İSTANBUL UNIVERSITY PRESS

A New Approach for Detection of Pathological Voice Disorders with Reduced Parameters

Haydar Ankışhan ⓘD

Biomedical Equipment Technology, Başkent University Vocational School of Technical Sciences, Ankara, Turkey

ABSTRACT

Voice data has demonstrated chaotic behavior in previous studies. Therefore, studying the linear properties alone does not yield successful results. This is valid for the examination of voice data as well. Therefore, conducting studies including chaotic features as well as existing technologies is inevitable. The main purpose of this study is to detect voice pathologies with fewer special features using new chaotic features. Both linear and nonlinear characteristics were used in this study. In this context, the largest Lyapunov exponents and entropy are preferred as chaotic properties because of their success in previous studies. Very few results with 100% accuracy were obtained in the experimental studies. In this study, multiple support vector machines (SVMs) were selected as a classifier because of their success in previous similar data types. Thus, the desired accuracy level was achieved using fewer features. Resultantly, the process complexity decreased and the system speed increased.

Keywords: Voice disorders, support vector machines, voice disorder detection

Introduction

Advanced evaluations in voice disorders are applied for diagnostic purposes in analysis laboratories with analyzing devices. Depending on the features, the results provide feedback for experts and patients, both visually and audiologically. It can be quite restrictive that human voice specialists, called foniatrists, work with a limited number of features to be examined on the axis of voice, frequency and time domain. This leads to inability to diagnose patients correctly in some special cases without going through the detailed investigations.

Although higher digital technologies have recently been used in acoustic signal analysis, it is still important to obtain meaningful and useful information. In signal processing studies, voice pathologies and problems have recently begun to be more carefully scrutinized [1,2]. Since the sounds have dynamic and complex characteristics, especially when different language constructions are taken into consideration, the complexity is increased depending on these different notations and emotions. For these reasons, the need for reliable software and analysis methods is increasing day by day. Language analysis can be done with the help of these methods and software, the diversity of pathologies can be examined and techniques can be investigated. Today, scientists are spending a lot of time developing tools that use different characters and traits and can predict different voice pathologies.

In clinical voice studies, patients' voice quality can be predicted by a sustained vocal phonetics or by mutual speaking. It is thought that sustainable vowels give useful results as they are avoided from linguistic artifacts and are observed to be sufficient in many sound estimations. Specialists examine whether or not it is possible for patients to maintain a sustainable vowel and their performances of voice analysis.

Voice quality analyzes can generally be performed in titles such as acoustics, aerodynamics, endoscopy, and a percentage of the patients' self-examination. It is observed that the

Address for Correspondence:
Haydar Ankışhan

E-mail:
hankishan@baskent.edu.tr

results of different methods are examined, in cases where the obtained results are uncertain [3]. It is shown from the literature that the acoustic properties of different features are evaluated together in the literature. Mostly preferred properties; fundamental frequency [4], jitter (frequency perturbation), shimmer (amplitude perturbation) [5-7], Harmonics to Noise Ratio [8] and Mel frequency celpstral coefficients (MFCC) [9,10].

According to [11], in vocal tract pathology, voice impairment is detected by interpreting the stimulation of the signal. It is observed that researchers devote a considerable amount of time to systems that can automatically calculate and group voice impairment with traditional diagnostic approaches. Some studies have shown that the correct diagnosis of the sound level increases the predictive accuracy of the noise impairment [1]. In addition to these studies, a universal method and approach is still not found for these kind of problems. Have obtained 100% accuracy with long-term average spectral properties, glottal noise measures and linear predictive modeling techniques [12]. Obtained 100% accuracy for diagnosed pathologies and 96.1% for ambiguous pathological conditions [13]. Has obtained 89.3% accuracy for Asthenia disease with MFCC, HNR, NNE, GTNE and PCA [14]. In [15] studies, adaptive time-frequency transform, octave max, octave mean, energy ratio, length ratio and frequency ratio together with Linear Discriminant Analysis (LDA) were used as properties and pathological classification was made. They obtained the 93.4% accuracy classification results. Again only used MFCC and obtained 96% accuracy with the help of the neural network – multi layer perceptron [16].

There are some other studies available in the literature, trying to detect vocal disorders by using different classification methods. For instance, [14] has helped to extract useful features. In their study, have shown that parameter reduction affects speech robustness in inner classificatios [17].

In some studies, acoustic signals and vocal cord images were combined for more precise analysis. Kymography and high-speed digital videoendoscopy (HSV) has played a major role in classifiying pathological disorders for clinicians [18]. In addition, [19], in their study, examined the correlation between the vocal cord vibration of the system and acoustic analysis of the voice.

Calculated the root mean square, delay spread and standard deviation of the voice signal. In the study, they have considered two situations; having disorders or not [11]. Additionally, 4 polyptical and 8 healthy person's wovel / a / or / i / were examined. Depending on the method, False alarm rate %0 and true negative 25% were calculated.

Some studies have used the Empirical mode decomposition (EMD) method to classify voice pathology. It was used by [20] at EMD Chemical plant controls. Could be classify into classes according to their pathology by using Maximum power spectral density of the intrinsic mode function as a feature [21]. For this operation used K-Nearest Neighbor (KNN) as a classifier. With the results of this process, they obtained 95.7% accuracy results.

In this study, fewer features were used than in previous studies, and features were calculated after voice modeling. In the aim of the study, classification of patients with 4 different pathologies was performed by using nonlinear, and linear time, frequency features with the help of the SVMs as a classifier. First, the sounds were recorded at specific phonetic intervals, then the LPC and MFCC coefficients were calculated and the voices were modeled with these coefficients. So, the modeled signals' features were then extracted, and pathological classification was performed with the SVMs depending on the characteristic values of the features. The flow chart of the events are given in Figure 1.

The study continuous as follows: second part is materials and methods, third part is results and discussions, and fourth part is conclusions.
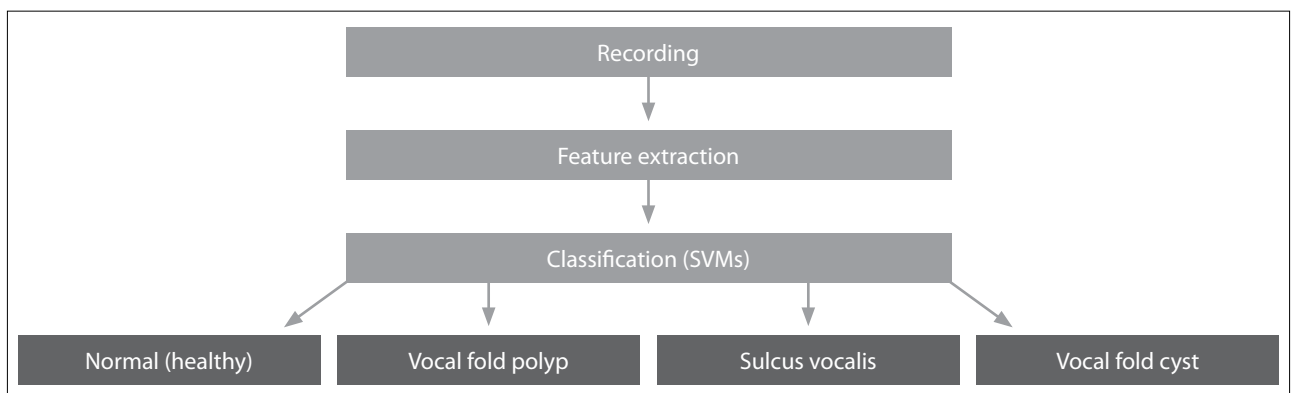
Figure 1. Block diagram of the proposed detection system

Materials and Method

Database

In this study, voice recordings were made through a dynamic one-way microphone (Shure SM58). EMU-Tracker Pre was preferred as the recording interface. The reason for this preference is that the recorded data can be used later in other programs. Recording was done in a sufficiently quiet room with a background noise of 45-60 dB. The sampling frequency of the recorded data is 44100 Hz and is in single channel PCM format with 16 bit resolution.

