



Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals

Ali, Z., Elamvazuthi, I., Alsulaiman, M., & Muhammad, G. (2016). Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals. *Journal of Medical Systems*, 40(1), 1-10. Article 20. <https://doi.org/10.1007/s10916-015-0392-2>

[Link to publication record in Ulster University Research Portal](#)

Published in:

Journal of Medical Systems

Publication Status:

Published (in print/issue): 01/01/2016

DOI:

[10.1007/s10916-015-0392-2](https://doi.org/10.1007/s10916-015-0392-2)

Document Version

Author Accepted version

General rights

The copyright and moral rights to the output are retained by the output author(s), unless otherwise stated by the document licence.

Unless otherwise stated, users are permitted to download a copy of the output for personal study or non-commercial research and are permitted to freely distribute the URL of the output. They are not permitted to alter, reproduce, distribute or make any commercial use of the output without obtaining the permission of the author(s).

If the document is licenced under Creative Commons, the rights of users of the documents can be found at <https://creativecommons.org/share-your-work/licenses/>.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk

Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals

Zulfiqar Ali^{1,2}, Irraivan Elamvazuthi², Mansour Alsulaiman¹, Ghulam Muhammad¹

¹Digital Speech Processing Group, Department of Computer Engineering
King Saud University
Riyadh 11543, Saudi Arabia.
{zuali, msuliman, ghulam}@ksu.edu.sa

²Centre for Intelligent Signal and Imaging Research, Department of Electrical and Electronic Engineering
Universiti Teknologi PETRONAS
Tronoh 31750, Perak, Malaysia.
zulfiqar_g02579@utp.edu.my, irraivan_elamvazuthi@petronas.com.my

Abstract

Voice disorders are associated with irregular vibrations of vocal folds. Based on the source filter theory of speech production, these irregular vibrations can be detected in a non-invasive way by analyzing the speech signal. In this paper we present a multiband approach for the detection of voice disorders given that the voice source generally interacts with the vocal tract in a non-linear way. In normal phonation, and assuming sustained phonation of a vowel, the lower frequencies of speech are heavily source dependent due to the low frequency glottal formant, while the higher frequencies are less dependent on the source signal. During abnormal phonation, this is still a valid, but turbulent noise of source, because of the irregular vibration, affects also higher frequencies. Motivated by such a model, we suggest a multiband approach based on a three-level discrete wavelet transformation (DWT) and in each band the fractal dimension (FD) of the estimated power spectrum is estimated. The experiments suggest that frequency band 1 - 1562 Hz, lower frequencies after level 3, exhibits a significant difference in the spectrum of a normal and pathological subject. With this band, a detection rate of 91.28% is obtained with one feature, and the obtained result is higher than all other frequency bands. Moreover, an accuracy of 92.45% and an area under receiver operating characteristic curve (AUC) of 95.06% is acquired when the FD of all levels is fused. Likewise, when the FD of all levels is combined with 22 Multi-Dimensional Voice Program (MDVP) parameters, an improvement of 2.26% in accuracy and 1.45% in AUC is observed.

Keywords: Voice pathology detection, wavelet transformation, fractal dimension, Katz algorithm, Higuchi algorithm, MDVP parameters

1 Introduction

The attributes of the elements of a speech production system of a person suffering from vocal folds disorder differ from those of a normal person. Phonation, resonance, and articulation are three important phases to produce speech. In a speech production system, expired air and pressure induce vocal folds vibration. Vocal folds directly affect phonation and resonance in the speech production system, and abnormal behavior of these folds makes the voice strained and harsh due to tight closure of the vocal fold, whereas excessive distance between the vocal folds make the voice weaker, whispering, and breathier. The lack of control regarding the length and tension of vocal folds distorts the voice due to uncontrolled pitch breaks and makes it flat and expressionless. Moreover, suppleness of the vocal folds' mucosa provides free vibration. The resonance is a modification of phonation and produces voice. The unmodified and phonated sound of the vocal folds is a strident and unpleasant noise [1]. Some of the more common

and most spreading voice disorders are vocal fold nodules, keratosis, vocal fold paralysis, and adductor spasmodic dysphonia [2], [3].

The use of subjective measures [4] for pathology detection depends on individual experience and area of expertise. On the other hand, an objective assessment that includes acoustical analysis is independent of human biasness and can assess the voice quality more reliably by relating certain parameters to vocal fold behavior. Various types of the automatic voice pathology detection (AVPD) systems are developed using different kinds of acoustic features [5], [6], [7], [8]. The researchers are trying to develop efficient detection systems with good accuracies by using fewer features and processing steps. The advantage of such a system is that they can be used in web-based applications to prevent patients from having to visit a hospital. It is also observed that few studies have introduced features from other scientific areas to detect voice pathology, such as MPEG-7 and local binary pattern (LBP) based features have been used in [9] and [10], respectively, for AVPD.

Like other features, fractal dimension (FD) have also been used to analyze biomedical signals. Fractal geometry was invented by Mandelbrot in 1977 and used for many applications in diverse fields of science and engineering [11]. Various algorithms, such as Katz [12], Higuchi [13], Petrosian [14], Maragos [15], and the amplitude scale method [16], were proposed to estimate the FD of time series/waveforms. Biomedical signals such as electrocardiogram (ECG), electroencephalogram (EEG), and speech can also be characterized as a time series, and hence, different FD algorithms were applied for the classification of biomedical signals [17], [18]. A comparison of some FD estimation algorithm is presented in [19]. Katz and Higuchi methods are compared in [20] to show the dependency of the algorithm on amplitude, frequency, and sampling frequency of the waveform.

FDs measure the complexity of a waveform [21], [22] and, based on this measure, a waveform can be characterized into different classes. Voice disorders affect the voice box and produce irregular vibrations in the vocal folds. Due to vocal fold irregularity, the speech signal becomes more transient in comparison with that of normal subjects, as shown in Figure 1. The Katz and Higuchi algorithms provide high FD fractal dimension for the signals that exhibit irregularities.

