Applied Computer Science, vol. 19, no. 4, pp. 14–25
doi: 10.35784/acs-2023-32

Submitted: 2023-09-05 | Revised: 2023-10-19 | Accepted: 2023-11-05

Keywords: Dysphonia, Machine learning, Multiclass classification, Voice signal

Miguel Angel BELLO-RIVERA [0009-0003-6641-3094]*,
Carlos Alberto REYES-GARCÍA [0000-0003-4773-9585]**,
Tania Cristal TALAVERA-ROJAS [0000-0001-7656-3115]***,
Perfecto Malaquías QUINTERO-FLORES [0000-0001-7651-4364]****,
Rodolfo Eleazar PÉREZ-LOAIZA [0000-0002-6500-258X]*****

AUTOMATIC IDENTIFICATION OF DYSPHONIAS USING MACHINE LEARNING ALGORITHMS

Abstract

Dysphonia is a prevalent symptom of some respiratory diseases that affects voice quality, even for prolonged periods. For its diagnosis, speech-language pathologists make use of different acoustic parameters to perform objective evaluations on patients and determine the type of dysphonia that affects them, such as hyperfunctional and hypofunctional dysphonia, which is important because each type requires a different treatment. In the field of artificial intelligence this problem has been addressed through the use of acoustic parameters that are used as input data to train machine learning and deep learning models. However, its purpose is usually to identify whether a patient is ill or not, making binary classifications between healthy voices and voices with dysphonia, but not between dysphonias. In this paper, harmonic-to-noise ratio, cepstral peak prominence-smoothed, zero crossing rate and the means of the Mel frequency cepstral coefficients (2-19) are used to make multiclass classification of voices with euphony, hyperfunction and hypofunction by means of six machine learning algorithms, which are: Random Forest, K nearest neighbors, Logistic regression, Decision trees, Support vector machines and Naive Bayes. In order to evaluate which of them presents a better performance to identify the three voice classes, bootstrap.632 was used. It is concluded that the best confidence interval ranges from 87% to 92%, in terms of accuracy for the K Nearest Neighbors model. Results can be implemented in the development of a complementary application for the clinical diagnosis or monitoring of a patient under the supervision of a specialist.

* Tecnológico Nacional de México, Campus Apizaco, Departamento de Sistemas Computacionales, México, podriaservirte@gmail.com
** Instituto Nacional de Astrofísica, Óptica y Electrónica, Departamento de Ciencias y Tecnologías Biomédicas, México, kargaxxi@inaoep.mx
*** Universidad Autónoma de Asunción, Facultad de Ciencias de la Salud, Departamento de Neuropsicología, Paraguay, ttalavera@uaa.edu.py
**** Tecnológico Nacional de México, Campus Apizaco, Departamento de Sistemas Computacionales, México, perfecto.qf@apizaco.tecnm.mx
***** Tecnológico Nacional de México, Campus Apizaco, Departamento de Sistemas Computacionales, México, rodolfo.pl@apizaco.tecnm.mx

1. INTRODUCTION

Dysphonia is a voice disorder characterized by the abnormal loss of typical voice quality due to functional or organic disturbances in the larynx. It can be classified into three primary categories: Functional, Organic, and Mixed (Behlau & Pontes, 1989).

Functional dysphonia, often associated with emotional stress, poor vocal habits, or excessive voice use in daily activities, is typically accompanied by symptoms such as intermittent hoarseness, vocal projection difficulties, and occasional throat discomfort or tension due to vocal fatigue.

Conversely, organic dysphonia results from physical laryngeal disorders, including injuries, infections, tumors, or other medical conditions that directly impact the laryngeal structure. It tends to present as persistent hoarseness, significant alterations in vocal quality, and occasional throat pain or discomfort.

Mixed dysphonias represent a combination of both organic and functional factors, resulting in diverse symptoms, including variable hoarseness, shifts in vocal quality, discomfort, occasional vocal fatigue, and more. Recognizing these distinctions is imperative in clinical practice, as they necessitate distinct therapeutic approaches.

