

RESEARCH

Open Access

A novel hybrid of genetic algorithm and ANN for developing a high efficient method for vocal fold pathology diagnosis

Vahid Majidnezhad

Abstract

In this paper, an initial feature vector based on the combination of the wavelet packet decomposition (WPD) and the Mel frequency cepstral coefficients (MFCCs) is proposed. For optimizing the initial feature vector, a genetic algorithm (GA)-based approach is proposed and compared with the well-known principal component analysis (PCA) approach. The artificial neural network (ANN) with the different learning algorithms is used as the classifier. Some experiments are carried out for evaluating and comparing the classification accuracies which are obtained by the use of the different learning algorithms and the different feature vectors (the initial and the optimized ones). Finally, a hybrid of the ANN with the 'trainscg' training algorithm and the genetic algorithm is proposed for the vocal fold pathology diagnosis. Also, the performance of the proposed method is compared with the recent works. The experiments' results show a better performance (the higher classification accuracy) of the proposed method in comparison with the others.

Keywords: Vocal fold pathology diagnosis; Wavelet packet decomposition; Mel frequency cepstral coefficient (MFCC); Principal component analysis (PCA); Artificial neural network (ANN); Genetic algorithm (GA)

1 Introduction

Early detection of vocal fold pathology by the use of non-invasive methods and employing different techniques of speech processing have attracted scientists' attention in recent years. Their aim is to develop new techniques for processing the speech signals of patients in order to decrease the treatment expenses and to increase the accuracy of diagnosis.

Nowadays, different medical techniques based on the direct examination of vocal folds such as laryngeal videoendoscopy [1,2], glottography [3], and stroboscopy [4] are used by medical specialists. But these methods have two main drawbacks. Firstly, they are invasive. They may cause patients feel uncomfortable and consequently to distort the actual signal. Secondly, they are expensive to buy and maintenance fees are high.

The best option for overcoming to the disadvantages related to the medical instruments is to employ acoustic analysis techniques. They let medical specialists to

examine vocal fold in short time with minimal discomfort. They also allow revealing the pathologies on early stages.

In recent years, a number of methods based on acoustic analysis were developed for vocal fold pathology classification [5-7]. These methods usually have two phases which are the feature extraction phase and the classification phase. The feature extraction phase involves the transformation of speech signal into some parameters or features. The second phase implies a choice of a variety of machine learning methods.

Traditionally, for the feature extraction phase, one deals with such parameters like jitter [8,9], shimmer [10,11], signal-to-noise ratio [12,13], and formants [14,15]. Also, some of the well-known classifiers for the classification phase in the previous works were used such as support vector machine (SVM) [16-19], Gaussian mixture model (GMM) [20-22], artificial neural network (ANN) [23-25], and hidden Markov model (HMM) [26-28].

In [29], the authors have investigated the role of two different datasets: the Massachusetts Eye and Ear Infirmary (MEEI) and the Principe de Asturias (PdA). Both datasets contain records of sustained vowel 'a.' They have reported

Correspondence: vahidmn@yahoo.com
Department of Computer Engineering, Shabestar Branch, Islamic Azad University, Shabestar, Iran

the classification accuracy of 96.37% for the MEEI and the classification accuracy of 87.85% for the PdA dataset by the use of the same method. But when they have developed a system based on the MEEI dataset and then they have used it for classifying of the PdA dataset as the test set, they have achieved the classification accuracy of 78.14%. Also, they have done it in the inverse manner. In other words, they have developed a system based on the PdA dataset and then they have used it for classifying of the MEEI dataset as the test set. For this case, they have reported the classification accuracy of 83.13%.

Therefore, in [29], it is proved that the classification accuracy of vocal fold pathology detection systems extremely depends on the dataset and its characteristics such as the volume of dataset. So, it is obvious that the reported accuracies of the previous works are not comparable due to the lack of the same conditions such as dataset. Even it is possible to use the same dataset with the different train and test sets and consequently, their reported accuracies cannot be compared.

In fact, the acoustic characteristics of vowels differ in different languages. There are some researches such as [30,31] in which the differences of vowels in the different languages have been investigated. That is why, each language needs its special technique for vocal fold pathology detection system. So, the existing methods for other languages such as English or Arabic or Korean cannot be used for Russian language. Of course our main aim is to develop a high-efficient method for vocal fold pathology detection based on Russian language.

Some of the previous works [9-12,17,20,21,23,24,27,29,32-53] for the vocal fold pathology classification problem have been analyzed from the dataset and classifier points of view [see Additional file 1]. Of course, due to the lack of the same conditions, the reported accuracies of the previous works cannot be compared in order to decide about the best classifier or method.

So, with respect to these differences, a same infrastructure for our research in Russian language should be established. Also, two good nominates [12,32] of the previous works under our conditions (e.g., the same datasets) should be implemented in order to compare their performance with ours. For this purpose, a dataset in Russian language was created by the experts of the Belarusian Republican Center of Speech, Voice and Hearing Pathologies. Its details will be presented in the Section 2. Also, the well-known MEEI dataset, which is employed by these nominates, is used.

As it can be seen in Additional file 1, another disadvantage of some of the previous works for the vocal fold pathology detection systems is that their reported classification accuracies are often about 90% or even less. These amounts of accuracies cannot be quite sufficient because this problem is related to the health of human beings and

it has vital role. But in this research, this problem is taken into account as the main goal of our research and it is tried to get classification accuracy up to a satisfiable amount by the use of feature dimensional reduction methods.

The rest of the paper is organized as follows. In Section 2, the datasets are described. In Section 3, the initial feature vector based on the combination of the MFCC and the WPD is presented. In Section 4, optimizing of the initial feature vector by the means of the feature reduction methods is investigated. In Section 5, the ANN as a classifier is described. Experimental results and analysis are summarized in Section 6. Section 7 concludes the paper.

2 Datasets

Our dataset (RusDS) was created by the specialists from the Belarusian Republican Center of Speech, Voice and Hearing Pathologies. It includes 500 healthy samples (related to 500 healthy persons) as well as 500 pathological samples (related to 500 patients with vocal fold paralysis). The information of subjects can be seen in Table 1. In recording, the utterers pronounce vowel 'a' for 1 s. Also, they read a given special text. All of the samples are the wave files in the PCM format, in mono mode, sample rate of 44,100 Hz, and bit-depth of 16 bit.

Also, the well-known MEEI dataset is used which was created by Massachusetts Eye and Ear Infirmary. It includes approximately 700 records. The acoustic samples are the sustained phonation of the vowel /ah/ (1 to 3 s long). The speech samples were collected in a controlled environment and sampled with a 50- or 25-kHz sampling rate and 16 bits of resolution. The subset taken from this dataset contains 53 normal and 173 pathological samples.

