET_{VT+} patients.

Our correlation analysis disclosed a positive correlation between power at 2–6 Hz and the FTM-v scale in ET_{VT+} patients both under and not under the best medical treatment. Hence, the higher was the power at 2–6 Hz, the greater was the severity of clinically overt voice tremor in ET_{VT+} patients. Furthermore, in ET_{VT+} patients under the best medical treatment, the lower was the power, the higher was the drug-induced symptomatic improvement of the voice. Our spectral analysis represents the first objective measurement of the main frequency component of voice tremor in ET_{VT+} patients both under and not under the best medical treatment.

Voice Analysis With SVM

In this study, we demonstrated for the first time that voice analysis based on SVM classifier may differentiate ET_{VT+} from HS, as shown by previously unreported highly significant outcome values. This finding reflects the ability of our voice analysis to objectively measure voice tremor in ET_{VT+} patients. The high statistical significance of our ROC curve analysis regarding the comparison between HS and ET_{VT+} patients suggests that clinically overt voice tremor in ET can be objectively recognized by techniques of artificial classifying. Furthermore, the high accuracy of the algorithm performance also points to a novel tool to support clinicians in the objective detection of voice tremor in ET.

Another relevant finding of the present study was the ROC curve analysis showing high accuracy in differentiating HS and ET_{VT-} patients. This finding was rather unexpected because ET_{VT-} patients do not manifest clinically overt voice tremor. The most likely hypothesis is that ET_{VT-} patients manifest a subclinical voice tremor that can be detected only by high-definition audio recording. It is also theoretically plausible that subclinical voice tremor in ET_{VT-} patients could reflect propagation to voice resonance organs of head tremor or even upper limb rest tremor.

The automatic analysis of voice tremor allowed us to discriminate the entire ET group and HS with high accuracy. ET diagnosis is currently based on standardized clinical criteria that require the presence of upper limb postural and action tremor that persists for at least 3 years without any associated neurologic signs.[6-8] Despite several relevant advances in the understanding of the clinical and pathophysiological bases, the diagnosis of ET is still based on qualitative clinical examination with the aid of several standardized clinical scales.[10-12] Overall, our findings suggest that voice analysis based on machine learning could be a helpful tool to assist clinicians in the diagnosis and follow-up of patients with ET and voice tremor.

Voice analysis with SVM classifier allowed us to objectively demonstrate for the first time the symptomatic effect of pharmacological treatment in ET, as shown by the high accuracy in discriminating between ET_{th+} and ET_{th-} patients. SVM classifier showed high accuracy in discriminating between HS and ET_{th+} patients, thus allowing us to conclude that current pharmacologic strategies do not restore voice emission in ET.

We found a negative correlation between VHI, FTM, and FTM-v scale scores and LR parameters calculated using our ANN approach, demonstrating that patients with higher voice tremor severity had lower LR values. Hence, LR values were reliable parameters to express the dynamic complexity of voice feature changes observed in voice tremor in ET. Therefore, we suggest that LR values may be considered an objective measure of voice tremor severity and may aid clinicians in the objective and quantitative evaluation of ET patients, including their response to pharmacological strategies.

The present study had several limitations. We did not record serial vocal samples in ET, and thus we cannot exclude possible variability because of daily vocal feature fluctuations. An additional source of variability may be the heterogeneous pharmacological compounds administered to ET patients. It should be considered that, in patients with ET, in addition to vocal cord oscillations, voice tremor might also reflect jaw and head tremor. Future studies will evaluate whether voice tremor in ET differs from other conditions including palatal tremor and will also clarify whether voice and upper limb tremor in ET reflects different pathophysiological mechanisms. Last, when considering our results, it should be taken into account that our ET patient cohort included a higher percentage of patients manifesting clinically overt voice tremor (58% of all cases) than that expected based on currently available epidemiological information (12% of all cases).[9] This finding reflects the selection criteria used according to the primary aim of the study.

