Academic Platform Journal of Engineering and Science

journal homepage: http://apjes.com/

Detection of Vocal Cyst Problem by Using High Order Moments and Support Vector Machines

*[1]Derya Yılmaz, [2]Haydar Ankışhan

[1]Department of Electrical and Electronics Engineering, Başkent University, Ankara, Turkey, derya@baskent.edu.tr,

[2]Vocational School of Technical Sciences, Başkent University, Ankara, Turkey, hankishan@baskent.edu.tr,

Reseach Paper Received Date: 17.01.2018 Accepted Date: 27.07.2018

Abstract

The voice disorders occurring due to the problems in the voice producing organs cause some changes in the intensity or tone of the voice. It is difficult to identify the diseased voice by the reason of its variable and different nature. One of the most popular voice disorder reasons is cyst which is located on the vocal cords. The aim of this study is to detect the vocal cyst problem by using acoustic voices data which were recorded from healthy people and patient with cyst diagnoses subjects by using high order statistics and support vector machines (SVMs) classifier. In this study, two experimental procedures were implemented for two different voice samples. In the first, /a/ vowel and the second, the Turkish word of “aydınlık” (mean in English “bright”) were investigated with skewness and kurtosis parameters which are third and fourth order cumulants (spectral moments), respectively. The obtained features values for healthy and cyst subjects were used as the SVMs’ inputs for classification. The experimental results show that the test accuracies of SVMs were found as 94.89% and 91.11% for /a/ vowel and “aydınlık” word, respectively. It is concluded from experimental studies that skewness provides more meaningful results than kurtosis in relation to distinguish into two voice groups as healthy and cyst. Additionally, it is assessed that the “aydınlık” is affective word for the pathological and normal acoustic voice discrimination as good as /a/vowel.

Keywords: Voice disorders, skewness, kurtosis, support vector machines

1. INTRODUCTION

Voice is aspecific identifier for each person and is an important tool that allows people to communicate with their environment.The voice characteristic of a healthy person is unique and may be having different features values from the other one. When distorted, the voice characteristics may change.Voice occurs by vibrating the vocal cords in larynx during speaking. Voice disorders may arise related to problems of voice producing organs which are inside and outside of the larynx [1]. There are many types of voice disorders such as vocal cord paralysis, chronic laryngitis, including polyps, cysts, nodules or sores on the vocal cords and neurological diseases [1, 2]. The invasive endoscopic device called as Stroboscopy/Laryngoscopy is generally used in the diagnosis of voice disorders by monitored the voice producing organs [1].These procedures need an equipped speech laboratory and an expert. Recently, there have been some voice analysis programs that uses non-invasive techniques developed for the investigating the voice quality and detecting the voice disorders and tracking the treatments process [2-7]. It is known that the characteristic of a healthy voice can change depending on many factors such as age, gender, spoken language, the structure of the mouth and tongue etc [2, 8-10]. Therefore, it is very difficult to determine invariant properties from any healthy sound. Hence, to distinguish the voice having pathology from the healthy one is quite difficult task.

The acoustic features used in clinical studies based on digital signal processing are commonly obtained by the analysis of vowels. Jitter, shimmer, harmonic-to-noiseratio (HNR), fundamental frequency, formant frequencies, Mel-frequency cepstral coefficients (MFCCs) are the most important and frequently used time and frequency parameters [5-7, 11, 12]. These parameters vary depending on voice disorders with respect to the healthy or pathological state. However, in recent years, some other methods such as wavelet, hidden Markov model (HMM), linear predictive coding (LPC) have been used for feature extracting from acoustic samples [12-17]. Additionally, many classifier methods such as quadratic discriminant analysis (QDA), self-organizing map (SOM), kth-nearest neighbours (KNN), artificial neural networks (ANN), support vector machines (SVMs), etc. have been also used in this kind of studies for the discrimination of voice

*[1]Corresponding author: Department of Electrical and Electronics Engineering, Başkent University, Bağlıca Campus, ESK. Road. 20. Km. Etimesgut/Ankara, derya@baskent.edu.tr, 0312 2466666/1303

Doi: 10.21541/apjes.380271

disorders [5-7, 11-18]. Classifier can easily make a decision according to parameters' values with respect to the training and test process.