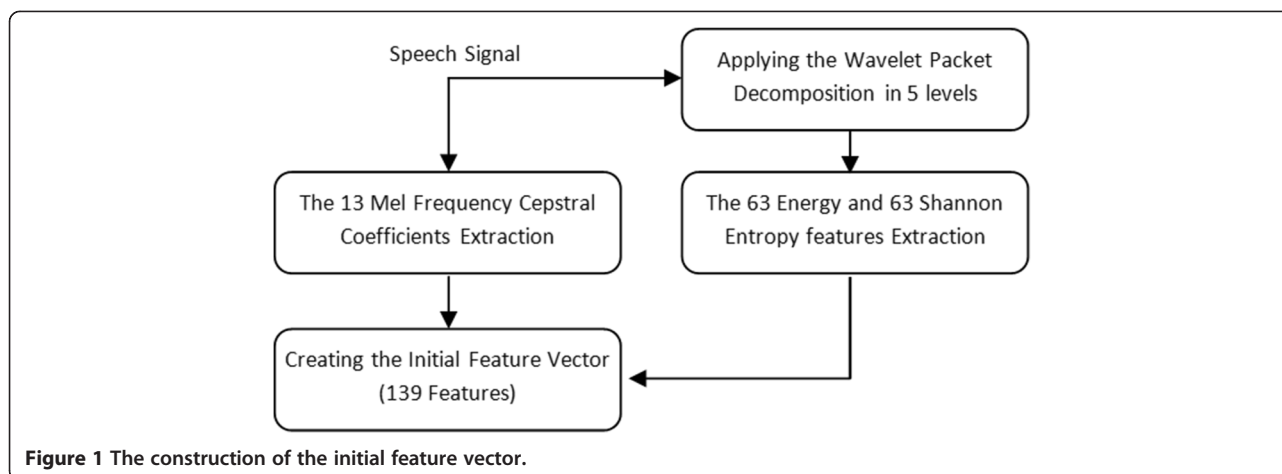
3 Initial feature vector

As it is shown in Figure 1, first, by the use of cepstral representation of input signal, 13 Mel frequency cepstral coefficients (MFCCs) are extracted. Then, the wavelet packet decomposition, in five levels, is applied on the input signal to make the wavelet packet tree. The structure of the obtained wavelet packet tree with the 63 nodes is illustrated in Figure 2. Then, from the nodes of the obtained wavelet packet tree, the 63 energy features as well as the 63 Shannon entropy features are extracted. Finally, by the combination of these features, the initial feature vector with the length of 139 features is constructed.

The Matlab code to calculate the MFCC features is adapted from the Auditory Toolbox (Malcolm Slaney). The speech signal is windowed with a Hamming window

Table 1 The details of the RusDS dataset

Subjects	Gender	Min, average, max ages
Healthy	250 males/250 females	18, 31.1, 79
Patient	290 males/210 females	29, 46.3, 83



in the time domain and by the use of the fast Fourier transform (FFT) converted into the frequency domain which gives the magnitude of the FFT. Then, the FFT data is converted into the filter bank outputs and the cosine transform is found to reduce dimensionality. The filter bank is made by using the 13 linearly spaced filters (133.33 Hz between center frequencies,) followed by the 27 log-spaced filters (separated by a factor of 1.0711703 in frequency). Each filter is made by the combination of the amplitude of FFT bin. Also, the tenth order of Daubechies wavelets ‘db10’ is used for the decomposition of signals.

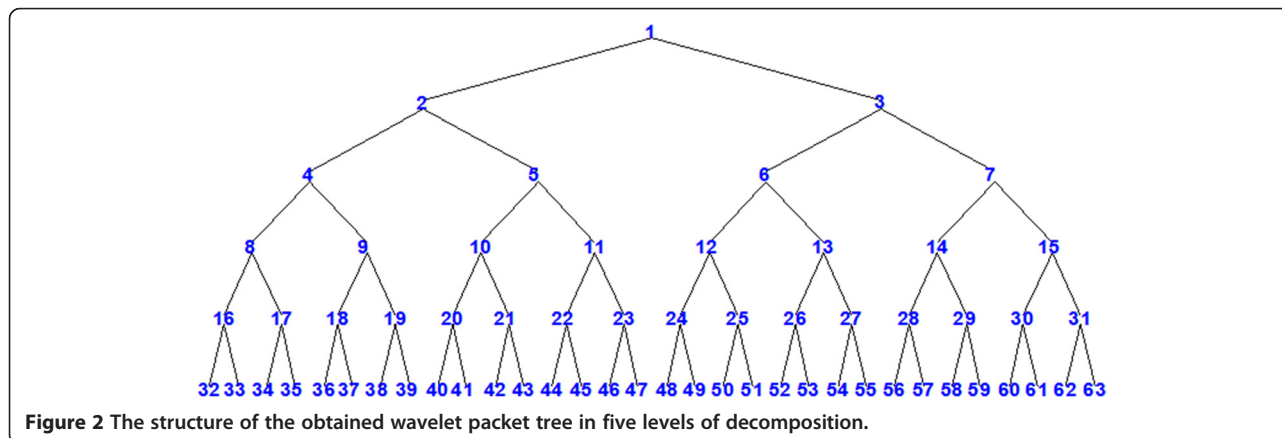
4 Optimized feature vector

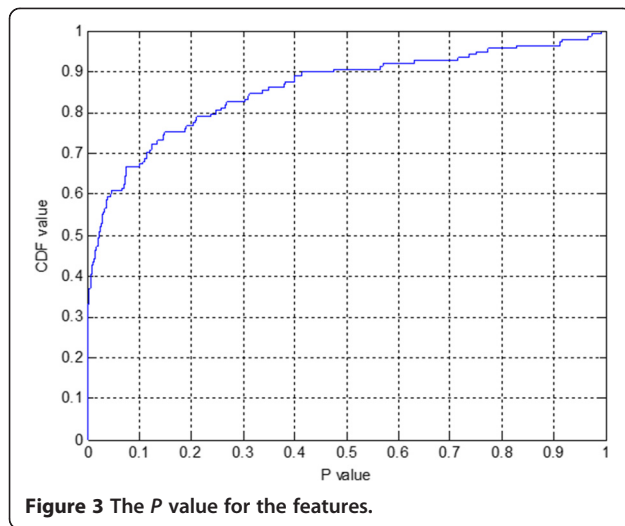
The discrimination power of the initial features can be evaluated by the use of *t* test. It can be used to investigate whether the means of two groups are statistically different from each other or not. For this purpose, it calculates a ratio between the difference of the two group's means and the variability. So, the *t* test is applied on each feature in our data set and compared to the *P* value for each feature as a measure of how effective it is at separating groups. The result is shown in Figure 3.

There are about 40% of features having *P* values close to 0, and 60% of features having *P* values smaller than 0.05 means that there are about 83 features among the original 139 features which have strong discrimination power. Also, 40% of the initial features (56 features) are not strong for the classification purpose. Therefore, feature reduction phase is necessary and important for our task. However, it is very difficult to know how many features are required unless someone has some domain knowledge or the maximum number of features has been dictated in advance based on outside constraints.

