Potential Feature Extraction using Meyer Wavelet Transform and Dimensionality Reduction using Sequential Forward Feature Selection for Voice Pathology Detection

P. Kokila[1](Corresponding Author), Dr. G. M. Nasira[2]
[1]*Research Scholar, Department of Computer Science,*
[2]*Professor & Head, Department of Computer Applications,*
Chikkanna Government Arts College, Tirupur-641602, Tamil Nadu, India.

Abstract - To assess the voice disorder, automatic voice pathology identification contributes more to diagnose them in earlier stages from which patients suffer and they indicate health-related complications. This paper addresses the feature extraction and feature selection process of voice data when they are voluminous features available in speech signals. This work aims at developing a precise and strong feature extraction using Meyer Wavelet Transform for detection of voice pathology by exploring different frequency band. In addition, from the extracted features, the potential features which contributes more in detection of voice pathology is determined by applying sequential forward feature selection approach which handles the voluminous dataset more accurately and overcomes certain weaknesses in existing methods. The simulation results are also proved the performance of these proposed approaches with voice dataset of sustained vowel a/o/u for both normal and pathological voices and produces better results in detection of voice disorders.

Keywords - voice disorder, Meyer Wavelet Transform, feature extraction, Sequential Forward Feature Selection and pathology

I. INTRODUCTION

An important tool for humans to communicate with other is voice. Any troubles in voice quality has great influence on an individual's social and personal life [1]. Generally, voice disorders are termed as any time the voice does not perform, work or sound as it normally does, or delays during communication and quality of life [2,3]. Voice disorder is considered as most common language and speech disorders, troubling nearly 6% of children below 14 years of age, and 3–9% of the adult people [4]. Complete prevalence of hoarseness in studies in India ranges from 0.51–0.7% [5, 6, 7].

As a result, working on digital processing of signal on speech has been found to offer a noninvasive diagnostic method that is measured to be an active supplementary tool to doctors when recognizing voice disorders, precisely in their initial stages. Voice pathologies distress the vocal folds in the phonation course. They make vocal folds making asymmetrical vibrations due to the faulty of diverse influences paying to vocal vibrations. Vocal folds are inversely exaggerated by vocal fold pathologies ensuing in disparity in the vibratory cycle of vocal folds since their aptitude to be closed appropriately is reduced. Voice disorders also disturb the form of the vocal tract (supra-glottal) and yield indiscretions in spectral possessions [8]. It is well known that there is no intralaryngeal consequence on the vocal tract throughout the making of a vowel if consider that the voicing source has immeasurable confrontation.

Though medicine and invasive interference can grip back the development of the voice disorder disease and improve some of the indications, there is no obtainable treatment. Thus, primary diagnosis is serious in order to progress the patient's eminence of life and extend it.

This paper aims at developing a potential feature extraction and feature selection approach which assist in the process of voice disorder detection in earlier stages more precisely. The detailed description of the proposed work is explained in the following sections.

II. RELATED WORK

Earlier, a substantial number of researchers had functioned on categorizing of voice disorders using data mining techniques. Maximum of those works were intended to rise the accurateness and recognize the finest approaches by applying a numeral of classification systems and statistical approaches.

Martinez et al. [9] presented a set of experiments on detection of voice pathology with SVD database using multifocal toolkit for discriminative standardization and synthesis. Outcomes were

related with the MEEI database. Meanwhile they used the data from the SVD dataset, continued vowel recordings of /a/, /i/, and /u/ were examined.

Little et al. [10,11] intensive on handling discerning healthy subjects from subjects with Parkinson's disease by perceiving dysphonia. They offered a new measure of dysphonia: pitch period entropy (PPE). The exploited data contained of 195 continued vowels from 31 subjects, of which 23 were diagnosed with Parkinson's disease. The mined features encompassed pitch period entropy, shimmer, jitter, fundamental frequency, pitch marks, HNR, etc.