There have been some studies, which tried to classify the acoustic voices into sub groups as pathological and other, available in the literature. Rosa et al., [5] have classified healthy and six kinds of voice disorders by using the time and frequency based features with ANN classifier. In their study, /a/, /e/ and /i/ vowels were used and they achieved the classification accuracy rate of 65% [5]. Arjmandi et al., [6] have used 22 features extracted from traditional voice analysis program to classify voice disorders and compared six classifier results with feature reduction methods such as principal component analysis (PCA) and LDA. According to their results, SVMs with LDA showed the best performance by getting 94.26% accuracy [6]. Wang et al., [10] have applied the thirty-six-dimensional MFCC parameters to the Gaussian mixture model (GMM) and the GMM-SVMs classifiers for detection of voice pathology, and obtained the best classification accuracy of 96.1±2.51% with the GMM-SVMs [11]. Hadjitodorov and Mitev [12] have used /a/ vowel with features such as jitter, shimmer, harmonics to noise ratio (HNR), degree of hoarseness (DH), first harmonic energy to the energy of the other harmonics ratio (NFHE) and turbulence noise index (TNI) for the classification of voices into normal and pathological. They showed that the accuracy rate with KNN can be increased to 96.1% by using NFHE and TNI [12]. Additionally, there have also been some studies available for extracting the new features such as higher order parameters, nonlinear characteristics etc. for increasing the success of the classifiers in the literature [17, 18, 20-22].

Higher order statistics (HOS) are an extension of second order measurements such as autocorrelation and power spectrums into higher orders. The signals which have a Gaussian distribution can be represented by second order statistics. Although some physiological signals cannot be represented by lower order statistics the cause of not having Gaussian distribution, some meaningful information can be obtained from higher order moments of non-Gaussian distributions by using the non-traditional signal processing methods [19]. So, there have been some studies available to be performed by using model-based approaches and spectral moments in the literature [18, 20-22]. Muhammad et al., [18] have studied on vocal tract area irregularity with five features (average, variance, variance ratio, skewness, kurtosis) and have classified pathological voices from two different databases by using SVMs [18]. For these databases, the accuracies were obtained 99.22±0.01% and 94.7±0.21% [18]. Tanner et al., [20] have also used four spectral moments (spectral mean, standard deviation, skewness, kurtosis) for investigating voice characteristics after therapy. Their results indicated that the changes in skewness and kurtosis were not significant, but spectral mean may be one acoustic marker sensitive to the improvement of the dysphonia severity [20]. Another study was achieved by Lowell et al., [21] who tried to determine whether all spectral moments and cepstral peak prominence (CPP) measures are significantly different from voice disorders of speakers or not. They found that significant differences are available in CPP, spectral mean, skewness and kurtosis between different groups ($P<0.001$) [21].

In this study, with consideration of previous studies, HOS parameters (skewness and kurtosis) were used as features. The earlier studies have generally made for the classification of the healthy and pathological voices which have many types of voice problems into two sub groups in the literature. Unlike the previous ones, patients having cyst problem was only tried to be detected from two voice group samples in this study. One of the voice samples selected for this study is /a/ vowel since previous studies showed that it gives less variable information about the acoustic features, that is, having minimum changes over time. The voice channel acoustic effects are quite stable in sufficiently long range for this vowel and hence the transfer function of /a/ vowel is relatively unaffected by small movements. The word of "aydınlık" is additionally used as other voice sample in this study. In here, as an additional purpose, in order to contribute to the literature for Turkish voice analysis, the "aydınlık" word was selected, because of "aydınlık" word having "ı" vowel that has the different voice characteristic in comparison with the other ones [10, 23]. The validation of estimated features was evaluated with statistical test and they were applied to the SVMs classifier. It is obtained from the experimental studies that the SVMs classifier provides the good results in order to see the benefits of expert systems approach for the decision-making process in voice disorder detection [6, 11].