The sound data is recorded in different settings. A record of 10 seconds is executed for wovel / a /, and also, a paragraph is read about this time period. Many studies in the literature use this vowel because of the stationary acoustic features of / a /. 120 patients were examined in this study. The details of the patients are given in Table 1.

With the software developed for the study, the features of the sounds can be extracted and can be divided into sub classes depending on the features.

Table 1. Patient database knowledge

Pathology	Total	Women	Men
Vocal fold polyp	30	18	12
Vocal fold cyst	34	18	16
Sulcus vocalis	24	14	10
Normal (Healthy)	34	12	22

Methodology

In this study, differently from the previous ones, the sounds were modeled before the feature extraction using two different models; linear predictive coding (LPC) and Mel frequency cepstral coefficients (MFCC). The modelling process was carried out in three stages as given in Figure 2.

In the scope of the study, two different coefficient types were taken as basis for modeling of signals. LPC and MFCC coefficients of the signals were calculated after the preliminary steps were completed. The sound data is re-modeled with the calculated coefficients, and so the recorded speech is divided into two different signal types, clean and residual signals. In the aimed study, it was investigated whether some information was extracted from the signals obtained from the models.

Preprocessing

The modeling of the sounds and the calculation of the attributes from the modeled data are performed after some preprocessing. First, the audio data is automatically divided into sub-segments at specific lengths. Discretized signal is shown in the low-order linear difference model for use in the preprocessing steps as follows:

$$\hat{s}(n) = s(n) - \alpha s(n-1) \qquad (1)$$

In this study, it is stated that it is appropriate to take α around 0.95 [17]. It is also assumed in this study that the characteristics of the voice data change slowly in time. In the scope of the study, the discrete M samples were taken and the sound windows with N samples were shown as follows;

$$x_l(n) = \hat{s}(Ml+n) \quad n=0,1,...,N-1, l=0,1,...,L-1 \qquad (2)$$
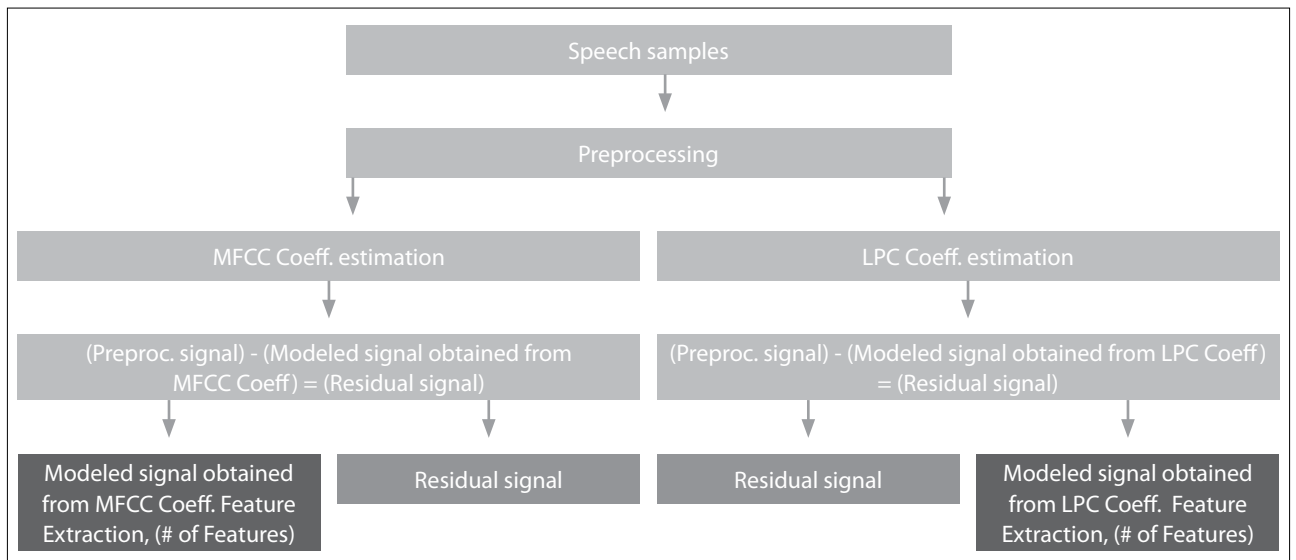
Figure 2. Flow chart of the modeled signal and residual signals from LPC and MFCC coefficients.

The purpose in the framing is to block the interruption of feature characteristics in the signal. Hamming windowing method is preferred for study by taking samples from similar studies. The window $w(n)$ is defined in interval $0 \leq n \leq N-1$, so the windowed signal is given as

$$\overline{x}_l(n) = x_l(n)w(n) \quad 0 \leq n \leq N-1 \tag{3}$$

Hamming windowing methods for this study is given as

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \quad 0 \leq n \leq N-1 \tag{4}$$

### Linear Predictive Coding (LPC)

Linear predictive coding is one of the most important analysis techniques [22]. In particular, it has become a superior method for estimating audio parameters such as pitch, formant, spectrum, and vocal tract, when speech signals are stored and transmitted. The main ability of LPC is to be able to obtain a sound sample by linear combination from previous sound samples [22]. The sound data given at time n is expressed in linear form with p coefficients:

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + ... + a_p s(n-p) \tag{5}$$

where a_1 a_2 $\cdots$ a_p denotes the parameter coefficients of the LPC. Once the autocorrelation of the window frames is established, these coefficients are obtained by the Levinson-Durbin method [23]. The signal modeled from the obtained coefficients is now cleaned from the recorded speech voice thanks to the coefficients.

Mel Frequency Cepstral Coefficients (MFCC)

The spectral coefficients of the windowed frames after the preprocessing are calculated using Fast Fourier Transform (FFT). In here, FFT gives the information of the amount of energy of each frequency band of the signal. The calculation of MFCC depending on the this procedure is given in Figure 3;

Figure 3. Flowchart of the MFCC coefficient estimations.
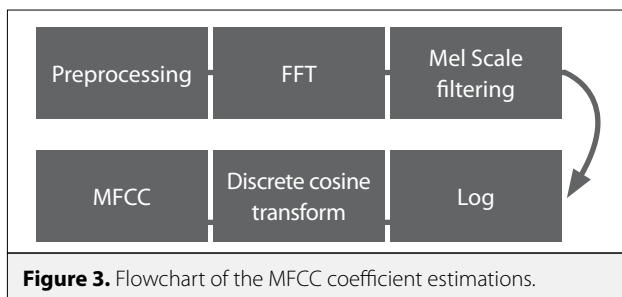
Human ears are not sensitive enough to detect sounds below 1000Hz [24]. Owing to these information, MFCC coefficient estimation is realized. Mel scaling frequency mapping isshown as linear below 1000Hz, logarithmic above 1000Hz [24]. Thus, at low frequencies, the corner and center frequencies are linear and logarithmic at high frequencies. Figure 4 gives information about this process.