Zuzana et al [12] aims to develop a machine learning approach to detect pathological speech, specifically dysphonia. They extracted 1560 speech features and performed training on the classification model. The three-classification used in their work are K-nn, random forests and SVM. They analyzed the performance of the classifier by considering or without considering the gender as a main parameter.

Sellam et al.[12], attempted to examine the voice pathology form normal voice of children using different classification models like SVM and Radial basis functional neural network. They used several acoustic parameters to perform this process.

Hugo Cordeiro [13] offered a set of trials to recognize the best set of features from the vocal tract and to classify them SVM and Gaussian Mixture Models is applied for the identification of pathologic voices. They also used Regression Trees to the pathological voice recognition based on formant analysis and harmonic-to-noise ratio with 95% of recognition rate.

In [14], the speech signal is examined by the acoustic parameters and the classification technique used is Support Vector Machine is used to classify the normal and pathology voice based on the features extracted in the previous phase.

Though there are several approaches involved in voice disorder detection, there are very few works on feature extraction and feature selection. So the ultimate aim of this research work is to discover potential features which are involved in determining the pathological voice in an effective manner.

III. PROPOSED METHODOLOGY OF POTENTIAL FEATURE EXTRACTION AND FEATURE SUBSET SELECTION

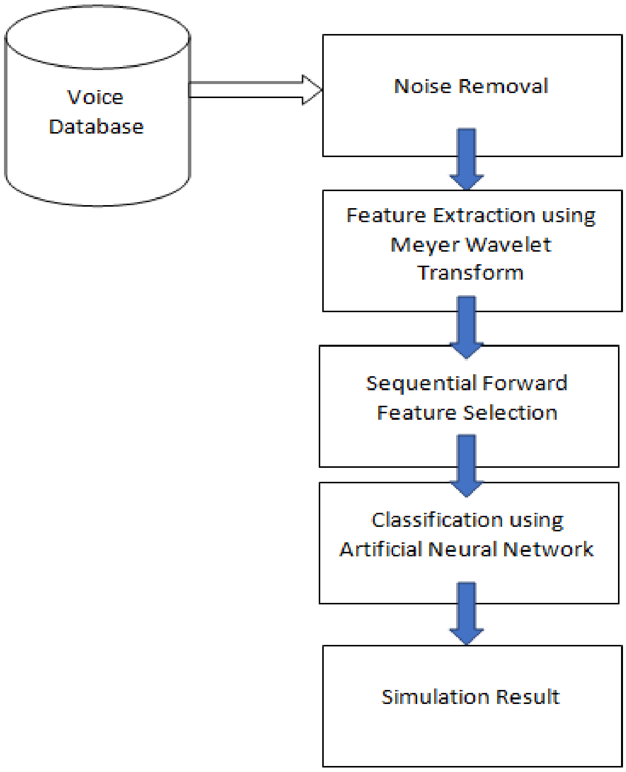

Figure 1: Overview of Proposed Methodology

To perform a potential feature extraction and feature selection to provide better results on voice disorder classification this research work develops a novel approach as depicted in the figure 1. The dataset used in this work is collected from Saarbrucken voice database (SVD). The collection of voice recording used in this work is 1040. The length of the audio clip with sustained vowels /a/, /i/, /u/ is 2 seconds. The collected raw voice samples are underwent for noise removal to improve the quality of the audio. Next the features of the voice are extracted using Meyer wavelet transform which extracts the significant features which involves in identification of pathological voice. With the extracted features to eliminate the irrelevant features sequential forward feature subset selection method is applied to select the potential features which contribute more in determining the pathological and normal voice. The simulation results proved the better achievement in voice disorder detection by this research methodology while comparing with existing approaches considered in this work. The detailed description of each stage is explained in the following subsections.