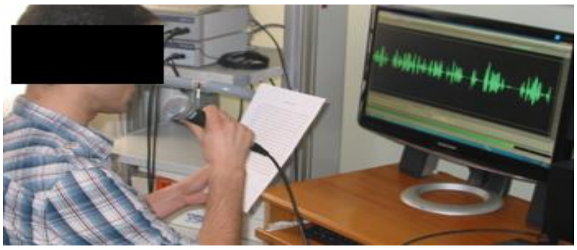

Figure 1. The recording setup.

2. MATERIALS AND METHODS

2.1. Data and Procedure of Study

In this study, data were obtained from University of Sütçü İmam at Faculty of Medicine in Department of Otolaryngology. 33 subjects (20-37 ages) data were recorded from 17 healthy and 16 patients who have cyst problem. As stated in [3], the recording system consists of a dynamic single direction microphone (Shure SM58), an interface (EMU Tracker Pre) which is suitable for especially recording voice samples and a recording and

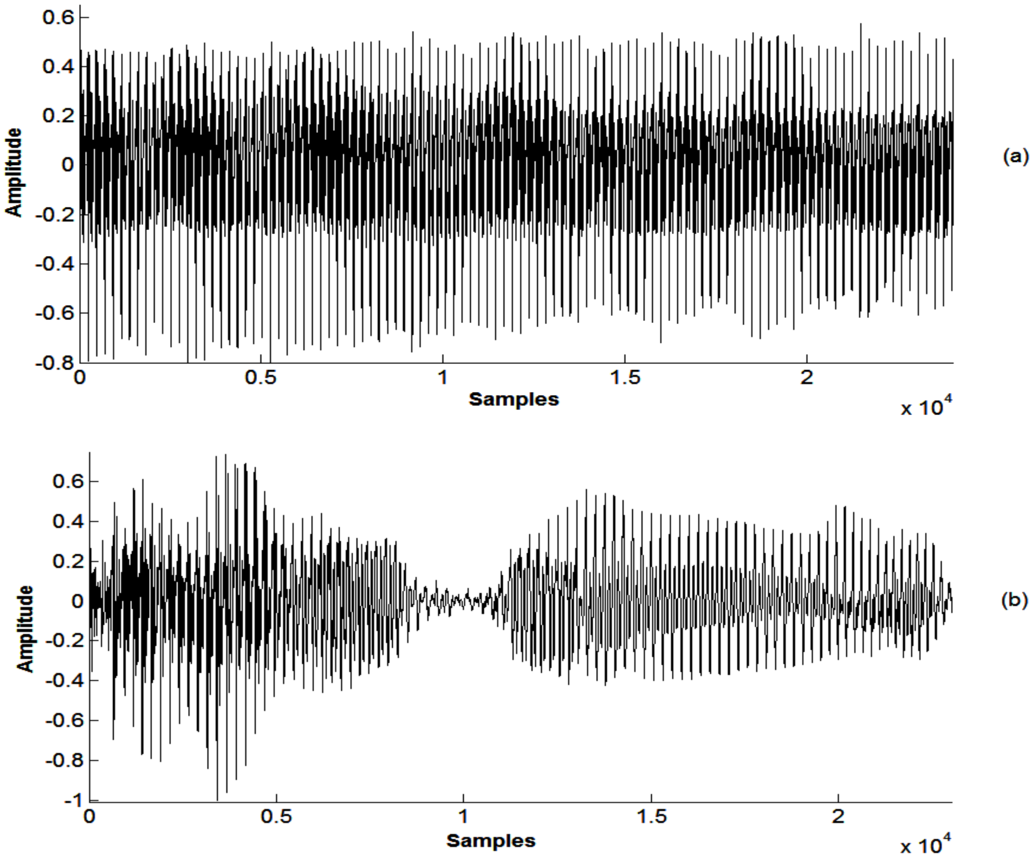

Figure 2. The voice signal samples (a) /a/ vowel (b) word of “aydınlık”.

