ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 3. Issue 2 (2023)
https://doi.org/10.62019/abbdm.v3i2.93

Automated Acoustic Evaluation of Voice Disorders: A Comprehensive Study on Parameter Analysis Using ANN

Hina Zameer*, Sidra Abid Syed, Marium Raziq, Muhammad Muzammil Khan, Sania Tanvir, Shahzad Nasim

Chronicle

Article history
Received: December 11,2023
Received in the **revised format:** January 2nd ,2024
Accepted: January 2nd, 2024
Available online: January 3, 2024

Hina Zameer, Sidra Abid Syed, Marium Raziq, Muhammad Muzammil Khan & Sania Tanvir are currently affiliated with Biomedical Engineering Department Sir Syed University of Engineering and Technology Karachi, Pakistan.
Email: Hinazameer0@gmail.com
Email: Sidra.agha@yahoo.com
Email: raziq@ssuet.edu.pk
Email:muzammilkhan_82@hotmail.com
Email: stanvir@ssuet.edu.pk

Shahzad Nasim is currently affiliated with Management Sciences & Technology Department, The Begum Nusrat Bhutto Women University Karachi, Pakistan.
Email: shahzad.nasim@bnbwu.edu.pk

***Corresponding Author**

Abstract

Analysis of the voice is an important diagnostic technique that may be used to identify anomalies in the voice. It provides a non-invasive alternative to treatments that are invasive. Within the scope of this research, a complete investigation of voice disorder evaluation approaches is investigated, with a particular emphasis placed on acoustic analysis and categorization. The research makes use of a number of metrics, including Jitter, Shimmer, and Harmonic-to-Noise Ratio (HNR), in conjunction with an Artificial Neural Network (ANN) classifier. Through the utilization of the Saarbruecken Voice Database (SVD), the research attempts to differentiate between voices that are healthy and voices that are dysphonic regardless of gender. In order to improve the accuracy of the model, Principal Component Analysis (PCA) is a useful tool for feature selection. Among the females, the best accuracy achieved when using all indications to differentiate between healthy and dysphonic persons was 87.9%. By using the produced output model, we were able to reduce the number of input parameters to 17. Consistent with all parameters, the highest obtained accuracy was 87.9%. Exhibiting neither loss nor gain of knowledge. The men's instance indicates significantly reduced levels of precision. The male group attained a maximum accuracy of 94% when classifying between healthy and dysphonic individuals using all measures. The generated output model enabled us to decrease the input parameters to 14. Once again, the highest level of accuracy achieved remained consistent at 70% across all parameters.The findings demonstrate that there are various degrees of accuracy in the male and female groups, demonstrating the efficacy of particular factors in categorization.

Keywords voice disorder, ANN, SVD, PCA.

© 2023 EuoAsian Academy of Global Learning and Education Ltd. All rights reserved

INTRODUCTION

Voice analysis methods are commonly employed to evaluate voice abnormalities. The efficacy of these treatments depends on their non-invasive nature, as opposed to more intrusive procedures like laryngoscopy tests (Syed et al., 2021). Voice abnormalities can be detected by auditory perceptual analysis, however the findings may vary depending on the practitioner's level of skill. Hoarseness in one's voice is a frequent complaint among individuals in primary care institutions (Teixeira et al., 2017). Dysphonia impacts 30% of individuals and 50% of elderly persons. This condition alters the quality of one's voice and

has substantial effects on overall quality of life. Furthermore, this also signifies a substantial financial hardship (Syed, Rashid, Hussain, Imtiaz, et al., 2021). For patients with a progressing disease, it is crucial to promptly diagnose the condition in order to have access to more effective treatment options and improve the prognosis (Abid Syed et al., 2020). The potential use of acoustic assessment in the treatment of disordered voice is the driving force behind this study. Voice disorder among specific populations, including primary teachers and salespeople, occurs at considerable rates. More than 40% of primary school teachers in Asia have suffered from voicing difficulties (Lee et al., 2010). Spokespersons for evaluation and rehabilitation are highly in demand. Sadly, voice therapist training is a long-term program. An assistant can be a reliable automatic acoustic evaluation system. A portable-devices can implement an automatic assessment system that can allow patients to conveniently carry out self-assessment (Alhussein & Muhammad, 2018). Authors must address real-world challenges not fully explored in earlier publications to design a viable automatic evaluation system.