Figure 4. Mel Scale Filter Bank.
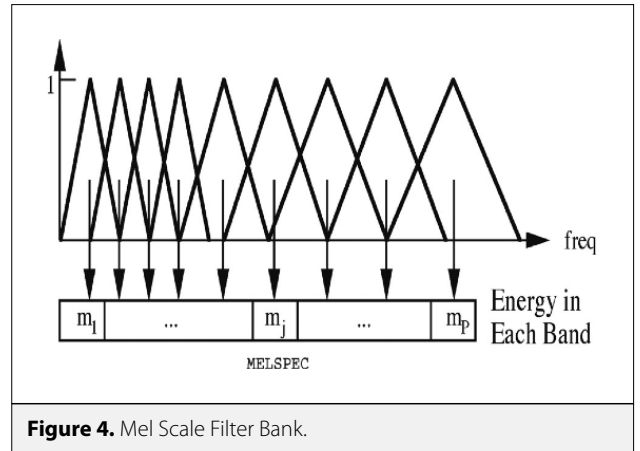
Mel scaling is given in (6),

$$Mel(f) = 2595\log_{10}\left(1 + \frac{f}{700}\right) \tag{6}$$

When this process occurs, the coefficients of each short time Fourier Transform (FT) are multiplied by the corresponding filter gain. So, the total is obtained by this process. Thus, DCT is applied to the log of Mel cesptral coefficients, so that MFCC is obtained.

Chaotic Model of System

The phase space must first be calculated before the chaotic model of a system can be obtained. According to Takens [25], if the size of the embedded dimension is chosen as large, the phase space of the time series can be constructed. One of the best techniques for the constructing phase space is the delay method.

An unknown dynamic system is defined as a time series $s(t)$. In this case, the m-dimensional $\tilde{s}(t)$ vector is obtained from the time delay of $s(t)$.

$$\tilde{s}(t) = \left[\begin{array}{cccc} s(t), & s(t+\tau), & ... & s(t+(m-1)\tau) \end{array} \right] \tag{7}$$

where τ and m represent the time delay and embedded dimension. These values are not known in practice and are calculated. In the study, it was observed that depending on the voice pathologies, the voice data have different characteristics, resulting in the reconstruction of phase space. The pathology related phase space representation is given in Figure 5.

When the figure 5 is examined, it is observed that the phase spaces obtained from the voices of the pathological signals also have different views.

Feature Extraction

In the study, features that can be evaluated in different characteristics, such as linear and nonlinear, were extracted. In here, before the features extraction, the voice signals were separated into frames, which were then used to calculate the features. In
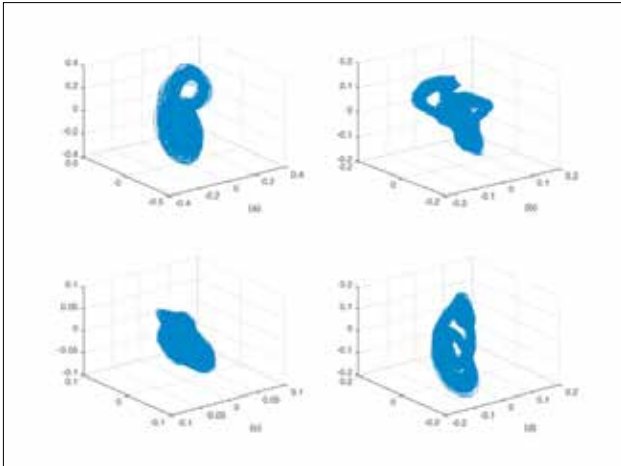
Figure 5. Phase space of the Cyst (a), Normal Person (b), Polyp (c), Sulcus (d) modeled signals.

this study, jitter, shimmer, skewness, kurtosis, entropy, and largest Lyapunov exponents (LLEs) were calculated.

Jitter
Jitter is known as the average approximate difference in divided time between consecutive periods [5-7].

$$jitter(relative) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i+}|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \qquad (8)$$

Where, T_i is the period length of the F_0 and N is the number of period of F_0 [5].

Shimmer
Shimmer is defined as the average approximate difference between the amplitudes of successive periods [5,6].

$$Shimmer(relative) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+}|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \qquad (9)$$

where, A_i represents the amplitude of the peak to peak, and N is number of the extracted fundamental frequency F_0.

Skewness and Kurtosis
The skewness of any variable or signal $\tilde{s}(t)$ is defined as the standardized third moment of this signals [26].

$$Skew[\tilde{s}(t)] = E\left[\left(\frac{\tilde{s} - \mu}{\sigma}\right)^3\right] \qquad (10)$$

where, E[.] represents the expected operator, μ represents average, and σ represents standard deviation.

The kurtosis of any variable or signals $\tilde{s}(t)$ is defined as the standardized fourth moment of the signals [27]. The kurtosis is given as,

$$Kurt[\tilde{s}(t)] = \frac{E[(\tilde{s} - \mu)^4]}{(E[(\tilde{s} - \mu)^2])^2} \qquad (11)$$

Entropy
Entropy, the probability distribution of a random variable or signal $\tilde{s}(t)$, or a level of uncertainty [28]. Shannon has defined the definition of entropy as the uncertainty of a probabilistic distribution [28]. Shannon has proposed a method, which is a self-knowledge, has a logarithmic function, the values between 0 and 1. If the relational probability of the events is defined as $X = \tilde{s}_i, (i = 1,2,...,n)$. Then, self probability $h(p_i)$,

$$h(p_i) = log_2(\frac{1}{p_i}) \qquad (12)$$

Entropy is derived as the weights of n number of self-information values.

$$H_k = \sum_{i=1}^{n} p_i log_2(\frac{1}{p_i}) \qquad (13)$$

Largest Lyapunov Exponent (LLE)
Dynamic systems have characteristic exponents that determine the precision of the initial conditions. Attractors determine the convergence, the measure of the average ratio, or the divergence of the trajectory. According to Rosenstein [29], after the time delay vectors are reconstructed, the closest negihbour of the state in the phase space orbital is searched. The nearest neighbor pairs by LLEs in a given position are expressed by the divergence formula,

$$d_j(i) \approx C_j e^{\lambda_1 (i, \Delta t)} \qquad (14)$$

where, C_j represents the initial separation, Δt represents the sampling period, and $d_j(i)$ represents the nearest neighbour of the jth pairs. After taking the logarithm of the two sides,

$$lnd_j(i) \approx lnC_j + (i.\Delta t) \qquad (15)$$

The LLEs are calculated as follows for the voice signals;

$$y(i) = \frac{1}{\Delta t}\langle lnd_j(i)\rangle \qquad (16)$$

where, $\langle....\rangle$ represents the average of the j values.