Another indicator frequently utilized to differentiate vocal disorders pertains to the biomechanics of phonation, which distinguishes between hyperfunction and hypofunction. Hyperfunctional dysphonia arises from excessive engagement of the muscles involved in glottic closure, affecting the ventricular bands, resulting in a tense, rough, high-frequency voice with increased intensity (López, 1997). Conversely, hypofunctional dysphonia is characterized by laryngeal muscle weakness, incomplete glottis closure, and a breathy, low-intensity voice (López, 2000). The former is typically associated with functional dysphonias, while the latter is linked to organic dysphonias (Farias, 2016). The recent COVID-19 pandemic has led to both hyperfunctional and hypofunctional voice alterations.

The causes of dysphonia are multifaceted and encompass factors such as viral invasion of the glottic epithelium, as observed in infections, which can result in damage. Research by Hoffmann et al. (2020) suggests that the ACE2 (Angiotensin-Converting Enzyme 2) protein, used as a receptor by certain viruses like SARS-CoV-2, plays a pivotal role in these vocal fold alterations. Moreover, Descamps et al. (2020) has demonstrated the expression of ACE2 in the epithelial cells of the vocal folds.

During the COVID-19 pandemic, dysphonia has been reported as a prevalent symptom, especially in patients with laryngeal inflammation, even in post-infection stages. Additionally, it can also arise in patients who have undergone tracheal intubation (Verdaguer et al., 2008), where structural changes in the larynx can impact vocal function. Prolonged intubation can lead to mild to severe damage.

Dysphonia requires therapy from an otolaryngologist or a speech therapist, who need objective tools to provide solutions (Behlau et al., 2005), which can be developed in the field of artificial intelligence, where machine learning algorithms are implemented to classify data for a specific purpose, and are frequently used in medicine to develop diagnostic or rehabilitation systems.

In clinical practice, it is essential to distinguish between hypofunctional and hyperfunctional dysphonia, as they require different rehabilitation schemes. Additionally, a patient may switch from one to another due to compensatory phonation mechanisms.

In this paper, the harmonic-to-noise ratio (HNR), cepstral peak prominence smoothed (CPPS), zero-crossing rate (ZCR) and the means of the Mel frequency cepstral coefficients (2-19) are used to perform a multiclass classification of euphonic, hyperfunctional and hypofunctional voices using six machine learning algorithms: Random Forest, K-Nearest Neighbors, Logistic Regression, Decision Trees, Support Vector Machines and Naive Bayes. The performance of these algorithms in identifying the three voice types through the mentioned acoustic parameters was evaluated using bootstrap.632. The results obtained from this research can be useful in developing a complementary application for clinical diagnosis or patient monitoring under specialist supervision.

The second section discusses some related works to observe how the problem has usually been addressed. The third section describes the methods and materials used in developing this project. The fourth section presents the implementation. The fifth section analyzes the results obtained. Finally, the sixth section presents the conclusions and future work.

2. RELATED WORKS

Several projects have been developed aiming to identify disorders that affect voice quality, even working with the same databases used in the present project. However, they focus on binary classification between healthy and pathological voices, that is, identifying whether a person is sick or not.

Verde et al. (2019) estimated F0, Jitter, Shimmer, and HNR from 2003 samples (796 healthy voices and 1207 pathological voices) as input to the classifier. The Boosted Trees algorithm is employed, using an 80-20 proportion for the training and testing phases, obtaining metrics of 83% and 86% for sensitivity and specificity, respectively. The authors clarify that the classification is binary. They use three databases: Massachusetts Eye and Ear Infirmary (MEEI), Saarbruecken Voice Database (SVD), and VOice ICar fEDerico II (VOICED).