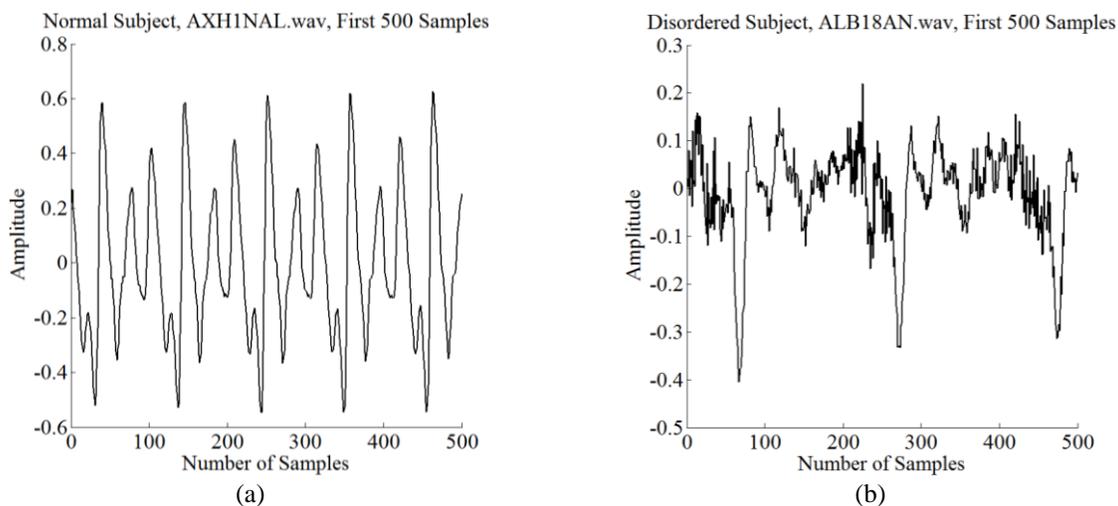


Figure 1. First 500 samples of (a) normal and (b) pathological subjects; the disordered subject is more transient in comparison with the normal subject

To the best of our knowledge, FD was used first by Accord in 1992 [23] for voice pathology detection when the author used short-term fractal dimension with two other parameters, zero crossing and the ratio between low- and high-frequency energy. Four normal and four dysarthria subjects were considered for the detection, and the results showed that when accompanied by other acoustic parameters, FD can differentiate voice pathology successfully. The FD was calculated from the method proposed by Maragos [15]. In 1997, Accord also estimated the FD of EEG signals using the Higuchi algorithm to classify different physio-pathological conditions [22].

Voice pathology detection was performed using FD and Mel-frequency cepstral coefficients (MFCC) in [21]. The FD was calculated directly from waveform in a time domain with three different methods: Katz algorithm [3], Higuchi algorithm, and Hurst exponent. A subset of the Massachusetts Eye & Ear Infirmary (MEEI) database [24] containing 173 pathological and 53 normal subjects was used for the experimental result. All samples are down sampled to 25 KHz, and different frame lengths with 50% overlapping were considered to extract the features. The maximum obtained accuracy rate obtained with FD was 87.72% using the Hurst exponent. The accuracy of 95.65% was obtained with MFCC. When the score level fusion of FD and MFCC was used, the obtained accuracy was 97.54%, which is only 1.89% better than the MFCC alone.

In a study [25], two new tools, recurrence and fractal scaling, were proposed and an accuracy of 91.8% was attained with 2.0% standard deviations, while obtained sensitivity and specificity were 95.4% and 91.5%, respectively. The comparison of the proposed features was done with combinations of different Multi-Dimensional Voice Program (MDVP) [24] parameters. The maximum accuracy provided by the MDVP was 81.4% with a combination of jitter and shimmer. Twenty-two MDVP parameters were used in [26] by Arjmandi and an accuracy of 89.29% was obtained, when 22 original features were used with a support vector machine (SVM) [27]. In [28], a short-term fractal dimension was computed to classify the normal and pathological subject. The highest obtained accuracy was 88.90%, and only two types of disorder, polyp and nodules, were considered in this study.

Different methods based on FD are proposed in this study to differentiate between disordered and normal subjects. By using the calculated FDs, which represent the complexity of the waveform, the signals are classified into two classes, normal and disordered. FDs are computed in both time and frequency domain. In time domain, the FD is estimated over a whole signal to produce a single value. In case of frequency domain, the signal is decomposed into multi-resolution frequency bands by applying a three-level discrete wavelet transformation (DWT), and the FD is calculated for each level. We implemented the Katz and Higuchi methods to estimate FD. MDVP parameters are also complemented with FD to enhance the system accuracy. SVM is implemented to differentiate between two types of subjects.

The rest of the paper is organized as follows; Section 2 presents the proposed methods and their major components. Section 3 gives brief information of the speech database. Section 4 provides the experimental results and a discussion, and Section 5 draws some conclusions.

2 Methodology

In this paper, different methods for pathology detection are proposed, and block diagrams of these methods are depicted in Figure 2. The first two methods use Katz and Higuchi algorithms, respectively, to compute FD in the time domain, as shown in Figure 2(a). The third and fourth methods decompose a speech signal into different frequency bands, and FD is calculated for each frequency band by using Katz and Higuchi algorithms, as depicted in Figure 2(b). In all methods, SVM is used to differentiate between normal and pathological subjects.

In terms of results, the approach depicted in Figure 3 is the best. The approach is a modification of the third method, where the FD, calculated by the Katz algorithm, of each frequency band is fused before being given to SVM. This approach provided better results than all other methods.

In the time domain, fractal algorithms are applied to the whole signal, and in the frequency domain the algorithms are applied to each level after decomposing the signal. No pre-processing, such as framing and windowing, is performed on the speech signal. The components of the proposed system are described in the following subsections. The major components are two FD algorithms, to capture the complexity of a signal, and DWT, to decompose a speech signal into frequency bands.