MultiClass Support Vector Machines (M-SVMs)
M-SVMs can be used in many different areas [30]. In the first stage, two different sets of data were used to separate each other under certain conditions. This process was done via binary SVMs. Later, it is also being used in multiple classification operations. According to Vapnik [30], if having training data $((x_i, y_i), i = 1,2,...,n \text{ and } x_i \epsilon R^d \text{ and } y_i \epsilon \{\pm 1\})$ and can transform it into a higher dimensional space, the feature space will be $\varphi(...).R^d \xrightarrow{\varphi(-)} R^H$.

$$\underset{w, \zeta, b}{Minimize} \quad \left\{ \frac{1}{2} ||w||^2 + C \sum_i \zeta_i \right\}$$

$$Subject\ to\ y_i(\varphi^T(x_i)w + b) \geq 1 - \zeta_i, \tag{17}$$

$$\forall_i = 1, ..., n, \quad \zeta_i \geq 0$$

Here, (w, b) defines a linear classifier in the features space. For forming two classes, SVMs force positive samples to +1 denominator and negative samples force -1 denominator lower. If these samples do not fulfill this condition, and if needing zero in the formulation, then there is a forcing/punishment in the objective function.

In (17), the included norm of w, checks whether margin is the maximum. The objective of $C \sum_i \zeta_i$ is to check the number of mis-classified samples. When optimization is applied to solving the problem, there is a degradation between the classes due to un-even density of feature space clusters. In multiple SVMs, binary SVMs use different weight vectors and bias is applied for each class (w^j and b^j for $j\epsilon\{1,2,...,k\}$). Classifier function of SVMs,

$$f(x) = arg.\max(\phi^T(x)w^i + b^j), j\epsilon\{1, ..., k\} \tag{18}$$

If equation(18) applies for each class,

$$\underset{w^j, b^j, \zeta_i^{j,m}}{Minimize} \quad \left\{ \frac{1}{2} \sum_{j=1}^{k} ||w^j||^2 + C \sum_{j=1}^{k} \sum_{\substack{m=1 \\ m \neq j}}^{k} \sum_{i=1}^{n^j} \zeta_i^{j,m} \right\}$$

$$Subject\ to\ (\phi^T(x_i^j)w^j + b^j) - (\phi^T(x_i^j)w^m + b^m) \tag{19}$$

$$\geq 2 - \zeta_i^{j,m}, \zeta_i^{j,m}$$

$$\geq 0, \forall_j = 1, ..., k, \forall m = 1, ..., k(m \neq j), \forall i = 1, ..., n^j$$

where, x_i^j represents the ith samples of the jth class, n^j also represents the number of samples of training data. We have focused our study on non-linearly separable problems. To extend this simple model to solve non-linear separable problems, researchers have developed kernel based methods. This type of methods apply mapping functions on the input features and carry them into a very high dimension space and then construct a hyperplane in that feature space. The properties of the feature space depend on the kernel functions of the SVMs. The generally used kernel functions are polynomial, Radial Basis Function (RBF), and sigmoidal functions. To find out more about kernel functions and implementation of SVMs, the reader is referred to [30,31].

Results and Discussion

In this work, the preprocessing step has been carried out since the recording of the sound data. After the preprocessing, the recorded data are modeled by calculating LPC and MFCC co-efficients. The modeled signals were first separated into sub-frames, then the characteristics of each frame were extracted and pathological classification was made depending on these features. When classification was performed, 30% of the data were randomly selected as test and 70% were randomly selected as training data.

The signals modeled from the coefficients are called clean data and are removed from the recorded data to generate the residual data. Thus, the main audio data is composed of two signal components, clean and residual. The signals split in two are analyzed in two dimensions. First, features extraction of the signals is performed. Features spaces of the clean signals owing to the different coefficients are given in Figure 6.

Figure 6 shows the spatial distribution of pathological data using 3 different features. When the entropy data obtained from the MFCC coefficients are analyzed, it is observed that the normal person and the patohological diagnosed person data can be distinguished from each other seriously. In the model based on LPC coefficients, this separation can be observed more clearly. Especially when the results of entropy, jitter and shimmer are examined, it is observed that the data can be separated from one another. Owing to the Student-T test results, the association of pathological and normal people with these traits can be distinguished, giving the following results.

When Table 2 is examined, it is observed that the signal obtained from the LPC-based model gives more meaningful results than the signal obtained from the MFCC-based model. The model features formed by LPC coefficients are seen from the results that pathological and normal person data can be separated from each other. Among the features, LLE is determined from the results that it does not give meaningful results in the analysis and separation of the data of pathological and normal person data. When the characteristics of the signals obtained from the MFCC coefficients are examined, it is observed that only the Shannon entropy feature values have distinctive meanings obtained from normal and pathological data.

Figure 7 shows histogram curves of voices based on feature values. By looking at the Figure 7, especially the entropy values, it can be said that their distributions are approximately separated from each other. This has also improved the sense of classification accuracy.

On the graph of entropy and skewness histogram, it is observed that the values are approximately separated from each other. Kurtosis, shimmer, and jitter, there is no clear separation. However, there is a small discrimination in the LLE values. When we use all the features together for the classifier, the differences are significant for the classifier.