Chen & Chen (2022) used a deep neural network for classification, extracting MFCC features (12 coefficients) from 114 samples of the VOICED database and obtaining metrics of sensitivity, specificity, precision, accuracy, and F1 with 97.8%, 99.4%, 99.4%, 98.6%, and 98.4%, respectively. It should be noted that the classification is binary.

Altayeb & Al-Ghraibah (2022) explored different voice feature extraction methods, including Mel frequency cepstral coefficients, zero-crossing rate, and discrete wavelet transform. The support vector machine algorithm was employed for classification, achieving 100% accuracy using a group of MFCC features (12 coefficients) and kurtosis, based on the VOICED database. The classification is also binary, as in previous projects, but the difference lies in conducting separate classifications for each disorder with a healthy voice, rather than distinguishing between pathological voices.

Hassan et al. (2020) used a dataset corresponding to 80 people (60 healthy and 20 with pathology), with audio samples of coughs, breathing, and voice. An LSTM-type RNN was used on the features: Spectral centroid, spectral roll-off, zero-crossing rate, and MFCC. A 70/30 ratio was adopted between training and testing. The area under the curve (AUC) reached around 97.4%, 98.8%, and 84.4% when classifying cough, breathing, and voice, respectively.

Radha et al. (2021) used the parameters chroma STFT, spectral centroid, spectral bandwidth, MFCC, roll-off, and zero-crossing rate for Parkinson's disease detection, as speech deterioration is an early indicator in this disease. A classification method based on convolutional neural networks, artificial neural networks, and hidden Markov model is used to distinguish samples of Parkinson's patients and healthy individuals. The artificial neural network-based model achieved a recognition rate of 96%.

3. METHODS AND MATERIALS

3.1. Databases

The dataset used throughout this project consists of samples from two databases: Saarbruecken Voice Data Base (Woldert-Jokisz, 2007) and Voice ICar fEDerico II (Cesari et al., 2018). Exactly 331 samples were taken, 208 from SVD and 123 from VOICED, corresponding to a total of 193 individuals, of whom 55 have hypofunction, 64 have hyperfunction, and 74 have euphony. All audios are sustained phonations of the vowel /a/. The samples were preprocessed since SVD samples have a sampling frequency of 50 kHz and a duration of 1 second, while VOICED samples have a sampling frequency of 8 kHz and a duration of approximately 4 seconds. To homogenize the samples, those with a longer duration were divided, setting all to 1 second, and the sampling frequency was modified to 8 kHz, equalizing the frequency range present. The number of samples for each class was standardized to 190, which is the maximum number of samples that could be obtained for the hypofunction class. Therefore, the dataset used is composed of 570 samples in total, 190 samples with euphony, 190 with hyperfunction, and 190 with hypofunction.

3.2. Acoustic Parameters

3.2.1. Means of the Mel Frequency Cepstral Coefficients

Mel Frequency Cepstral Coefficients (MFCC) are a representation of the audio signal that is commonly used in speech processing and voice recognition. To obtain these coefficients, a Short-Time Fourier Transform (STFT) is applied to a voice signal. Then, the resulting power is mapped onto a logarithmic Mel-frequency scale, and the cepstral coefficients are calculated from the resulting spectrum. Since the MFCC are calculated on a logarithmic scale, it fits better with human auditory perception, which is more sensitive to differences in frequency in the low-frequency range than in the high-frequency range (Rivera et al., 2022). In general, MFCC are used as an efficient representation of a voice signal, which may be relevant for identifying voices with euphonia, hyperfunction, and hypofunction. It is worth noting that only the means of the coefficients are used, and not the complete vectors of the coefficients, as using the means reduces the dimensionality of the data and eliminates redundancy. This is due to the fact that the samples belong to sustained phonations of the vowel /a/ (Flórez-Gómez et al., 2022), so it would be expected that a voice with euphonia (which is characterized by being balanced and harmonic) possibly presents more constant or similar values throughout the vectors of each coefficient, while voices with hypofunction and hyperfunction possibly present more outliers and variability that would raise the values of the means, which is considered relevant for classification.

