Journal of Advances in Computer Research
Quarterly ISSN: 2008-6148
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 4, No. 2, May 2013), Pages: 53-62
www.jacr.iausari.ac.ir

A Hybrid of Genetic Algorithm and Gaussian Mixture Model for Features Reduction and Detection of Vocal Fold Pathology

Vahid Majidnezhad✉, **Igor Kheidorov**

The United Institute of Informatics Problems, National Academy of Science, Minsk, Belarus

vahidmn@yahoo.com; ikheidorov@sakrament.com

Received: 2013/01/18; Accepted: 2013/03/04

Abstract

Acoustic analysis is a proper method in vocal fold pathology diagnosis so that it can complement and in some cases replace the other invasive, based on direct vocal fold observation, methods. There are different approaches and algorithms for vocal fold pathology diagnosis. These algorithms usually have three stages which are Feature Extraction, Feature Reduction and Classification. In this paper initial study of feature extraction and feature reduction in the task of vocal fold pathology diagnosis has been presented. A new type of feature vector, based on wavelet packet decomposition and Mel-Frequency-Cepstral-Coefficients (MFCCs), is proposed. Also a new GA-based method for feature reduction stage is proposed and compared with conventional methods such as Principal Component Analysis (PCA). Gaussian Mixture Model (GMM) is used as a classifier for evaluating the performance of the proposed method. The results show the priority of the proposed method in comparison with current methods.

Keywords: *Vocal Fold Pathology Diagnosis; Wavelet Packet Decomposition(WPD); Mel-Frequency-Cepstral-Coefficient (MFCC); Principal Component Analysis (PCA); Genetic Algorithm (GA); Gaussian Mixture Model (GMM).*

1. Introduction

Voice signal information often plays an important role for specialists to understand the process of vocal fold pathology formation. In some cases vocal signal analysis can be the only way to analyze the state of vocal folds. Nowadays diverse medical techniques exist for direct examination and diagnose of pathologies. Laryngoscopy, glottography, stroboscopy, electromyography, videokimography are most frequently used by medical specialists. But these methods possess a number of disadvantages. Human vocal tract is hardly-accessible for visual examination during phonation process and that makes it more problematic to identify pathology. Moreover, these diagnostic means may cause the patients feel much discomfort and distort the actual signal so that it may lead to incorrect diagnosis as well [1-4].

Acoustic analysis as a diagnostic method has no drawbacks, peculiar to the above mentioned methods. It possesses a number of advantages. First of all, acoustic analysis is a non-invasive diagnostic technique that allows pathologists to examine many people in short time period with minimal discomfort. It also allows pathologists to reveal the pathologies on early stages of their origin. This method can be a great interest for medical institutions. In recent years a number of methods were developed for segmentation and classification of speech signals with pathology.

Different parameters for feature extraction are used. Traditionally, one deals with such parameters like pitch, jitter, shimmer, amplitude perturbation, pitch perturbation, signal to noise ratio, normalized noise energy [5] and others [6-9]. Feature extraction, using the above mentioned parameters, has shown its efficiency for a number of practical tasks [8]. These parameters are frequently used in systems for automatic vocal fold pathology diagnosis, in speaker identification systems or in multimedia database indexing systems. In the proposed method, the Mel-Frequency-Cepstral-Coefficients (MFCCs), Energy and Shannon Entropy parameters have been used for creating the features vector.

Also different approaches for feature reduction are used such as Principal Component Analysis (PCA) [10-13] and Linear Discriminant Analysis (LDA) [14]. In the proposed method, the GA-based method has been used and the results of experiments show its better performance in comparison with the PCA.

Finally, the reduced features are used for speech classification into the healthy and pathological class. Different machine learning methods such as Support Vector Machines [10], Artificial Neural Networks [15], Hidden Markov Model [9], etc can be used as a classifier. In the proposed method, the GMM has been used for the classification purpose. In table 1, some pervious methods in the vocal fold pathology diagnosis are summarized.

Table 1. Summary of some pervious works.