BoxPlot graphs based on the values of the features are shown in Figure 8. Figure 8, information about the property-specific distribution of pathological and healthy people. Again, based on previous findings, similar results were obtained on the basis of the evaluations. In entropy values, it differs from the normal and pathological patients in distinguishing qualities. Is is ob-
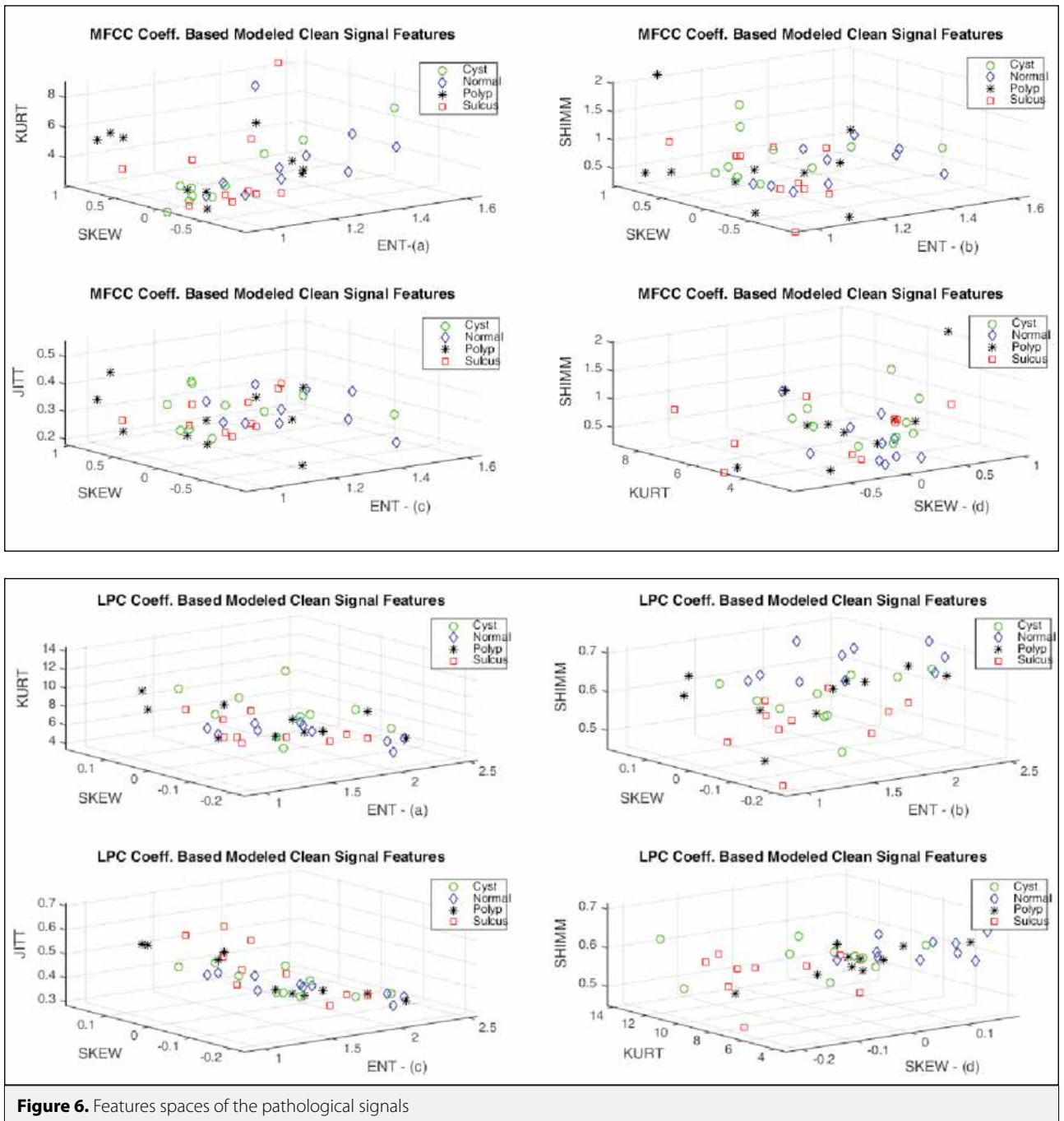
Figure 6. Features spaces of the pathological signals

served from the Figure 8 that while skewness and kurtosis are not determinant for cystic pathology, entropy, LLE and shimmer values are determinant. For sulcus pathologies, shimmer is the most defining feature.
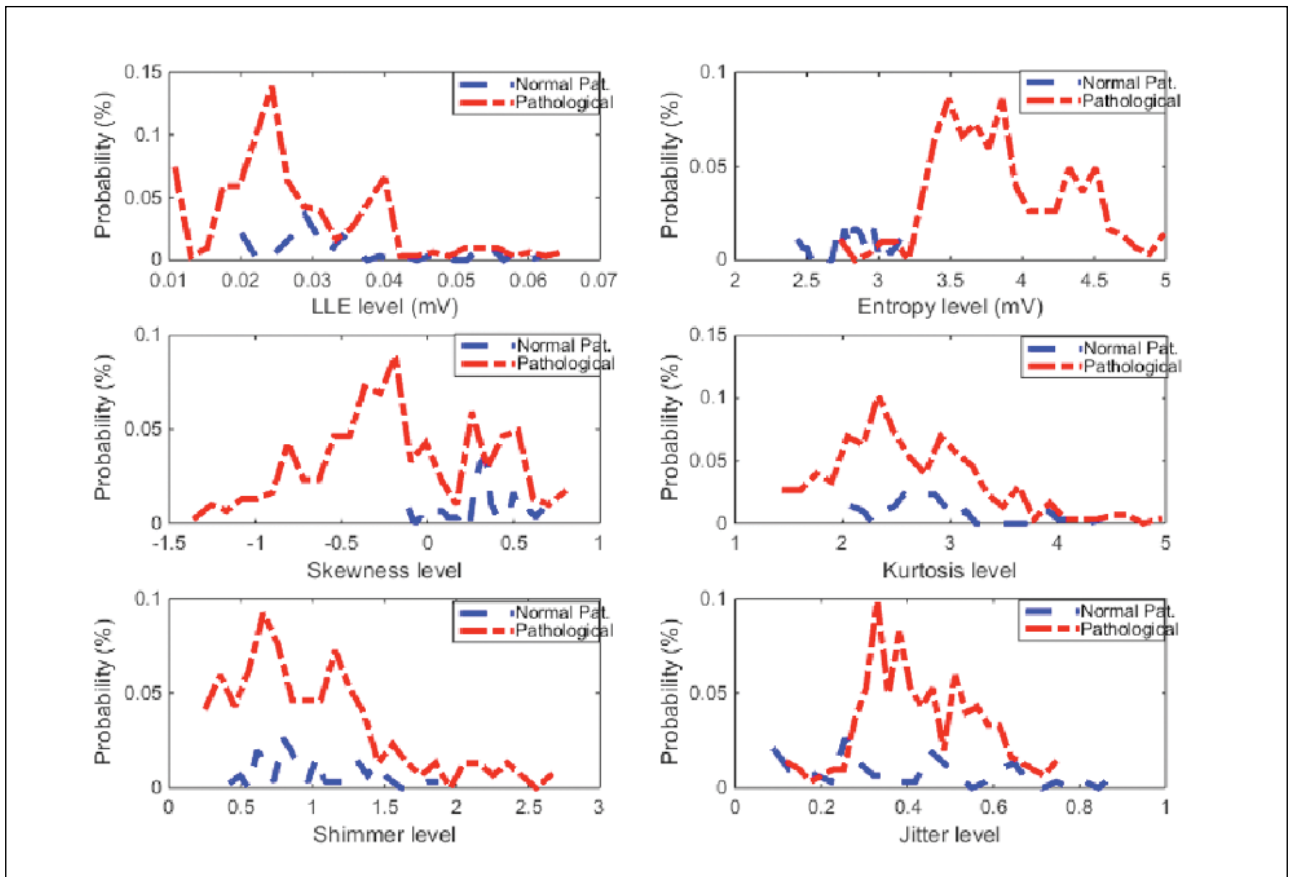
Classifier Results

In the scope of the study, SVMs with multiple classification features were preferred. Especially in multiple classifications, the ability of SVMs to be high is the reason for selecting the classification and training accuracy in the same conditions to be high compared to some other classifiers [32].

Throughout this paper, it is carried out experiments using the one-against-all SVMs design scheme. Given a five-class problem, there are five binary SVMs and each is trained to separate one class of samples, which are positive in the experiments different than the others. The decision is made according to the maximal output among these binary classifiers [30]. It has been conducted the experiments using SVMs tool, which efficiently handles large-scale SVMs learning problems [33].