3.2.2. Cepstral Peak Prominence-Smoothed (CPPS)

Cepstral Peak Prominence-Smoothed represents the difference between the most prominent cepstral peak, which corresponds to the first rahmonic, and the point with the same quefrency on the regression line through the smoothed cepstrum. Studies indicate that this parameter reflects the degree of dysphonia (Núñez-Batalla et al., 2019). Currently, in clinical practice, CPPS is the most widely used acoustic parameter for measuring vocal quality levels.

3.2.3. Zero Crossing Rate (ZCR)

Zero Crossing Rate (ZCR) is a value between 0 and 1 that indicates an average number of times the signal crosses the X-axis in defined intervals (Celdrán, 2015). Zero Crossing Rate can be useful in identifying differences between voices with euphonia and those with hyperfunction and hypofunction, as they may have a different zero crossing rate due to changes in vocal fold vibration.

3.2.4. Harmonic-to-Noise Ratio (HNR)

Harmonic-to-Noise Ratio is an evaluation of the ratio between the periodic and non-periodic components comprising a segment of vocal speech. The periodic component comes from vocal fold vibration, while the non-periodic component comes from glottal noise, expressed in dB. The evaluation of the ratio between these components reflects speech efficiency; that is, the greater the flow of air expelled from the lungs in vibrational energy of the vocal folds, the higher the HNR. A low HNR denotes an asthenic voice and dysphonia.

3.3. Machine learning models

The six machine learning models that have been employed to address the task of multiclass classification of the three voice classes in question are described below.

3.3.1. Random Forest

Random Forest is an ensemble model that uses multiple decision trees to perform classification. Each tree is trained with a random subsample of the data and a random selection of features. To classify an input object *x*, its position is evaluated in each tree and a class label is assigned based on the results from all the trees. The final classification is determined by voting the results of the individual trees (Schonlau & Zou, 2020).

3.3.2. K Nearest Neighbors

It is a classification model based on distance. The idea is that an object is classified based on the labels of its k nearest neighbors in the feature space. The value of k is chosen beforehand and can vary depending on the problem.

For an input object *x*, the distance between *x* and each of the training points is calculated, and the k closest points are selected. The class of the input object is determined by majority voting among the k classes of the nearest neighbors (Taunk et al., 2019).

3.3.3. Logistic Regression

Logistic regression is a linear classification model that uses the logistic function to predict the probability of belonging to a class. It can be extended for multiclass classification using techniques such as One-vs-Rest or SoftMax. In the One-vs-Rest approach, a logistic regression model is trained for each class, while in the SoftMax approach, a SoftMax activation function is used to assign probabilities to each class. Multiclass logistic regression uses a loss function such as Categorical Cross-Entropy to train the model and adjust the parameters (Daniels & Minot, 2019).

3.3.4. Decision Trees

It is a classification model based on creating a tree of decision rules. Each node of the tree represents a decision rule based on a feature, and each branch represents a possible outcome of that rule. The tree is constructed from the training data, and the final classification is performed by traversing the tree from top to bottom until reaching a leaf that contains the class label (Charbuty & Abdulazeez, 2021).

3.3.5. Support Vector Machines

Support vector machines are classification models that attempt to find the best separation between classes by creating a hyperplane that maximizes the distance between the closest samples of each class. A kernel function is used to project the data into a higher-dimensional feature space to find the optimal hyperplane (James et al., 2013).

3.3.6. Naive Bayes

Naive Bayes is a probabilistic classification model based on Bayes' theorem and the naive assumption that all features are independent of each other. Using training data, the model calculates the probability of an object belonging to each class using Bayes' theorem (Murphy, 2006).

3.4. Model Evaluation Using Bootstrap.632

Bootstrap.632 is a technique for evaluating machine learning models that uses bootstrap sampling to generate multiple random samples with replacement from the training dataset and trains a model on each of them. The model is then evaluated on a test dataset that was not used to train the model, and the evaluation of the models is adjusted using a correction based on the probability of a particular data point being in the training or test set in each bootstrap sample. This technique produces a more accurate estimate of the model's performance on new and unknown data (Efron, 1983).