Reference	Feature set	Feature Reduction Approach	Classifier
[10]	25 Acoustic parameters given by MDVP	PCA	Support Vector Machine
[11]	Spectral perturbation	PCA	K-Means clustering
[12]	Acoustic feature, noise	PCA	Threshold
[13]	Linear prediction coefficients	PCA	K-nearest Neighbours
[14]	Mel-frequency-cepstral-coefficients	LDA	Gaussian Mixture Model
[15]	Spectral	-	Artificial Neural Network

The main motivation of this article is to develop a method for detecting vocal fold pathology. This method should have higher accuracy in comparison with the current methods. For this propose, some novel approaches have been proposed. These approaches have been used in the different phases of the proposed method especially in the feature reduction phase. In the feature reduction phase, the accuracy of the system by the use of genetic algorithm has been increased incredibly.

2. Methodology

The Block diagram of the proposed method is illustrated in Fig. 1. In the first stage, by the use of MFCC and Wavelet Packet Decomposition, feature vector containing 139 features is made. In the second stage, by the use of the proposed GA-based method, the dimension of feature vector is reduced. In the last stage, by the use of Gaussian Mixture Model (GMM), the speech signal classified into two classes: pathological or healthy.

As it is shown in Fig. 1, first, by the use of cepstral representation of input signal, 13 Mel-Frequency-Cepstral-Coefficients (MFCC) are extracted. Then the wavelet packet decomposition in 5 levels is applied on the input signal to make the wavelet packet tree. Then, from the nodes of resulting wavelet packet tree, 63 energy features along with 63 Shannon entropy features are extracted. Finally, by the combination of these features, the initial feature vector with the length of 139 features is created.

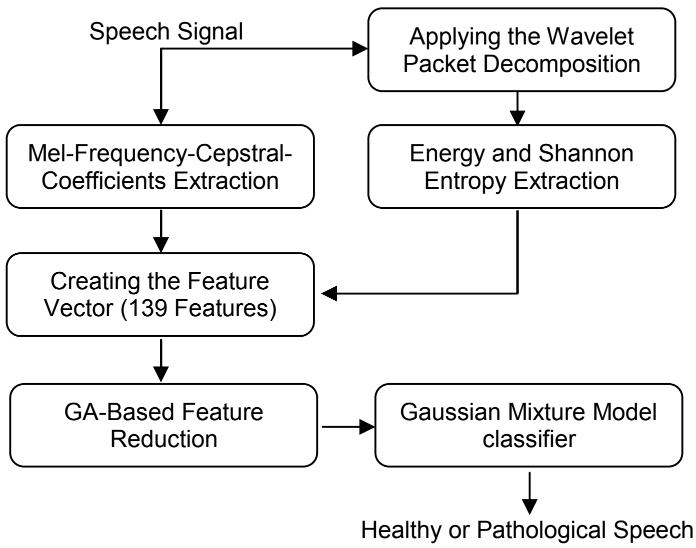

Figure 1. The scheme of the proposed method for detection of vocal fold pathology

2.1 Feature Extraction

2.1.1. Mel-Frequency-Cepstral-Coefficients (MFCCs): MFCCs are widely used features to characterize a voice signal and can be estimated by using a parametric approach derived from linear prediction coefficients (LPC), or by the non-parametric discrete fast Fourier transform (FFT), which typically encodes more information than the LPC method. The signal is windowed with a Hamming window in the time domain and converted into the frequency domain by FFT, which gives the magnitude of the FFT. Then the FFT data is converted into filter bank outputs and the cosine transform is found to reduce dimensionality. The filter bank is constructed using 13 linearly-spaced filters (133.33Hz between center frequencies,) followed by 27 log-spaced filters (separated by a factor of 1.0711703 in frequency.) Each filter is constructed by combining the amplitude of FFT bin. The Matlab code to calculate the MFCC features was adapted from the Auditory Toolbox (Malcolm Slaney). The MFCCs are used as features in [14] to classify the speech into pathology and healthy class. The MFCC information has been reduced by averaging the sample's value of each coefficient.