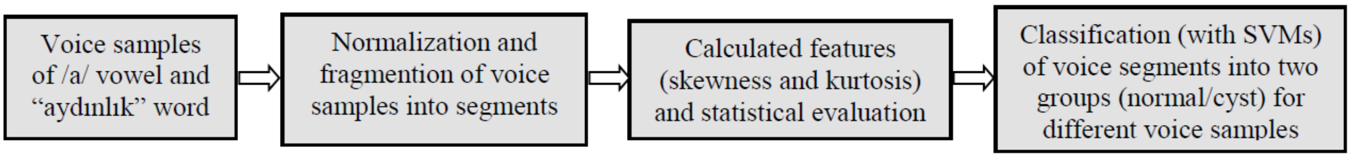

Figure 3. The procedure of study.

editing program (Adobe Audition CS5.5) with Windows 7 Ultimate ultra-noiseless computer. The record was made in a silent room which has 45-50 dB(A) background noise level, with 44100 sampling rate and 16 bits resolution level in single channel pulse code modulation (PCM) format [3]. Fig.1. shows the recording setup.

Each subject has two kinds of recorded samples, which are /a/ vowel and a paragraph that includes different words. /a/ vowel was recorded average 3 seconds, and the sentence (“Dar kapısından başka aydınlık girecek hiçbir yeri olmayan dükkânında, tek başına, gece gündüz kıvılcımlar saçarak çalışan Koca Ali, tıpkı kafese konmuş terbiyeli bir arslanı andırıyordu.” Diyet, Ömer Seyfettin) which contains “aydınlık” word was recorded average 12 seconds. After restoring data, the “aydınlık” word is removed from reading sentences and saved for each subject. The two samples of data used in this study are given in Fig. 2.

The procedure of study is given in Fig. 3. In this study, data (voice sample of /a/ vowel and word of “aydınlık”) were sub-divided into the segments after the normalization; each having 17640 samples (400ms). For all segments, skewness and kurtosis values were calculated and statistically evaluated. These features values obtained for two different voice samples /a/ vowel and “aydınlık” word were separately applied to the SVMs’ inputs for classification into two groups as normal/cyst. For different voice samples, the performances of SVMs with two features were evaluated.

2.2. Calculated Features

The important step is to extract meaningful features from the voice segments in this study, after the segmentation. In here, skewness and kurtosis values of samples were used to determine the features of voice segments.

Skewness: Skewness that is meaningful features for the analyzing of voice sample distributions measures the asymmetry of the probability distribution of voice. Karl Pearson first suggested measuring the skewness by standardizing the difference between the mean and mode [24]. The skewness of a voice sample (*X*) is the third standardized moment, which is given in equation (1) [22];

$$\gamma_1 = \frac{\Sigma(X-\mu)^3}{n\sigma^3} \tag{1}$$

where n is the number of samples, μ is mean, and σ standard deviation of *X*.

Kurtosis: Karl Pearson defined the kurtosis as a measure of how flat the top of a symmetric distribution is when compared to a normal distribution of the same variance [25]. Kurtosis of a voice sample *X* is the fourth standardized moment, which is given in equation (2) [22];

$$\eta = \beta_2 - 3 \tag{2}$$

where

$$\beta_2 = \frac{\Sigma(X-\mu)^4}{n\sigma^4} \tag{3}$$

where β_2 is often referred to as "Pearson's Kurtosis".

2.3. Support Vector Machines (SVMs)

Suppose we have a set of training data $((x_i, y_i)$, for i=1,…,n and xi Є Rd and yi Є {±1}) and a nonlinear transformation to a higher dimensional space, the feature space $(\Phi(.,.), R^d \xrightarrow{\Phi(.,.)} R^H)$, the SVMs solves

Minimize (w, ζi, b) $$\left\{ \frac{1}{2}\|w\|^2 + C\sum_i \zeta_i \right\} \tag{4}$$

Subject to $$y_i(\phi^T(x_i)w + b) \geq 1 - \zeta_i \quad \forall_i = 1,\dots,n \;\; \zeta_i \geq 0 \tag{5}$$

where (w, b) defines the linear classifier in the feature space. The SVMs try to enforce positive samples to present an output greater than +1, and the negative samples present an output less than -1 for two class. The details of the algorithm are referred to [26].