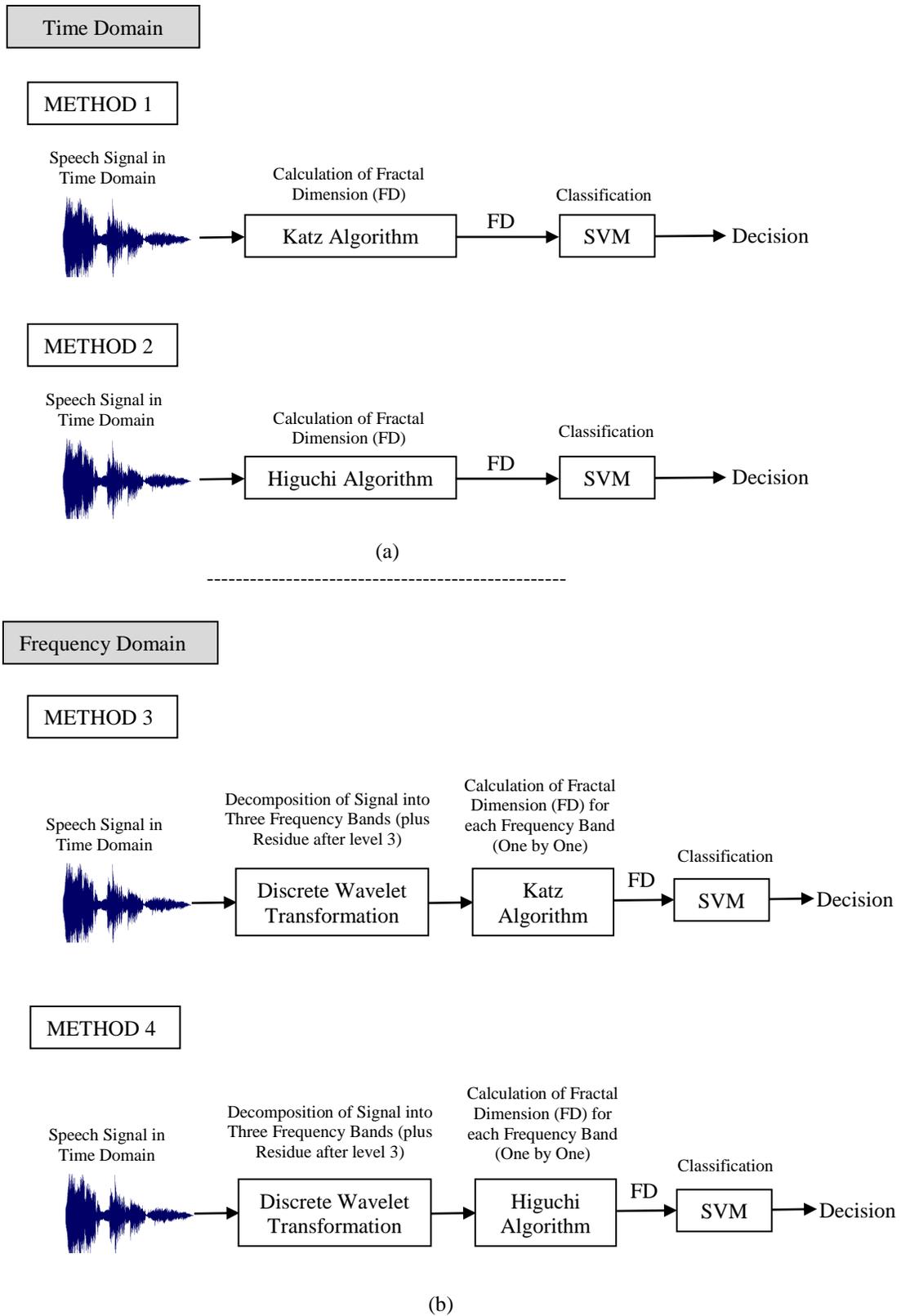


Figure 2. A block diagram of proposed methods

Best Approach in
term of Results

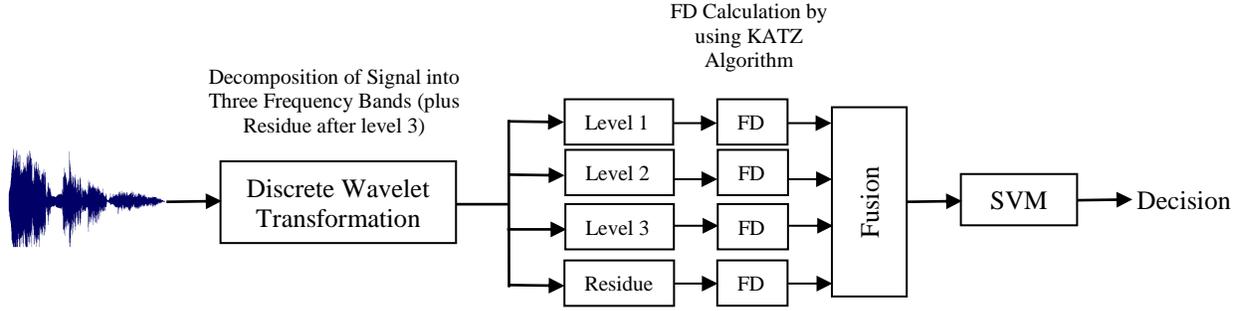


Figure 3. The best approach (a modification of the third method) in term of results

2.1 Katz Algorithm

The FD with Katz algorithm can be calculated as a ratio between the length of the curve, represented by L , and the maximum distance from the first point of the curve to any point on it, represented by d and also called planar extent. The length of the curve is a summation of Euclidean distances of all consecutive points of the curve. The roughness of the curve results in a larger length, which reveals the transient behavior of the curve. The FD for a curve C of n points with the Katz method is given by Eq. (1).

$$FD = \frac{\log_{10}\left(\frac{L}{\bar{a}}\right)}{\log_{10}\left(\frac{d}{\bar{a}}\right)}, \quad \bar{a} \text{ represents averaged length between two consecutive points} \quad (1)$$

where

$$L = \sum_{i=1}^{n-1} \left[(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 \right]^{1/2}$$

and

$$d = \max \left(\left[(x_{i+1} - x_1)^2 + (y_{i+1} - y_1)^2 \right]^{1/2} \right) \quad \forall i = 1, 2, 3, \dots, n-1$$