So, using every feature for classification process is not good idea, and it may cause to increase the misclassification error rate. Therefore, it is better to select the proper features from the whole features. This process is called ‘Feature Reduction’ or ‘Feature Selection.’ 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.

It is possible to categorize the feature reduction algorithms into two categories: filter methods and wrapper methods. Filter methods focus on the general characteristics of the data to evaluate and to select the feature subsets without





involving the chosen learning algorithm or the classifier. But wrapper methods use the performance feedbacks of the chosen learning algorithm or the classifier to evaluate each candidate feature subset. Wrapper methods search for a subset of features which has better fit for the chosen learning algorithm or the classifier, but they can be significantly slower than filter methods if the learning algorithm takes a long time to learn.

In this article, a well-known approach of the filter methods, which is called the principal component analysis (PCA), is employed for the feature reduction phase. It is frequently used in the previous works. Also, a novel approach, the generic algorithm (GA)-based method, is proposed for the feature reduction phase. This method belongs to the wrapper methods.

The main limitation of the PCA is that it 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 misclassification error rate of the classifier in its fitness function and tries to minimize it. For this purpose, the chromosomes are defined as the vectors of integers (from 1 to 139) with the length equal to the expected length for the reduced feature vector. The value of each gene represents the feature's number (from the initial feature vector) which should be taken part in the reduced feature vector. Also, a fitness function f is defined which shows the misclassification error rate of the ANN classifier for the train set.

$$f = \frac{\sum_{i=1}^n |a_i - r_i|}{n}$$

The a_i is the result of classifier and the r_i is the real class for i th sample. The n is the number of samples in

the train set. The aim of the proposed GA-based method is to find the subset of the initial features so that they minimize the f .

Also, some of the parameters which are used in MATLAB for developing the proposed optimized feature vector are shown in Table 2. The uniform mutation function is a two-step algorithm. First, the algorithm selects a fraction of the vector entries of an individual for mutation, and then in the second step, the algorithm replaces each selected entry by a random number selected uniformly from the range for that entry. The scattered crossover function creates a random binary vector, selects the genes where the vector is a 1 from the first parent and the genes where the vector is a 0 from the second parent, and combines the genes to form the child.

5 Classifier

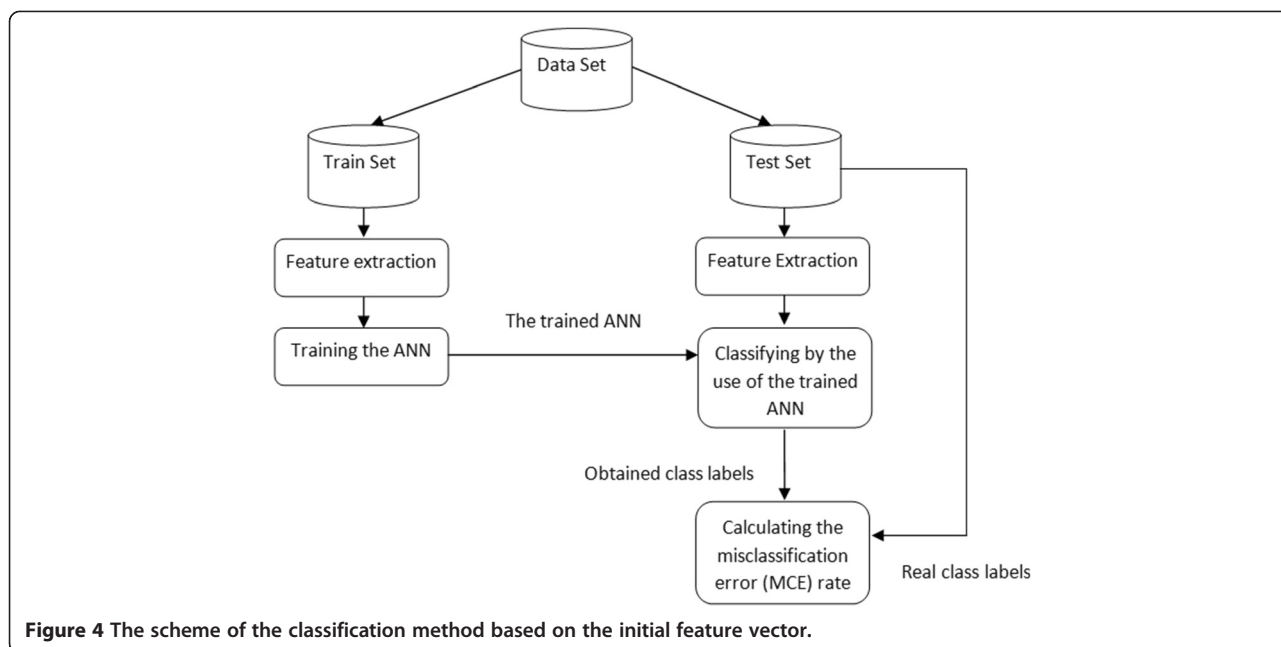
Artificial neural networks generally consist of several layers of interconnected nodes, each node generating a non-linear function of its inputs. The inputs to a node may come directly from the input data or from other nodes. Also, some nodes are considered as the output of the network. In this article, the ANN is used in the term of supervised learning so that a neural network is trained by giving a target output to a certain input group. A simplest way is to use a feed-forward neural network. The inputs form the input nodes of the network; the outputs are taken from the output nodes. The middle layer of nodes, visible to neither the inputs nor the outputs, is termed the hidden layer, and unlike the input and output layers, its size is not fixed.

6 Experiments and results

In this section, five experiments have been designed. These experiments are simulated in the MATLAB. In the experiments, the performances of three famous learning algorithms for the ANN are evaluated which are resilient backpropagation ('trainrp' in MATLAB), scaled conjugate gradient backpropagation ('trainscg' in MATLAB), and gradient descent with momentum and adaptive learning

Table 2 The implementation's parameters

Parameter	Description	Value
Population size	Indicating the number of individuals	60
Generations	Maximum number of generations allowed	100
Stall time limit	Maximum time over which change in fitness function value is less than zero	600
Time limit	Maximum time (in seconds) allowed	1,800
Crossover fraction	The fraction of genes swapped between individuals	0.5
Crossover function	Function used to do crossover	Crossover scattered
Mutation function	Function used in mutating genes	Mutation uniform



