CLASSIFICATION OF NORMAL AND ABNORMAL VOICES USING SPEECH PARAMETERS

Dishi Jain[1], Sattwik Sahoo[2], Debabiswa Panda[3] Dr. Balakrishnan P[4]

[1,2,3]U. G STUDENTS, School of Computer Science and Engineering, Vellore Institute of Technology-Vellore, Tamil Nadu, India

[4]Dr. Balakrishnan P: Professor, Dept. of Analytics, Vellore Institute of Technology-Vellore, Tamil Nadu, India

---***---

Abstract - *The examination of the human voice has developed as a significant zone of study for its different applications in medication just as building sciences. Voice analysis essentially manages the extraction of certain parameters from voice signals for processing of voice in respective applications by utilizing appropriate techniques. Disease detection utilizing voice analysis is an indispensable research topic in clinical science. It additionally assists with identifying the illness at its previous stage. Voice features contain data on the wellbeing of the voice track and of the organs coordinating in speech creation. These features present a particular voice and might be used to differentiate the voice of people who are in good health or bad. At first, the voice features are mined utilizing time-domain analysis or some other and then these features are classified into groups using various classification techniques. In this paper, the voice analysis is done using the speech parameter by KNIME and PRAAT and decision tree is predicted after classification of Parkinson dataset.*

Key Words: pathological voice analysis, voice feature extraction, time domain, frequency domain, decision tree, machine learning approach

1. INTRODUCTION

Voice of an individual gives them a unique identity. Though there are billions of people yet each one of them has a different voice. Voice make us understand others feelings and also helps to convey one's ideas. Communication skills are one of the most required skill in any field for earning livelihood. Those who have speech disorder have to come across many problems in their daily lives.

The voice pathologic clutters are related to respiratory, nasal, neural and larynx sicknesses [13]. In this manner, investigation and analysis of vocal issues have become a significant clinical strategy. This has inspired a great deal of research in voice analysis measures for developing models to evaluate human verbal communication capabilities. Investigation of the voice signals is typically performed by the extraction of acoustic parameters utilizing computerized signal processing techniques. These parameters are investigated to decide the specific attribute of the voice.
Likewise, voice is a period fluctuating sign that has all the signal data present in now is the right time-variable characteristics. Voice scanning includes the change of voice signal into a lot of signals or a lot of parameters with a goal to disentangle the voice sign to extricate includes legitimately appropriate for various applications and to smother repetitive parts of the signal. These strategies help speech pathologists and clinical experts to consider and determine the voice signs to have ease. Voice examination should be possible either in the time-space or in the frequency area. The time-space waveform of voice signal conveys all the sound-related data like phonemes, syllables, prosody levels, and so forth while the frequency area speaks to the phantom characteristics of the voice, for instance, frequency, pitch, etc.

Table 1 shows data that can be extracted from voice parameters by time-frequency based analysis. Speech Recognition innovation is a prestigious clinical use of voice analysis [13].

Table -1: Time frequency based extracted feature [13]

PARAMETERS	FEATURES	INFORMATION EXTRACTED
Frequency	Jitter (local, absolute, ppq5, ddp)	Variation in the pitch of voice
Pulse	Standard deviation of period.	Speech rate of speaker
Amplitude	Shimmer (local, db, dda, apq11, apq5, apq3)	Variation in loudness of voice.
Voice	Unvoiced frames, voiced frames.	Structure of voice.
Harmonic	Noise- to- harmonic, Harmonic-to-noise	Relative highs or lows of voice
Pitch	Maximum or Minimum pitch, Mean or Median pitch	Peaks of the sound spectrum of voice

2. LITERATURE SURVEY

While studying the field of pathological voice investigation, we find numerous strategies. In the paper [1], a new approach for voice screening has been proposed. Nonetheless, to examine a lot of information is a tough task. Consequently, computer-based analytical instruments for investigation and characterization of information can be helpful in diagnostics [3]. This research work manages the classification of diseases utilizing the neural network and the classifier introduced has 80–85% of precision. This paper [4] assisted with discovering the dataset that we can use for our characterization, for example, Parkinson's sickness dataset. Using the Parkinson dataset utilizing remarkable ML devices proceeded with vowels that are found to convey more PD-discriminative information. In the wake of examining the dataset, proceeded with vowels were found to convey more PD-discriminative information in contrast with brief sentences or words do. To evaluate how good the central tendency and scattering measurements be an extraordinary alternative, all things considered, few blends of these measurements have been attempted and have found that addressing the instances of a subject with the customary mean and standard deviation progresses the theory of prescient model.