2.1.2. Wavelet Packet Decomposition: Recently, wavelet packets (WPs) have been widely used by many researchers to analyze voice and speech signals. There are many out-standing properties of wavelet packets which encourage researchers to employ them in widespread fields. The most important, multi resolution property of WPs is helpful in voice signal synthesis [16-17].

The hierarchical WP transform uses a family of wavelet functions and their associated scaling functions to decompose the original signal into subsequent sub-bands. The decomposition process is recursively applied to both the low and high frequency sub-bands to generate the next level of the hierarchy. WPs can be described by the following collection of basic functions:

$$W_{2n}(2^{p-1}x-1)=\sqrt{2^{1-p}}\sum_{m}h(m-2l)\sqrt{2^{p}}\,W_n(2^{p}x-m) \tag{1}$$

$$W_{2n+1}(2^{p-1}x-1)=\sqrt{2^{1-p}}\sum_{m}g(m-2l)\sqrt{2^{p}}\,W_n(2^{p}x-m) \tag{2}$$

where p is scale index, l the translation index, h the low-pass filter and g the high-pass filter with

$$g(k)=(-1)^{k}h(1-k) \tag{3}$$

the WP coefficients at different scales and positions of a discrete signal can be computed as follows:

$$C^{p}_{n,k}=\sqrt{2^{p}}\sum_{m=-\infty}^{\infty}f(m)W_n(2^{p}m-k) \tag{4}$$

$$C^{p-1}_{2n,l}=\sum_{m}h(m-2l)C^{p}_{n,m} \tag{5}$$

$$C^{p-1}_{2n+1,l}=\sum_{m}g(m-2l)C^{p}_{n,m} \tag{6}$$

for a group of wavelet packet coefficients, energy feature in its corresponding sub-band is computed as

$$Energy_n=\frac{1}{N^2}\sum_{k=1}^{n}\left|C^{p}_{n,k}\right|^2 \tag{7}$$

The entropy evaluates the rate of information which is produced by the pathogens factors as a measure of abnormality in pathological speech. Also, the measure of Shannon entropy can be computed using the extracted wavelet-packet coefficients, through the following formula

$$Entropy_n=-\sum_{k=1}^{n}\left|C^{p}_{n,k}\right|^2\log\left|C^{p}_{n,k}\right|^2 \tag{8}$$

In this study, mother wavelet function of the tenth order Daubechies has been chosen and the signals have been decomposed to 5 levels. The mother wavelet used in this study is reported to be effective in voice signal analysis [18-19] and is being widely used in many pathological voice analyses [17]. Due to the noise-like effect of irregularities in the vibration pattern of damaged vocal folds, the distribution manner of such variations within the whole frequency range of pathological speech signals is not clearly known. Therefore, it seems reasonable to use WP rather than DWT or CWT to have more detail sub-bands.

2.2 Feature Reduction

Using every feature for classification process is not good idea and it may be causes to the increasing the rate of misclassification. Therefore, it is better to choose the proper features from the whole features. This process is called as “Feature Reduction”. In other words, the goal is to reduce the dimension of the data by finding a small set of important features which can give good classification performance. One way for feature reduction is Principal Component Analysis (PCA) which is used frequently in pervious works such as [10-13]. In this section also a novel approach, the GA-based method, is proposed for the feature reduction stage.

2.2.1 Principal Component Analysis

This method searches a mapping to find the best representation for distribution of data. Therefore, it uses a signal-representation criterion to perform dimension reduction while preserving much of the randomness or variance in the high-dimensional space as possible [20]. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. PCA involves the calculation of the eigenvalues decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each attribute. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate, called the first principal component, the second greatest variance on the second coordinate, and so on. The main limitation of PCA is that it does not consider class separately, since it does not take into account the class label of the feature vectors.