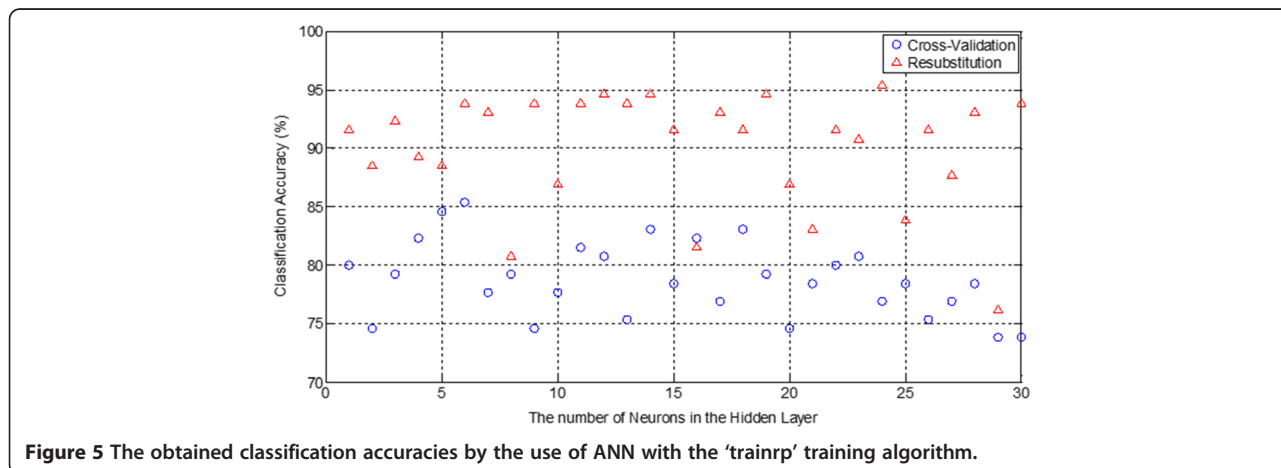
rate backpropagation ('traingdx' in MATLAB). Also, the different numbers of neurons in the hidden layer of the ANN is investigated. In the experiments, the tenfold cross-validation scheme has been adapted to assess the generalization capabilities of the system in the obtained results. Also, in the first experiment, the resubstitution error has been calculated.

In the first experiment, the classification has been done based on the initial feature vector which contains all the 139 features. As it is shown in Figure 4, first, for each sample in the train set, the 139 features are extracted according to the initial feature vector (13 MFCC, 63 energy, and 63 entropy features). Then, they are fed to the ANN for training it. After that, for each sample in the test set, the 139 features are extracted according to the initial feature vector. Then, they are fed to the trained ANN for classifying them. Finally,

the real class labels and the obtained class labels of the test set samples are compared to calculate the misclassification error rate.

The experiment result for the resilient backpropagation training algorithm, in the term of the classification accuracy, is illustrated in Figure 5. It is obvious that the tenfold cross-validation MCE is a good error estimation because it uses the samples in the test set that they do not take part in the training phase. So, from the tenfold cross-validation point of view, it is clear that the ANN with the resilient backpropagation training algorithm achieves a better performance (the accuracy of 85.4%) with the six neurons in its hidden layer.

The experiment result for the scaled conjugate gradient backpropagation training algorithm, in the term of the classification accuracy, is illustrated in Figure 6. From the



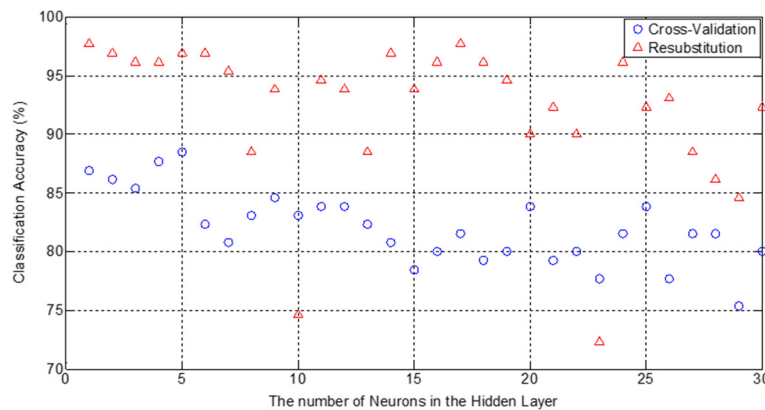


Figure 6 The obtained classification accuracies by the use of ANN with the 'trainscg' training algorithm.

tenfold cross-validation point of view, it is clear that the ANN with the scaled conjugate gradient backpropagation training algorithm achieves a better performance (the accuracy of 88.5%) with the five neurons in its hidden layer.

The experiment result for the gradient descent with momentum and adaptive learning rate backpropagation training algorithm, in the term of the classification accuracy, is illustrated in Figure 7. From the tenfold cross-validation point of view, it is clear that the ANN with the gradient descent with momentum and adaptive learning rate backpropagation training algorithm achieves a better performance (the accuracy of 88.5%) with the ten neurons in its hidden layer.

Finally, the results show a better performance of the 'traingdx' in comparison with the others in terms of classification accuracy. It has a better performance (the accuracy of 100%) in the resubstitution case and also a better performance in the tenfold cross-validation case (the accuracy of 88.5%) in comparison with others.

In the second experiment, the classification has been done based on the optimized feature vector which is obtained by the use of the PCA-based method. As it is

shown in Figure 8, first, for each sample in the train set, the 139 features are extracted according to the initial feature vector (13 MFCC, 63 energy, and 63 entropy features). Then, by the use of PCA-based method, the length of initial feature vector is reduced and the final feature vector is constructed. Then, according to the final feature vector, the selected feature values of the samples are fed to the ANN for training it. After that, for each sample in the test set, the 139 features are extracted according to the initial feature vector. Then, according to the final feature vector, the selected features values are fed to the trained ANN for classifying them. Finally, the real class labels and the obtained class labels of the test set samples are compared to calculate the misclassification error rate. Also, five neurons in the hidden layer of ANN are used. The different feature vectors with the length of 1 to 83 features are evaluated. The selected features as the final optimized feature vector, and the obtained classification accuracies are shown in Table 3.

In the third experiment, the classification has been done based on the optimized feature vector which is obtained by the use of the proposed GA-based method. As it is