A. Description of Database - The voice samples are collected from the Saarbrucken voice database. This database has been made available freely online. It is a collection of voice recordings from more than 2000 persons, where a session is defined as a collection of, recordings of vowels /a/, /i/, /u/ produced at normal, high, low and low-high-low pitch. In addition, the Electro Glotto Graph (EGG) signal is also stored for each case in a separate file. The length of the audio clips with sustained vowels is 2 seconds. All recordings are sampled at 50 kHz and their resolution is 16-bit. For these experiments' files with sustained vowels and people whose age in the range of 30- 35 are used. Both male and female voice is taken for four different diseases like laryngoceles, dysphonia, diplophonia, and chorditis are used. A database consists of 400 voice samples such as 200 voice samples for healthy and 200 voice samples for pathology is created.

B. Audio Noise Removal - Almost entire audio recordings comprise some volume of noise. This noise may link audio signal through recording process or due to long media storage, which is not adequate by sound engineers. To yield finest quality audio soundtracks these unsolicited audio noises must be detached to the utmost degree conceivable. Sound integrally starts and ends as a referend signal. Few years ago, eliminating noise from the audio was a very problematic job for the reason that of the high cost tangled which was not reasonable for a typical user. The sound digitization and digital signal processing technologies transformed this situation intensely. The new generation high-speed PCs fortified with excellence sound cards and software submissions has made audio noise removal work even more reasonable. Any records surface is focus to micro fissures, scratches, and soiling. Over a time period, the consequence is a continuous worsening in the eminence of the audio. This weakening shows the aforementioned in unsolicited noise, clicks, and crackles. Digitization process itself does not eliminate any type of the misrepresentation or noise, but it permits us to overpower or eliminate such distortions or noise by spread over special soft Noise Removal algorithms to a digital sound.

C. Features Extraction using Meyer Wavelet Transform - Meyer wavelet transform (MVT) is applied to solve the issue of non-stationary feature of ECG signals. The Meyer wavelet is an orthogonal wavelet introduced by Yves Meyer [16]. It is a kind of continuous wavelet and mainly used in multi-fault classification, fractal random fields and adaptive filters. This MVT is substantially differentiable with unlimited provision and well-defined in domain of frequency domain in terms as

$$\psi(w) = \begin{cases} \frac{1}{\sqrt{2\pi}}\sin\left(\frac{\pi}{2}\gamma\left(\frac{3|w|}{2\pi}-1\right)\right)e^{\frac{jw}{2}} & if\ \frac{2\pi}{3} < |w| < \frac{4\pi}{3} \\ \frac{1}{\sqrt{2\pi}}\cos\left(\frac{\pi}{2}\gamma\left(\frac{3|w|}{4\pi}-1\right)\right)e^{\frac{jw}{2}} & if\ \frac{4\pi}{3} < |w| < \frac{8\pi}{3} \\ 0 & otherwise \end{cases}$$

$$\gamma(x) = \begin{cases} 0 & if\ x < 0, \\ x & if\ 0 < x < 1 \\ 1 & if\ x > 1 \end{cases}$$

The reason to uses MVT wavelet is that it has additional beneficial features compared to the standard DWT and FFT wavelet transforms. Because it has the ability to automatically adjust to the different ECG signal features, by adapting large size window to look for long lived low frequency signals aspects and it uses small window to analyze the short time high frequency components.

This wavelet transform is used in ECG for decomposing the signal into various different frequency scales where the characteristics waveforms are signified by Zero Crossings. The R peak is precisely discovered using MWT. In general ECG signal comprised of many data but for efficient energy consumption in mobile computing based offloading the essential features are extracted using MWT with the aid of spectral analysis. These features alone characterize the behavior of the ECG signals. And these information's are composed as data packets, to transfer in mobile adhoc network to the base station(Experts Location).