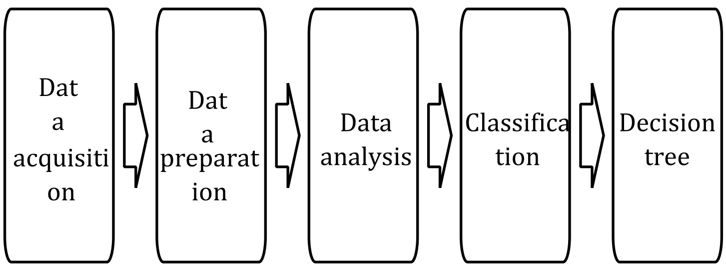
Then to check the presence of a relationship between age, time of device use, voice discovery limit, hearing class score, and language classification score with acoustic information of voices of cochlear embedded children. Fifty-one kids extending in age from 3 years to 5 years and 11 months who singularly utilized cochlear implants took part. Acoustic analysis of the supported vowel/a/, consecutive speech and unconstrained speech was performed. The outcomes were connected with segment information and hearing test outcomes. Kids with a more terrible voice recognition threshold demonstrated higher recurrence in the continued vowel ($p \leq 0.001$) and in the unconstrained discourse ($p \leq 0.005$). For the evaluated cochlear implanted kids, a relationship between voice recognition threshold and the rate of recurrence values for the continued vowel and furthermore in the unconstrained speech was there. This clearly shows the more awful the voice detection threshold is, the shriller is the kid's voice [8]. Other acoustic parameters didn't associate with the other evaluated factors. The acoustic analysis evaluates the sound sign and gives enough documentation to plot the benchmark of the person's voice; the acoustic parameters of the basic recurrence and their unsettling influence records, alongside estimations of noise, have significant clinical implications. Among all investigated acoustic parameters, just the central frequency demonstrated a relationship with the surveyed demographics and hearing viewpoints. Seeing the value of p, this correlation stands out. The acoustic voice analysis framework (VArt) was created utilizing a handheld device running Android OS [9]. The progressions were done to the principal frequency extraction calculation and planned an instinctive UI representing a new hoarseness index, which is a subordinate of the harmonics-to-noise proportion created by Kojima and Shoji (Ra2). The strategy includes VA running on a Windows PC and VArt running on two sorts of handheld devices in a sound-treated room or in a clinical diagnostic room. Intraclass correlation were determined for the two frameworks under the two conditions and the outcome shows all comparisons were exceptionally correlated. Estimations got utilizing this recently created VArt were profoundly consistent with those utilizing VA, demonstrating high reliability. Additionally, the new framework expands the clinical plausibility of acoustic voice analysis.

The literature on acoustic measures of voice in depression has also been reviewed [10]. Writers have isolated outcomes gotten from studies of programmed speech, for example, tallying or perusing, from free speech. Free discourse requires subjective activity like word-finding plus the motor activity of automatic speech. Likewise, results would be less unclear if consistent groups of retarded depressed patients were measured. 10 women and 12 men aged depressed patients and an age-matched similar patient were studied. Patients were then grouped. The outcome from measures of fluency and prosody are defined. Thus, the depressed patients showed less prosody than normal ones. Enhancement in the retarded group was mirrored in short-term pauses but no extended words. Clinical imprints are generously identified with acoustic parameters. Temporary changes related with depression seem to show the discouraged state though prosodic highlights have shown depressed characteristic. Acoustic measures of the patient's discourse give objective dealings to help in the assessment of depression. Another article that presents similar study in the field of pathological voice analysis focusing more on parametrization techniques. As a unique work, this paper [11] presents 36 totally new pathological voice measures. To test the importance of these characteristics according to sensitivity, accuracy and specificity, the 3 databases in English, Czech and Spanish which contains data related to pathological voice were used. The method based on composite feature extraction and strong testing performed better than all works that are published already in this field. Our work also follows the same idea of feature extraction from the voice for analysis.

3. ALGORITHM FOR PROPOSED ARCHITECTURE

Using Machine learning we will be able to classify the data by determining the coefficients on which the property of voice depends.

Here, we follow the same rule as machine learning.



