Voice Disorders Identification Based on Different Feature Reduction Methodologies and Support Vector Machine

Meisam Khalil Arjmandi, Mohammd Pooyan, Hojat Mohammdnejad, Mansour Vali
Biomedical Engineering
Shahed University
Tehran, Iran
msmarjmandi@gmail.com

Abstract— Identification of voice disorders has been a vital role in our life nowadays. Acoustic analysis can be useful tool to diagnose voice disorders as a complementary technique to other medicine methods such as Laryngoscopy and Stroboscopy. In this paper, we scrutinized feature reduction techniques such as principal component analysis (PCA) and linear discriminant analysis (LDA) as feature subset extraction methods and individual feature selection (IFS), forward feature selection (FFS), backward feature selection (BFS) and branch and bound feature selection (BBFS) as feature subset selection procedures. Performance of each method is evaluated by different classifiers. Between feature selection methods, individual feature selection followed by SVM classifier shows the best recognition rate of 91.55% and AUC of 95.80% among these methods. The experimental results demonstrated that highest performance could be achieved by recognition rate of 94.26% and AUC of 97.94% using linear discriminant analysis along with support vector machine as a classifier. Also this mixture has lowest order of computational complexity in comparison with other architectures.

Keywords- voice disorders identification; features subset reduction; support vector machine; linear discriminant analysis; features subset selection.

I. INTRODUCTION

In modern communities, voice disorders problem due to severe daily activities becomes a major issue. Invasive techniques such as Stroboscopy, Laryngoscopy and Endoscopy are employed by physicians to diagnose of voice impairments; especially disorders disrupt vocal cord mechanism [1]. Rate of health of the vocal folds affect quality of the voice. If the vocal folds become inflamed, some growths may develop on them or they become paralyzed and therefore suffer speech production process. The common disorders are vocal fold paralysis, vocal fold edema, adductor spasmodic dysphonia, A-P squeezing, etc. Disorders usually show up in speech signal in the form of acoustic perceptual measures like hoarseness, breathiness and harshness [2]. The speech results from three components of voice production; voiced sound, resonance and articulation [2]. Voiced sound produced by vocal fold vibration then amplified and modified by the vocal tract resonators (the throat, mouth cavity, and nasal passage) and finally vocal tract articulators (the tongue, soft palate, and lips) modify the voiced sound, therefore produce recognizable words [3]. In this context, we are countered with several types of organic voice disorders such as abductor spasmodic dysphonia, A-P squeezing, cancer, arytenoids dislocation, cyst, and etc [4]. Digital processing of voice signal has proved can be used as a non-invasive technique and an objective diagnosis to assess voice disorders in research setting. The aims of this research could be the evaluation of the performance of laryngitis treatment, surveying pharmacological treatment and rehabilitation effects. In general, short-time and long-time methods are applied as two categories for feature extraction from speech signal. In the previous researches, many long-time parameters such as fundamental frequency (F_0), jitter, shimmer, harmonics to noise ratio (HNR), pitch perturbation quotient (PPQ), amplitude perturbation quotient (APQ), normalized noise energy (NNE), voice turbulence index (VTI), frequency amplitude tremor (FATR), glottal to noise ratio (GTR) ([2],[5],[6],[7],[8],[9],[10]) are used in order to evaluate voice system health. Through this study the automatic detection of voice impairments is carried out by means of two feature reduction methods and four feature selection techniques. Then efficiency of each method is examined by different classifiers. This paper presents a novel approach to detect the pathological voice sample from normal one. Efficiency of feature extraction methods, as a feature reduction procedure [12], and feature selection techniques are evaluated.

This research attempted to investigate different algorithms for disorders classification and finally introduce the efficient structure to improve recognition rate and computational complexity. 22 features, include long-time acoustic parameters developed by the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Labs [11], are used to identification task. Effect of principal component analysis and linear discriminant analysis-as feature extraction methods- [12] and individual feature selection, forward feature selection, backward feature selection and branch & bound feature selection -as feature selection methods- are examined by different classifiers. The details of long-time features are mentioned in [11].