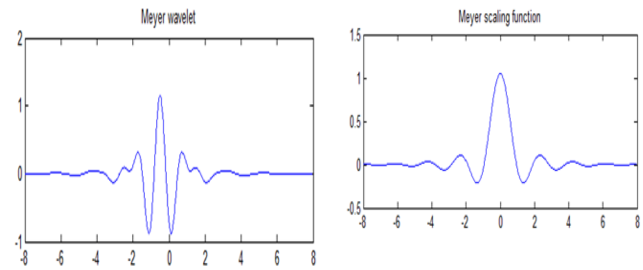

Figure 2: Meyer Wavelet Transform mother wavelet and its corresponding scaling function for feature extraction

D. Sequential Forward Feature Selection (SFFS) on Voice data - Feature selection is considered as a prominent problem which involves in searching for possible feature subsets, in a given search space. They choose the subset that is optimal with respective to a specific objective function.

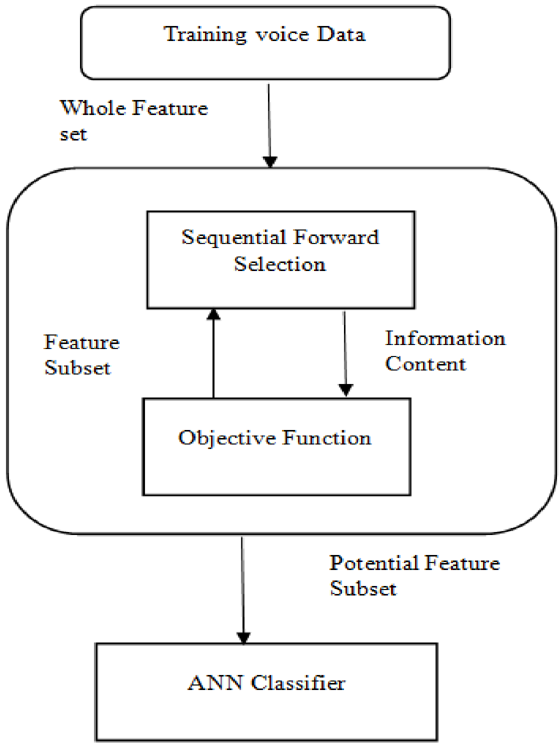

Figure 3: Working of Sequential Forward Feature Selection on Voice Dataset

In this work sequential forward feature selection algorithm as shown in the figure 3, is used for voice feature selection, this method is a kind of greedy search algorithms that are used to reduce an initial *d*-dimensional feature space to a *k*-dimensional feature subspace where $k < d$. The inspiration behind feature selection algorithms is to automatically select a subset of features that is most relevant to the problem. The goal of feature selection is two-fold: To improve the computational efficiency and reduce the generalization error of the model by removing irrelevant features or noise. Starting from the empty set, sequentially add the feature x^+ that maximizes $J(Y_k + x^+)$ when united with the features Y_k that have previously been nominated. Consequently, the SFFS earnings animatedly increasing and decreasing the number of features till the anticipated d is reached.

Algorithm for Feature selection using SFFS

Input: $Y=\{y_1,y_2,...,y_d\}$

The ***SFS*** algorithm takes the whole d-dimensional feature set as input.

Output: $X_k=\{x_j \mid j=1,2,...,k;\ x_j \in Y\}$ where, $k=(0,1,2,...,d)$

- SFS returns a subset of features; the number of selected features k, where k<d, has to be specified *a priori*.

Initialization: X0=∅, k=0

- We initialize the algorithm with an empty set ∅ ("null set") so that k=0(where k is the size of the subset).

Step 1 (Inclusion):

$x^{+}= \text{argmax } J(x_k+x)$, where $x \in Y-X_k$

$X_{k+1}=Xk+x^{+}$

$k=k+1$

Go to Step 1

- in this step, add an additional feature, x^{+}, to our feature subset X_k.
- x^{+} is the feature that maximizes our criterion function, that is, the feature that is associated with the best classifier performance if it is added to X_k.
- Repeat this procedure until the termination criterion is satisfied.

Termination: k=p

- Add features from the feature subset X_k until the feature subset of size kk contains the number of desired features pp that we specified *a priori*.