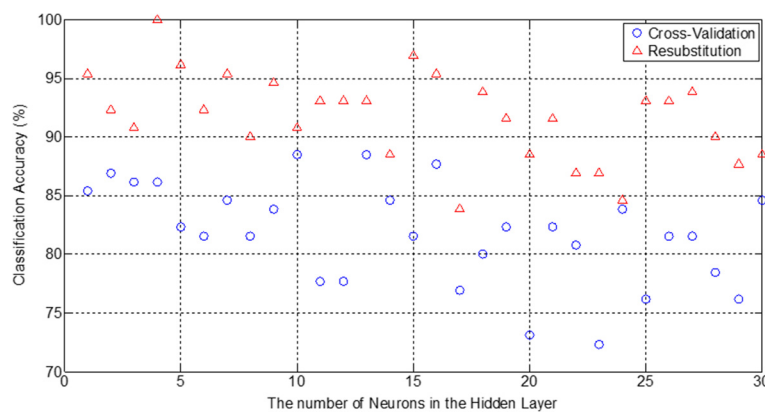
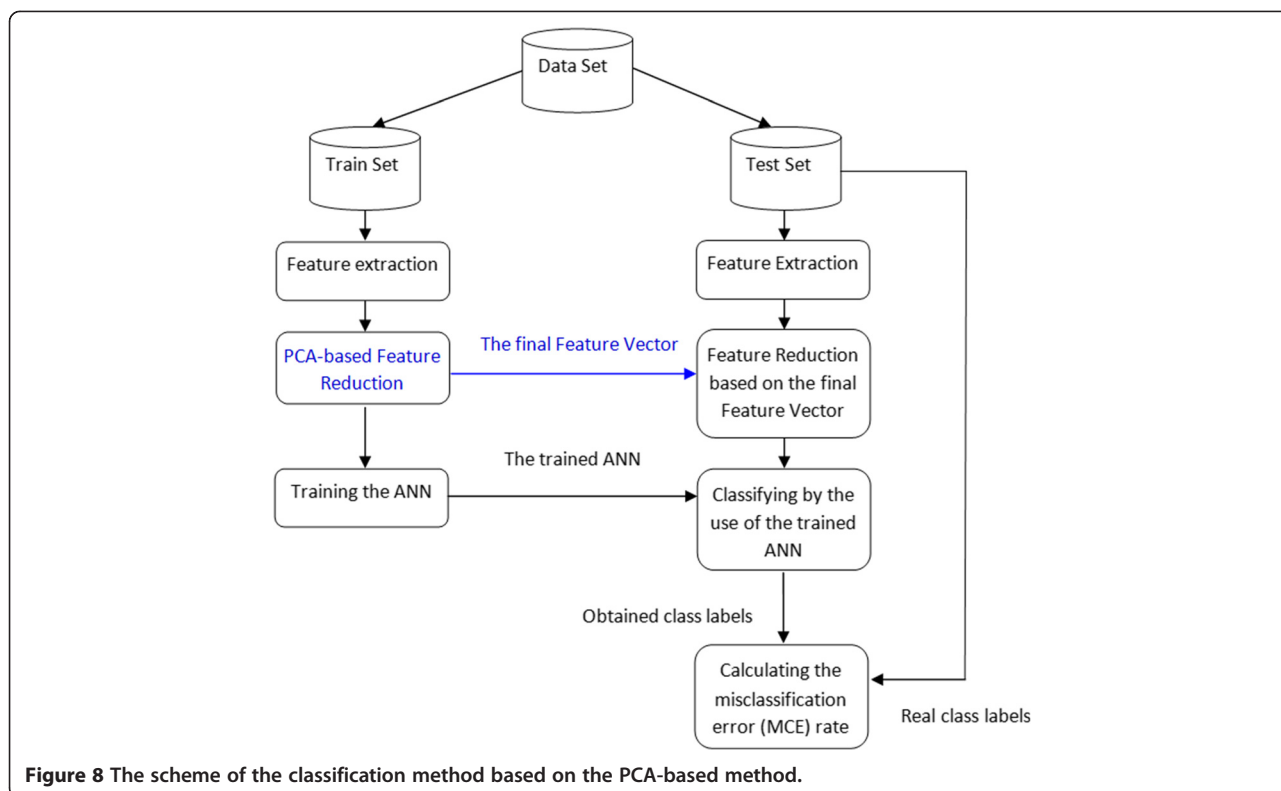


Figure 7 The obtained classification accuracies by the use of ANN with the 'traingdx' training algorithm.



shown in Figure 9, first, for each sample in the train set, the 139 features are extracted according to the initial feature vector (13 MFCC, 63 energy, and 63 entropy features). Then, by the use of GA-based method, the length of initial feature vector is reduced and the final feature vector is constructed. Then, according to the final feature vector, the selected feature values of the samples are fed to the ANN for training it. After that, for each sample in the test set, the 139 features are extracted according to the initial feature vector. Then, according to the final feature vector, the selected features values are fed

to the trained ANN for classifying them. Finally, the real class labels and the obtained class labels of the test set samples are compared to calculate the misclassification error rate. Also, five neurons in the hidden layer of ANN are used. The different feature vectors with the length of 1 to 83 features are evaluated. The selected features as the final optimized feature vector, and the obtained classification accuracies are shown in Table 4.

The comparative results are shown in Table 5. As it is obvious in Table 5, both PCA-based and GA-based methods can lead to the increasing of the classification

Table 3 The selected features and the obtained classification accuracies by the use of the PCA-based method

Training algorithm	Selected features	Classification accuracy (%)
'trainrp'	The first coefficient of MFCCs Energy at the eight, 18th, 19th, 32nd, 33th, and 35th nodes of WP tree Entropy at the first, fourth, eight, 16th, 17th, 32nd, 35th, and 39th nodes of WP tree (Feature vector length = 15)	89.9
'trainscg'	The first and second coefficients of MFCCs Energy at the eight, 18th, 19th, 32nd, 33th, and 35th nodes of WP tree Entropy at the first, fourth, eight, 16th, 17th, 32nd, 35th, and 39th nodes of WP tree (Feature vector length = 16)	90.7
'traingdx'	The first, second, fourth, and fifth coefficients of MFCCs Energy at the eight, 18th, 19th, 32nd, 33th, and 35th nodes of WP tree Entropy at the first, fourth, eight, 16th, 17th, 32nd, 35th, and 39th nodes of WP tree (Feature vector length = 18)	89.9

