iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: V Month of publication: May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5087

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Pathological Voice Detection from Speech Signal using SVM

Manoj Patil[1], Rushikesh Rane[2], Aadesh Thange[3], Pratik Gite[4], Mrs. Dhanashree Phalke[5], Mr. Jayant Shimpi[6]
[1, 2, 3, 4, 5, 6]*Dept. of Computer engineering D. Y. Patil College of engineering Pune, Maharashtra, India*

***Abstract:** Speech processing (Discourse preparing) is one of the hotly debated issues for the specialist as it is utilized in the Identification and Classification of voice pathology. Discourse is unquestionably the most normal correspondence mode which people use to interface with one another. The nonappearance or disintegration of such a media will represent an extraordinary risk to appropriate comprehension between people. Identifying the variation from the norm from voice signals is as yet a difficult territory of research. Efforts are being made to plan a programmed, fast, practical and non-obtrusive acknowledgment framework which will separate such pathologic examples from those of the ordinary ones. This paper gives the various endeavors made by the scientists to recognize acoustic highlights of discourse principally to segregate typical voices from obsessive voices.*
***Keywords:** Glottal Pathology, Glottal Features, Support Vector Machines (SVM), 12 Mel-Frequency Filter Bank Cepstral Coefficients (MFCC), Zero-Crossing Rate (ZCR).*

I. INTRODUCTION

Speech is one of the fundamental mediums of communicating known to mankind since the beginning of mankind. The absence or deterioration of such a media will pose a great threat to proper understanding between individuals. Such difficulty may arise due to the malfunction of the human speech production system or auditory system. In this work, we center around the issue caused in vocal folds, as their development which contributes towards enormous, in the creation of voice. Any pathology that occurs to alter the periodic movements of the vocal folds affect speech produced. Different methods are used by researches to find voice pathology with different features like speech signal is analyzed by the acoustic parameters like Signal Energy, pitch, Silence removal, Windowing, Mel frequency Cepstrum, and Jitter. Some of the used techniques are mentioned in section II.

Here an attempt is made to analyze and to differentiate pathological voice from normal voice using data mining techniques like Support Vector Machine (SVM). We conducted cross-validation experiments on The Saarbruecken Voice Database using support vector machines (SVM) for the classification of normal and pathological voices. The speech signal is analyzed to extract the acoustic parameters such as 12 Mel-Frequency Filter Bank Cepstral Coefficients (MFCC) and zero-crossing rate (ZCR). The system gives promising accuracy in the detection of Glottal Pathology.

II. LITERATURE REVIEW

Mythili J et al.[1] use the discriminative features in the voice signals to detect the pathologically affected voices. Here Ensemble learning technique is used to find the types of disorder invoices. The first component is the extraction of feature vectors using Mel-frequency cepstral coefficients (MFCC), Linear predictive coding (LPC), Wavelet Packet Decomposition (WPD), Cepstral Analysis (CA), Jitter, Shimmer, Pulse, Pitch, Hormonicity, Intensity, Energy, and entropy methods. The classification of feature vectors is done using ensemble learning methods.

Pranav S. Deshpande et al. [2] introduced as a method for reliable detection of glottal instants and EGG parameters from an electroglottographic (EGG) signal composed of voiced and non-voice segments. An adaptive variation mode decomposition-based algorithm is used or suppressing low-frequency artifacts and additive high-frequency noises. Depending on the center frequency criterion, the proposed system first constructs a candidate EGG feature signal for the determination of glottal closure and opening instants.

In the second stage, the candidate glottal instants are determined by detecting the positive and negative zero crossings in normalized candidate EGG feature signals, respectively. Finally, an autocorrelation features based post-processing algorithm is presented to reject non-glottal instants from the non-speech production segments.

N.A. Sheela Selvakumari et al.[3] analyze conversation signal by the acoustic guidelines and variables like transmission energy, pitch, Silence removal, Windowing, Mel consistency and occurrence Cepstrum, and Jitter. In the end, the classification strategy i.e Support Vector Machine is employed to classify the standard and pathology speech, predicated on the features extracted in the last phase. A successful computerized Voice Pathology Detector was designed, which can help the speech therapist to investigate and detect the patient's tone of voice.