Gaussian kernel as kernel and 1e-5 as lambda value are selected. In case of experimental studies, the obtained experimental

Table 2. Student-t Test results

	Separation Group	LPC coeff. Based model signal	MFCC coeff. Based model signal
Jitter	N-P	0.005987885	0.651831
	N-S	7.43402E-07	0.248628
	N-C	0.000549804	0.159287
Entropy	N-P	0.0165838	0.006515
	N-S	2.27914E-06	0.01639
	N-C	0.007125662	0.033019
Kurtosis	N-P	0.545246	0.073383
	N-S	0.589144	0.189005
	N-C	0.587493	0.465145
Shimmer	N-P	0.005434	0.182185
	N-S	0.000626	0.122752
	N-C	0.021906	0.002425
Skewness	N-P	0.368412	0.945534
	N-S	0.001053	0.250572
	N-C	0.082233	0.349198
LLE	N-P	0.593211	0.173884
	N-S	0.839202	0.210503
	N-C	0.583968	0.433694

Figure 7. Histogram of the feature values
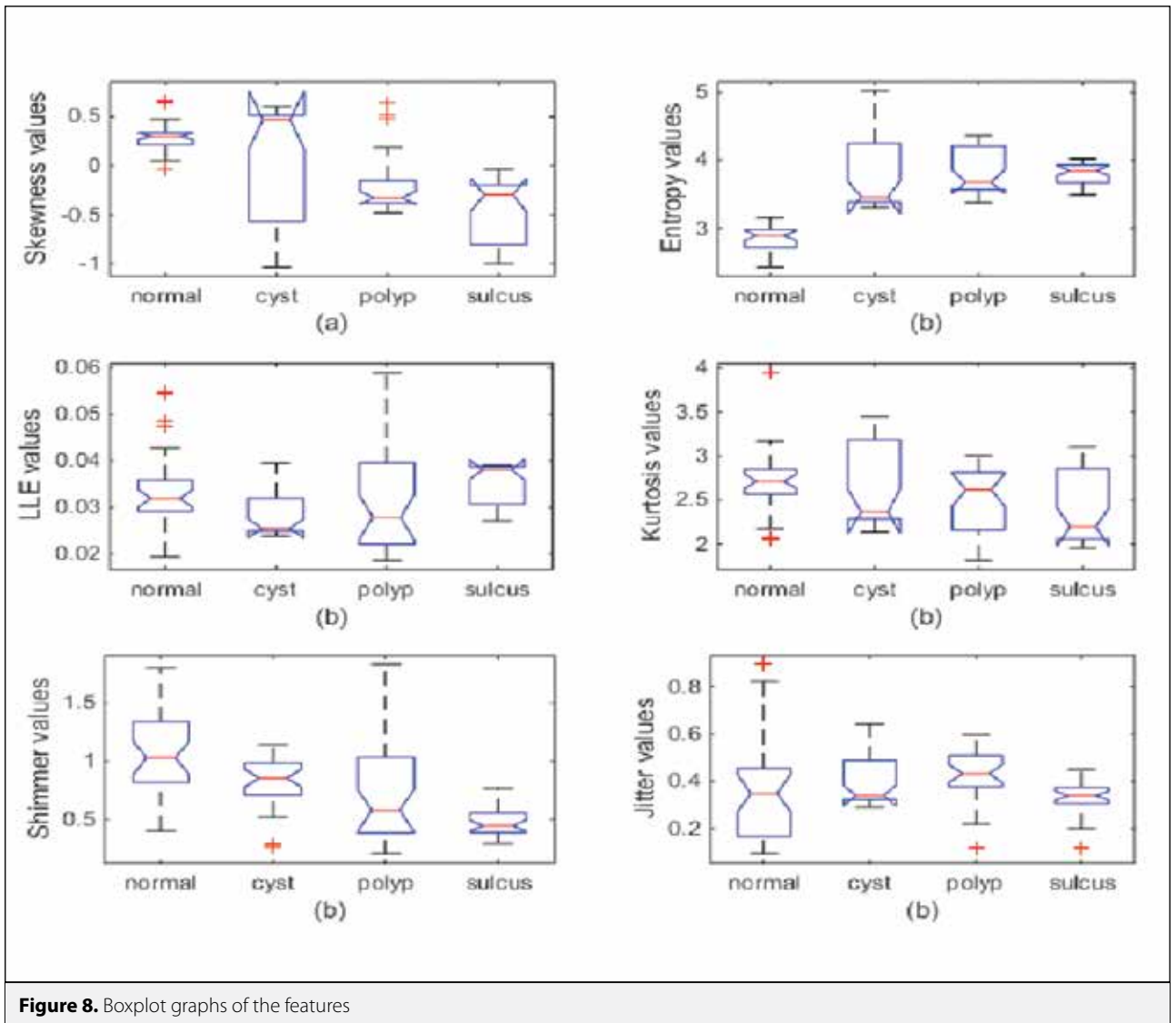
Figure 8. Boxplot graphs of the features

results have been compared with the other kernel functions (polynomial and sigmoid functions). However, the best (maximum classification results) results have been obtained with the Gaussian RBF. Classification results are given in Figure 8.

Entropy values play a critical role when looking at the classification results. Effect on the results of all classification, and other features, together with significant results. Entropy as well as the jitter that contributed most to the accuracy of the classification. It is observed from the results that the LLE is the least contribution feature. It can be seen from the student-t test the reason why we have to reach this results already. The effects that multiple features have added to the classifier are still observed on the Table 3. It can be seen from the Table 3 that each feature increased the classification accuracy of the test data results. Previous studies on the pathologic classification of voice data and a comparison of our study are given in Table 4.

When Table 4 is examined, it seems that working has some advantages. The number of features is reduced first. The process complexity is reduced and the system speed is increased. Especially in lesser numbers of reatures, the accuracy rate seems to be increased. On this count, the need for less data storage arises. In addition, chaotic analysis of the sounds was performed as a result of the operation of the proposed model. As well as pathological classification based on the characteristics of the data was performed. Thus, the study shows that pathological data can be separated less number of features.