As indicated above, numerous speech tasks' perceptual and auditory contributions to voice assessment have been explored (Pylypowich & Duff, 2016). As mentioned, continuous voice calculation features are unlikely because of variations in voice and other non-language content. Over the past decade, a breakthrough in deep learning has occurred, for instance, the development of the large-scale ASR system (Teixeira et al., 2020). The development of ASR techniques has benefitted from the research on pathological voice assessment. Parkinson's condition and aphasia were detected using non-conventional ASR features from distinct speech samples. In order to measure voice irregularities, authors are motivated to offer phoneme-based attributes on ASR output. Author are interested in carefully examining the efficacy of different voice phonemes. Furthermore, an auditory perceptual utterance level rating for comparison is recommended.

This research aims to comprehensively assess voice abnormalities through the meticulous evaluation of acoustic parameters, including Jitter, Shimmer, and Harmonic-to-Noise Ratio (HNR). The primary aim is to discern a clear distinction between healthy and dysphonic voices across different genders using the Saarbruecken Voice Database (SVD). Through the utilization of an Artificial Neural Network (ANN) classifier, this study seeks to establish a robust framework for accurate classification, capitalizing on varied tones and vowels present in voice recordings. Additionally, employing Principal Component Analysis (PCA) facilitates the identification of key features, enhancing the efficacy of classification models.

LITERATURE REVIEW

Analyzing voice signals yields a number of parameters. For the purpose of diagnosing dysphonia, Teixeira and Fernandes (Teixeira & Fernandes, 2015) investigated how reliable Jitter, Shimmer, and HNR features were. Three factors were statistically examined for the vowels /a/, /i/, and /u/ in high, low, and normal tones. According to this study, Jitter and Shimmer are useful parameters for an automated dysphonia diagnosis system. To assess this study, a complex tool and numerous dimension reduction and variable selection procedures are needed. The best predictor subset is found by variable selection. When dealing with huge datasets that contain redundant information and unexpected variables, the problem of variable selection becomes apparent. Optimal selection of

input variables improves training using intelligent approaches on a restricted subset. The efficacy of six chaotic measures derived from nonlinear dynamics theory in distinguishing between normal and pathological voice quality was investigated by Henríquez et al. (Henriquez et al., 2009). Rényi entropies of the first and second orders, correlation dimension, and correlation entropy are among the metrics that are taken into consideration. There was also an investigation into Shannon entropy and the shared data function's initial optimal. A commercial database called MEEI Voice Disorders and a multi-quality dataset were used to evaluate the measures' usefulness. A standard neural network classifier was used to evaluate the proposed measures. Ran a 99.7 percent commercial database and an 82.5 percent multi-quality database on a worldwide scale. To differentiate between nodules, unilateral paralysis, and healthy voices, Forero et al. (Forero M. et al., 2016) used many components of the glottal signal. We contacted a speech pathologist who recorded the sounds of twelve patients with nodules, eight patients with paralysis of the vocal folds, and eleven healthy controls. With each speaker recording eight voices, a total of 248 recordings were produced. We utilised ANN, SVM, and HMM classifiers. Glottal signal characteristics, MFCCs, and an SVM classifier were used to achieve the greatest accuracy of 97.2%.

For the purpose of diagnosing and differentiating voice diseases, Markaki et al. (Markaki & Stylianou, 2011) investigated the use of modulation spectrum, a representation that combines acoustic and modulation frequency information. The initial approximation is reduced to a space with fewer dimensions via higher-order singular value decomposition. The SVM classifier had a 94.1% accuracy rate in detecting voice illnesses. In their investigation, Panek et al. (Panek et al., 2015) looked at a set of 28 sound characteristics. In order to identify four different disorders—hyperfunctional dysphonia, functional dysphonia, laryngitis, and vocal cord paralysis—this vector is evaluated using PCA, kPCA, and an auto-associative neural network (NLPCA).

High, low, and normal pitches of the vowels /a/, /i/, and /u/ are the primary targets of the study. Optimal efficiency values of around 100% are shown by the results. Across a range of frequency ranges, Al-Nasheri et al. (Al-nasheri et al., 2017) examined correlation functions. After obtaining the greatest peak values and lag values from each frame of the spoken signal, we utilised correlation functions to find and label samples that may not be reliable. Three datasets were utilised: AVPD, SVD, and MEEI. Here, a support vector machine classifier was employed. Precision in pathology categorization ranged from 91.1% to 99.8% across all three datasets. For the three databases, the accuracy for pathologic occupational categories was 99.2%, 98.9%, and 95.1%.