Conclusion

Voice analysis based on spectral analysis allowed us to detect the mean frequency components of voice tremor in ET patients under and not under the best medical treatment. Furthermore, voice analysis through SVM classifier with high accuracy objectively classified HS and ET patients as under or not under the effect of the best medical treatment. Thus, we propose voice analysis as a novel methodological tool to support clinicians in the evaluation of voice tremor and the assessment of pharmacological response in ET patients. Future studies are needed to confirm our observations and also clarify whether voice analysis with spectral analysis and machine-learning algorithms may support clinicians in the objective differential diagnosis of voice tremor in patients with ET and dysphonia.[16,17]

References

1. Clark LN, Louis ED. Essential tremor. Handb Clin Neurol 2018; 147:229–239.
2. Espay AJ, Lang AE, Erro R, et al. Essential pitfalls in "essential" tremor. Mov Disord 2017;32:325–331.
3. Jankovic J. Essential tremor: a heterogenous disorder. Mov Disord 2002;17:638–644.
4. Louis ED. The evolving definition of essential tremor: what are we dealing with? Parkinsonism Relat Disord 2018;46:87–91.
5. Louis ED, Ferreira JJ. How common is the most common adult movement disorder? Update on the worldwide prevalence of essential tremor. Mov Disord 2010;25:534–541.
6. Bhatia KP, Bain P, Bajaj N, et al. Consensus statement on the classification of tremors. From the task force on tremor of the International Parkinson and Movement Disorder Society: IPMDS task force on tremor consensus statement. Mov Disord 2018;33:75–87.
7. Deuschl G, Bain P, Brin M. Consensus statement of the Movement Disorder Society on tremor. Mov Disord 2008;13:2–23.
8. Elble RJ. Diagnostic criteria for essential tremor and differential diagnosis. Neurology 2000;54:S2–S6.
9. Chen W, Hopfner F, Szymczak S, et al. Topography of essential tremor. Parkinsonism Relat Disord 2017;40:58–63.
10. Elble R, Bain P, Forjaz MJ, Haubenberger D, Testa C, Goetz CG, et al. Task force report: scales for screening and evaluating tremor: critique and recommendations. Mov Disord 2013;28:1793–1800.
11. Louis ED, Gerbin M. Voice handicap in essential tremor: a comparison with normal controls and Parkinson's disease. Tremor Other Hyperkinet Mov 2013;3.1–6.
12. Ondo W, Hashem V, LeWitt PA, et al. Comparison of the Fahn-Tolosa-Marin clinical rating scale and the essential tremor rating assessment scale. Mov Disord Clin Pract 2018;5:60–65.
13. Schindler A, Ottaviani F, Mozzanica F, et al. Cross-cultural adaptation and validation of the voice handicap index into Italian. J Voice 2010;24:708–714.
14. Heman-Ackah YD, Sataloff RT, Laureyns G, et al. Quantifying the cepstral peak prominence, a measure of dysphonia. J Voice 2014;28: 783–788.
15. Hlavnička J, Tykalová T, Ulmanová O, et al. Characterizing vocal tremor in progressive neurological diseases via automated acoustic analyses. Clin Neurophysiol 2020;131:1155–1165.
16. Suppa A, Asci F, Saggio G, et al. Voice analysis in adductor spasmodic dysphonia: objective diagnosis and response to botulinum toxin. Parkinsonism Relat Disord 2020;73:23–30.
17. Suppa A, Marsili L, Giovannelli F, et al. Abnormal motor cortex excitability during linguistic tasks in adductor-type spasmodic dysphonia. Eur J Neurosci 2015;42:2051–2060.
18. Hegde S, Shetty S, Rai S, Dodderi T. A survey on machine learning approaches for automatic detection of voice disorders. J Voice 2019; 33:947.
19. Asci F, Costantini G, Di Leo P, et al. Machine-learning analysis of voice samples recorded through smartphones: the combined effect of ageing and gender. Sensors 2020;20:5022.
20. Deo RC. Machine learning in medicine. Circulation 2015;132: 1920–1930.
21. Van Calster B, Wynants L. Machine learning in medicine. N Engl J Med 2019;380:2588.
22. Vu M-AT, Adalı T, Ba D, et al. A shared vision for machine learning in neuroscience. J Neurosci 2018;38:1601–1607.
23. Zaharchuk G, Gong E, Wintermark M, Rubin D, Langlotz CP. Deep learning in neuroradiology. AJNR Am J Neuroradiol 2018;39: 1776–1784.
24. 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;12:189–1198.
25. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967;6:278–296.
26. Jacobson BH, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger MS, et al. The voice handicap index (VHI): development and validation. Am J Speech Lang Pathol 1997;6:66–70.
27. Ferreira JJ, Mestre TA, Lyons KE, et al. MDS evidence-based review of treatments for essential tremor. Mov Disord 2019;34:950–958.
28. Tripoliti E. Voice tremor and acoustic analysis: finding harmony through the waves. Clin Neurophysiol 2020;131:1144–1145.
29. Boersma P. Praat: doing phonetics by computer [Computer program]. http://www.praat.org/. 2011.
30. Brückl M, Ghio A, Viallet F. Measurement of tremor in the voices of speakers with Parkinson's disease. Procedia Computer Science 2018;128:47–54.
31. Shao J, MacCallum JK, Zhang Y, Sprecher A, Jiang JJ. Acoustic analysis of the tremulous voice: assessing the utility of the correlation dimension and perturbation parameters. J Commun Disord 2010;43:35–44.
32. Alpaydin E. Introduction to Machine Learning. 2nd ed. Cambridge, MA: MIT Press; 2010:537.