Shweta et al., [5] present an LDB algorithm for audio classification. This helps to achieve high classification accuracy. LDB also uses simple dissimilarity measures for selecting the nodes and features. A database of 213 audio signals was used.LDB performs well for all signals while MFCC works well for music. A combination of MFCC and LDB also gives promising results.

A. Support Vector Machine (SVM)

Anis Ben Aicha et al.[6] proposed an automatic detection method of premalignant lesions based on human voice production theory. The non-invasive process based on the recorded speech is used for experiments that get an accuracy of 92%. The source signal is first extracted namely the glottal flow signal from the acoustic speech using the IAIF technique. The relevance of used descriptors using boxplot and PCA analysis is also done. SVM module is used to classify and discriminate premalignant lesions from normal voices.

In [7] the classification techniques like Primary Component Evaluation i.e PCA and Support Vector Machine are used to classify the standard and pathological speech predicated on the 27 features confirm extracted from the tone of voice or speech dataset.

V. Srinivasan et al., [8] explored a method of finding the ability of acoustic parameters in discrimination of normal voices from pathological voices that were analyzed and classified.

The classification of pathological voice from a normal voice was implemented using a support vector machine (SVM) [10] and the classifiers were trained and tested. The dataset was recorded by speech utterances of a set of Tamil phrases containing speech samples of 10 distinct subjects (5 normal, 5 pathological children). The speech signals were analyzed and were extracted. A Genetic Algorithm (GA) based feature selection has improved the classification accuracy of this work. The support vector machine shows better performance in terms of classification accuracy.

B. Radial Basis Functional Neural Network (RBFNN)

V. Sellam, J. Jagadeesan, et al.[11] provides a classification of pathological voice from normal voice using Support Vector Machine (SVM) and Radial Basis Functional Neural Network (RBFNN) with the dataset of Tamil phrases. The voice features like Signal Energy, pitch, formant frequencies, Mean Square Residual signal, Reflection coefficients, Jitter and Shimmer are taken into consideration to detect voice disorders in children.

Ashish et al.[12] build a Glottal pathology detection system speech processing technology. Here Multilayer Neural Network is used to classify the cancerous voice from that of normal voice with various combinations of features like jitter local, jitter local absolute, jitter PPQ, shimmer local absolute, shimmer relative, shimmer in dB and pitch periods. A speech database collected from a prominent cancer hospital is used for training in a supervised learning model for the classification of cancerous voice from that of a normal voice.

C. Mean- Area-Peak-Value (MAPV)

L Gavidia-Ceballos and J H Hansen [14] used speech production parameters as complete glottal closure is very hard to obtain in vocal fold pathology. The method proves advantageous in the estimation of Enhanced Spectral Pathology Component (ESPC) instead of glottal flow waveform, which varies considerably between pathology and health conditions. The meaningful measures Mean- Area-Peak-Value (MAPV) and Weighted-Slope (WSLOPE) were derived from the ESPC feature.

A voice contour of the voice signal was used to discriminate between normal and pathological samples in [15], because pathologies in the vocal folds cause the voice to weaken and fluctuate.

This method may not be successful if the recording of the voice is done in a different environment or by a diverse microphone. In [16], the authors used an information theory-based technique, called correntropy, to distinguish between normal and pathological voices. The accuracy was around 97%.

Zhong et al. proposed a vocal fold damage detection method using a type-2 fuzzy classifier, where the input voice was transformed by a short-time Fourier transform and singular value decomposition [17]. The whole method was deployed in a heterogeneous sensor network.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

III. PROPOSED SYSTEM

Pathology voice classification model which detects the pathologically defective voice signals accurately. The implementation of statistical methods through the ensemble learning concept is used to identify and predict the normal and abnormal voices precisely. A difficulty may arise due to the malfunction of the human speech production system or auditory system. Below figure shows the block diagram of the proposed system.