IV. EXPERIMENTS AND RESULT'S

The proposed work is simulated using MATLAB tool. The dataset for voice dataset feature subset selection is done using SVD. The number of instances used in this work is 1040. The performance comparison is done on both feature extraction and feature subset selection. The meyer wavelet transform is used for extracting the features form voice signals and from the extracted features, only potential attributes are selected using Sequential Forward Subset Selection. The features extracted using Meyer Wavelet transform is given in the table 1. The first 5 features are related to frequency parameters, the features from 6 to 9 belongs to pulse parameters, features from 10 to 15 belongs to amplitude parameters, 16 to 18 related to voicing parameters, 19 to 23 related to pitch parameters and remaining parameters 24 to 26 related to harmonicity parameters.

Table 1: Feature Extracted by Meyer Wavelet Transform

S.No	Features Extracted	Description
1	Jitter(local)	Frequency Parameters
2	Jitter(local,absolute)	
3	Jitter(rap)	
4	Jitter(ppq5)	
5	Jitter(ddp)	
6	Number of pulses	Pulse Parameters
7	Number of periods	
8	Mean period	
9	Standard deviation of period	
10	Shimmer(local)	Amplitude Parameters
11	Shimmer(local,dB)	
12	Shimmer(apq3)	
13	Shimmer(apq5)	
14	Shimmer(apq11)	
15	Shimmer(dda)	
16	Fraction of locally unvoiced frames	Voicing Parameters
17	Number of voice breaks	
18	Degree of voice breaks	
19	Median pitch	Pitch Parameters
20	Mean pitch	
21	Standard deviation	
22	Minimum pitch	
23	Maximum pitch	
24	Autocorrelation	Harmonicity Parameters
25	Noise-to-harmonic	
26	Harmonic-to-noise	

Detailed Description of the features extracted by Meyer wavelet transform are as follows:

Jitter (local) - This is the average absolute difference between consecutive periods, divided by the average period.

Jitter (local, absolute) - This is the average absolute difference between consecutive periods, in seconds

Jitter (rap) - This is the Relative Average Perturbation, the average absolute difference between a period and the average of it and its two neighbours, divided by the average period.

Jitter (ppq5) - This is the five-point Period Perturbation Quotient, the average absolute difference between a period and the average of it and its four closest neighbours, divided by the average period.

Jitter (ddp) - This is the average absolute difference between consecutive differences between consecutive periods, divided by the average period.

Shimmer (local)

This is the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.

Shimmer (local, dB) - This is the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20

Shimmer (apq3) - This is the three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours, divided by the average amplitude.

Shimmer (apq5) - This is the five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbours, divided by the average amplitude.

Shimmer (apq11) - This is the 11-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its ten closest neighbours, divided by the average amplitude

Shimmer (ddp) - This is the average absolute difference between consecutive differences between the amplitudes of consecutive periods

HNR (Harmonic to Noise Ratio) - The HNR is an assessment of the ratio between periodic components and non periodic component comprising a segment of voiced speech, as Murphy and Akande [6]. The first component arises from the vibration of the vocal cords and the second follows from the glottal noise, expressed in dB.

Table 2: Performance Comparison of feature extraction on Voice dataset using three different Wavelet Transform

Performance	Accuracy
Fast Fourier Transform	89.40%
Haar Wavelet Transform	92.60%
Meyer Wavelet Transform	98.50%

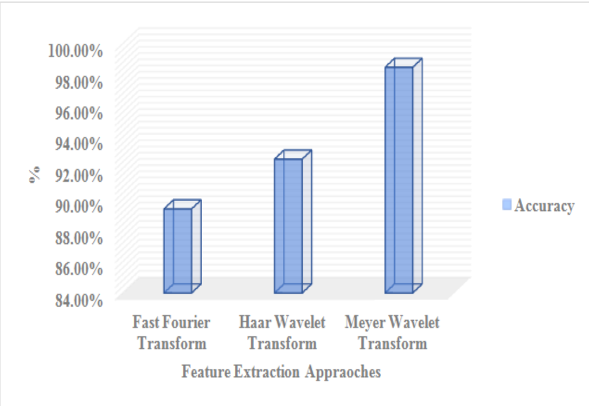

Figure 4: Performance Comparison of feature extraction on Voice dataset using three different Wavelet Transform

The table 2 and the figure 4 show the performance comparison of three different wavelet transforms based on their achieved accuracy. The result shows that meyer wavelet transform achieves more accuracy in extraction of voice signal features in a precise manner.