Conclusions

In this paper, we showed multi-class SVMs whether can classify the pathological and normal voices into sub classes. It is concluded from the experimental results that multi-class SVMs provide good performance for classification of these voice data
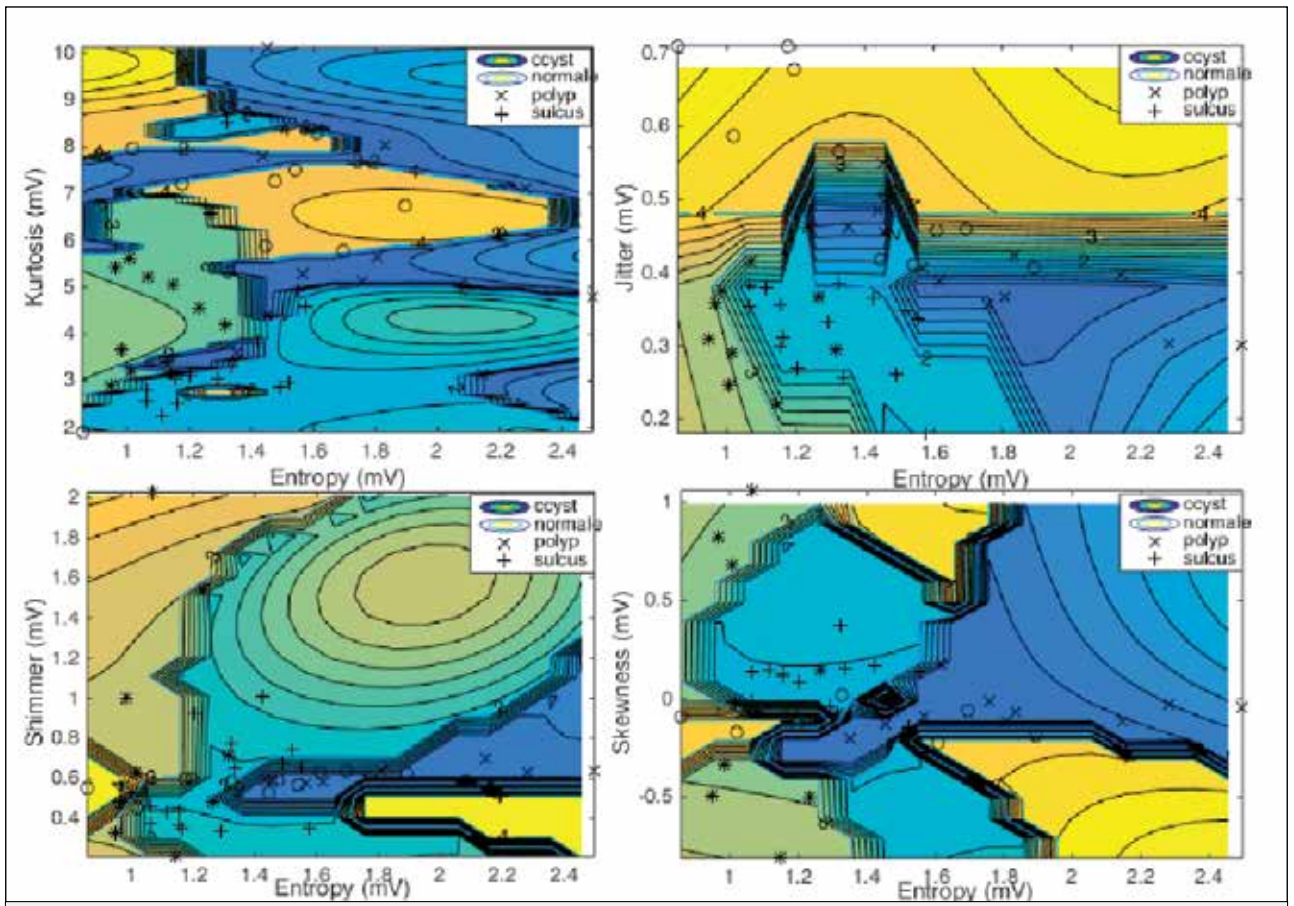
Figure 9. M-SVMs classification results depending on the related features

Table 3. M-SVMs classification results for the features

Features	Data	(%)
Entropy-Kurtosis	Training	100.00
	Test	93.45
Entropy-Jitter	Training	85.76
	Test	80.79
Entropy-Shimmer	Training	77.89
	Test	70.56
Entropy-Skewness	Training	72.67
	Test	66.60
Entropy-LLE	Training	78.87
	Test	67.00
Entropy-Kurtosis-Jitter	Training	100.00
	Test	94.57
Entropy-Kurtosis-Jitter-Shimmer	Training	100.00
	Test	96.60
Entropy-Kurtosis-Jitter-Shimmer-Skewness	Training	100.00
	Test	97.05
Entropy-Kurtosis-Jitter-Shimmer-Skewness-LLE	Training	100.00
	Test	97.20
10 features	Training	100.00
	Test	99.56

Table 4. Speech data classification studies and accuracy results

Methods			Database			
Post-Processing	Classifier	Feat.No	Normal	Pathological	Accuracy (%)	Study
-	M-SVMs	10	68	172	100.00	Proposed Study
GA and LDA	SMVs	18	53	240	98.3	[34]
LDA	SVMs	40	53	240	97.01	[35]
LDA and PCA	SVMs	64	53	67	100.00	[36]
LDA and PCA	SVMs	22	50	50	94.26	[37]
-	SVMs	26	53	173	98.23	[38]
-	ANN and LVQ	19	53	82	96	[10]

depending on used the features. For one-against-all SVMs, we proved that in associated with LLE, entropy, skewness, kurtosis, shimmer, jitter and LPC, MFCC based model coefficients, the SVMs classification performance is very good. Using the ten features, the accuracy rate of training data is 100 % for optimal C values and chossing Gaussian kernel, however, estimated accuracy rate for testing data is 99.56 %.

From the previous studies, [34-38] that showed that the related voices could be classified into normal / pathological as depending on sounds' characteristic features. On the other hand, their used classifier and methods, in [34] as accuracy rate 98.3%, in [35] as accuracy rate 97.01%, in [36] as accuracy rate 100%, in [37] as accuracy rate 94.26%, and in [38] as accuracy rate 98.23%. However, in this study we not only have increased accuracy rate of correct class for pathological and normal classification, but also are able to classify the related voices as four different classes, which are important for diagnosying speech voice' analysis. Additionally, with less number of features, obtaining the accuracy rate. So, the database storage could be reduced, not only processing complexity has gone to reduce, but also system performance being increased.

Additionally, we can easily see that gaussian RBF kernel provides good performance for classifying data. Finally, we demonstrated by computer experiment that multi-class SVMs can provide good result and can detect normal, voice sound segments.

Acknowledgements

References

1. J. R. O. Arroyave, J.F.V. Bonilla, E. D. Trejos, "Acoustic analysis and non linear dynamics applied to voice pathology detection: A review", *Recent Patents on Signal Processing,* vol. 2, pp. 1–11, 2012.

2. S. Fong, K. Lan, R. Wong, "Classifying human voices by using hybrid SFX time-series preprocessing and ensemble feature selection", *BioMed Research International,* vol. 1–27, 2013.

3. J. I. Godino-Llorente, N. S´aenz-Lech´on, V. Osma-Ruiz, S.Aguilera-Navarro, P. G´omez- Vilda, "An integrated tool for the diagnosis of voice disorders," *Medical Engineering & Physics,* vol. 28, pp. 276–289, 2006 [CrossRef]

4. C. Manfredi, M. D'Aniello, P. Bruscaglioni, A. Ismaelli, "A comparative analysis of fundamental frequency estimation methods with application to pathological voices," *Medical Engineering & Physics,* vol. 22, pp. 135–147, 2000. [CrossRef]

5. M. Farrus, J. Hernando, P. Ejarque, "Jitter and shimmer measurements for speaker Recognition," *Annual Conference of the International Speech Communication Association (Interspeech 2007),* Antwerp, Belgium, pp. 778–781, 2007.