33. Russell SJ, Norvig P, Davis E. Artificial Intelligence: A Modern Approach. 3rd ed. Upper Saddle River, NJ: Prentice Hall; 2010:1132.

34. Schuller B, Steidl S, Batliner A, et al. The INTERSPEECH 2016 Computational Paralinguistics Challenge: Deception, Sincerity and Native Language. 2016: 2001–2005. http://www.isca-speech.org/archive/Interspeech_2016/abstracts/0129.html. Accessed March 16, 2020.

35. Specht D. A general regression neural network. IEEE Trans Neural Netw 1991;2:568–578.

36. 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 [Internet]. Firenze, Italy: ACM Press; 2010:1459. http://dl.acm.org/citation.cfm?doid=1873951.1874246. Accessed March 16, 2020.

37. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD Explorat Newslett 2009;11:10.

38. Hall M. Correlation-based feature selection for machine learning. Dep Comput Sci 2000;19:1–198.

39. Karegowda AG, Manjunath A, Jayaram M. Comparative study of attribute selection using gain ratio and correlation based feature selection. Int J Inform Technol Knowled Manage 2010;2:271–277.

40. Quinlan JR. Induction of decision trees. Mach Learn 1986;1: 81–106.

41. Fayyad UM, Irani KB. On the handling of continuous-valued attributes in decision tree generation. Mach Learn 1992;8:87–102.

42. Gupta A, Mehrotra KG, Mohan C. A clustering-based discretization for supervised learning. Stat Probab Lett 2010;80:816–824.

43. Lavangnananda K, Chattanachot S. Study of discretization methods in classification. In conference: Proceedings of the 9th International Conference on Knowledge and Smart Technology (KST), Chonburi, Thailand, February 1-4 2017;50–55.

44. Kumar A, Zhang D. Biometric recognition using entropy-based discretization. IEEE 2007;2:125.

45. Hassan A, Damper RI. Multi-class and hierarchical SVMs for emotion recognition; In conference: INTERSPEECH 2010, Japan; 2010: 2354-2357.

46. Costantini G, Casali D, Todisco M. An SVM Based Classification Method for EEG Signals; In conference: Proceedings of the 14th WSEAS international conference on Circuits, Corfu Island, Greece, July 22–24 2010; Volume 2224.

47. Costantini G, Todisco M, Perfetti R, et al. Svm Based Transcription System with Short-Term Memory Oriented to Polyphonic Piano Music; In conference: MELECON 2010–2010 15th IEEE Mediterranean Electrotechnical Conference, La Valletta, Malta, April 26–28 2010:196–201.

48. Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv Large Margin Classif 1999;10:61–74.

49. Frank E, Hall M, Holmes G, et al. Weka-a machine learning workbench for data mining. Data Mining and Knowledge Discovery Handbook. Berlin: Springer; 2009:1269–1277.

50. Brin MF, Blitzer A, Stewart C. Laryngeal dystonia (spasmodic dysphonia): observations of 901 patients and treatment with botulinum toxin. Adv Neurol 1998;78:237–252.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

SGML and CITI Use Only
DO NOT PRINT

Authors' Roles

1. Research project: A. conception — A.S., F.A.; B.organization — A.S., F.A., P.D.L., G.S.; C. execution — A.S., F.A., P.D.L., G.S., G.F., Z.Z., G.R.; 2. Statistical analysis: A. design — F.A., P.D.L., G.C.; B. execution — F.A., P.D.L., Z.Z., G.C.; C. review and critique — A.S., G.S.; 3. Manuscript preparation: A. writing of the first draft — A.S., F.A., P.D.L.; B. review and critique — G.S., A.B., G.C.

Financial Disclosures of All Authors

The authors have received no funding from any institution, including personal relationships, interests, grants, employment, affiliations, patents, inventions, honoraria, consultancies, royalties, stock options/ownership, or expert testimony for the last 12 months.