Table 3: Performance Comparison of three different Feature selection Approaches on Voice Signal

Feature Selection Algorithm	No of Attributes Selected	Selected Attributes	List of Attributes selected
Sequential Forward Feature Selection	9	4,7,9,10,14, 17,19,20,22	Jitter (ppq5),Shimmer (localdb), Shimmer (apq5), Shimmer (apq11), HTN, Standard deviation, Maximum pitch, Number of pulses, Mean period
Information Gain	11	4,5,7,9,10,12,14,17,19, 21,22	Jitter (ppq5), Jitter (ddp),Shimmer (localdb), Shimmer (apq5), Shimmer (apq11), AC,HTN, Standard deviation, Maximum pitch, Number of periods, Mean Period
Principal component Analysis	10	1,2,3,4,5,6, 7,8,9,10	Jitter (local), Jitter (localabs), Jitter (rap), Jitter (ppq5), Jitter (ddp), Shimmer (local), Shimmer (localdb), Shimmer (apq3),Shimmer (apq5), Shimmer (apq11)

From the table 3 it is observed that the proposed sequential forward feature selection method chooses 9 attributes which belongs to frequency, amplitude, pitch, pulses and harmonicity parameters are more involved in recognizing voice disorder while comparing the other two feature selection methods information gain and principal component analysis. In this both these algorithms uses frequency and amplitude as the prime features for voice disorder classification by selecting 11 and 10 attributes respectively.

Table 4: Performance Analysis based on Time taken to build model

Feature Subset Selection	Time Taken (sec)
Sequential Forward Feature Selection	0.67
Information Gain	13.6
Principal component Analysis	20.4

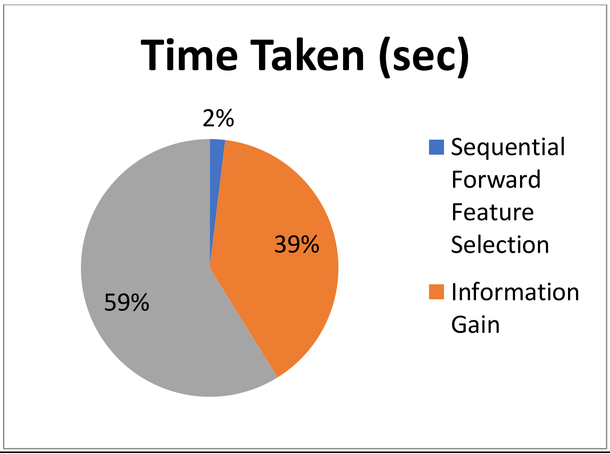

Figure 6: Performance comparison based on time taken

The table 4 and the figure 6 show the performance comparison by SSFS, IG and PCA. It is observed that the time taken by PCA is 59%, information gain is 39% and the sequential forward selection consumes only 2%.

Table 5: Performance Evaluation using various Metrics

Evaluation Metrics	ANN + Information Gain	ANN+ SFSS
Correctly Classified Instances	62.7885 %	96.4%
Incorrectly Classified Instances	37.2115 %	4.6%
Kappa statistic	0.2558	0.055
Mean absolute error	0.4084	0.3284
Root mean squared error	0.5329	0.4229
Relative absolute error	81.6731 %	61.02 %

The table 5 shows the performance comparison of information gain and SFSS by validating them using Artificial neural network. The results revealed that SFSS with ANN produces 96.4% accuracy while comparing the ANN with information gain which obtains 62.789% as its accuracy. Thus, the classification of ANN using SFSS produces optimal result in voice pathology and normal voice recognition more precisely. The error rate is also lesser while comparing the existing method information gain.