2.2 Higuchi Algorithm

The Higuchi method creates new curves from the original waveform by selecting different samples as the starting point, and delay factor k . For a given waveform X of n points: $X = x(1), x(2), x(3), \dots, x(n)$

X_k^m represents new curves, with k indicating the delay factor and m indicating the starting point for each new curve given by Eq. (2),

$$X_k^m = x(m), x(m+k), x(m+2k), \dots, x\left(m + \left\lfloor \frac{n-m}{k} \right\rfloor \cdot k\right), \text{ where } m=1,2,3,\dots,k \quad (2)$$

Then, the length of each generated curve can be calculated as

$$L_m^k = \left(\sum_{i=1}^{\left\lfloor \frac{n-m}{k} \right\rfloor} |x(m+ik) - x(m+(i-1)k)| \right) \cdot \frac{(n-1)}{\left\lfloor \frac{n-m}{k} \right\rfloor \cdot k} \quad (3)$$

where $\frac{(n-1)}{\left\lfloor \frac{n-m}{k} \right\rfloor \cdot k}$ is a normalization factor. The lengths of the generated curves are averaged as

$$L_k = \sum_{m=1}^k L_m^k \quad (4)$$

To calculate the FD, a least square linear best fit graph is plotted between $\ln(L_k)$ and $\ln\left(\frac{1}{k}\right)$. The slope of the graph refers to the FD. For a larger delay factor, which involves bypassing many samples, the curves will become smoother and their lengths will be expected to decrease. The slope of the graph will also decrease and results in a larger FD. The waveform of a pathological sample is more irregular and transient than the waveform of a normal sample. Therefore, the FD of the pathological samples will be larger than that of normal samples.

2.3 Discrete Wavelet Transformation

Wavelet transformation has an advantage of localization over Fourier transformation. Fourier transformation is localized over frequency but not in time. Therefore, Fourier transformation produce change everywhere in the time domain when small change occurs in frequency. Wavelets produce change in both time and frequency via translations and dilations, respectively. The wavelet is a small wave, and many waves can be produced from it by dilation and translation [29]. The wavelet function is defined as

$$W_{kl} = w\left(2^k t - l\right) \quad (5)$$

The wavelet function W_{kl} is obtained from the main wavelet function by shifting l samples and compressing by a factor of 2^k . A time signal can be represented in terms of wavelet functions as given by Eq. (6)

$$h(t) = \sum_{l=0}^{2^0} c_{0l} W_{0l} + \sum_{l=0}^{2^1} c_{1l} W_{1l} + \sum_{l=0}^{2^2} c_{2l} W_{2l} + \dots + \sum_{l=0}^{2^m} c_{ml} W_{ml} \quad (6)$$

The wavelet coefficients c_{kl} at level k are given by

$$c_{kl} = \int_0^T h(t) \cdot W_{kl} dt \quad (7)$$

where T represents the frame duration. Eq. (7) is valid for orthogonal wavelets. Every summation in Eq. (6) represents the whole signal in different frequency bands.

3 Material

Sustained vowels /a/ in the MEEI database [24] have different average lengths for normal and disordered subjects: three seconds for the former and one second for the latter. Before applying fractal algorithms, the length of normal samples was reduced to one second to make the duration equal to disordered samples. As the MEEI database is recorded at two sampling frequencies, 25 KHz and 50 KHz, all samples of both types of subject are down sampled to one frequency, i.e., 25 KHz. A subset of the MEEI database as mentioned in [30] and [31] is used for the evaluation of the proposed system for the automatic disorder detection. The subset contains 173 pathological and 53 normal samples. The distribution of the normal and pathological samples having different types of disorders is depicted in Table 1.

Table 1: Distribution of normal and pathological samples

Subjects	Disorder Types	No. of Samples	Total
Pathological	Adductor	22	173
	Vocal Nodules	20	
	Keratosiis	26	
	Vocal Fold Polyp	20	
	Paralysis	85	
Normal	---	53	53

Normal subjects include 21 males and 32 females, while the number of men and women for pathological subjects are 70 and 103, respectively. For normal subjects, the age range of male speakers is 26-59 years, and for female speakers the range is 22-52 years. The average age for men and women in case of normal subject is 38.81 and 34.1 years, respectively. For pathological subjects, the age of male patients is within the range of 26-58 years, and the range for female patients is 21-51 years. The average age for men is 41.71 years and for women is 37.58 years. A summary of statistics of the MEEI database used in this study is provided in Table 2.

Table 2: Statistics of the MEEI database subset with respect to gender and age

Subject	Number		Mean age		Age range		Standard Deviation for age	
	Male	Female	Male	Female	Male	Female	Male	Female
Normal	21	32	38.8	34.2	26-59	22-52	8.5	7.9
Pathological	70	103	41.7	37.6	26-58	21-51	9.4	8.2

*Age cited in years

4 Experiments and Discussion

In this paper, MDVP parameters are appended with fractal dimensions in few experiments. Five samples in the MEEI database out of 173 pathological samples do not have MDVP parameters; hence, the database used in the study includes a total of 221 samples containing 168 disorders and 53 normal cases. The five samples that do not have MDVP parameters are listed in Appendix A. The dataset was divided into five distinct test sets. Each time one of the sets was used to evaluate the system, the remaining four sets were used for training. The five-fold approach was repeated four times, which provides 20 trials for the system evaluation and helps to draw a smooth receiver operating characteristic (ROC) curve to exhibit the classifier performance graphically. Classification of pathological and normal samples was carried out by using LIBSVM [32] with a radial basis function as kernel, given by Eq. (8).

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right) \quad (8)$$

where x is the training sample, x' is the testing sample, and γ is a free parameter. SVM is a linear classifier, however, in most of the cases, data is not linearly separable. Therefore, kernel function is implemented to map the original input space to higher dimensional space, where features are linearly separable.