Using a variety of classification schemes, Sellam et al. (Sellam & Jagadeesan, 2014) aim to examine and differentiate between healthy and sick voices in children. Using SVM and RBFNN for classification, we are able to distinguish between normal voice and speech affected by illness. Signal energy, pitch, formant frequencies, reflection coefficients, Jitter, and Shimmer were among the acoustic properties obtained. With a score of 91%, the RBFNN outperformed the SVM, which only managed 83% accuracy. For classification tasks like these, artificial neural networks are commonly used.

Researchers Hugo Cordeiro (Cordeiro et al., 2015) looked at several vocal tract features and classifiers to find the ones that worked best for identifying problematic voices. These included Mel-Line Spectral Frequencies, first peak of the spectral envelope, SVM, and

LSF. He accurately identified between healthy individuals, those with vocal fold nodules and edoemas, and those with unilateral vocal fold paralysis and other neurological laryngeal diseases with 84.4% of the time. To further pinpoint ill voices with 95% accuracy, he employed formant analysis, Regression Trees, and the harmonic-to-noise ratio (Cordeiro, Meneses, et al., 2015).

METHODOLOGY

Datasets

SVD stands for Saarbruecken Voice Database. It is one of the very common voice disorder data base. A voice recording collection by more than 2,000 people (Barry et al., 2007). There is voice registration "[/i/, /a/, /u/] in standard high and low pitches. In a recording session, the truth was recorded increasing vocal pitch documentation [/i/, /a/, /u/]. Recording the sentence, 'Wow do you like it'('How are you, good morning?'). What do you like?' In each file for the specified components, the voice signal and EGG signal were stored. The text file in the database contains all relevant data collection information. Such features make the experimenters a good choice. All recorded SVD voices have been sampled at 50 kHz with a 16-bit resolution. Some recording sessions do not include all vowels depending on their recording quality in each version. This web interface provides the 'Saarbruecken Voice Server' and available on http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4.

Table 1.
Statistics of SVD dataset

	Subject		**Age**	
	Male	*Female*	*Male*	*Female*
Normal	29	41	20-69	19-56
Pathologic	29	41	11-77	18-73

Artificial Neural Network (ANN)

Multilayer Perceptron (MLP) structures were trained using back-propagation for the ANN classifier. Different topologies with different numbers of hidden layer neurons were tested to get the best adaptation performance (Han et al., 2018). The dataset was divided into testing, validation, and training sets. Each set of data included 70%, 15%, and 15% proportions. Weights and bias alter the ANN to produce the desired output. The model has one output neuron. The output objective consisted exclusively of binary digits. The output produced by the ANN does not consistently yield binary values of zero or one (Argatov, 2019). Therefore, we needed to apply additional processing to ensure that the output is constrained to zero or one. An experimental threshold of 0.5 was determined for the output in this technique. Each model will have a distinct number of input nodes and hidden nodes.

PARAMETERS

Jitter: It quantifies fluctuations in the pitch and is defined as variations in the frequency of consecutive vibration cycles. The prevalence of problematic voices is much greater than that of normal voices. For example, the wave form of the prolonged vowel phonation from the standard speaker and the other with seriously disordered voice. Normal voice

waveform has good periodicity while severe voice periodicity is poor. In jitter, it is higher than the normal voice which is 0.2% whereas severe distorted voice is 2.2%. The jitter (local jitter) has been calculated using Praat computer software (Boersma, 2021).