4. IMPLEMENTATION

The data analysis and machine learning models training were programmed in Python, using various libraries such as NumPy, Pandas, and Scikit-learn. Additionally, the

Parselmouth library was used for the analysis of acoustic signals and the extraction of acoustic parameters. All the code was executed on Google Colab.

In order to obtain better results, hyperparameter optimizations were performed on the machine learning models, using the grid search method, which resulted in the following configurations:

1. For the Random Forest model, the entropy criterion was chosen to evaluate the quality of the partitions at each node. The maximum depth of the trees in the random forest was set to 30 levels. All available features were considered when searching for the best partition at each node. A minimum of 2 samples were required to split an internal node. The random forest consisted of a total of 500 decision trees,
2. For the K Nearest Neighbors model, the ball tree algorithm was used to calculate the distances between instances. Three nearest neighbors were considered for classification, and the Manhattan distance (L1) was used as the distance metric. During classification, the nearest neighbors have more influence, as they are weighted inversely proportional to the distance,
3. For the Decision Trees model, a split quality measure based on the 'log_loss' criterion was used. The tree was configured with a maximum depth of 20 levels to prevent overfitting. All features were considered, with the number of maximum features set to the logarithm with a base of 2 of the total number of features. Moreover, a minimum of 2 samples were required in a node to perform an additional split. The 'best' strategy was employed to select the optimal split based on the defined quality criterion.
4. For the Support Vector Machines model, a combination of hyperparameters with C=10, gamma=0.2, and kernel='rbf' achieved the best results in terms of accuracy. This indicates that a high value of C was used, prioritizing precise classification even with a narrower separation margin. Additionally, a moderate value of gamma (0.2) was chosen to consider the influence of nearby training points on the decision boundary. The 'rbf' kernel was used to capture nonlinear relationships between the data,
5. For the Logistic Regression model, the optimal combination of hyperparameters was found to be C=10, penalty='l2' (Ridge regularization), and solver='newton-cg'. This indicates that a relatively weak regularization was chosen to prevent overfitting by adding the square of the coefficients. The 'newton-cg' solver was used, which is suitable for multiclass optimization problems and relies on the conjugate Newton method to optimize the model's objective function,
6. For the Naive Bayes model, it was found that the optimal configuration includes a smoothing term for the covariance matrix of the Gaussian distributions equal to $1e^{-09}$.

5. RESULTS ANALYSIS

This section presents the results of the machine learning models, which have undergone hyperparameter optimization and bootstrap.632 evaluation (see Table 1). Confidence intervals for 500 iterations based on the accuracy metric of each model were obtained for the multiclass classification task of voices with euphonia, hypofunction, and hyperfunction.

Tab. 1. Confidence intervals of each machine learning model

Model	Confidence interval
Random Forest	86-90%
K Nearest Neighbors	87-92%
Support Vector Machines	84-91%
Decision Trees	79-85%
Logistic Regression	58-65%
Naïve Bayes	53-61%

The distribution histograms obtained after applying the Bootstrap.632 method for the three best performing models is shown below.

5.1. Random Forest Results

The confidence interval obtained suggests that the model has a high probability of obtaining an accuracy between 86% and 90% on new data sets, which is an indicator that the model may be useful for the task of classifying voices with euphony, hypofunction and hyperfunction. Figure 1 shows that in more than 120 of the 500 iterations performed there is a concentration in accuracies of 88%. Table 2 shows the densities of the Random Forest model accuracies.