II. MATERIAL AND METHODS

A. Database

In this study, the voice samples are selected from the disordered voice samples [11], model 4337, developed by the Massachusetts Eye and Ear Infirmary Voice and Speech Lab. In this research, database contains 50 normal voice samples (32 males and 18 females) and 50 pathological voice samples (35 males and 15 females) with ages from 16 to 85 years' old. Data have been divided into two subset which 70% used for training and 30% for validation. Data include 33 parameters were calculated by the voice data sampled from a vowel /a/ for each subject. These acoustic parameters were derived by traditional Multi-Dimensional Voice program (MDVP^{TM}) [11]. Totally we used 22 long-time acoustic parameters of each sample since other parameters do not have further information and do not included intrinsic specification of voice sample.

B. Feature reduction methodology

From the viewpoint of pattern recognition, there are two general approaches for performing dimensionality reduction; feature extraction and feature selection [13]. Feature extraction performs transforming existent feature space to a lower space, whereas feature selection methods select features that result in an efficient feature subset. Necessity for using feature extraction and feature selection methods are demonstrated by effect of them on improving the "curse of dimensionality" and also complexity reduction of pattern recognition algorithm [13]. In this context, we studied extensively effect of two feature extraction algorithms and four feature selection methods in order to improve disordered voice recognition performance.

1) Feature extraction methods

The purpose of feature extraction methods is creating a subset of new features by combining the existing features [12]. In practice, a mapping must be founded to represent original features in a new space in order to reduce "curse of dimensionality" problem. The problem of feature extraction can be stated as; given a feature space $x_i \in R^N$ find a mapping $y = f(x) : R^N \to R^M$ with $M < N$ such that the transformed feature vector $Y_i \in R^N$ preserves most of the information or structure in R^N, where N is existence dimension and M is dimension result from mapping [13]. An objective function evaluates optimization degree of feature selection. Depending on the criterion used by objective function, feature extraction techniques are grouped into two categories: signal representation and signal classification [12].

a) Principal component analysis

This method searches a mapping to find the best representation for distribution of data. Therefore, it uses a signal representation criterion to perform dimensionality reduction while preserving as much of the randomness or variance in the high-dimensional space as possible [14].

The main limitation of PCA is that it does not consider class separability since it does not take into account the class label of the feature vector [14]. Therefore, we propose linear discriminant analysis that its objective is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible.

b) Linear discriminant analysis

Linear discriminant analysis uses a signal classification criterion in order to find a good projection vector. We seek to obtain a scalar Y by projecting the samples X onto a line, $Y=W^TX$. In a two-class problem, assume the mean vector of each class in X and Y feature space as μ_1 and μ_2, respectively. We define $\tilde{\mu}_i$ such as [12]:

$$\tilde{\mu}_i = \frac{1}{N_i}\sum_{Y\in\omega_i} Y = \frac{1}{N_i}\sum_{X\in\omega_i} W^T X = W^T \mu_i \tag{1}$$

We could then choose the distance between the projected means as our objective function [12],

$$J(W) = \left|\tilde{\mu}_1 - \tilde{\mu}_2\right| = \left|W^T(\mu_1 - \mu_2)\right| \tag{2}$$

However, the distance between the projected means is not a very good measure since it does not take into account the standard deviation within classes. The solution proposed by Fisher is to maximize a function that represents the difference between the means of classes, normalized by a measure of the *within-class* scatter. The Fisher linear discriminant is defined as the linear function W^TX that maximizes the criterion $J(W)$ function [12],

$$J(W) = \frac{\left|\tilde{\mu}_1 - \tilde{\mu}_2\right|}{\tilde{S}_1^2 + \tilde{S}_2^2} \tag{3}$$

Where, $\tilde{S}_i^2 = \sum_{y\in\omega_i}(y - \tilde{\mu}_i)^2$ is an equivalent of the variance for each class. Finally Fisher's Discriminant is expressed as follow [12],

$$W^* = \arg\max_W \left\{\frac{W^T S_B W}{W^T S_W W}\right\} = S_W^{-1}(\mu_1 - \mu_2) \tag{4}$$

Where, S_B, between-class scatter, and S_W, within-class scatter, are defined as follow [12],

$$S_W = S_1 + S_2 \tag{5}$$

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \tag{6}$$

LDA attempts to map distribution of original feature set to a 1-D space with highest severance between two classes.