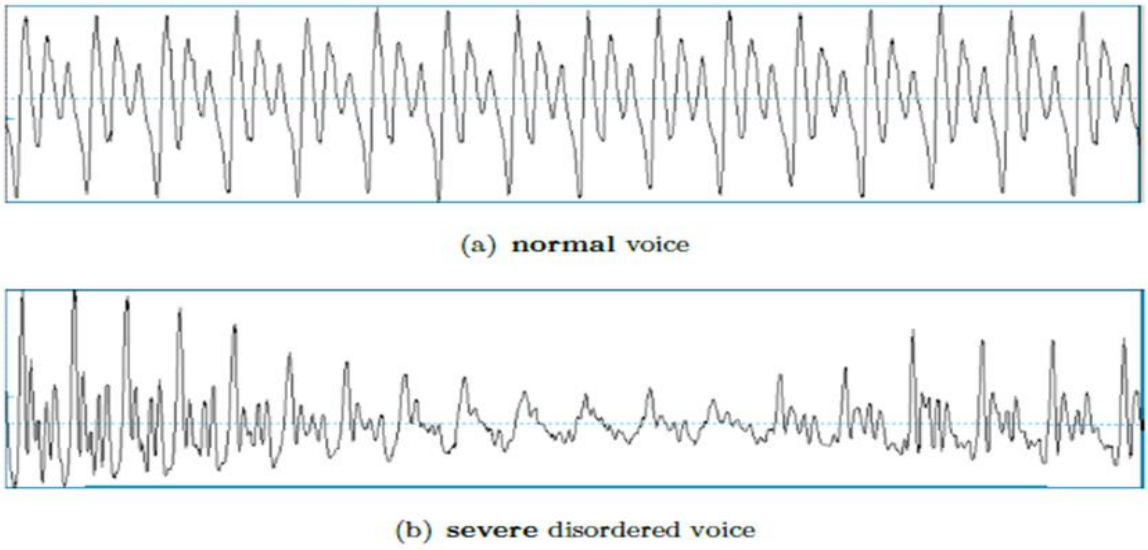

Figure 1.
Sustained phonation vowel /a/ waveform.

- *Shimmer:* The glow is a disturbance of the vocal sound, in contrast to jitter. Shimmer measures the intensity variation of adjacent vocal fold vibration cycles. Shimmer measures as jitter, a higher shimmer percentage is generally seen in pathological voices. As shown in Figure 2.5, the waveform of the serious distorted voice can show significant amplitude changes. Shimmer is calculated by Praat, with a normal voice value of 4.1% and a severely impaired voice value of 14.4% (Boersma, 2021).

- *HNR (Harmonic-to-Noise Ratio):* It is a ratio of harmonic to noise. Aperiodic and periodic components compose the acoustic signals produced by the vibration of the vocal pliers. Irregular vibrations and/or poor vocal fold closure can lead to aperiodic noise. The ratio of voice signal addiction in decibels is quantified by HNR. Lower HNR means greater voice noise. HNR is usually used as a voice disorder indicator, such as heartburn (Bilal et al., 2019).

Experimental Setup

Principal Components Analysis was the tool of choice for the writers. Mathematical concepts including eigenvalues, eigenvectors, standard deviation, and covariance are utilised in this statistical technique. To begin, we need to take the average out of every data dimension. This generates a dataset, referred to as "data adjusted," in which the average value is zero. Subsequently, the eigenvectors and eigenvalues are computed based on the covariance matrix. It is necessary to determine the number of primary components to select. The primary components were computed using the Matlab function "princomp". The eigenvalues are already presented in an orderly fashion in the output of this function. We need to add together all of these data and find the percentage. Accordingly, the first set of eigenvectors is selected so that they correspond to either 90% or 95% of the total percentage. Thus, 90% or 95% of the data is accounted

for by the original collection of eigenvectors. The last step is to multiply the updated data by the inverse of the matrix containing the 17 eigenvectors that were chosen. Finding values that are near approximations required computing the mean for the training set alone and then excluding it from the validation and test sets.

RESULTS AND DISCUSSION

An attempt was undertaken to locate dysphonic sounds by utilizing recordings from SVD. A categorization was conducted separately for women and men, distinguishing between healthy and sick conditions. The sample size of the control group was equivalent to that of the diseased group being examined. An algorithm was used which was developed in (J. P. Teixeira & Gonçalves, 2014) consisting of 9 parameters, including 4 parameters for Jitter, 4 parameters for Shimmer, and HNR. This vector was then multiplied by three tones (High, Low, and Normal) and three vowels (/a/, /i/, and /u/). Dimension reduction and variable selection procedures were used once the feature vector, which included 81 variables, was constructed. The processing time is reduced, and the features that distinguish healthy individuals from those with dysphonia are brought to light. The categorization task was carried out by use of a neural network. In order to find the best model for generalization, we tested several various topologies with varying numbers of neurons in the hidden layer. In this situation, just the test set's precision is being considered, even if the training and validation sets' precision was also computed. This means that the dataset used to report the precision was not really used for training.

Among the females, the best accuracy achieved when using all indications to differentiate between healthy and dysphonic persons was 87.9%. By using the produced output model, we were able to reduce the number of input parameters to 17. Consistent with all parameters, the highest obtained accuracy was 87.9%. Exhibiting neither loss nor gain of knowledge. The men's instance indicates significantly reduced levels of precision. The male group attained a maximum accuracy of 94% when classifying between healthy and dysphonic individuals using all measures. The generated output model enabled us to decrease the input parameters to 14. Once again, the highest level of accuracy achieved remained consistent at 70% across all parameters.