3. RESULTS AND DISCUSSION

The estimated probability distribution functions of the calculated features for all segments were given in Fig. 4.

Fig. 4 gives information about histograms of the skewness (a) and kurtosis (b) feature values for /a/ vowel.To look at the figure, it is seen that the feature values of skewness can be easily separated for two voice groups. However, kurtosis values do not give specific and meaningful information about discrimination of the subjects.

The descriptive statistics of voice signal demonstrated that the average values of the skewness and kurtosis values of /a/ vowel and "aydınlık" word are different for both groups. The mean, median, minimum, maximum, first quartile (Q1), and third quartile (Q3) values of the features are given in Fig. 5. To compare the means of skewness and kurtosis values for the cyst and normal subjects, we can see the results in Fig. 5 after the transformation of variable response according to the box plot. Where the central horizontal mark represents the median, the horizontal edges of the box represent the 25th (Q1) and 75th (Q3) percentiles, vertical discrete mark in on above and below the box is the whisker, and plus marks are the outside values.

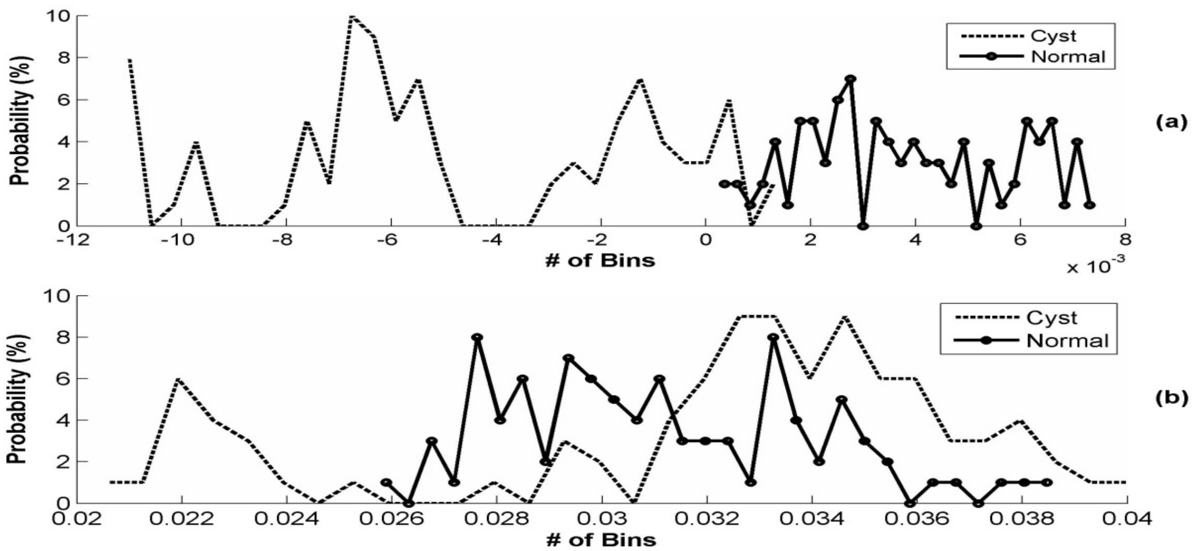

Figure 4. Feature histograms of /a/ vowel parts: a) Skewness and b) Kurtosis.