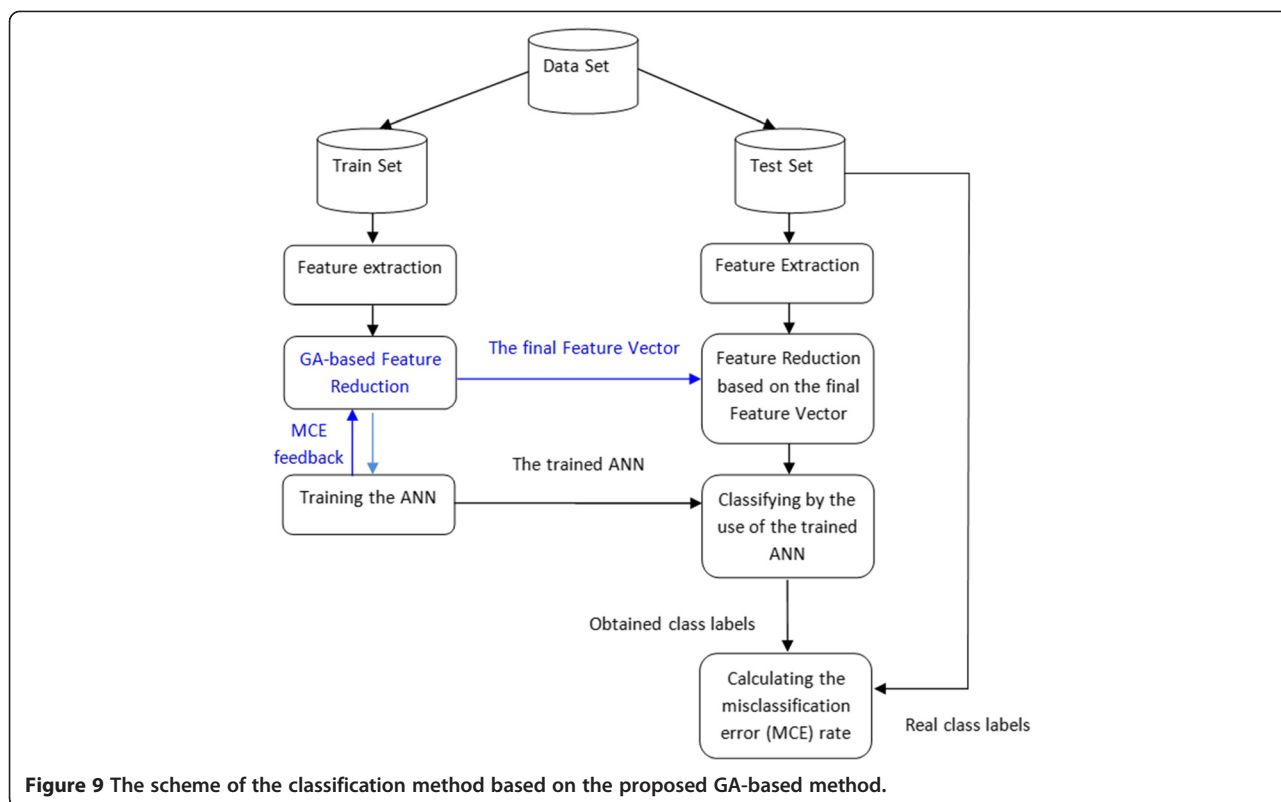


Table 4 The selected features and the obtained classification accuracies by the use of the GA-based method

Training algorithm	Selected features	Classification accuracy (%)
'trainrp'	The first, fifth, and eight coefficients of MFCCs Energy at the seventh, ninth, 36th, 40th, 42nd, 48th, 49th, 51st, 61st, and 62nd nodes of WP tree Entropy at the first, seventh, eight, 16th, 27th, 28th, 29th, 37th, 45th, 46th, 53rd, and 61st nodes of WP tree (Feature vector length = 25)	89.9
'trainscg'	The first and fifth coefficients of MFCCs Energy at the 11th, 17th, 22nd, 25th, 29th, 35th, 40th, 41st, 45th, 46th, 49th, 55th, 61st, and 63rd nodes of WP tree Entropy at the first, 13th, 15th, 19th, 26th, 35th, 36th, 42nd, 43rd, 44th, 46th, 52nd, 53rd, and 55th nodes of WP tree (Feature vector length = 30)	95.3
'traingdx'	The first, third, and fifth coefficients of MFCCs Energy at the eight, 20th, 21st, 43rd, 60th, and 62nd node of WP tree Entropy at the 11th, 13th, 35th, 52nd, 54th, 58th, 59th, and 62nd nodes of WP tree (Feature vector length = 17)	93.8

accuracies. Of course, in terms of classification accuracy, the proposed GA-based method has a better performance in comparison with the PCA-based method. Also, from the training algorithm point of view, the 'trainscg' shows better results than the others.

Finally, the experiments' results show a better performance of the proposed method which is based on the hybrid of the ANN with the 'trainscg' algorithm as the classifier and the GA-based method as the feature reduction approach. It provides the best classification accuracy (95.3% of accuracy) in comparison with others. It also leads to reduce the length of feature vector from 139 to 30 features. So, the response time of the vocal fold pathology classification system based on the initial feature vector (with the length of 139) and the reduced feature vector (with the length of 30) should be different.

The fourth experiment is carried out to compare the response time of the vocal fold pathology classification system based on the initial feature vector and the reduced feature vector. This experiment has carried out on a

Table 5 The obtained classification accuracies (%)

Feature reduction method	'trainrp'	'trainscg'	'traingdx'
-	85.4	88.5	88.5
PCA-based method	89.9	90.7	89.9
Proposed GA-based method	89.9	95.3	93.8

Table 6 The comparison of the proposed method with recent works

Reference	Feature vector	Feature selection approach	Classifier	Accuracy based on the MEEI (%)	Accuracy based on the RusDS (%)
The proposed method	WP energy, WP entropy, MFCC	Genetic algorithm	Artificial neural network	94.24	95.3
[32]	WP entropy	Genetic algorithm	Support vector machine with radial basis function	94.69	92.1
[12]	Harmonic-to-noise ratio, normalized noise energy, - glottal-to-noise excitation ratio, energy, MFCC	-	HMM with two states	93.8	90.3

personal computer which is equipped by the processor of Intel dual-core 2.13 GHz and the memory of 2 GB. The response time in the case of initial feature vector (139 features), 8.7 ms is reported. The response time in the case of reduced feature vector (30 features), 4.6 ms is reported. Therefore, using the reduced feature vector leads to the decreasing of the response time of the program of vocal fold pathology classification in comparison with the non-reduced feature vector.

The fifth experiment is done in order to compare the performance of the proposed method with the recent works [12,32]. Of course, the effects of two different datasets (the MEEI and the RusDS) are investigated. The results are shown in Table 6. The results show the higher classification accuracy of the proposed method in comparison with recent works.

7 Conclusions

In this article, an initial feature vector based on the combination of the wavelet packet decomposition and the Mel frequency cepstral coefficients (MFCCs) is proposed. The performances of the ANN with the three kinds of training algorithms ('trianrp', 'trainscg', and 'traingdx') in the task of vocal fold pathology diagnosis are investigated. Also, the performance of the three kinds of feature vector (the initial feature vector, the optimized feature vector by means of the PCA-based method, and the optimized feature vector by means of the proposed GA-based method) in the task of vocal fold pathology diagnosis is evaluated. The experiments' results show the priority of the optimized feature vector by means of the proposed GA-based method in comparison with others. This better performance is due to taking into consideration of the ANN classifier in the feature reduction phase. In other words, the proposed GA-based method tries to optimize the initial feature vector with the aim of decreasing the misclassification error rate of the ANN. But the PCA-based method just focuses on the data without any attention on the misclassification error rate of the ANN classifier.

Finally, the proposed method is proposed based on the hybrid of the ANN with the 'trainscg' algorithm as the classifier and the GA-based method as the feature reduction

approach. It is concluded that the proposed method has the higher accuracy (95.3% of accuracy) and the lower response time in comparison with others. Also, the performance of the proposed method is compared with recent works [12,32]. For this purpose, the effects of two different datasets (the MEEI and the RusDS) are investigated. Finally, it is observed that the proposed method shows the higher classification accuracy in comparison with recent works.