4. TOOLS USED

PRAAT:

'Praat' is a computer software with the help of which phoneticians can analyze, synthesize, and manipulate speech, and create high- quality pictures for articles and theses. It consists of functions for speech analysis, speech synthesis, learning algorithms, speech manipulation, listening experiments, labelling and segmentation and many more. For this paper we have used praat software to create a database containing jitter, shimmer, voice breaks, Average pitch.

There are many Praat manuals which are designed for software documentation and assumes a strong phonetics or programming background of readers. The current manual is compiled from a variety of detailed manuals with a special focus on those most-frequently used functions and techniques for acoustic analysis. The target readers of this are those beginners who are not equipped with a strong phonetics or programming background but want to do some phonetic analysis of speech sounds. The clear visual presentation of operational procedures and introduction to acoustic knowledge are provided to facilitate the use of Praat in linguistic research.

Praat has a user friendly and easy to use interface, various default options which you can learn from search manual and you can analyse, manipulate and label data.

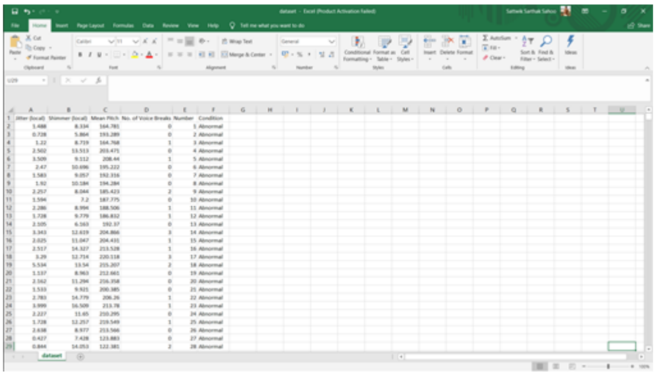


Fig -1: database with labelled values

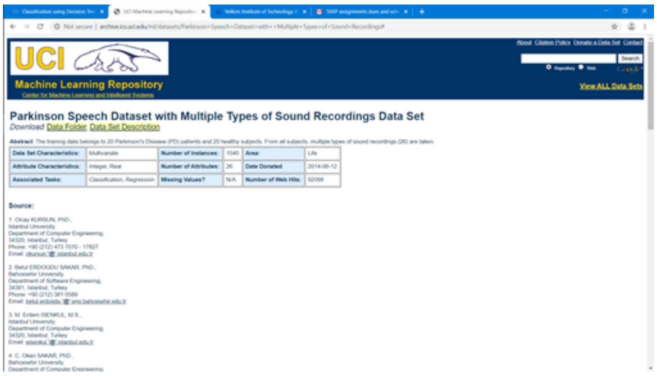


Fig -2: parkinson speech dataset.csv

KNIME:

KNIME, stands for Konstanz Information Miner which is free and open-source data analytics, reportage and integration platform. KNIME assimilates several components for data mining and machine learning through its integrated data pipelining concept. A graphical user interface and using the database JDBC allows assembly of nodes combining unlike data sources, together with preprocessing which includes Extraction, Transformation, for modeling, Loading, data analysis and data visualization with minimal, programming. KNIME can also be used as a SAS substitute.

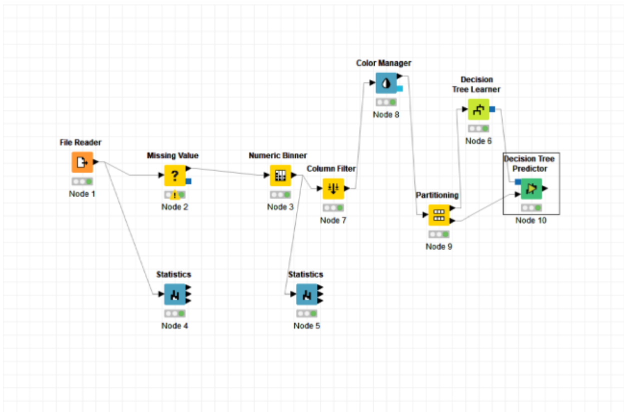


Fig -3: The workflow model

5. RESULTS AND DISCUSSION

Data acquisition: Here, the data is acquired and selected using the file reader node and used for further analysis.