6. P. Lieberman, "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges", *The Journal of the Acoustical Society of America,* vol. 35, pp. 344–353, 1963. [CrossRef]

7. Y. Horii, "Vocal shimmer in sustained phonation," *Journal of Speech, Language, and Hearing Research,* vol. 23, pp. 202–209, 1980. [CrossRef]

8. E. Yumoto, W. J. Gould, T. Baer, "Harmonics-to-noise ratio as an index of the degree of hoarseness," *The Journal of the Acoustical Society of America,* vol. 71, pp: 1544–1550, 1982. [CrossRef]

9. L. R. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition,"Vol. 14, PTR Prentice Hall, Englewood Cliffs, NJ.

10. J. I. Godino-Llorente, P. Gomez-Vilda "Automatic detection of voice impairmentsby means of short-term cepstral parameters and neural network based detectors," *IEEE Trans. Biomed. Eng.,* vol. 51, pp. 380–384, 2004. [CrossRef]

11. Y. Zhang, et al. "Nonlinear dynamic analysis of voices before and after surgicalexcision of vocal polyps," *J. Acoust. Soc. Am.,* vol. 115, pp. 2270–2277, 2004. [CrossRef]

12. T. L. Eadie, P. C. Doyle, "Classification of dysphonic voice: Acoustic and auditory-perceptual measures," *Journal of Voice,* vol. 19, pp. 1–14, 2005. [CrossRef]

13. S. Hadjitodorov and P. Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening," Medical Engineering & Physics, vol. 24, p. 419–429, 2002. [CrossRef]

14. N. Saenz-Lechon, J. I. Godino-Llorente, V. Osma-Ruiz, M. Blanco-Velasco, F. Cruz-Roldan, "Automatic assessment of voice quality according to the GRBAS scale," 28th Annual International Con-

ference of the IEEE Engineering in Medicine and Biology Society, EMBS'06, New York, NY, USA, pp. 2478–2481, 2006. [CrossRef]

15. K. Umapathy, S. Krishnan, V. Parsa, D. G. Jamieson, "Discrimination of pathological voices using a time-frequency approach," *IEEE Transactions on Biomedical Engineering*, vol. 52, p. 421–430, 2005. [CrossRef]

16. J. I. 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 Transactions on Biomedical Engineering,* vol. 53, p. 1943–1953, 2006. [CrossRef]

17. S. Jothilakshmi, "Automatic system to detect the type of voice pathology," *Applied Soft Computing,* 2014. [CrossRef]

18. A. Skalski, T. Zielinski, D. Deliyski, "Analysis of vocal folds movement in high speed videoendoscopy based on level set segmentation and image registration," *International Conference on Signals and Electronic Systems,* ICSES' 08, Krak´ow, Poland, pp. 223–226, 2008. [CrossRef]

19. D. D. Mehta, D. D. Deliyski, S. M. Zeitels, T. F. Quatieri, R. E. Hillman, "Voice production mechanisms following phonosurgical treatment of early glottic cancer," T*he Annals of Otology, Rhinology, and Laryngology,* vol. 119, p. 1, 2010. [CrossRef]

20. R. Srinivasan, R. Rengaswamy, R. Miller, "A modified empirical mode decomposition (EMD) process for oscillation characterization in control loops," *Control Engineering Practice* vol. 15, pp. 1135-1148, September, 2007. [CrossRef]

21. J. C. Saldanha, T. Ananthakrishna, R. Pinto, "Vocal fold pathology assessment using mel- frequency cepstral coefficients and linear predictive cepstral coefficients features," *Journal of Medical Imaging and Health Informatics,* vol. 4, pp. 168–173, 2014. [CrossRef]

22. L. Deng and O. Douglas O. "Speech processing: a dynamic and optimization-oriented approach," *Marcel Dekker,* pp. 41–48, 2003.

23. N. Levinson, "The Wiener RMS error criterion in filter design and prediction," *J. Math. Phys.,* vol. 25, pp. 261–278, 1947. [CrossRef]

24. M. Xu, et al. "HMM-based audio keyword generation," In Kiyoharu Aizawa; Yuichi Nakamura; Shin'ichi Satoh. Advances in Multimedia Information Processing – PCM 2004: 5th Pacific Rim Conference on Multimedia (PDF). Springer. [CrossRef]

25. Takens F. "Detecting strange attractors in turbulence", Dynamical Systemsand Turbulence, Warwick 1980, Springer, 1981, pp. 366–381.

26. N. L. Johnson, S. Kotz, N. Balakrishnan "Continuous Univariate Distributions," 1994, vol.1, 2nd Edition Wiley.

27. P. H. Westfall, " Kurtosis as Peakedness," 1905–2014. R.I.P., The American Statistician 68, 2014, p. 191–195. [CrossRef]

28. C. E. Shannon, "A mathematical theory of communication", *Bell System Technical Journal,* vol. 27, pp. 623–656, 1948. [CrossRef]

29. M. T. Rosenstein, J. J. Collins, C. J. De Luca. "A practical method for calculating largest Lyapunov exponents from small data sets," *Physica D,* vol. 65, pp. 117–134, 1993. [CrossRef]

30. V. Vapnik. "Statistical Learning Theory", New York, NY, USA: John Wiley&Sons, 1998.

31. T. Joachims, "Making large-scale SVM learning practical in Advances in Kernel Methods-Support Vector Learning", B. Schlkopf, C. J. C. Burges, and A. J. Smola, Eds., 1999, pp. 169–184, MIT Press, Cambridge, Mass, USA.

32. H. Ankışhan and D. Yılmaz, "Comparison of SVMs and ANFIS for Snore Related Sounds Classification by Using the Largest Lyapunov Exponents," *Computational and Mathematical Methods in Medicine,* vol. 2013, 2013.

33. C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transaction on Neural Networks,* vol. 13, pp. 415-426, May 2002. [CrossRef]

34. G. Hamzeh, T. K. Mehdi, K. A. Meisam, P. Mohammad "Detection of vocal disorders based on phase space parameters and Lyapunov spectrum," *Biomedical Signal Processing and Control*, vol. 22, p. 135–145, 2015. [CrossRef]

35. A. Akbari, M. K. Arjmandi, "Employing linear prediction residual signal of waveletsub-bands in automatic detection of laryngeal pathology," *Biomed. Signal Processing and Control* vol. 18, 293–302, 2015. [CrossRef]

36. M. K. Arjmandi, M. Pooyan, "An optimum algorithm in pathological voice qualityassessment using wavelet-packet-based features, linear discriminant analysisand support vector machine," *Biomedical Signal Processing and Control,* vol. 7, pp. 3–19, 2012. [CrossRef]

37. M. K. Arjmandi, et al. "Identification of voice disorders using long-time featuresand support vector machine with different feature reduction methods," *J. Voice,* vol. 25, pp. 275–289, 2011. [CrossRef]

38. J. D. Arias-Londono, et al. "Automatic detection of pathological voices using com-plexity measures, noise parameters, and mel-cepstral coefficients," *IEEE Trans. Biomed. Eng.* vol. 58, pp. 370–379, 2011. [CrossRef]

Haydar Ankışhan was born in Ankara, Turkey, 1979. He received th B.S. degree in electrical and electronics engineering from the Gaziantep University, M.S. and Ph.D. degrees in electronics/ electrical and electronics engineering from the Ankara University, Turkey, 2003, 2007, 2015., respectively. Currently, he works as an assistant professor since 2015 in Vocational School of Technical Sciences in Baskent University, Turkey. His main research interests are acoustic signal processing, biomedical signal processing and implementation, medical devices, simultaneous localization and mapping.