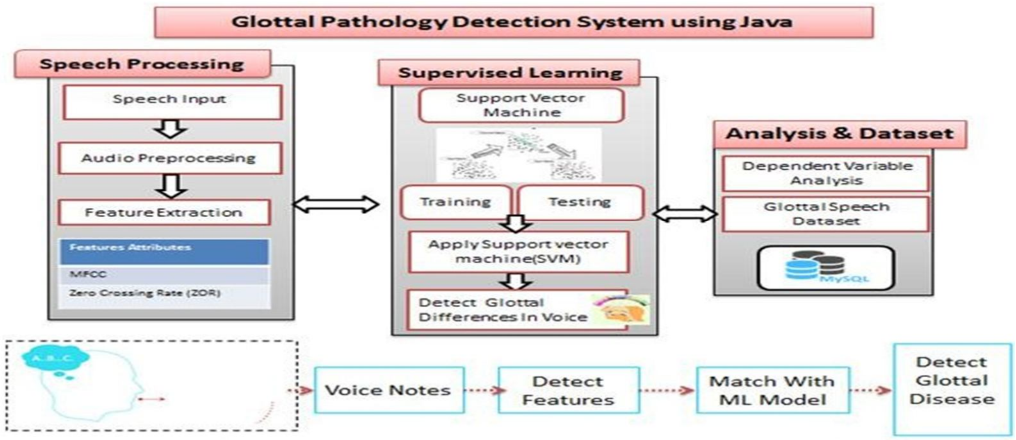

Figure 1: - System Architecture Diagram

The fundamental point of this framework is to assist patients with neurotic issues for checking their advancement throughout voice treatment. As of now, patients are required to routinely visit an expert to catch up their advancement. Besides, the current approaches to distinguish the voice pathology are abstract, intrusive techniques, for example, the immediate review of the vocal folds and the perceptions of the vocal creases by endoscopic instruments are finished. Also the current frameworks are typically founded on data identified with the vocal tract arrangement, the air-flow going through the vocal creases, and called glottal flow. These systems are costly, dangerous, tedious, distress to the patients and require exorbitant assets, for example, unique light sources, endoscopic instruments and concentrated camcorder hardware. So as to maintain a strategic distance from the above issues, a strong framework is executed to distinguish vocal crease pathology at a beginning period from set of highlights like 12 Mel-Frequency Filter Bank Cepstral Coecient (MFCC) and zero intersection rate (ZCR) got from basic voice test. The framework helps the clinicians and language teachers for early location of vocal overlay pathology and can improve the exactness of the evaluations utilizing Saarbruecken Voice Database. The proposed framework can break down voice source includes in discourse information for identification of glottal pathology utilizing information mining methods. SVM Classier is created for different element blends to characterize the glottal pathologic voice from ordinary voice. The framework is executed utilizing 12 Mel-Frequency Filter Bank Cepstral Coecients (MFCC) and zero intersection rate (ZCR).

IV. MATHEMATICAL MODEL

Let us consider S be a Systems such that , S= {U,FT,P S,T,Ss,Dsg,}
Where,

A. U= {U1, U2, U3 ....Un |`U' is a Set of all Users} There may be number of users for making use of system. So this is the Innite Set.
B. I = {I | is the speech input of the system }
C. FT= {F1, F2,F3 ...... Fn | F are the features extracted }

FT is the feature extracted from speech input.

D. { T=f T is the technique used to process Input Voice}
E. P is the technique used to preprocess the input image.
F. T is the technique used for detecting glottal difference from voice.
G. SS = {S REG, S LOGIN,| S Feature Dataj SS is a Set of Storage Service} STORAGE SERVER will provide four services like Registration, Login, and values of P,Q,R,S,T parameters As this set also has nite attributes, so this is also Finite Set.
H. DS = {Train Data | DS is a Set of data table for permanent storing of data on server }

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

1) *Event 1:* User will make registration on SYSTEM & Storage Server. Let f(U) be a function of User
   Thus, f(U)$\rightarrow$ fSs}
2) *Event 2:* Speech Input is given to the system. Let f(U) be a function of User.
   Thus, {(U) $\rightarrow$ {I1,I2,I3... In} ε S }
3) *Event 3:* Pre Processing is performed on Input to remove noise. Let f(P) be a Audio processing function.
   Thus, {(S) $\rightarrow$ fP | I} ε P
4) *Event 4:* Let f(FT) be the feature extraction function of the system. Let f(S) is a function of the System
   Thus, f(S) = {FT1, FT2, FT3. . . . . . FTn} $\rightarrow$ I
5) *Event 5:* Let f(E) be an function used to classify glottal pathologic voice using SVM classier.
   Let f(R) be a Detection function Thus, f(S) $\rightarrow$ fEg ε T