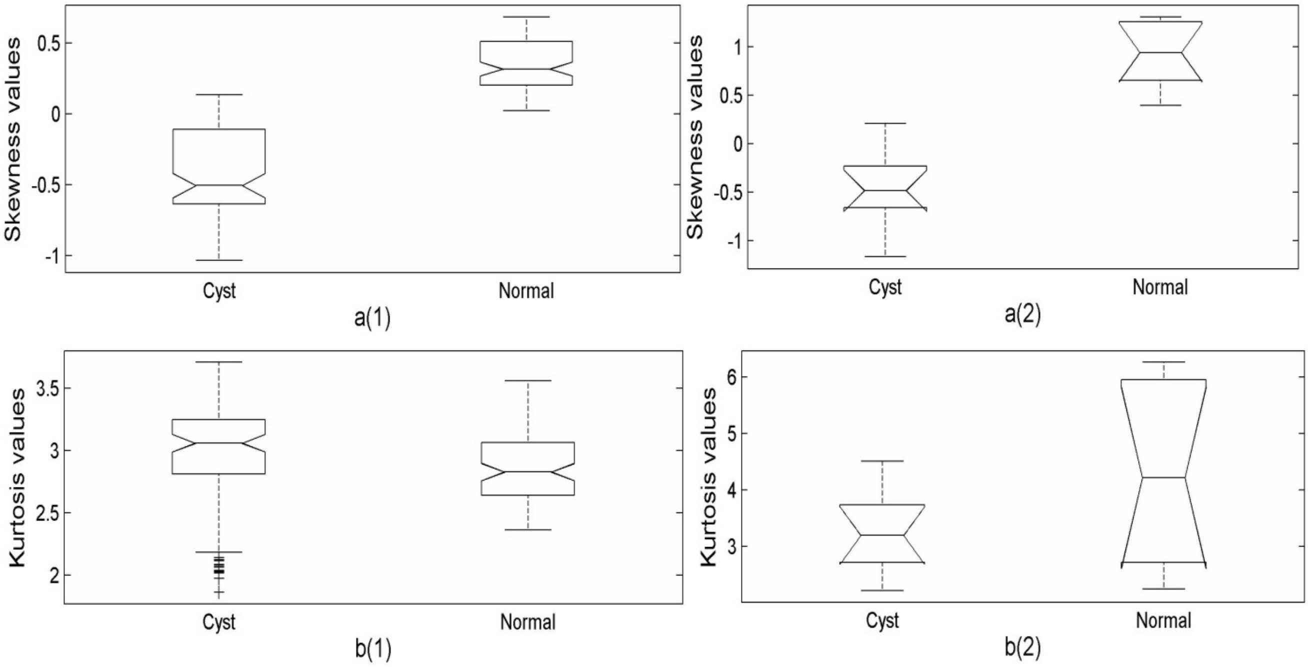

Figure 5. Box plot of skewness and kurtosis values for /a/ vowel (a(1), b(1)) and word of "aydınlık" (a(2), b(2)) for cyst and normal subjects.

Table 1. Descriptive statistics of /a/ vowel and word of "aydınlık".

Groups	Feature	Voice	Max.	Min.	Mean
Cyst	Skewness	/a/	-0.08	-0.98	-0.47
		"aydınlık"	0.29	-0.69	-0.27
	Kurtosis	/a/	3.41	2.07	2.93
		"aydınlık"	3.41	2.91	3.18
Normal	Skewness	/a/	0.62	0.22	0.37
		"aydınlık"	1.13	0.57	0.89
	Kurtosis	/a/	3.14	2.60	2.88
		"aydınlık"	5.14	2.71	4.18

The details of features obtained from the subject's data are given in Table 1. The Table 1 shows that /a/ vowel and "aydınlık" word presented smaller values for cyst subjects than the normal ones for skewness levels, whereas they did not give specific information about levels of kurtosis.

Table 2. Results of Student-t tests.

Groups	Feature	Voice	P value
Cyst/Normal	Skewness	/a/	**<0.0001**
		"aydınlık"	**<0.0001**
	Kurtosis	/a/	0.241
		"aydınlık"	**0.021**

To perform the statistical analysis of skewness and kurtosis values, the student-t test was used with a significant level of $P<0.05$ and the results were given in Table 2. The obtained significant results for the kurtosis values of the word of "aydınlık" is noticeable, it shows that the speech samples may give many effective results than vowel samples.