The results of the developed system are expressed in terms of following performance parameters: sensitivity (SEN), specificity (SPE), accuracy (ACC), and the area under the ROC curve (AUC). These parameters are defined as follows: SEN is the ratio between truly identified pathological samples and the total number of pathological samples, SPE is the ratio of truly classified normal samples and the total number of normal samples, and ACC is the ratio between correctly detected samples and the total number of samples,. The performance parameters SEN, SPE, and ACC are calculated by using following relationships:

$$SEN = \frac{TP}{TP + FN} \times 100 \quad (9)$$

$$SPE = \frac{TN}{TN + FP} \times 100 \quad (10)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (11)$$

where TN stands for true negative and means that the system detects a normal subject as a normal subject, TP stands for true positive and means that the system detects a pathological subject as a pathological subject, FN stands for false negative and means that the system detects the pathological subject as a normal subject, and FP stands for false positive and means that the system detects the normal subject as a pathological subject.

4.1 Fractal Dimension in Time Domain

To calculate the FD in time domain, neither windowing nor any pre-processing was applied to the signal. The FD was determined by considering the whole signal, and a single value was obtained for the entire signal. The SVM performed the classification among the two classes. The results of the experiments in time domain by using the Katz and Higuchi algorithms are provided in Table 3, and they are explained in terms of performance parameters SEN, SPE, ACC and AUC. Positive class represents the pathological samples and negative class is for the normal samples. Standard deviation, represented by STD, is calculated over 20 trials.

Table 3: Detection rate for fractal dimension in time domain

Method	No. of Features	SEN \pm STD	SPE \pm STD	ACC \pm STD	AUC
Katz	1	71.52 \pm 3.1	34.88 \pm 7.2	62.77 \pm 1.6	55.45
Higuchi	1	96.24 \pm 0.6	26.54 \pm 2.3	79.53 \pm 0.4	62.96

SEN, SPE, ACC, and AUC are provided in percentage.

4.2 Fractal Dimension in Frequency Domain

Speech signal is transformed into frequency domain by applying DWT. The signal is decomposed into three levels, where level 1 (L1) signifies a frequency band of 6251 – 12500 Hz, level 2 (L2) indicates a band of 3126 – 6250 Hz, and level 3 (L3) represents the frequencies of 1563 – 3125 Hz. During decomposition of the signal, lower frequencies are decomposed at each level into lower and higher frequencies. Residue indicates the lower frequencies achieved after decomposition of level 3.

For each level, after applying an algorithm for FD, a single feature is obtained. The calculated features (FDs) are then provided to the SVM for classification. Different experiments are performed for the pathology detection with L1, L2, L3, and Residue. Some experiments by concatenating the FDs of L1, L2, L3 and Residue are also performed and these

experiments are represented by L1-L3+R in Table 4 and 5. Moreover, the experiments known as Synthesis are also performed by doing the synthesis of all levels. Results of all experiments with Katz and Higuchi algorithms are listed in Tables 4 and 5, respectively.

Figure 4 shows the spectrum of L1, L2, L3, and Residue. The spectrum of the Residue for the pathological subject (ALB18AN) can be observed in Figure 4(b), which shows an irregular pattern in comparison with the spectrum of the normal subject (AXH1NAL) exhibited in Figure 4(a). This is the reason that maximum accuracy to discriminate between the normal and disordered subject is obtained with Residue, which is 91.28% with the Katz algorithm. The accuracy with Residue is better than that of L1, L2, and L3 by 15%, 4%, and 4%, respectively. The irregular pattern in the spectrum of pathological subject might be due to being a more transient signal than a normal subject, which is captured by the FD and leads to a differentiation among the two classes significantly.