V. RESULT ANALYSIS

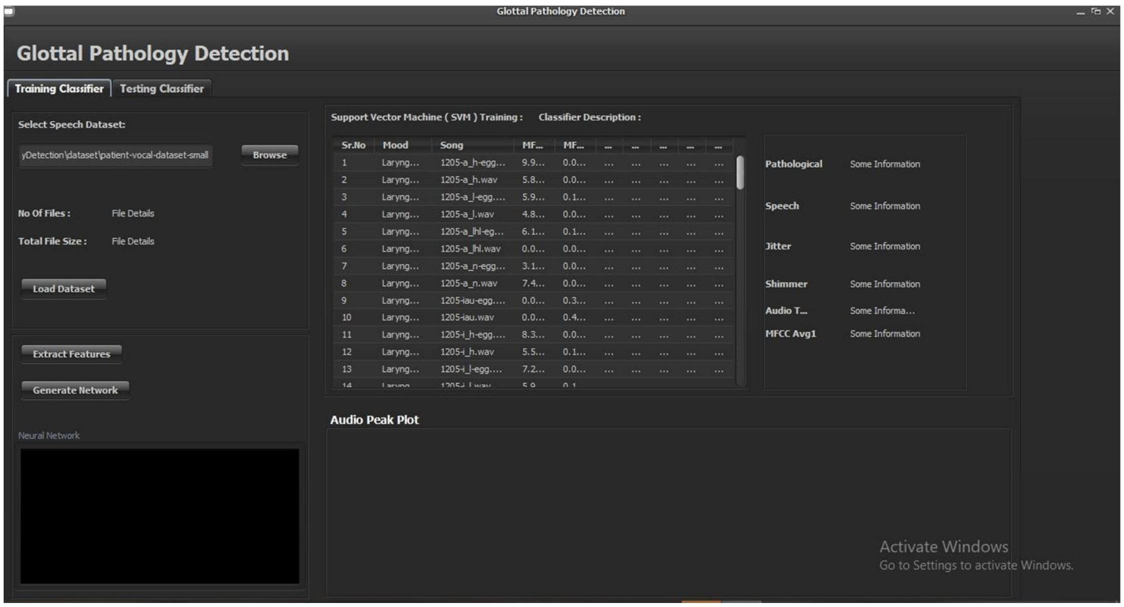

Figure 2: Training Classifire

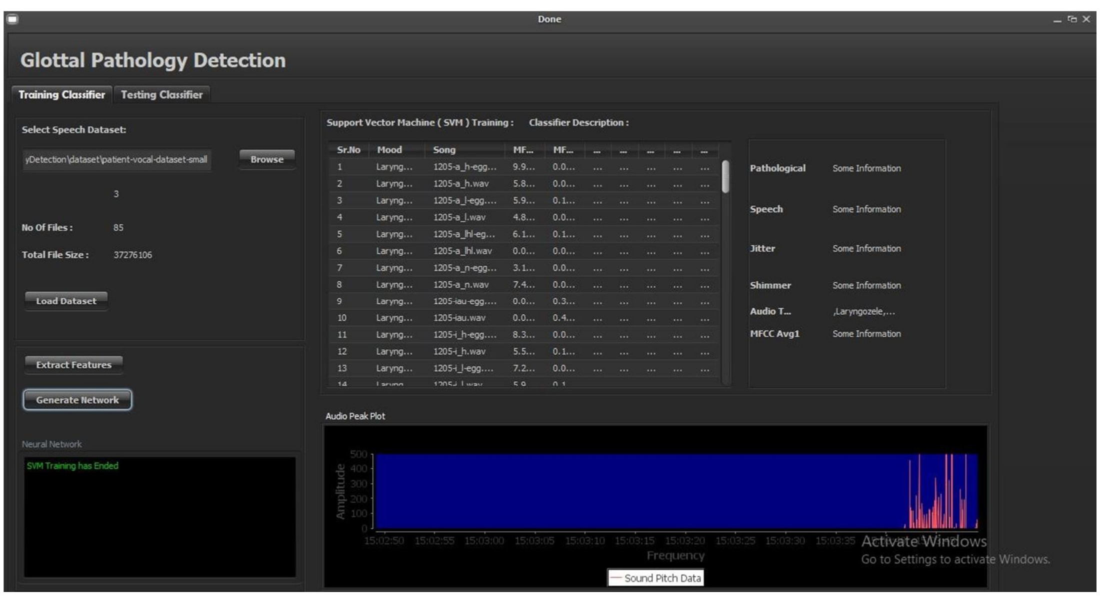

Figure 3: Trained SVM

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

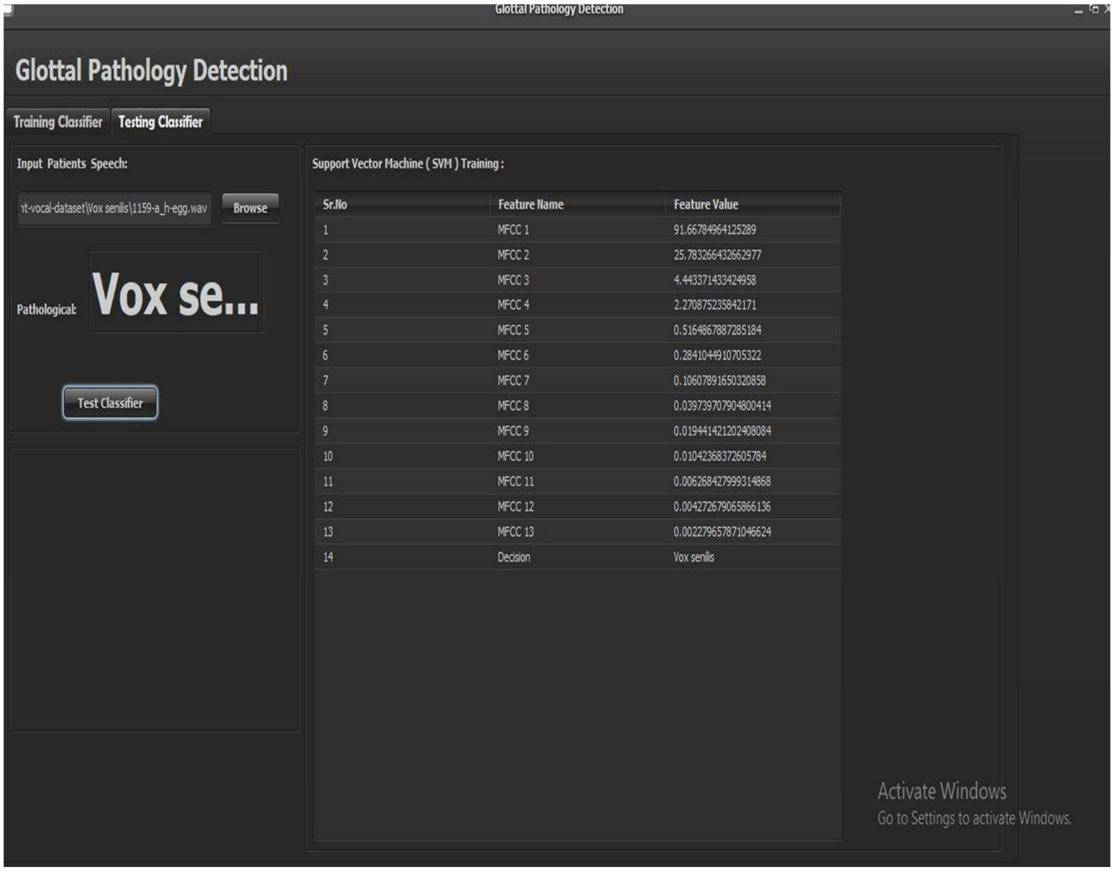

Figure 4. Glottal Result {Testing Classifire}.

VI. CONCLUSION

Short time energy adds considerable value in making correct decision as it is the reflection of amplitude variations. This paper gives the survey of different techniques used for tracking the objects and abstract view of the system that we are going the implement to solve the issued related to the existing system.

An attempt is made to implement a system that can give a low cost and quick solution by analyzing voice source features like 12 Mel-Frequency Filter Bank Cepstral Coefficients (MFCC) and zero-crossing rate (ZCR) in speech data for detection of glottal pathology

REFERENCES

[1] Mythili J, Vijaya MS,"Pathology Voice Detection and Classification Using Ensemble Learning",International Journal of Engineering Science Invention (IJESI),2018.