In this paper, we trust on the previous studies that we used SVMs to classify voice data automatically into two groups with the help of the HOS features. The subjects in training and testing datasets are different; that is, the training set was composed of 18 subjects (9 normal and 9 cysts). The other 15 subjects (8 normal and 7 cysts) were used as test set. The proposed algorithm for SVMs classifier, optimal C value is chosen as 10^6 and regularization parameter (*Lambda*) is 10^{-7}. So the classification results are given in Table 3. In Table 3, performance of classifier was evaluated in terms of accuracy (Acc), which is defined as 100x(TP+TN)/(TP+FN+FP+TN), and positive predictive value (PPV), which is defined as 100xTP/(TP+FP), and sensitivity (Sens), which is defined as 100xTP/(TP+FN). It can be seen from Table 3 that the results are good enough to classify normal and cyst problem subjects into two groups. The classifier can easily classify subjects with used two features.

4. CONCLUSION

In this study, it is focused on the identification of cyst disorder from acoustic voices with a specific words and vowels using certain features. Within the scope of the study, /a/ vowel and "aydınlık" word that from the recorded paragraph were used for cyst disorder detection. Along with the minimum number of features (skewness and kurtosis) used, experimental studies have shown that significant differences occurred between the voice groups concerning skewness values of /a/ vowel and word of "aydınlık" and

Table 3. Results of SVMs classifier.

Voice samples	TP	FN	TN	FP	Sens (%)	PPV (%)	Acc (%)	
/a/	82	7	74	0	92.13	100	**95.71**	**Training**
"aydınlık"	29	3	24	0	90.63	100	**94.64**	
/a/	74	6	56	1	92.5	98.67	**94.89**	**Testing**
"aydınlık"	24	4	17	0	85.71	100	**91.11**	

kurtosis values of word "aydınlık", whereas kurtosis values for /a/ vowel did not give significant difference than normal ones. In this study, we used SVMs for classification of the groups as normal/cyst and evaluate the performance of the accuracy with two features. Classifier has provided feasible performance and its best testing accuracy is 94.89% for vowel /a/. However the 91.11% testing accuracy result of "aydınlık" word is as good as vowel /a/ result.

In the previous studies in the literature, different datasets in which have many voice problems (cyst, polyp, sulcus vocalis, etc.) were collected to one disordered group, many features, models and some additional techniques were used. In this study only two features were used and only cyst problem was studied. The study showed that, in the comparison between cyst and normal subjects, skewness is more distinctive parameter than kurtosis for both process. When the results of this study were evaluated with the previous one especially which using the same features [18, 20, 21, 22], it seems to be quite considerable. Additionally, the results of Turkish word of "aydınlık" are promising for the following studies in Turkish literature of voice analysis.

Acknowledgement

The authors would like to thank Prof. Dr. M. Akif Kılıç who provided the data used in this study.