Table 2.
Accuracy Table

	Cross Validation Score	Accuracy	Mean Square Error	Mean Absolute Error	Root2 Score
Female	0.879996	0.879995	0.120005	0.120005	0.517516
Male	0.937284	0.945885	0.054115	0.054115	0.782430

In order to determine the shim, jitta, and HNR parameters for distinct vowels across tones, the given output model uses these metrics. The word "shim" is often pronounced with the vowel /i/ and the Low (L) tone. Most people agree that the L tone is the best one. Nevertheless, the stressed vowel in 4 is pronounced as /a/. The jitta vowel system emphasized nearly every tone. This pertains to the examination conducted in (Teixeira & Fernandes, 2015), which highlights a distinction between normal and pathological conditions in terms of Jitter characteristics across three different tones and vowels. No emphasis was placed on any certain pitch or vowel for the jitter settings. The HNR

parameter is linked to a low tone and the /i/ vowel. Only the HNR for vowel /a/ at L tone and the shim for vowel /i/ at N tone remain unchanged in the second approach model, compared to the first technique. This study focuses on the apq3 and apq5 parameters, which are distinct from the first technique. Nevertheless, apq3 and apq5 represent certain types of shimmer characteristics. This analysis demonstrates a high occurrence of the vowel /a/ and the N tone. The first strategy emphasized the shim, jitta, and HNR factors in the female group. Shim exhibits a high occurrence of the /a/ vowel and the N tone. The vowel /i/ and the L tone are selected for the jitta sound. The first technique continues to priorities the selection of HNR, even in cases when there is an equal occurrence of two vowels (/a/ and /u/) and a preference for the L tone.

CONCLUSION

The findings of this extensive study underscore the significance of acoustic analysis in diagnosing voice disorders. By carefully exploring parameters like Jitter, Shimmer, and Harmonic-to-Noise Ratio (HNR) across diverse tones and vowels, this research elucidates the nuanced distinctions between healthy and dysphonic voices. The tailored approach in feature selection, facilitated by Principal Component Analysis (PCA), not only refines classification models but also reveals gender-specific trends in voice disorders. The notable divergence in classification accuracy between male and female groups prompts further investigation into gender-based characteristics within voice pathology. This underscores the importance of personalized diagnostic tools that consider individual differences, potentially revolutionizing clinical assessments and treatment strategies. Moreover, the exploration of an Artificial Neural Network (ANN) classifier highlights the potential for automated systems to streamline voice disorder evaluations. The development of such systems could mitigate the subjectivity associated with auditory perceptual analysis, offering standardized and efficient diagnostics.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the supervisors and their different affiliations.

Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.

Availability of data and material: In the approach, the data sources for the variables are stated.

Authors' contributions: Each author participated equally to the creation of this work.

Conflicts of Interests: The authors declare no conflict of interest.

Consent to Participate: Yes

Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

Alhussein, M., & Muhammad, G. (2018). Voice pathology detection using deep learning on mobile healthcare framework. IEEE Access: Practical Innovations, Open Solutions, 6, 41034–41041. https://doi.org/10.1109/access.2018.2856238

Al-nasheri, A., Muhammad, G., Alsulaiman, M., & Ali, Z. (2017). Investigation of voice pathology detection and classification on different frequency regions using correlation functions.

Journal of Voice: Official Journal of the Voice Foundation, 31(1), 3–15. https://doi.org/10.1016/j.jvoice.2016.01.014

Argatov, I. (2019). Artificial neural networks (ANNs) as a novel modeling technique in tribology. Frontiers in Mechanical Engineering, 5. https://doi.org/10.3389/fmech.2019.00030

Barry, W. J., P�tzer, M., Caradonna, R., Dressler, C., Erriquez, A., Hausmann, T., Jarmut, S., Moos, A., Tengowski, S., Trifsik, O., Walz, M., & Woldert-Jokisz, B. (2007). Saarbruecken voice database. http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4

Bilal, N., Selcuk, T., Sarica, S., Alkan, A., Orhan, İ., Doganer, A., Sagiroglu, S., & Kılıc, M. A. (2019). Voice acoustic analysis of pediatric vocal nodule patients using ratios calculated with biomedical image segmentation. Journal of Voice: Official Journal of the Voice Foundation, 33(2), 195–203. https://doi.org/10.1016/j.jvoice.2017.11.010

Boersma, P. (2021). Praat: doing phonetics by computer (Version 4.6.33). https://www.academia.edu/49651969/Praat_doing_phonetics_by_computer_Version_4_6_33

Cordeiro, H., Fonseca, J., Guimaraes, I., & Meneses, C. (2015). Voice pathologies identification speech signals, features and classifiers evaluation. 2015 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA).