2) Feature selection methods

In feature selection procedure, the aim is selecting a subset of the existing features without a transformation. Feature selection, is also called Feature Subset Selection in literature [12]. First, we must justify this question; why feature subset selection? With feature selection we can simply project a high-dimensional feature onto a low-dimensional space by selection of the best features. Feature subset selection requires two main steps: 1) a search strategy to select candidate subset, and 2) an objective function to evaluate these candidates [13]. Filter and

wrappers are two groups of objective function [12]. In filters, the objective function evaluates feature subset by their information contents, typically interclass distance, statistical dependence or information theoretic measures. Wrapper objective function is a pattern classifier, which evaluates feature subsets by their recognition rate on test data by statistical resembling or cross-validation [12].

a) Individual feature selection

In individual feature selection (IFS), each individual feature is evaluated separately and features with highest score are selected [12]. This method usually is not efficient since it does not take into account feature dependency.

b) Forward feature selection

Forward feature selection (FFS) is simplest grasping search algorithm. Forward feature selection algorithm can expressed as [12],

a. Start with the empty set, $Y_0 = \{\varnothing\}$.

b. Select the next best feature, $X^+ = \arg\max_{X \in Y_K} \left[J\left(Y_K + X\right) \right]$

c. Update, $Y_{K-1} = Y_K + X^+; K = K + 1$.

d. Go to b.

Where, *J (.)* is objective function.

Forward feature selection is efficient when the optimal subset has a small number of features [12].

c) Backward feature selection

Backward feature subset selection (BFS) works in the opposite direction of FFS. The BFS algorithm is as follow [12]:

a. Start with the full set, $Y_0 = X$.

b. Remove the worst feature, $X^- = \arg\max_{X \in Y_K} \left[J\left(Y_K - X\right) \right]$.

c. Update, $Y_{K+1} = Y_K - X^-; K = K + 1$.

d. go to b.

BFS works best when the optimal feature subset has a large number of features, since BFS spends most of times to visit large subsets. The main limitation of BFS is its inability to re-evaluate the usefulness of a feature after it has been discarded.

d) Branch & Bound feature selection

The branch and bound feature subset selection (BBFS) is an experimental method that is ensured to find the optimal feature subset under the monotonicity assumption [12]. BBFS begins from the full set and removes features using a depth-first strategy [12] and nodes whose objective functions are lower than the current best are not explored, since the monotonicity assumption ensures that their children will not contain a better solution.

C. Performance evaluation

In order to classification of pathological voice samples from normal voice samples, Quadratic Discrimination classifier, Nearest Mean classifier, Parzen classifier, K-Nearest Neighbor classifier, Support Vector Machine and Multilayer Back Propagation Neural Network are applied to each feature reduction methods and therefore performance of routines are evaluated. In this work, efficiency of any admixture is surveyed by accuracy, sensitivity, specificity and AUC measures. For each classifier optimum parameters are investigated in order to find the best voice disorders recognition rate for each classification procedure.

III. EXPERIMENTAL RESULTS

As previously mentioned, two feature reduction categories are implemented in order to optimize the feature set. Principal component analysis and linear discriminant analysis -as feature extraction techniques- in comparison with individual feature selection, forward feature selection, backward feature selection and branch & bound feature selection -as feature selection strategy- have been examined. We have defined a performance matrix in order to simplify the representation of the results as shown in Fig. 1.

Table 1 shows experimental results for two feature reduction methods. In these results, classifiers are examined in different condition and the best results for their ability in discrimination of disordered voice from normal one are illustrated. Results in table 1 demonstrates that the best recognition is achieved by LDA as a feature extraction procedure mixed with support vector machine as a classifier tool with accuracy of 94.26% and AUC of 97.94%. IFS, FFS, BFS and BBFS are applied to selected database in order to investigate effect of each method on classification improvement. IFS, FFS and BFS pursue a sequential search strategy and BBFS utilize an exponential search method [14]. As explained before, FFS is optimum when the optimal subset has a small number of features but BFS is efficient when the optimal feature subset has a large number of features, since BFS spends most of its time to visit large subsets. This fact is confirmed by results of classification using feature selection methods in Fig. 2. Also, results show the best optimization is achieved when dimension of feature subset vector is high in all methods. Therefore, as figure shows, BFS has better performance in comparison with FFS and BBFS. There is interesting point; IFS method shows the best performance in optimization task. Our results indicate that, in IFS, FFS, BFS and BBFS, optimum size of feature vector is 16, 9, 16 and 7. Also in all cases, SVM improves performance measures. The results of our simulation showed that, in all feature selection methods, 1-nearest neighbor (leave-one out) has highest ability to evaluate each feature subset in order to improve recognition rate. Table 1 show that we must have a trade-off in order to achieve high disorders identification rate with low complexity simultaneously. Totally, table 1 demonstrated that feature extraction methods have higher efficiency than feature selection procedures.