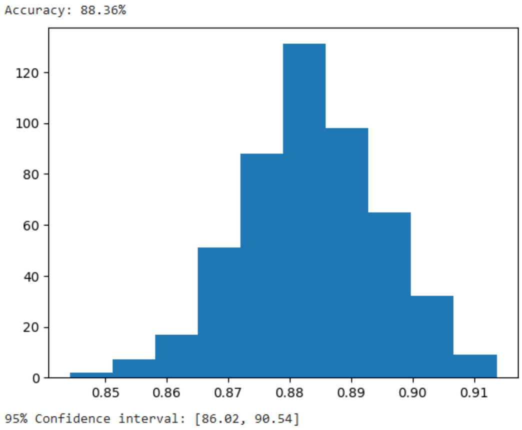

Fig. 1. Histogram of predictions and confidence interval of the Random Forest model

Tab. 2. Densities of the Random Forest model accuracies

Accuracy	Density
84.35%	2
85.06%	7
85.77%	18
86.48%	53
87.19%	89
87.90%	131
88.61%	97
89.32%	64
90.03%	31
90.74%	8

5.2. K-Nearest Neighbors Results

The 87% to 92% confidence interval indicates that if multiple random samples are taken from the data set, the accuracy of the model is expected to be within this range in most cases. Furthermore, the fact that the range is narrow indicates that the model is robust and that its performance is unlikely to vary significantly across different data sets. Figure 2 shows that much of the accuracies fall between 89% and 91% of the 500 iterations performed, with its highest concentration at 90% in more than 130 of the total iterations. Table 3 shows the densities of the K nearest neighbors model accuracies.

The results obtained suggest that the K nearest neighbors model is a good candidate for voice classification in the three classes mentioned above, and that its accuracy is high enough to be considered useful in practical applications.

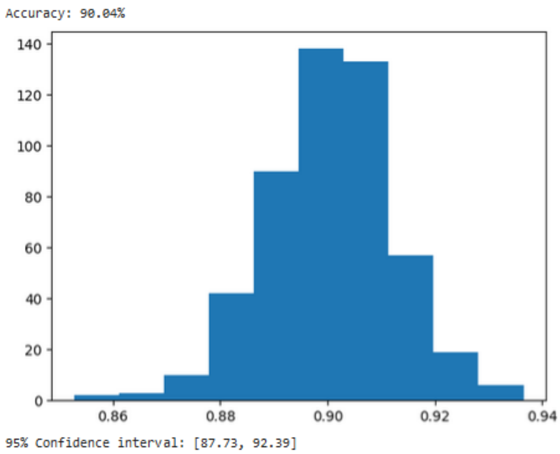

Fig. 2. Histogram of predictions and confidence interval of the K-Nearest Neighbors model

Tab. 3. Densities of the K-Nearest Neighbors model accuracies

Accuracy	Density
85.25%	2
86.10%	3
86.96%	9
87.81%	42
88.67%	91
89.52%	138
90.38%	134
91.23%	57
92.09%	18
92.94%	6

5.3. Support Vector Machine Results

The confidence interval obtained for this model, ranging from 84% to 91%, may indicate that the model's performance is consistent and stable. It suggests that the model is able to accurately classify most of the samples. Figure 3 shows that there are a large number of accuracies that fall around 87%; however, many points fall within a higher percentage range, going from 88% to 90%. The model is able to classify all three speech classes with acceptable accuracy. Table 4 shows the densities of the Support Vector Machine model accuracies.

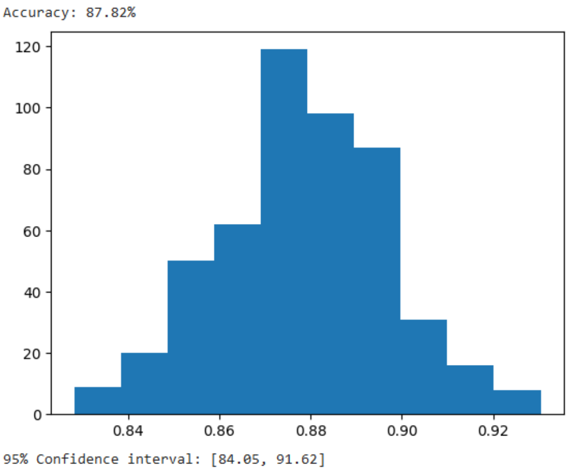

Fig. 3. Histogram of predictions and confidence interval of the Support Vector Machine model