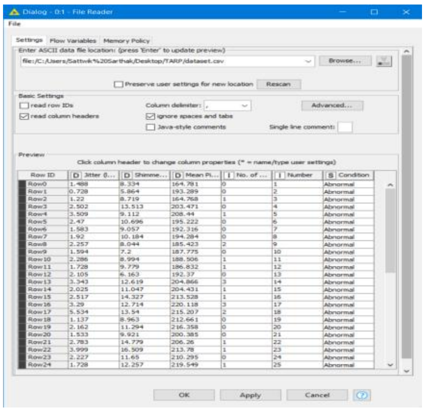


Fig -4: Acquired data

Data preparation: Here after acquiring the data it is checked for missing values then passed onto the numeric binner, here we bin the data into two parts, normal and abnormal as you can see from the excel sheet values till 5 are abnormal. Values after 5 are normal hence using this we created two bins.

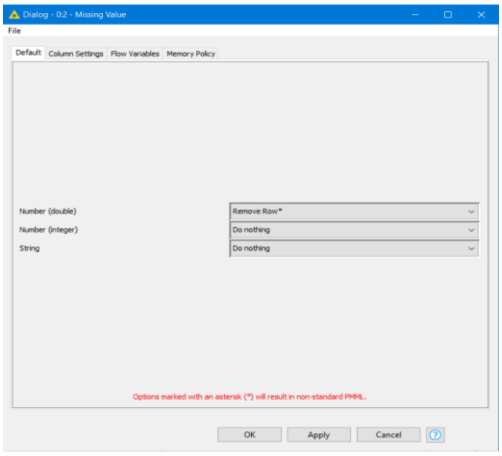


Fig -5: missing value

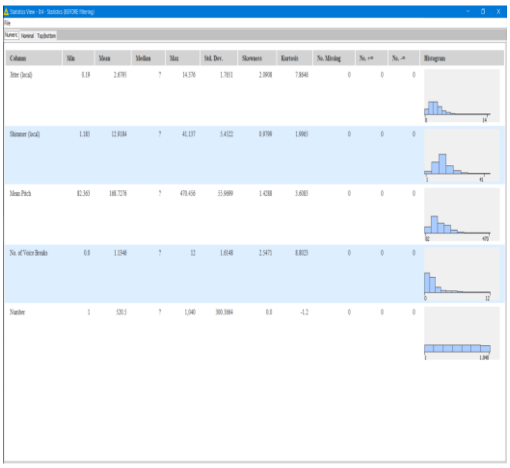


Fig -6: statistics view (before filtering)

Data Analysis: The prepared data is analyzed based on the parameters.

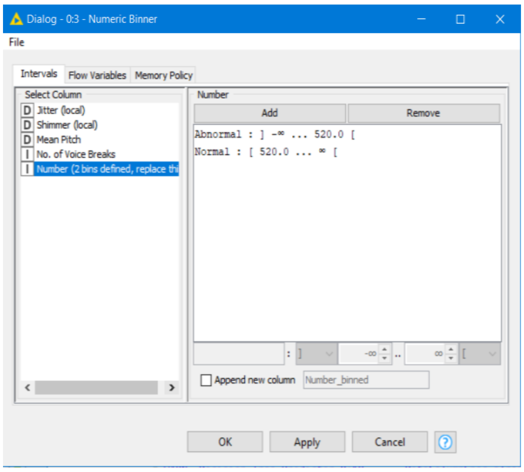


Fig -7: Name of the figure

Classification: Classification of data is done by passing it through Partitioning and Decision Tree Learner Node. The input table is split into two partitions (i.e. row-wise), e.g. train and test data. The two partitions are presented at the two output ports.

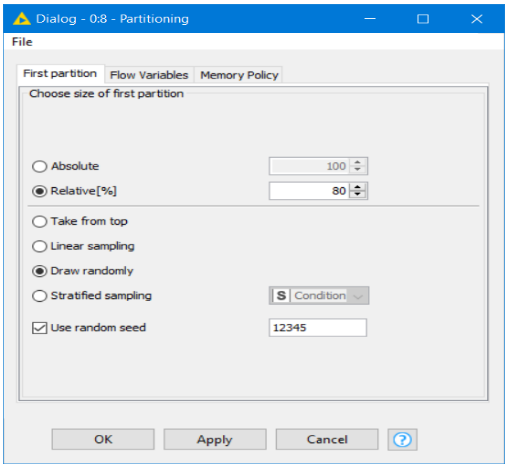


Fig -8: partitioning (training and testing data)