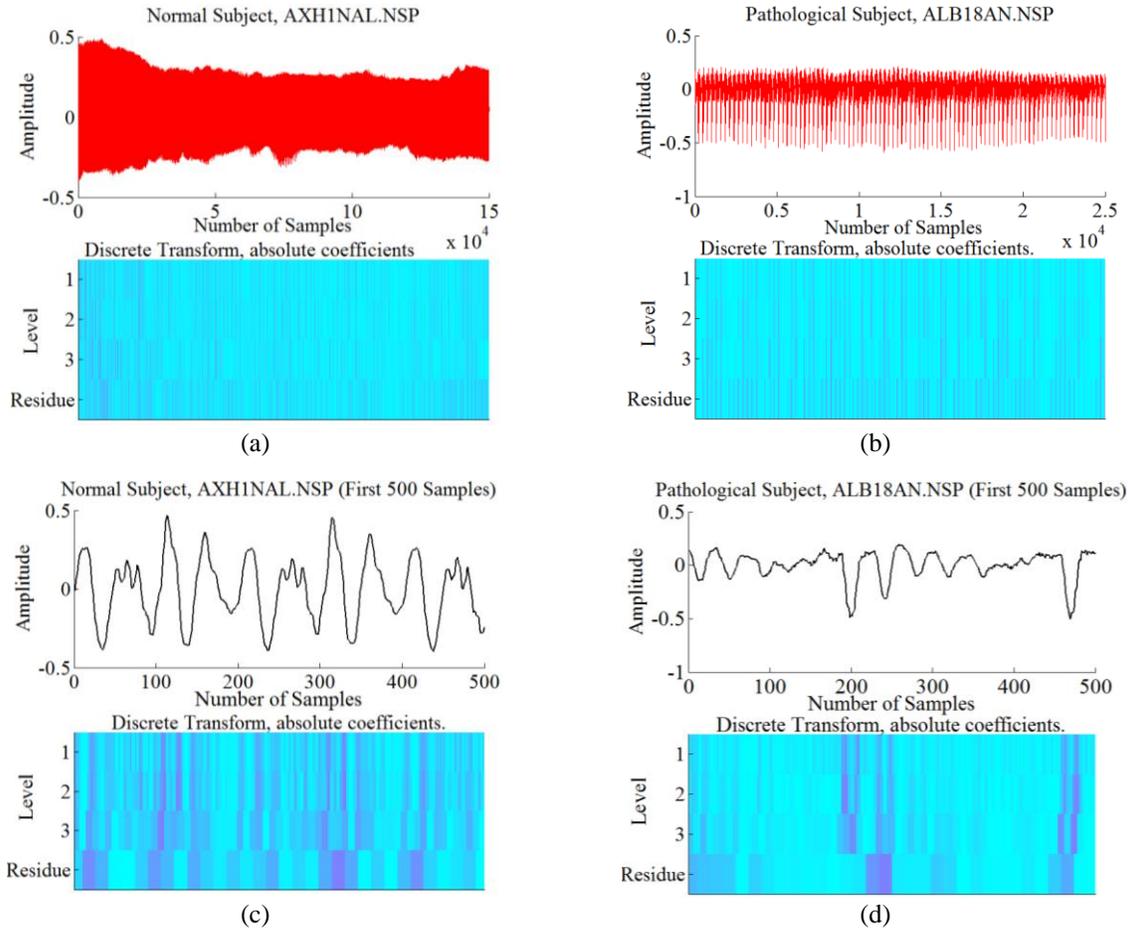


Figure 4. Discrete Wavelet Transformation of (a) Normal subject, (b) Disordered subject, (c) Normal subject (first 500 samples), and (d) Disordered subject (first 500 samples)

The maximum obtained recognition rate for the individual frequency band is 91.28%, which implies that the contribution of the frequency band 1 – 1562 Hz is more than the bands represented by L1, L2, and L3. The overall maximum accuracy is found with L1-L3+R and it is 92.45% by using four features. AUC with these four features is 95.06%.

Table 4: Detection rate for fractal dimension in frequency domain with Katz Algorithm

Level	Frequency band	No. of Features	SEN \pm STD	SPE \pm STD	ACC \pm STD	AUC
1	6251 – 12500 Hz	1	98.96 \pm 1.4	4.67 \pm 5.9	76.36 \pm 0.6	52.14
2	3126 – 6250 Hz	1	93.01 \pm 0.9	68.41 \pm 3.1	87.10 \pm 0.6	85.89
3	1563 – 3125 Hz	1	95.45 \pm 1.3	61.15 \pm 3.9	87.22 \pm 0.4	84.59
Residue	1 – 1562 Hz	1	93.98 \pm 0.9	82.74 \pm 2.9	91.28 \pm 0.4	89.79
L1-L3+R	---	4	93.99 \pm 0.9	87.61 \pm 2.7	92.45 \pm 0.6	95.06
Synthesis	---	1	93.71 \pm 1.0	82.73 \pm 2.3	91.06 \pm 0.6	88.85

SEN, SPE, ACC, and AUC are provided in percentage.

Table 5: Detection rate for fractal dimension in frequency domain with Higuchi Algorithm

Level	Frequency band	No. of Features	SEN \pm STD	SPE \pm STD	ACC \pm STD	AUC
1	6251 – 12500 Hz	1	98.42 \pm 0.8	3.62 \pm 2.2	75.69 \pm 0.4	51.98
2	3126 – 6250 Hz	1	98.34 \pm 1.2	7.2 \pm 4.2	76.50 \pm 0.6	54.31
3	1563 – 3125 Hz	1	94.75 \pm 1.4	14.0 \pm 4	75.40 \pm 0.3	59.37
Residue	1 – 1562 Hz	1	100 \pm 0	0 \pm 0	76.02 \pm 0	50.40
L1-L3+R	---	4	95.58 \pm 1.6	13.82 \pm 5.4	76.0 \pm 0.7	62.57
Synthesis	---	1	93.73 \pm 1.0	59.94 \pm 3.9	84.85 \pm 0.6	87.13

SEN, SPE, ACC, and AUC are provided in percentage.

The ROC curves for both algorithms are plotted and depicted in Figure 5. False positive rate ($1 - \text{SEN}$) is along the X-axis and true positive rate (SPE) is along the Y-axis. In Figures 5(a) and 5(b), the ROC curves for the FD in time and frequency domain are plotted. For frequency domain, the ROC curve for each level is plotted separately. Figure 5 shows that, for both algorithms, the performance of the classifier is best when the FD of all four levels was concatenated.