Forero M., L. A., Kohler, M., Vellasco, M. M. B. R., & Cataldo, E. (2016). Analysis and classification of voice pathologies using glottal signal parameters. Journal of Voice: Official Journal of the Voice Foundation, 30(5), 549–556. https://doi.org/10.1016/j.jvoice.2015.06.010

Han, S.-H., Kim, K. W., Kim, S., & Youn, Y. C. (2018). Artificial neural network: Understanding the basic concepts without mathematics. Dementia and Neurocognitive Disorders, 17(3), 83. https://doi.org/10.12779/dnd.2018.17.3.83

Henriquez, P., Alonso, J. B., Ferrer, M. A., Travieso, C. M., Godino-Llorente, J. I., & Diaz-de-Maria, F. (2009). Characterization of healthy and pathological voice through measures based on nonlinear dynamics. IEEE Transactions on Audio, Speech, and Language Processing, 17(6), 1186–1195. https://doi.org/10.1109/tasl.2009.2016734

Lee, S. Y.-Y., Lao, X. Q., & Yu, I. T.-S. (2010). A cross-sectional survey of Voice Disorders among Primary School Teachers in Hong Kong. Journal of Occupational Health, 52(6), 344–352. https://doi.org/10.1539/joh.l10015

Markaki, M., & Stylianou, Y. (2011). Voice pathology detection and discrimination based on modulation spectral features. IEEE Transactions on Audio, Speech, and Language Processing, 19(7), 1938–1948. https://doi.org/10.1109/tasl.2010.2104141

Pylypowich, A., & Duff, E. (2016). Differentiating the symptom of dysphonia. The Journal for Nurse Practitioners: JNP, 12(7), 459–466. https://doi.org/10.1016/j.nurpra.2016.04.025

Sellam, V., & Jagadeesan, J. (2014). Classification of normal and pathological voice using SVM and RBFNN. Journal of Signal and Information Processing, 05(01), 1–7. https://doi.org/10.4236/jsip.2014.51001

Syed, S. A., Rashid, M., & Hussain, S. (2020). Meta-analysis of voice disorders databases and applied machine learning techniques. Mathematical Biosciences and Engineering, 17(6), 7958-7979.

Syed, S. A., Rashid, M., Hussain, S., & Zahid, H. (2021). Comparative analysis of CNN and RNN for voice pathology detection. BioMed Research International, 2021, 1–8. https://doi.org/10.1155/2021/6635964

Syed, S. A., Rashid, M., Hussain, S., Imtiaz, A., Abid, H., & Zahid, H. (2021). Inter classifier comparison to detect voice pathologies. Mathematical Biosciences and Engineering: MBE, 18(3), 2258–2273. https://doi.org/10.3934/mbe.2021114

Teixeira, J. P., & Gonçalves, A. (2014). Accuracy of Jitter and Shimmer Measurements. Procedia Technology, 16, 1190–1199. https://doi.org/10.1016/j.protcy.2014.10.13

Teixeira, João Paulo, & Fernandes, P. O. (2015). Acoustic analysis of vocal dysphonia. Procedia Computer Science, 64, 466–473. https://doi.org/10.1016/j.procs.2015.08.544

Teixeira, João Paulo, Alves, N., & Fernandes, P. O. (2020). Vocal acoustic analysis: ANN versos SVM in classification of dysphonic voices and vocal cords paralysis. International Journal of E-Health and Medical Communications, 11(1), 37–51. https://doi.org/10.4018/ijehmc.2020010103

Teixeira, João Paulo, Fernandes, P. O., & Alves, N. (2017). Vocal acoustic analysis – classification of dysphonic voices with artificial neural networks. Procedia Computer Science, 121, 19–26. https://doi.org/10.1016/j.procs.2017.11.004

2023 by the authors; EuoAsian Academy of Global Learning and Education Ltd. Pakistan. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).