REFERENCES

[1]. Lions Voice Clinic. University of Minnesota, Department of Otolaryngology, P.O. Box 487, 420 Delaware St., SE, Minneapolis, MN 55455, USA. URL:http://www.lionsvoiceclinic.umn.edu (Access time: October 2017)
[2].R.H.G. Martins, H.A. Amaral, E.L.M. Tavares, M.G. Martins, T.M. Gonçalves, N.H. Dias, "Voice disorders: Etilogy and diagnosis", Journal of Voice, 30(6):761-769, (2016).
[3].M.A. Kılıç, "Ses problem olan hastanın objektif ve subjektif yöntemlerle değerlendirilmesi", CurrPract ORL, 6(2):257-265, (2010).
[4].A. Lovato, W. De Colle, L. Giacomelli, A. Piacente, L. Righetto, G. Marioni, C. Filippis, "Multi-dimensional voice program (MDVP) vs Praat for assessing euphonic subjects: A preliminary study on the gender-discriminating power of acoustic analysis software", Journal of Voice, 30(6):765e1-765e5, (2016).
[5].M.O. Rosa, J.C. Pereira, M. Grellet, and A.C.P.L.F. Carvalho, "Signal processing and statistical procedures to identify laryngeal pathologies", IEEE International Conference on Electronics, Circuits, and Systems, 1: 423-426, (1999).
[6].M.K. Arjmandi, M. Pooyan, M. Mikaili, M. Vali, A. Moqarehzadeh, "Identification of voice disorders using long-time features and support vector machine with different feature reduction methods", Journal of Voice, 25(6), e275-e289, (2011).
[7].S.W. Sonu, R.K. Sharma, "Disease detection using analysis of voice parameters", International Journal of Computing Science and Communication Technologies, 4(2), (2012).
[8].M.A. Kılıç, "Ses bozuklukları: Yeni bir sınıflandırma sistemi", Ç. Ü. Tıp Fakültesi Arşiv Kaynak Tarama Dergisi, 8(3):321-337, (1999).
[9].M. Gerçeker, İ. Yorulmaz, A. Ural, "Ses ve konuşma", K.B.B. ve Baş Boyun Cerrahisi Dergisi, 8(1):71-78, (2000).
[10]. O. Türk, Ö. Şayli, A.S. Özsoy, L.M. Arslan, "Türkçe'de ünlülerin formant analizi", Proceedings of the 18th National Conference in Turkish Linguistics, Ankara, Turkey, (2004).
[11]. X. Wang, J. Zhang, Y. Yan, "Discrimination between pathological and normal voices using GMM-SVM approach", Journal of Voice, 25(1):38-43, (2011).
[12]. S. Hadjitodorov, P. Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening", Med. Eng. Phys., 24:419–429, (2002).
[13]. J.D. Arias-Londono, J.I. Godino-Llorente, N. Saenz-Lechon, V. Osma-Ruiz, G. Castellanos-Dominguez, "An improved method for voice pathology detection by means of a HMM-based feature space transformation", Pattern Recognition, 9(43):3100-3112, (2010).
[14]. I.A. Mahmoud, S.A. Hanaa, "Wavelet-based Mel frequency cepstral coefficient for speaker identification using hidden Markov models", Journal of Telecommunications, 1 (2):16-21, (2010).
[15]. M.K. Arjmandi, M. Pooyan, "An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine", Biomedical Signal Processing and Control, 7:3-19, (2012).

[16]. Ö. Eskidere, Ö. Aktaş, C. Ünal, "Voice disorders identification using discrete wavelet based features", Medical Technologies National Conference (TIPTEKNO), 15-18 Oct., Bodrum, Turkey, (2015).
[17]. J.B. Alonso, F. Díaz-de-Maria, C.M. Trivieso, M.A. Ferrer, "Using nonlinear features for voice disorder detection," 3rd Int. Conf. Nonlinear Speech Process, Barcelona, Spain, Apr. pp 94-106, (2005).
[18]. G. Muhammad, G. Altuwaijri, M. Alsulaiman, Z. Ali, T.A. Mesallam, M. Farahat, K.H. Malki, A. Al-nasheri, "Automatic voice pathology detection and classification using vocal tract area irregularity", Biocybernetics and Biomedical Engineering, 36, 309 – 317, (2016).
[19]. B. Porat, B. Friendlander, "Direction finding algorithms based on high-order statistics", IEEE Transaction on Signal Processing, 39(9):2016-2024, (1991).
[20]. K. Tanner, N. Roy, A. Ash, E.H. Buder, "Spectral moments of the long-term average spectrum: sensitive indices of voice change after therapy?", Journal of Voice, 19(2):211-22, (2005).
[21]. S.Y. Lowell, R.H. Colton, R.T. Kelley, Y.C. Hahn, "Spectral- and cepstral-based measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker", Journal of Voice, 25(5) :223-232, (2011).
[22]. E. Nemer, R. Goubran, S. Mahmoud, "Robust voice activity detection using higher-order statistics in the LPC residual domain", IEEE Transaction on Speech and Audio Processing, 9(3):217-231, (2001).
[23]. E. Malkoç, "Türkçe ünlü formant frekans değerleri ve bu değerlere dayalı ünlü dörtgeni", Dil Dergisi, 146:71-85, (2009).
[24]. K. Pearson, "Contribution to the mathematical theory of evolution, II. Skew variation in homogeneous material", Philosophical Transactions of the Royal Society of London, 91:343-414, (1895).
[25]. A.M. Fiori, M. Zenga, "Karl Pearson and the origin of kurtosis", International Statistical Review, 77(1):40-50, (2009).
[26]. V. Vapnik, "Statistical learning theory", John Wiley&Sons, New York, (1998).