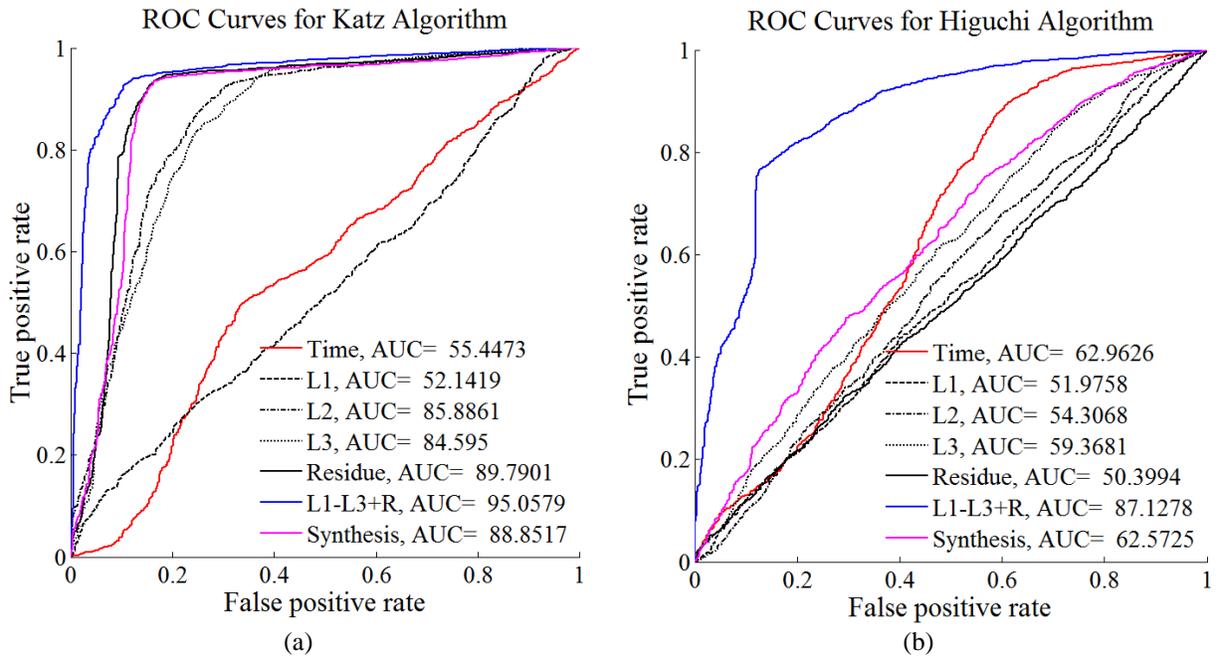


Figure 5. The ROC curves for (a) Katz and (b) Higuchi algorithms

4.3 MDVP Parameters and Fractal Dimension in Frequency Domain

An accuracy of 89.29% was achieved with 22 original MDVP parameters, as reported in [26]. In this paper, we attained an accuracy of 88.4%, as depicted in Table 6, with 22 MDVP parameters. An improvement in accuracy is observed when 22 MDVP parameters are concatenated with FDs of four levels, referred as L1-L3+R in Table 4, estimated by the Katz' method. The detection rate of 94.71% with standard deviation of 0.5% is obtained and is comparable with that of [26]. Three ROC curves are plotted in Figure 6, showing the performance of MDVP parameters only, MDVP with Katz, and MDVP with Higuchi algorithm.

Table 6: Detection rate for 22 MDVP parameters and fractal dimension in frequency domain

Features	No. of Features	SEN \pm STD	SPE \pm STD	ACC \pm STD	AUC
MDVP only	22	97.43 \pm 1.3	62.00 \pm 5.3	88.94 \pm 1.1	89.96
Katz + MDVP	4 + 22	97.48 \pm 0.8	85.94 \pm 3.2	94.71 \pm 0.5	96.56
Higuchi + MDVP	4 + 22	96.60 \pm 0.9	68.0 \pm 5.0	89.75 \pm 0.9	91.91

SEN, SPE, ACC, and AUC are provided in percentage.

ROC Curves for Combinations of MDVP with Katz and Higuchi Algorithm

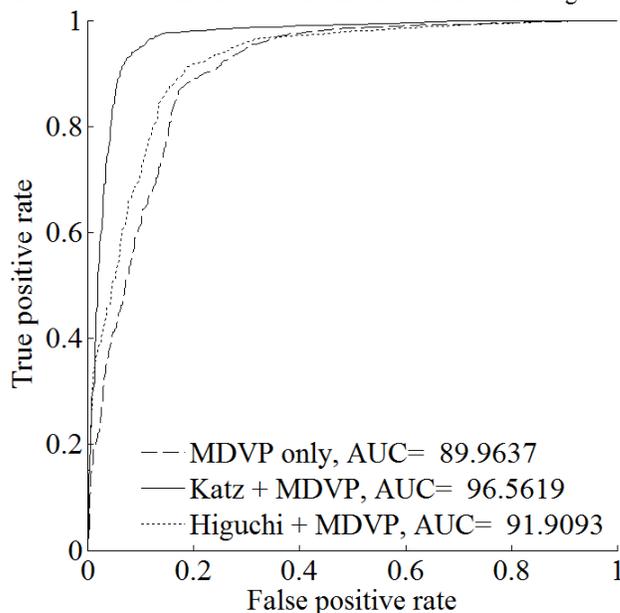


Figure 6. ROC curves for combinations of MDVP parameters with Katz and Higuchi algorithms

When the FDs of different levels are concatenated with MDVP features, multi-dimensional features arise that are difficult for human interpretation. Studies based on these types of multi-dimensional features need a machine-learning stage to make a decision for a test utterance. In the proposed system, the output of the SVM classifier, which is a decision value for a test utterance, might be considered as a discriminant measurement to differentiate between the samples of two classes. The significance of the results depends on the discriminative power of the computed decision values. The Mann-Whitney U-test is performed to check the discriminant power of the decision values at the 5% significance level.

A two-sided p-value using the Mann-Whitney U-test is determined in the case of best accuracy (94.71%). The obtained p-value is 0.0001E-15, which is less than 0.05 and therefore rejects the null hypothesis that the decision values of normal and pathological classes are the result of continuous distributions with equal medians. A smaller p-value for

decision values shows that the proposed system can discriminate between normal and pathological samples successfully.

5 Conclusion

The contribution of different frequency bands for pathology detection through multiresolution analysis has been examined in this study. The speech signal is decomposed into three levels by DWT. The FDs are calculated with the Katz and Higuchi algorithms for each level, and are provided to the SVM for classification. The multiresolution analysis of the speech signal signifies an irregular pattern in pathological samples, specifically in a spectrum of the frequency band of 1 - 1562 Hz. This band also provides the highest detection accuracy ($91.28\% \pm 0.4$) than any other frequency bands. The irregular pattern is due to the more transient nature of the abnormal waveform than the normal waveform, occurring as a result of abnormal vibration of vocal folds. The experimental results show that the FD successfully captures the transient behavior of the disordered speech and leads to a characterization of the signal into normal and pathological classes. To enhance the accuracy of the system, the FD of all levels is concatenated with MDVP parameters that provided an accuracy of 94.71%. The obtained accuracy is higher than the accuracy reported in [30] and [31], which is 94.07% and 94.1%, respectively. The FD did not perform well in a time domain as the best accuracy achieved is 79.53%.

Acknowledgments

This project was funded by the National Plan for Science, Technology, and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Award Number (12-MED-2474-02).