V. CONCLUSION

The aim of this research work is to build an optimal feature selection model by introducing Meyer wavelet transform for extracting features from the voice signal and form the obtained features sequential forward selection is used for choosing potential features. This approach greatly helps to handle the voluminous voice dataset more precisely. The discovery of voice disorder samples in the earlier stages will avoid to serious cause in later stages. The simulation results proved the performance of the proposed model achieves better results.

VI. REFERENCES

[1]. Soni S, Chouksey S. A study of clinicopathological profile of patients of hoarseness of voice presenting to tertiary care hospital. Ind J Otolaryngol Head Neck Surg. 69(2):244–7, 2017

[2]. Roy N, Stemple J, Merrill RM et al. Epidemiology of voice disorders in the elderly: Preliminary findings. The Laryngoscope, 117(4):628-33, 2007

[3]. Stachler RJ, Francis DO, Schwartz SR, Damask CC, Digoy GP, Krouse HJ, et al. Clinical Practice Guideline: Hoarseness (Dysphonia) (Update) Executive Summary. Otolaryngol Head Neck Surg, 158(3):409-26, 2018

[4]. Kiakojoury K, Dehghan M, Hajizade F, Khafri S. Etiologies of dysphonia in patients referred to ENT clinics based on video laryngoscopy. Iranian J Otorhinolaryngology, (3):169–174, 2014;26

[5]. Rao SV, Sharma VS, Subbarao MV, Keerthi GS, Deta SI, Seepana PT, et al. Etiopathology of hoarseness of voice:A clinical study. Int J Phonosurg Laryngol, 6(2):64–7, 2016

[6]. Sindhu BS, Jayaprabha S. A descriptive study of aetiopathological evaluation of hoarseness: our experience. Int J Otorhinolaryngol Head Neck Surg. 4(2):356-361, 2018;

[7]. Banjara H, Mungutwar V, Singh D, Gupta A. Hoarseness of voice: a retrospective study of 251 Cases. Int J Phonosurg Laryngol. 1(1):21–7, 2011

[8]. Pylypowich, A., Duff, E. Differentiating the Symptom of Dysphonia. The Journal for Nurse Practitioners. Elsevier, Vol. 12, Issue 7, 459–466, (2016).

[9]. Martinez, D. Voice Pathology Detection on the Saarbrücken Voice Database with Calibration and Fusion of Scores Using MultiFocal Toolkit. In Communications in Computer and Information Science; Springer: Berlin, Germany, Volume 328, ISBN 978-3-642-35291-1, 2012

[10]. Little, A.M. Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. IEEE Trans. Biomed. Eng, 56, 1015–1022, 2009

[11]. Tsanas, A.; Little, M.A.; McSharry, P.E.; Ramig, L.O. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J. R. Soc. Interface, 8, 842–855, 2011

[12]. Zuzana Dankovicova Dávid Sovák, Peter Drotár and Liberios Vokorokos, Machine Learning Approach to Dysphonia Detection, MDPI, Appl. Sci, 8, 1927, pp 1-12, 2018

[13]. Sellam, V., Jagadeesan, J. Classification of Normal and Pathological Voice Using SVM and RBFNN. Journal of Signal and Information Processing, 2014, 5, 1-7.

[14]. Hugo Cordeiro, Carlos Meneses Ribeiro, Speaker Characterization with MLSFs, The Journal for Nurse Practitioners. Elsevier, 2016

[15]. S. C. Punitha¸ N. A. Sheela Selvakumari, Voice Pathology Identification System using SVM Classifier, International Journal of Advance Research in Computer Science and Management Studies, Volume 5, Issue 1, January 2017

[16]. I. Meyer, Wavelets and operators. Vol.1. Cambridge University Press, 1995