Accuracy (%)	**Sensitivity (%)**
Specificity (%)	**AUC (%)**

Figure 1. Performance matrix.

TABLE 1. PERFORMANCE EVALUATION OF DIFFERENT FEATURE REDUCTION METHODS: PCA, LDA, IFS, FFS, BFS AND BBFS, BY DIFFERENT CLASSIFIERS.

Classifier	22 Original Features	Feature Selection Method				Feature Extraction Method	
		Individual	Forward	Backward	Branch & Bound	PCA	LDA
Quadratic Discriminant Classifier	78.90 | 88.00 66.00 | 37.00	83.40 | 88.43 75.93 | 36.36	83.00 | 78.53 85.5 | 49.75	85.64 | 91.33 77.57 | 30.10	86.27 | 80.90 89.50 | 44.24	85.20 | 77.83 89.90 | 57.82	92.20 | 93.67 89.97 | 96.27
Nearest Mean Classifier	87.20 | 70.90 97.60 | 94.30	87.25 | 69.70 97.83 | 95.35	82.81 | 56.30 97.46 | 92.99	86.46 | 67.70 97.76 | 94.72	85.40 | 68.77 95.63 | 91.52	87.20 | 70.90 97.60 | 94.30	92.27 | 93.70 90.13 | 97.94
Parzen classifier	85.50 | 73.35 93.85 | 31.80	87.27 | 77.15 94.30 | 32.36	86.14 | 71.76 95.60 | 52.21	87.08 | 74.23 95.76 | 27.37	86.48 | 70.83 96.43 | 39.52	86.32 | 74.70 94.10 | 37.44	92.30 | 93.70 90.13 | 96.47
K-Nearest neighbor Classifier	88.86 | 78.30 96.17 | 93.77	89.48 | 81.67 95.00 | 94.00	86.62 | 80.10 91.30 | 92.13	89.56 | 79.76 96.36 | 93.59	85.41 | 75.20 92.40 | 89.58	88.57 | 80.24 94.33 | 94.15	90.25 | 91.10 88.03 | 96.27
Support Vector Classifier	89.29 | 82.25 94.30 | 94.50	91.55 | 83.17 95.80 | 95.80	88.75 | 68.23 98.13 | 94.34	91.00 | 80.76 96.60 | 95.59	87.10 | 69.46 98.46 | 93.91	90.52 | 84.27 94.93 | 95.15	94.26 | 93.70 90.13 | 97.94
BP Neural Network	88.79 | 83.00 85.10 | 88.77	84.79 | 84.75 86.73 | 89.76	87.00 | 86.26 86.20 | 92.25	85.68 | 82.80 87.20 | 89.74	86.10 | 83.43 87.17 | 92.40	88.93 | 88.00 88.66 | 93.81	92.97 | 91.66 88.53 | 96.92