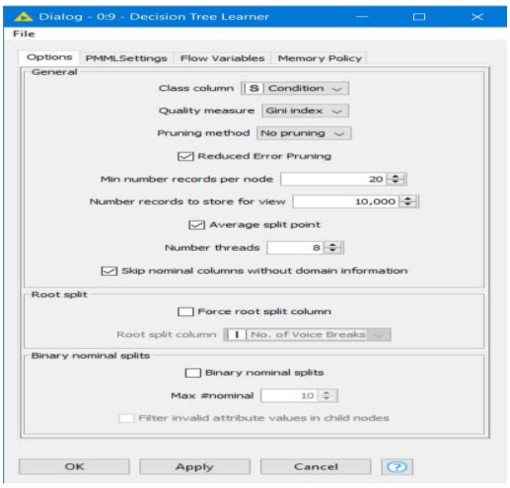


Fig -9: decision tree learner

Training data: Here, the data is passed through the decision tree learner to train the working model to give the appropriate results. Here the partitioning allows only a certain amount of data into decision tree learner and predictor.

Decision tree predictor: The Decision tree predictor gives the most suitable results to classify data depending upon the amount of input give so on and so forth.

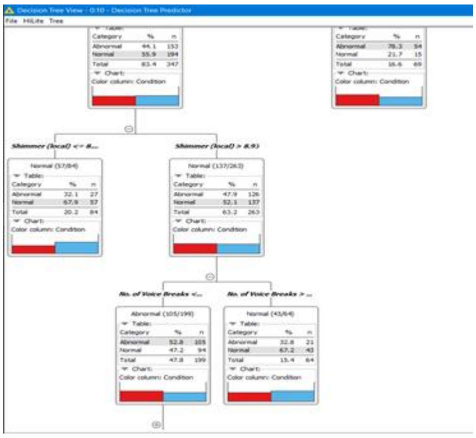


Fig -10: predicted decision tree

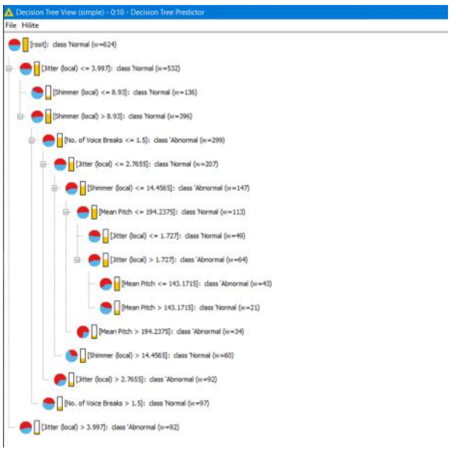


Fig -11: predicted decision tree

Fig -12: predicted decision tree

OUTPUT:

After each learning we can see the most suitable values with which we can classify the given data using KNIME.

For relative 40%:

Here as per the algorithm breaks have been used to classify the following data. The more the data the more variables come into play.

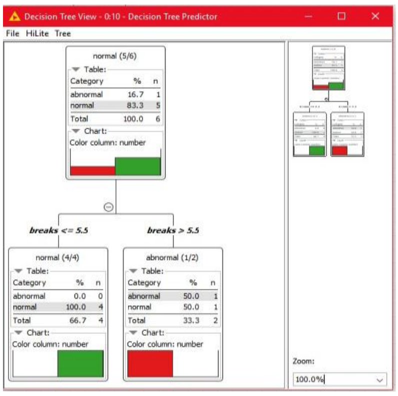


Fig -13: classified data based on breaks in voice

For relative 80%:

Here depending upon the algorithm, the data has been classified using shimmer as the more suitable way possible at the value of 19.983.

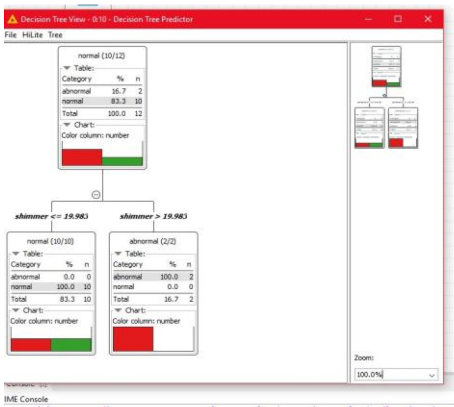


Fig -14: classified data based on shimmer

Hence after large amount of data and a greater number of iterations the values play a role in classification of the human voice to be normal or abnormal. Hence, the decision tree prediction.