[2] Pranav S. Deshpande ; M. Sabarimalai Manikandan ,"Effective Glottal Instant Detection and Electroglottographic Parameter Extraction for Automated Voice Pathology Assessment", IEEE Journal of Biomedical and Health Informatics (Volume: 22 , Issue: 2 , March 2018.

[3] N.A. Sheela Selvakumari ; V. Radha ,“ A voice activity detector using SVM and Naïve Bayes classification algorithm ”, 2017 International Conference on Signal Processing and Communication (ICSPC)

[4] D. Pravena, S. Dhivya and A. Durga Devi, “Pathological Voice Recognition for Vocal Fold Didease,” International Journal of Computer Applications (0975-888), Vol. 47, No. 13, 2012.

[5] Shweta Vijay Dhabarde, P.S.Deshpande, “Feature Extraction and Classification of Audio Signal using Local Discriminant Bases”, International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-3, Issue-5, May-2015.

[6] Anis Ben Aicha,"Noninvasive Detection of Potentially Precancerous Lesions of Vocal Fold Based on Glottal Wave Signal and SVM Approaches",International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2018, 3-5 September 2018, Belgrade, Serbia.

[7] C. Pend, Q. J. Xu, B. K. Wan and W. X. Chen, “Pathological Voice Classification Based on Features Dimension Optimization,” Transactions of Tianjin University, Vol. 13, No. 6, 2007

[8] V.Srinivasan, V.Ramalingam, V.Sellam, “Classification of Normal and Pathological Voice using GA and SVM”, International Journal of Computer Applications (0975 –8887) Volume 60–No.3, December 2012.

[9] M.Farra, J. Hernando and P. Ejarque, “Jitter and Shimmer Measurements for Speaker Recognition,” TALP Research Center, Department of Signal Theory and Communications, Universitat Politecnica de Catalunya, Barcelona.

[10] Maria Markaki,Student Member, IEEE,and Yannis Stylianou,Member, IEEE,"Voice Pathology Detection and Discrimination based on Modulation Spectral Features",october, 2010. Aicha Bouzid and Noureddine Ellouze,"GLOTTAL OPENING INSTANT DE- TECTION FROM SPEECH SIGNAL".
[11] V. Sellam, J. Jagadeesan,"Classification of Normal and Pathological Voice Using SVM and RBFNN" ,Journal of Signal and Information Processing, 2014, 5, 1-7, 2014.
[12] Ashish Gajanan Nanoti , Jagannath Nirmal ,"Glottal Pathology Detection with Vocal Features for Developing Laryngeal Cancer Screening Tool Using Neural Network", Conference: International Symposium on Acoustics for Engg. Applications Organised by Acoustical Society of India (ASI) and KIIT College of Engineering, Gurgaon, At Gurgaon,
[13] Volume: NSA031
[14] CHRISTOPHE DALESSANDRO and NICOLAS STURMEL, "Glottal closure instant and voice source analysis usingtime-scale lines of maximum amplitude",Sadhan VOL. 36, Part 5, October 2011, pp. 601–622.c, Indian Academy of Sciences.
[15] L Gavidia-Ceballos and J H Hansen,"Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection". IEEE transactions on bio-medical engineering, 43(4):373–83, April 1996.
[16] Ali Z., Alsulaiman M., Elamvazuthi I., Muhammad G., Mesallam T.A., Farahat M., Malki
[17] K.H. Voice pathology detection based on the modified voice contour and SVM. Biol. Inspir. Cogn. Archit. 2016;15: 10–18. doi: 10.1016/j.bica.2015.10.004.
[18] Fontes A.I.R., Souza P.T.V., Neto A.D.D., Martins A.D.M., Silveira L.F.Q. Classification System of Pathological Voices Using Correntropy. Math. Probl. Eng. 2014;2014:924786. doi: 10.1155/2014/924786.
[19] Zhong Z., Zhang B., Durrani T.S., Xiao S. Nonlinear signal processing for vocal folds damage detection based on heterogeneous sensor network. Signal Process. 2016;126:125–133. doi: 10.1016/j.sigpro.2015.08.019.