2.2.2 GA-Based Feature Reduction

Genetic Algorithm (GA) is a heuristic optimization method which acts on the basis of evaluation in nature and search for the final solution among a population of potential solutions. It has three basic operations which are selection, mutation and crossover. Recently, it is used widely in optimizations problems such as [21]. As it is mentioned before, the main limitation of PCA is that it does not take into account the class labels and it just focus on the sample’s value. In other words, the PCA searches for the features which their sample’s value have bigger variance in comparison with others and it does not collaborate with the classifier. So, for overcoming this disadvantage, by using genetic algorithm a GA-based method is proposed which considers the error rate of the classifier in its fitness function and tries to minimize it. For this purpose, a fitness function f is defined which shows the error rate of the GMM classifier for the train set.

$$f = \sum_{i=1}^{n} \left| a_i - r_i \right| \tag{9}$$

The a_i is the result of classifier and the r_i is the real class for i^{th} speech signal. The n is the number of speech files in the “train” dataset. The aim of the GA-based method is to find the subset of features so that they minimize the f.

2.3 Gaussian Mixture Model

Let $x \in R^n$ be a random vector that has an arbitrary distribution. The distribution density of x is modeled as a Gaussian mixture density, a mixture of *Q* component densities, given by [14]

$$p(\frac{x}{\lambda}) = \sum_{i=1}^{Q} c_i . p_i(x), \sum_{i=1}^{Q} c_i = 1, c_i \geq 0 \qquad (10)$$

where , $p_i(x)$, $i=1,\ldots,Q$ are the component densities, and c_i, $i=1,\ldots,Q$ are the component weights. Each component density is an n-variate Gaussian function of the form

$$p_i(x) = \frac{1}{(2\pi)^{n/2} |C_i|^{1/2}} \exp[-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)] \qquad (11)$$

with μ_i the *n*1* mean vector and C_i the n*n covariance matrix.

The main motive for using the GMM as a representation of the acoustic space is that it has been demonstrated that a linear combination of Gaussian basis functions has a capacity to represent a large class of sample distributions [14]. In this article, two GMMs have been trained for the healthy and pathological speeches. And for each speech in the test set, the probabilities of its generation by the means of each GMM are calculated. Then, the maximum probability specifies the class of that speech.

3. Experiments and Results

In this section, three experiments have been designed. These experiments are simulated in the Matlab 7.11.0. The whole scheme of the proposed method is illustrated in Fig. 1. 10 folds cross-validation scheme has been adapted to assess the generalization capabilities of the system in our experiments.

3.1 Dataset Description

The dataset was created by specialists from the Belarusian Republican Center of Speech, Voice and Hearing Pathologies. 75 pathological speeches and 55 healthy speeches have been selected randomly which are related to sustained vowel “a”. All the records are wave files in the PCM format.

3.2 Results

In first experiment, the t-test has been applied on each feature and compare p-value for each feature as a measure of how effective it is at separating groups. The result is shown in Fig. 2. There are about 40% of features having p-values close to zero and 55% of features having p-values smaller than 0.05, means that there are about 76 features among the original 139 features which have strong discrimination power. One can sort these features according to their p-values (or the absolute values of the t-statistic) and select some features from the sorted list. However, it is usually difficult to decide how many features are needed unless one has some domain knowledge or the maximum number of features that can be considered has been dictated in advance based on outside constraints.

One quick way to decide the number of needed features is to plot the MCE (misclassification error, i.e., the number of misclassified observations divided by the number of observations) on the test set as a function of the number of features.

In second experiment, the PCA approach is applied for the feature reduction. Since the total number of our observations is 130, so it is better to use the lower number of features for our classification's purpose. Therefore, the MCE has computed for various numbers of features between 1 and 30. The result is shown in Fig. 3 with circular marks.

In third experiments, the proposed GA-based method is applied for the feature reduction. The MCE has computed for various numbers of features between 1 and 30. The result is shown in Fig. 3 with triangular marks.