Additional file

Additional file 1: Analysis of some of the previous works. Some of the previous works [9-12,17,20,21,23,24,27,29,32-53] for the vocal fold pathology classification problem have been analyzed from the dataset and classifier points of view.

Competing interests

The author declares that he has no competing interests.

Acknowledgements

The author wishes to thank the Belarusian Republican Center of Speech, Voice and Hearing Pathologies and also the faculty of Literature and Human Science at University of Tehran by their support in providing the RusDS and the MEEI datasets, respectively.

Received: 1 July 2014 Accepted: 19 December 2014

Published online: 21 January 2015

References

- G Chen, J Kreiman, A Alwan, The glottaltopogram: a method of analyzing high-speed images of the vocal folds. *Comput Speech Lang* **28**(5), 1156–1169 (2014). doi:10.1016/j.csl.2013.11.006
- DD Mehta, M Zaňartu, TF Quatieri, DD Deliyki, RE Hillman, Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy. *J Acoust Soc Am* **130**(6), 3999–4009 (2011). doi:10.1121/1.3658441
- P Kitzing, Glottography, the electrophysiological investigation of phonatory biomechanics. *Acta Otorhinolaryngol Belg* **40**(6), 863–878 (1986)
- DM Bless, M Hirano, RJ Feder, Videostroboscopic evaluation of the larynx. *Ear Nose Throat J* **66**(7), 289–296 (1987)
- C Manfredi, Adaptive noise energy estimation in pathological speech signals. *IEEE Trans Biomed Eng* **47**(11), 1538–1543 (2000). doi:10.1109/10.880107
- JB Alonso, JD Leon, I Alonso, MA Ferrer, Automatic detection of pathologies in the voice by HOS based parameters. *EURASIP J Appl Signal Process* **2001**(4), 275–284 (2001). doi:10.1155/S1110865701000336
- MDO Rosa, JC Pereira, M Grellet, Adaptive estimation of residue signal for voice pathology diagnosis. *IEEE Trans Biomed Eng* **47**(1), 96–104 (2000). doi:10.1109/10.817624
- C Jo, T Li, J Wang, Estimation of harmonic and noise components from pathological voice using iterative method. Paper presented at the 27th