Appendix A

The list of files that do not contain MDVP parameters in the MEEI database:

JFG08AN.RES
KXH30AN.RES
LES15AN.RES
TAB21AN.RES
WPB30AN.RES

References

- [1] B. Mohan, *Diseases of Ear, Nose and Throat: Head and Neck Surgery*, 1st ed., Jaypee Brothers Medical Publishers, New Delhi, India, 2013.
- [2] M. H. L. Hecker and E. J. Kreul, "Descriptions of the Speech of Patients with Cancer of the Vocal Folds. Part I: Measures of Fundamental Frequency," *The Journal of the Acoustical Society of America*, vol. 49, pp. 1275-1282, 1971.
- [3] G. Muhammad, T. A. Mesallam, K. H. Malki, M. Farahat, A. Mahmood, and M. Alsulaiman, "Multidirectional regression (MDR)-based features for automatic voice disorder detection," *J Voice*, vol. 26, pp. 817 e19-27, Nov 2012.

- [4] R. J. Baken and R. Orlikoff, *Clinical Measurement of Speech and Voice*, 2nd ed. San Diego, CA: Singular, 2000.
- [5] J. W. Lee, H. G. Kang, J. Y. Choi, and Y. I. Son, "An investigation of vocal tract characteristics for acoustic discrimination of pathological voices," *BioMed Research International*, vol. 2013, pp. 1-11, 2013.
- [6] A. I. R. Fontes, P. T. V. Souza, A. D. D. Neto, A. d. M. Martins, *et al.*, "Classification System of Pathological Voices Using Correntropy," *Mathematical Problems in Engineering*, vol. 2014, pp. 7, 2014.
- [7] L. Jung-Won, S. Kim, and K. Hong-Goo, "Detecting pathological speech using contour modeling of harmonic-to-noise ratio," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5969-5973, 2014.
- [8] D. Panek, A. Skalski, and J. Gajda, "Quantification of Linear and Non-linear Acoustic Analysis Applied to Voice Pathology Detection," *Information Technologies in Biomedicine*, vol. 4, *Advances in Intelligent Systems and Computing*, vol. 284, pp. 355-364, 2014.
- [9] G. Muhammad and M. Melhem, "Pathological voice detection and binary classification using MPEG-7 audio features," *Biomedical Signal Processing and Control*, vol. 11, pp. 1-9, May 2014.
- [10] G. Muhammad, Z. Ali, M. Alsulaiman, and K. Al-Mutib, "Vocal Fold Disorder Detection by Applying LBP Operator on Dysphonic Speech Signal," *Proc. of Recent Advances in Intelligent Control, Modelling and Simulation*, pp. 222-228, 2014.
- [11] R. Lopes and N. Betrouni, "Fractal and multifractal analysis: A review," *Medical Image Analysis*, vol. 13, pp. 634-649, 2009.
- [12] M. J. Katz, "Fractals and the analysis of waveforms," *Computers in Biology and Medicine*, vol. 18, pp. 145-156, 1988.
- [13] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, pp. 277-283, 1988.
- [14] A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns," *Proc. of the Eighth IEEE Symposium on Computer-Based Medical Systems*, pp. 212-217, 1995.
- [15] P. Maragos, "Fractal aspects of speech signals: Dimension and interpolation," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 417-420, 1991.
- [16] T. R. Senevirathne, E. L. J. Bohez, and J. A. Van Winden, "Amplitude scale method: New and efficient approach to measure fractal dimension of speech waveforms," *Electronics Letters*, vol. 28, pp. 420-422, 1992.
- [17] Y. W. Kim, K. K. Kriebel, C. B. Kim, J. Reed, and A. D. Rae-Grant, "Differentiation of alpha coma from awake alpha by nonlinear dynamics of electroencephalography," *Electroencephalography and Clinical Neurophysiology*, vol. 98, pp. 35-41, 1996.
- [18] A. K. Mishra and S. Raghav, "Local fractal dimension based ECG arrhythmia classification," *Biomedical Signal Processing and Control*, vol. 5, pp. 114-123, 2010.
- [19] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, "A comparison of waveform fractal dimension algorithms," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol. 48, pp. 177-183, 2001.
- [20] B. S. Raghavendra and D. Narayana Dutt, "A note on fractal dimensions of biomedical waveforms," *Computers in Biology and Medicine*, vol. 39, pp. 1006-1012, 2009.
- [21] P. N. Baljekar and H. A. Patil, "A comparison of waveform fractal dimension techniques for voice pathology classification," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4461-4464, 2012.
- [22] A. Accardo, M. Affinito, M. Carrozzini, and F. Bouquet, "Use of the fractal dimension for the analysis of electroencephalographic time series," *Biological Cybernetics*, vol. 77, pp. 339-350, 1997.

- [23] A. Accardo, F. Fabbro, and E. Mumolo, "Analysis of normal and pathological voices via short-time fractal dimension," *Proc. of 14th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society*, pp. 1270-1271, 1992.
- [24] Massachusetts Eye & Ear Infirmiry Voice & Speech LAB, "Disordered Voice Database Model 4337 (Ver. 1.03) ", ed. Boston, MA: Kay Elemetrics Corp, 1994.
- [25] M. Little, P. McSharry, S. Roberts, D. Costello, and I. Moroz, "Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection," *BioMedical Engineering OnLine*, vol. 6, p. 23, 2007.
- [26] M. K. Arjmandi, M. Pooyan, M. Mikaili, M. Vali, and A. Moqarehzadeh, "Identification of voice disorders using long-time features and support vector machine with different feature reduction methods," *J Voice*, vol. 25, pp. e275-89, Nov 2011.
- [27] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [28] G. Vaziri and F. Almasganj, "Pathological Assessment of Vocal Fold Nodules and Polyp via Fractal Dimension of Patients' Voices," *Proc. of the 2nd International Conference on Bioinformatics and Biomedical Engineering*, pp. 2044-2047, 2008.
- [29] M. H. Farouk, *Application of Wavelets in Speech Processing*: Springer, 2014.
- [30] J. I. Godino-Llorente, P. Gómez-Vilda, and M. Blanco-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*, vol. 53, pp. 1943-1953, 2006.
- [31] M. Markaki and Y. Stylianou, "Voice Pathology Detection and Discrimination Based on Modulation Spectral Features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 1938-1948, 2011.
- [32] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 1-27, 2011.