Tab. 4. Densities of the Support Vector Machine model accuracies

Accuracy	Density
82.99%	9
83.99%	20
84.99%	53
85.99%	62
86.99%	118
87.99%	97
88.99%	88
89.99%	29
90.99%	17
91.99%	7

6. CONCLUSIONS AND FUTURE WORK

According to the presented results, it can be concluded that the K Nearest Neighbors model performs best for the multiclass classification task of voices with euphony, hypofunction, and hyperfunction, and could be considered for the development of complementary tools that may be useful in the objective evaluation of patients under the supervision of a specialist, by identifying the three voice classes and assigning an appropriate rehabilitation scheme. The KNN model presents a narrower confidence interval and a higher accuracy (87-92%) than other models such as Random Forest (86-90%) and Support Vector Machines (84-91%). Decision trees (79-85%) are above the 70% threshold that is considered acceptable for many classification problems; however, it is not the best option. On the other hand, logistic regression, and Naive Bayes showed less satisfactory performance with confidence intervals that do not exceed 65%, indicating that these models may not be the most suitable for the task at hand. It should be noted that the classifications performed are multiclass, which is relevant considering that hypofunction and hyperfunction are two common types of dysphonia, and although they have individual characteristics, they also share some features.

It is important to note that these results are valid for the hyperparameter settings and the dataset used in this study, and further experiments should be conducted to validate these results for different configurations and datasets.

Data availability

The set of voice samples with euphonia, hypofunction, and hyperfunction used in this study, which belong to the VOICED database, are publicly available under an Open Data Commons Attribution License on PhysioNet's website at https://physionet.org/content/voiced/1.0.0/. Additionally, voice samples from the Saarbruecken Voice Database are freely accessible to the public, provided by the Institute of Phonetics at the University of Saarland, and can be found at https://www.stimmdatenbank.coli.uni-saarland.de.

Acknowledgments

The authors thank the reviewers for their helpful suggestions that helped improve the quality of this article.

REFERENCES

Altayeb, M., & Al-Ghraibah, A. (2022). Classification of three pathological voices based on specific features groups using support vector machine. *International Journal of Electrical and Computer Engineering (IJECE)*, *12*(1), 946-956. https://doi.org/10.11591/ijece.v12i1.pp946-956

Behlau, M., & Pontes, P. (1989). *Avaliação Global da Voz*. Editora Paulista Publicações Médicas.

Behlau, M., Madazio, G., Feijó, D., Azevedo, R., Gielow, I., & Rehder, M. (2005). Perfeccionamiento vocal y tratamiento fonoaudiológico de las disfonías. In M. Behlau (Eds.), *Voz: O livro do especialista*. Thieme Revinter.

Celdrán, E. M. (2015). Naturaleza fonética de la consonante 'ye'en español. *Normas: revista de estudios lingüísticos hispánicos*, *5*, 117-131. https://doi.org/10.7203/Normas.5.6825

Cesari, U., De Pietro, G., Marciano, E., Niri, C., Sannino, G., & Verde, L. (2018). A new database of healthy and pathological voices. *Computers & Electrical Engineering*, *68*, 310-321. https://doi.org/10.1016/j.compeleceng.2018.04.008

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20-28. https://doi.org/10.38094/jastt20165

Chen, L., & Chen, J. (2022). Deep neural network for automatic classification of pathological voice signals. *Journal of Voice*, *36*(2), 288.e15-288.e24. https://doi.org/10.1016/j.jvoice.2020.05.029

Daniels, L., & Minot, N. (2019). *An introduction to statistics and data analysis using Stata®: From research design to final report*. Sage Publications.