- annual conference on IEEE Engineering in Medicine and Biology, Shanghai, China, 17–18 Jan. 2006, p. 4678–4681
9. P Gomez, F Diaz, C Lazaro, K Murphy, R Martinez, V Rodellar, A Alvarez, Spectral perturbation parameters for voice pathology detection. Paper presented at the International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, Romania, 14–15 July 2005, p. 299–302
 10. W Xu, H Zhiyan, W Jian, Pathological speech deformation degree assessment based on integrating feature and neural network. Paper presented at the 27th Chinese Control, Kunming, China, 16–18 July 2008, p. 441–444
 11. Y Wei, H GholamHosseini, A Cameron, MJ Harrison, A Al-Jumaily, Voice analysis for detection of hoarseness due to a local anesthetic procedure. Paper presented at the 3rd International Conference on Signal Processing and Communication Systems (ICSPCS 2009), Omaha, NE, 28–30 September 2009, p. 1–7
 12. M Sarria-Paja, G Castellanos-Domínguez, E Delgado-Trejos, A new approach to discriminative HMM training for pathological voice classification. Paper presented at the 32nd annual international conference of the IEEE EMBS, Buenos Aires, Argentina, 31 August–4 September 2010, p. 4674–4677
 13. T Li, C Jo, Discrimination of severely noisy pathological voice with spectral slope and HNR. Paper presented at the 7th International Conference on Signal Processing, (ICSP '04), Beijing, China, 31 August– 4 September 2004, p. 2218–2221
 14. JHL Hansen, L Gavidia-Ceballos, JF Kaiser, A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment. *IEEE Trans Biomed Eng* **45**(3), 300–313 (1998). doi:10.1109/10.661155
 15. OG Fetisova, DV Lamtyugin, VK Makukha, EM Voronin, Spectrum analysis of vocalization application for voice pathology detection. Paper presented at the international conference on computer as a tool (EUROCON2007), Warsaw, Poland, 9–12 September 2007, p. 2725–2728
 16. V Majidnezhad, I Kheidorov, A hybrid of genetic algorithm and support vector machine for feature reduction and detection of vocal fold pathology. *Int J Image Graph Signal Process* **5**(9), 1–7 (2013). doi:10.5815/ijigsp.2013.09.01
 17. M Markaki, Y Stylianou, Voice pathology detection and discrimination based on modulation spectral features. *IEEE Trans Audio Speech Lang Process* **19**(7), 1938–1948 (2011). doi:10.1109/TASL.2010.2104141
 18. V Majidnezhad, I Kheidorov, A novel method for feature extraction in vocal fold pathology diagnosis. Paper presented at the 3rd International Conference MobilHealth2012, Paris, France, 21–23 November 2012, p.96-105
 19. V Majidnezhad, I Kheidorov, The SVM-based feature reduction in vocal fold pathology diagnosis. *Int J Future Generation Commun Netw* **6**(1), 45–55 (2013)
 20. JI Godino-Llorente, P Gomez-Vilda, M Blanco-Velasco, Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. *IEEE Trans Biomed Eng* **53**(10), 1943–1953 (2006). doi:10.1109/TBME.2006.871883
 21. JD Arias-Londono, JI Godino-Llorente, N Saenz-Lechon, V Osma-Ruiz, G Castellanos-Domínguez, Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. *IEEE Trans Biomed Eng* **58**(2), 370–379 (2011). doi:10.1109/TBME.2010.2089052
 22. V Majidnezhad, I Kheidorov, A novel GMM-based feature reduction for vocal fold pathology diagnosis. *Res J Appl Sci Eng Technol* **5**(6), 2245–2254 (2013)
 23. RTS Carvalho, CC Cavalcante, PC Cortez, Wavelet transform and artificial neural networks applied to voice disorders identification. Paper presented at the 3rd World Congress on Nature and Biologically Inspired Computing (NaBiC), Salamanca, Spain, 19–21 October 2011, p. 371–376
 24. T Drugman, T Dubuisson, T Dutoit, Phase-based information for voice pathology detection. Paper presented at the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011, p. 4612–4615
 25. V Majidnezhad, I Kheidorov, An ANN-based method for detecting vocal fold pathology. *Int J Comput Appl* **62**(7), 1–4 (2013)
 26. L Gavidia-Ceballos, JHL Hansen, Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection. *IEEE Trans Biomed Eng* **43**(4), 373–383 (1996). doi:10.1109/10.486257
 27. JD Arias-Londono, JI Godino-Llorente, G Castellanos-Domínguez, N Saenz-Lechon, V Osma-Ruiz, Complexity analysis of pathological voices by means of hidden markov entropy measurements. Paper presented at the 31st annual international conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, Minnesota, USA, 3–6 September 2009, p. 2248–2251
 28. V Majidnezhad, I Kheidorov, A HMM-based method for vocal fold pathology diagnosis. *Int J Comput Sci Issues* **9**(6), 135–138 (2012)
 29. M Markaki, Y Stylianou, JD Arias-Londono, JI Godino-Llorente, Dysphonia detection based on modulation spectral features and cepstral coefficients. Paper presented at the International Conference on Acoustics Speech and Signal Processing (ICASSP 2010), Dallas, 14–19 March 2010, p. 5162–5165
 30. K Tsukada, An acoustic comparison of vowel length contrasts in Arabic, Japanese and Thai: durational and spectral data. *Int J Asian Lang Process* **19**(4), 127–138 (2009)
 31. J Vaisiere, On the acoustic and perceptual characterization of reference vowels in a cross-language perspective. Paper presented at the 17th International Congress of Phonetic Sciences (ICPhS XVII), Hong Kong, August 2011, p. 52–59
 32. R Behroozmand, F Almasganj, Comparison of neural networks and support vector machines applied to optimized features extracted from patients' speech signal for classification of vocal fold inflammation. Paper presented at the IEEE International Symposium on Signal Processing and Information Technology, Athens, 21 December 2005, p. 844–849
 33. MDO Rosa, JC Pereira, ACPLF Carvalho, Evaluation of neural classifiers using statistic methods for identification of laryngeal pathologies. Paper presented at the 5th Brazilian Symposium on Neural Networks, Belo Horizonte, Brazil, December 1998, p. 220–225
 34. JP Papa, AA Spadotto, AX Falcao, JC Pereira, Optimum path forest classifier applied to laryngeal pathology detection. Paper presented at the 15th International Conference on Systems, Signals and Image Processing, (IWSSIP 2008), Bratislava, 25–28 June 2008, p. 249–252
 35. G Muhammad, M Alsulaiman, A Mahmood, Z Ali, Automatic voice disorder classification using vowel formants. Paper presented at the IEEE International Conference on Multimedia and Expo (ICME), Barcelona, 11–15 July 2011, p. 1–6
 36. Z Mahmoudi, S Rahati, MM Ghasemi, V Asadpour, H Tayaran, Classification of voice disorder in children with cochlear implantation and hearing aid using multiple classifier fusion. Paper presented at the 10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010), Kuala Lumpur, Malaysia, 10–13 May 2010, p. 304–307
 37. ES Fonseca, RC Guido, AC Silvestre, JC Pereira, Discrete wavelet transform and support vector machine applied to pathological voice signals identification. Paper presented at the Seventh IEEE International Symposium on Multimedia (ISM2005), Brazil, 12–14 December 2005, p. 785–789
 38. RC Guido, JC Pereira, ES Fonseca, CD Maciel, LS Vieira, FLSMBA Guilermo, S Barbon, Support vector machines and wavelets for voice disorder sorting. Paper presented at the 38th Southeastern IEEE Symposium on System Theory, Tennessee Technological University, Cookeville, USA, 5–7 March 2006, p. 434–438
 39. R Behroozmand, F Almasganj, MH Moradi, Pathological assessment of vocal fold nodules and polyp using acoustic perturbation and phase space features. Paper presented at the IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP 2006), Toulouse, 14–19 May 2006, p.1056- 1059
 40. G Vaziri, F Almasganj, Pathological assessment of vocal fold nodules and polyp via fractal dimension of patients' voices. Paper presented at the 2nd IEEE International Conference on Bioinformatics and Biomedical Engineering, (ICBBE 2008), Shanghai, China, 16–18 May 2008, p. 2044–2047
 41. J Lohscheller, Towards evidence based diagnosis of voice disorders using phonovibrograms. Paper presented at the 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies, Bratislava, 24–27 November 2009, p. 1–4
 42. J Wang, C Jo, Vocal folds disorder detection using pattern recognition methods. Paper presented at the 29th Annual International Conference of the IEEE EMBS, Lyon, France, 22–26 August 2007, p. 3253–3256
 43. L Nayak, PH Bhat, Identification of voice disorders using speech samples. Paper presented at the Conference on Convergent Technologies for the Asia-Pacific Region, (TENCON 2003), 15–17 October 2003, p. 951–953
 44. CE Martinez, HL Rufiner, Acoustic analysis of speech for detection of laryngeal pathologies. Paper presented at the 22nd Annual International Conference on EMBS, Chicago, USA, 23–28 July 2000, p. 2369–2372
 45. MP Paulraj, S Yaacob, M Hariharan, Diagnosis of vocal fold pathology using time-domain features and systole activated neural network. Paper presented at the 5th International Colloquium on Signal Processing & Its Applications, (CSPA 2009), Kuala Lumpur, Malaysia, 6–8 March 2009, p. 29–32
 46. AA Dibazar, TW Berger, SS Narayanan, Pathological voice assessment. Paper presented at the 28th IEEE Annual International Conference on EMBS, New York City, USA, 30 August–3 September 2006, p. 1669–1673
 47. EJ Wallen, JHL Hansen, A screening test for speech pathology assessment using objective quality measures. Paper presented at the Fourth

- International Conference on Spoken Language Proceedings, (ICSLP 96), Philadelphia, 3–6 October 1996, p. 776–779
48. SC Costa, BGA Neto, JM Fechine, Pathological voice discrimination using cepstral analysis, vector quantization and hidden Markov models. Paper presented at the 8th IEEE International Conference in Bioinformatics and BioEngineering, (BIBE 2008), Athens, 8–10 October 2008, p. 1–5
 49. JY Lee, M Hahn, Automatic assessment of pathological voice quality using higher-order statistics in the LPC residual domain. *EURASIP J Adv Signal Process* **2009**(748207), 1–8 (2009). doi:10.1155/2009/748207
 50. JY Lee, S Jeong, M Hahn, Pathological voice detection using efficient combination of heterogeneous features. *IEICE Trans Inf Syst* **E91-D**(2), 367–370 (2008)
 51. T Ananthakrishna, K Shama, UC Niranjana, k-Means nearest neighbor classifier for voice pathology. Paper presented at the IEEE India Annual Conference, (INDICON 2004), India, 20–22 December 2004, p. 352–354
 52. M Hariharan, MP Paulraj, S Yaacob, Identification of vocal fold pathology based on mel frequency band energy coefficients and singular value decomposition. Paper presented at the International IEEE Conference on Signal and Image Processing Applications, (ISCIAP 2009), Kuala Lumpur, Malaysia, 18–19 November 2009, p. 514–517
 53. MK Arjmandi, M Pooyan, H Mohammadnejad, M Vali, Voice disorders identification based on different feature reduction methodologies and support vector machine. Paper presented at the 18th Iranian Conference on Electrical Engineering (ICEE 2010), Isfahan, Iran, 11–13 May 2010, p. 45–49

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