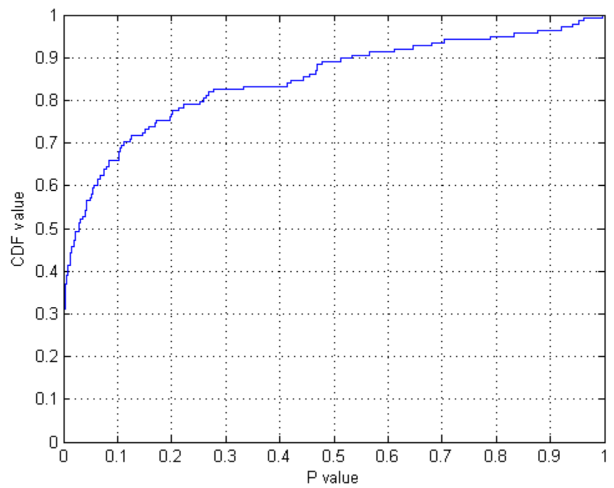

Figure 2. The P-value for features

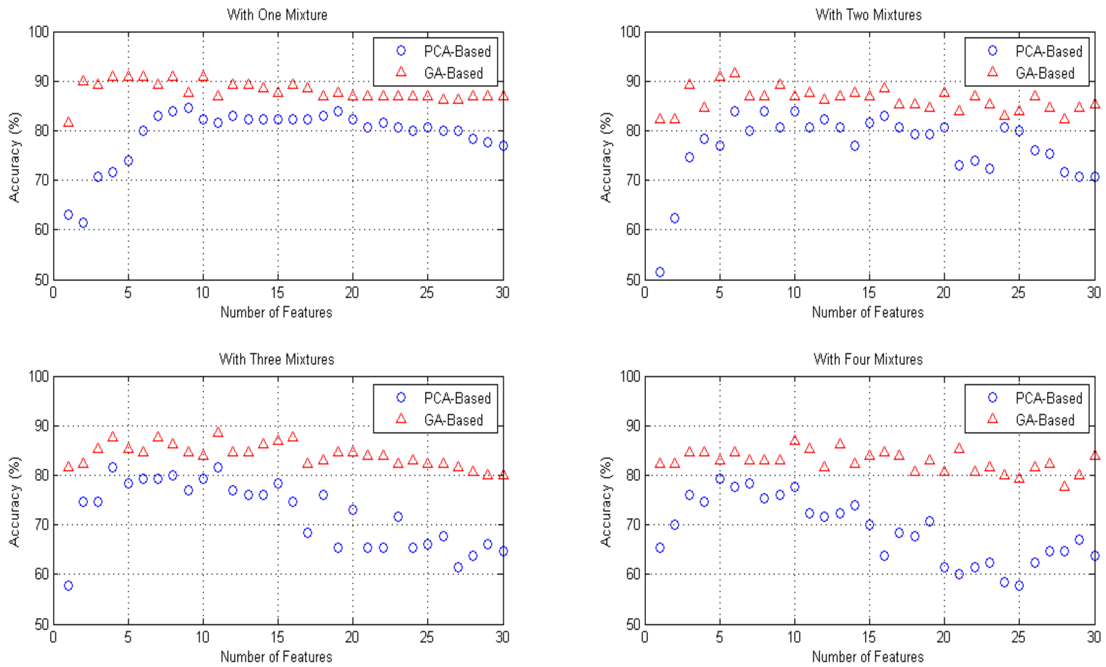

Figure 3. The obtained MCE by means of the PCA-Based and the GA-Based Methods

As it is obvious in Fig. 3, from the MCE point of view, the performance of the proposed GA-based approach is better than the PCA-based approach. In the best cases, the PCA-based method leads to 84.62% of accuracy with 9 selected features and 1 mixture while the GA-based method leads to 91.54% of accuracy with just 6 selected features and 2 mixtures. The selected features by means of these methods are shown in table 2.