Descamps, G., Verset, L., Trelcat, A., Hopkins, C., Lechien, J. R., Journe, F., & Saussez, S. (2020). ACE2 protein landscape in the head and neck region: the conundrum of SARS-CoV-2 infection. *Biology*, *9*(8), 235. https://doi.org/10.3390%2Fbiology9080235

Efron, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association*, *78*(382), 316-331. https://doi.org/10.2307/2288636

Farias, P. (2016). *Guía clínica para el especialista en laringe y voz*. Librería Akadia Editorial.

Flórez-Gómez, A. F., Orozco-Arroyave, J. R., & Roldán-Vasco, S. (2022). Correlación entre espacios de características acústicas del habla y trastornos clínicos de la voz en pacientes con disfagia. *TecnoLógicas*, *25*(53), e2220. https://doi.org/10.22430/22565337.2220

Hassan, A., Shahin, I., & Alsabek, M. B. (2020). COVID-19 detection system using recurrent neural networks. *2020 International conference on communications, computing, cybersecurity, and informatics (CCCI)* (pp. 1-5). IEEE. https://doi.org/10.1109/CCCI49893.2020.9256562

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, *181*(2), 271-280.e8. https://doi.org/10.1016/j.cell.2020.02.052

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

López, J. A. P. (1997). *Los trastornos de la voz en el personal docente de logroño. Estudio de la voz en los profesionales de la enseñanza*. (Doctoral dissertation, Universidad de Navarra).

López, J. A. P. (2000). Estudio de la prevalencia de los trastornos de la voz en el personal docente de Logroño. *Zubía*, *12*, 111-145.

Murphy, K. P. (2006). Naive bayes classifiers. *University of British Columbia*, *18*(60), 1-8.

Núñez-Batalla, F., Cartón-Corona, N., Vasile, G., García-Cabo, P., Fernández-Vanes, L., & Llorente-Pendás, J. L. (2019). Validez de las medidas del pico cepstral para la valoración objetiva de la disfonía en sujetos de habla hispana. *Acta Otorrinolaringológica Española*, *70*(4), 222-228. https://doi.org/10.1016/j.otoeng.2018.04.005

Radha, N., Sachin Madhavan, R. M., & Sameera holy, S. (2021). Parkinson's Disease detection using Machine Learning Techniques. *International Journal of Early Childhood Special Education (INT-JECSE)*, *30*(2), 543. https://doi.org/10.24205/03276716.2020.4055

Rivera, M. A. B., Flores, P. M. Q., Loaiza, R. E. P., & Rivera, L. G. (2022). Analysis of audio signals using deep learning algorithms applied to COVID diagnostic systems. *2022 IEEE Mexican International Conference on Computer Science (ENC)* (pp. 1-6). IEEE. https://doi.org/10.1109/ENC56672.2022.9882932

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, *20*(1), 3-29. https://doi.org/10.1177/1536867X20909688

Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A brief review of nearest neighbor algorithm for learning and classification. *2019 International Conference on Intelligent Computing and Control Systems (ICCS)* (pp. 1255-1260). IEEE. https://doi.org/10.1109/ICCS45141.2019.9065747

Verdaguer, J. M., Górriz, C., Prim, M. P., del Palacio, A. J., Gavilán, J., & de Diego, J. I. (2008). Análisis de los cambios en el espectrograma tras la intubación endotraqueal. *Acta Otorrinolaringológica Española*, *59*(5), 217-222. https://doi.org/10.1016/S0001-6519(08)73298-9

Verde, L., De Pietro, G., Alrashoud, M., Ghoneim, A., Al-Mutib, K. N., & Sannino, G. (2019). Leveraging artificial intelligence to improve voice disorder identification through the use of a reliable mobile app. *IEEE Access*, *7*, 124048-124054. https://doi.org/10.1109/ACCESS.2019.2938265

Woldert-Jokisz, B. (2007). *Saarbruecken voice database*. Computer Science.