6. CONCLUSION

The purpose of writing this paper was to express the importance of voice analysis frameworks in the pre-analysis of certain ailments which later may change into deadly or serious ailments. Besides, another technique this voice analysis technique guarantees high precision, patients ease, affordable and less tedious in foreseeing side effects at a beginning period of any disease is the need of great importance. Despite the fact that this is a small branch of biomedical science, which manages an assortment of techniques for sickness related voice findings, it is presently pulling in the consideration of numerous scientists and health care services. It opens new skylines for the progression of voice diagnosis frameworks. As it appears to be present, the procedure of advancement of biomedical innovation has just started and the manner in which it is sustaining mankind looks encouraging for the people in the future.

ACKNOWLEDGEMENT

The authors would like to thank Prof. Balakrishnan P for his guidance and support throughout the project and successfully completing the manuscript.

REFERENCES

[1] Verde, Laura et al. "A Methodology for Voice Classification Based on the Personalized Fundamental Frequency Estimation". Biomedical Signal Processing and Control, vol 42, 2018, pp. 134-144. doi:10.1016/j.bspc.2018.01.007.

[2] Boyanov, B., and S. Hadjitodorov. "Acoustic Analysis of Pathological Voices. A Voice Analysis System for The Screening of Laryngeal Diseases". IEEE Engineering In Medicine And Biology Magazine, vol 16, no. 4, 1997, pp. 74-82. doi:10.1109/51.603651.

[3] Nayak, J. et al. "Classification and Analysis of Speech Abnormalities". ITBM-RBM, vol 26, no. 5-6, 2005, pp. 319-327. doi:10.1016/j.rbmret.2005.05.002.

[4] Sakar, Betul Erdogdu et al. "Collection and Analysis of A Parkinson Speech Dataset With Multiple Types Of Sound Recordings". IEEE Journal of Biomedical and Health Informatics, vol 17, no. 4, 2013, pp. 828-834. doi:10.1109/jbhi.2013.2245674.

[5] Saloni, N.A. et al. "Disease Detection Using Voice Analysis: A Review". International Journal of Medical Engineering and Informatics, vol 6, no. 3, 2014, p. 189. doi:10.1504/ijmei.2014.063173

[6] Umapathy, K. et al. "Discrimination of Pathological Voices Using A Time-Frequency Approach". IEEE Transactions on Biomedical Engineering, vol 52, no. 3, 2005, pp. 421-430. doi:10.1109/tbme.2004.842962.

[7] Demircan, S., and H. Kahramanlı. "Feature Extraction from Speech Data for Emotion Recognition". Journal of Advances in Computer Networks, vol 2, no. 1, 2014, pp. 28-30. doi:10.7763/jacn.2014.v2.76.

[8] Coelho, Ana Cristina et al. "Hearing Performance and Voice Acoustics of Cochlear Implanted Children". Brazilian Journal of Otorhinolaryngology, vol 82, no. 1, 2016, pp. 70-75. doi:10.1016/j.bjorl.2015.11.002.

[9] Fujimura, Shintaro et al. "Real-Time Acoustic Voice Analysis Using A Handheld Device Running Android Operating System". Journal of Voice, 2019. Elsevier BV, doi:10.1016/j.jvoice.2019.05.013.

[10] Alpert, Murray et al. "Reflections of Depression in Acoustic Measures of The Patient'S Speech". Journal of Affective Disorders, vol 66, no. 1, 2001, pp. 59-69. Elsevier BV, doi:10.1016/s0165-0327(00)00335-9.

[11] Mekyska, Jiri et al. "Robust and Complex Approach of Pathological Speech Signal Analysis". Neurocomputing, vol 167, 2015, pp. 94-111. Elsevier BV, doi:10.1016/j.neucom.2015.02.085.

[12] Teixeira, João Paulo et al. "Vocal Acoustic Analysis – Classification of Dysphonic Voices with Artificial Neural Networks". Procedia Computer Science, vol 121, 2017, pp. 19-26. Elsevier BV, doi:10.1016/j.procs.2017.11.004.

[13] Dixit, Dixit et al. "Voice Parameter Analysis for The Disease Detection". IOSR Journal of Electronics and Communication Engineering, vol 9, no. 3, 2014, pp. 48-55. IOSR Journals, doi:10.9790/2834-09314855.

[14] Hammami, I. et al. "Voice Pathologies Classification and Detection Using EMD-DWT Analysis Based on Higher Order Statistic Features". IRBM, 2020. Elsevier BV, doi:10.1016/j.irbm.2019.11.004.

[15] Al-Nasheri, Ahmed et al. "Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions". IEEE, vol 6, 2018, pp. 6961-6974. doi:10.1109/access.2017.2696056.