3.3 Discussion

In the first experiment, it is shown that about 45% of initial features (63 features) are not strong for the classification purpose. So, feature reduction phase is necessary and important. Feature reduction algorithms can be roughly grouped into two categories: filter methods and wrapper methods. Filter methods rely on general characteristics of the data to evaluate and to select the feature subsets without involving the chosen learning algorithm. Wrapper methods use the performance of the chosen learning algorithm to evaluate each candidate feature subset. Wrapper methods search for features better fit for the chosen learning algorithm, but they can be significantly slower than filter methods if the learning algorithm takes a long time to run. The concepts of "filters" and "wrappers" are described in [22].

Table 2. The selected features for the construction of feature vector

Feature Reduction Method	The Selected Features	Number of Mixtures	Accuracy (%)
The Proposed GA-Based (feature vector length=6)	The 1st and 5th coefficient of MFCCs. Energy at the 60th node of WP Tree. Entropy at the 17th, 52nd and 60th nodes of WP Tree.	2	91.54
PCA-Based (feature vector length=9)	The 1st coefficient of MFCCs. Energy at the 9th, 17th, 33rd and 35th nodes of WP Tree. Entropy at the 8th, 16th, 17th and 32nd nodes of WP Tree.	1	84.62

In this article, the GA-based method for the feature reduction stage has been proposed which belongs to the wrapper methods. On the other side, the Principal Component Analysis (PCA) as the one of famous filter methods has been used. Also this method is used frequently in pervious works such as [10-13].

In the experiments 2 and 3, the performances of the proposed method as a wrapper method and the PCA-based method as a filter method have been compared. In the Fig. 3, it is clear that the GA-based method has better performance in comparison with the PCA-based method. This better performance is due to take into consideration of the accuracy of GMM classifier in the feature reduction phase. In other words, the GA-based method tries to reduce the initial feature vector with the aim of increasing the GMM classifier accuracy. But the PCA-based method just focuses on the data without any attention on the classifier accuracy.

4. Conclusion

In this article, it is shown that features based on wavelet transformation have potential for detection of vocal fold pathology. So, in the proposed scheme, Mel-Frequency-Cepstral-Coefficients (MFCC) along with the wavelet packet decomposition are used for the feature extraction phase.

Also a novel approach, GA-based method, for the feature reduction phase in the vocal fold pathology diagnosis is proposed. Three experiments are designed to investigate the efficiency of the proposed method. The results of experiments show the priority of the GA-based method in comparison with the conventional PCA-based method.

In this article, the GMM is used as the classifier. One of the main advantages of GMM is its ability to classify results correctly even when classes are similar. This methodology requires a shorter time for training than other approaches such multilayer perceptron (MLP) or learning vector quantization (LVQ). Furthermore, the GMM approach displays comparable accuracy with respect to LVQ or MLP.

Although it may be possible to try to build a complete multiclass classification system with a hierarchy of GMM so that detection of different type of pathological speech will be possible. For this propose, it is suggested to do further researches on more sophisticated feature extraction phase.

5. Acknowledgements

This work was supported by the speech laboratory of the United Institute of Informatics Problems of NASB in Belarus. The authors wish to thank the Belarusian Republican Center of Speech, Voice and Hearing Pathologies by its support in the speech database.

6. References

[1] J.B. Alonso, J.D. Leon, I. Alonso and M.A. Ferrer. *"Automatic Detection of Pathologies in the Voice by HOS Based Parameters"*. EURASIP Journal on Applied Signal Processing, 2001(4): 275-284, 2001.

[2] L.G. Ceballos, J. Hansen and J. Kaiser. *"A Non-Linear Based Speech Feature Analysis Method with Application to Vocal Fold Pathology Assessment"*. IEEE Trans. Biomedical Engineering, 45(3): 300-313, 2005.

[3] L.G. Ceballos, J. Hansen and J. Kaiser. *"Vocal Fold Pathology Assessment Using AM Autocorrelation Analysis of the Teager Energy Operator"*. ICSLP-1996 Proc., pp: 757-760, 1996.

[4] C. Adnene, and B. Lamia. *"Analysis of Pathological Voices by Speech Processing. Signal Processing and Its Applications"*, 2003 Proc., 1(1): 365-367, 2003.

[5] C. Manfredi. *"Adaptive Noise Energy Estimation in Pathological Speech Signals"*. IEEE Trans. Biomedical Engineering, 47(11): 1538-1543, 2000.

[6] J.I.G. Llorente and P.G. Vilda. *"Automatic Detection of Voice Impairments by Means of Short-Term Cepstral Parameters and Neural Network Based Detectors"*. IEEE Trans. Biomedical Engineering, 51(2): 380-384, 2004.

[7] M.D.O. Rosa, J.C. Pereira and M. Grellet. *"Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis"*. IEEE Trans. Biomedical Engineering, 47(1): 96-104, 2000.

[8] V. Majidnezhad and I. Kheidorov. *"A Novel Method for Feature Extraction in Vocal Fold Pathology Diagnosis"*. Proceeding of the 3rd International Conference on Wireless Mobile Communication and Healthcare, LNICST 61, pp: 96-105, 2013.

[9] V. Majidnezhad and I. Kheidorov. *“A HMM-Based Method for Vocal Fold Pathology Diagnosis”*. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2. pp: 135-138, 2012.
[10] W. Chen, C. Peng, X. Zhu, B. Wan and D. Wei. *“SVM-based identification of pathological voices”*. Proceedings of the 29th Annual International Conference of the IEEE EMBS, 2007.
[11] P. Go´mez, F. Dı´az, A. A´lvarez, K. Murphy, C. Lazaro, R. Martinez and V. Rodellar. *“Principal component analysis of spectral perturbation parameters for voice pathology detection”*. Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems, pp: 41-46, 2005.
[12] D. Michaelis, M. Frohlich and H.W. Strube. *“Selection and combination of acoustic features for the description of pathologic voices”*. Journal of the Acoustical Society of America, 103(3): 1628-1639, 1998.
[13] M. Marinaki, C. Kotropoulos, I. Pitas and N. Maglaveras. *“Automatic detection of vocal fold paralysis and edema”*. Proceedings of Eighth International Conference on Spoken Language Processing-ICSLP, 2004.
[14] J.I.G. Llorente, P.G. Vilda and M.B. Velasco. *“Dimensionality Reduction of a Pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters”*. IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, 53(10): 1943-1953, 2006.
[15] R.T. Ritchings, M.A. McGillion and C.J. Moore. *“Pathological voice quality assessment using artificial neural networks”*. Medical Engineering & Physics, 24(8): 561-564, 2002.
[16] H.K. Herisa, B.S. Aghazadeh and M.N. Bahrami. *“Optimal feature selection for the assessment of vocal fold disorders”*. Computers in Biology and Medicine, 39(10): 860-868, 2009.
[17] E. Fonseca, R.C. Guido, J.C. Pereira, P.R. Scalassarsa, C.D. Maciel and J.C. Pereira. *“Wavelet time frequency analysis and least squares support vector machines for identification of voice disorders”*. Computers in Biology and Medicine, 37(4): 571-578, 2007.
[18] R.C. Guido, J.C. Pereira, E. Fonseca, F.L. Sanchez and L.S. Vierira. *“Trying different wavelets on the search for voice disorders sorting”*. Proceedings of the 37th IEEE International Southeastern Symposium on System Theory, pp: 495-499, 2005.
[19] K. Umapathy and S. Krishnan. *“Feature analysis of pathological speech signals using local discriminant bases technique”*. Medical and Biological Engineering and Computing, 43(4): 457-464, 2005.
[20] M.K. Arjmandi and M. Pooyan. *“An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine”*. Biomedical Signal Processing and Control, 7(1): 3-19, 2012.
[21] V. Majidnezhad, H.M. Gader and E. Efimov. *“A Novel Hybrid Algorithm for Task Graph Scheduling”*. International Journal of Computer Science Issues, Vol. 8, Issue 2, pp: 32-38, 2011.
[22] R. Kohavi and G. John. *“Wrappers for feature subset selection”*. Artificial Intelligence, 97(1-2): 272-324, 1997.