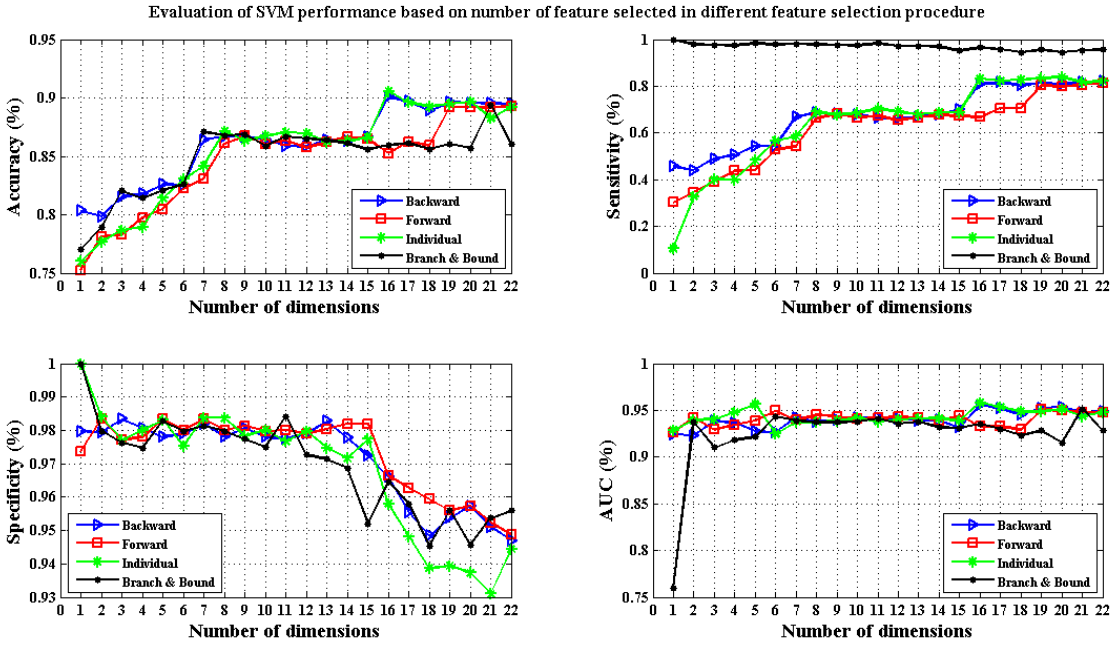

Figure 2. IFS, FSS, BFS and BBFS evaluation based on length of feature subset vector. SVM is used for classification.

Fig. 3 illustrates effectiveness of all feature reduction algorithms. It is also obvious that, BFS and FFS show higher performance between feature subset selection methods. As mentioned before, BFS is more efficient when the optimal feature set has a large length with regard to original feature space size. Experimental results demonstrate that optimal feature set occurred in high dimension in comparison with original feature vector length.

IV. DISCUSSION

Speech production is a complicated process that involves synchronization of a lot of organs in order to accomplish a healthy voice. Speech researchers pursue diagnosis of speech disorders in early stage of affect since it significantly prevents occurrence critical conditions. In speech processing tasks, as a pattern recognition process, two considerations are important; improving recognition rate and reducing of implementation costs by decreasing the complexity at the same time. In this

paper, linear discriminant analysis provides a feature reduction procedure with appropriate voice disorders identification rate and low algorithm complexity.

Also, it has been shown that the idea of using feature selection routines in order to decrease implementation time and complexity problem cannot result in significant improvements in identification of disordered speech. In best state, individual feature selection method together with SVM classifier illustrates accuracy of 91.55% and AUC of 95.80%, whereas linear discriminant analysis shows accuracy of 94.26% and AUC of 97.94% using similar classifier. It has been demonstrated by experiments, the SVM spend shorter time for training than any other techniques such as BP Neural Network, Linear Vector Quantization and K-Nearest Mean classifier. Simultaneously, it shows noticeable improvement in voice disorders classification rate.

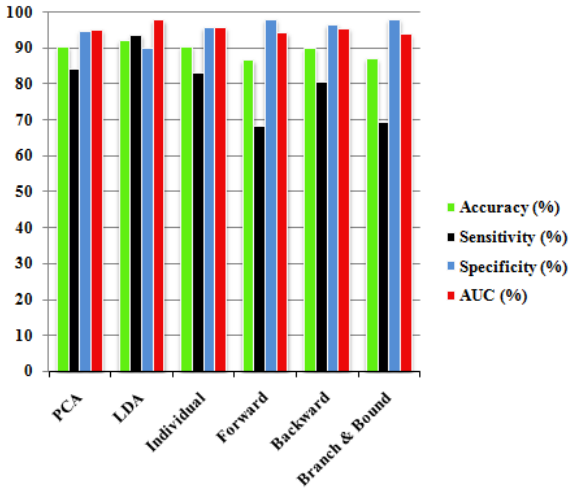

Figure 3. Comparison of performance of different feature reduction methods in voice disorders recognition task.

V. CONCLUSION

This paper scrutinized feature extraction methods in comparison with feature selection routines and finally demonstrated that linear discriminant analysis as a feature extraction algorithm leads to optimal experimental results by accuracy of 94.26% and AUC of 97.94%. In feature selection methods, IFS and BFS with efficiency of 91.55% and 91%, respectively, based on how effective compromise between the number of efficient features and classifier performance for selected features, are more reliable. However, in voice disorders identification task, feature extraction methods are more reliable than feature selection methods and complexity of these algorithms are appropriate compared to other feature reduction methods. In addition, SVM has established an adequate result of generalization to distinguish between normal and pathologically affected voices.

REFERENCE

[1] F. Planet, H. Kessler, B. Cheetham, J Eans, "Speech Monitoring of Infective Laryngitis", *Proceeding of ICSLP96*, Philadelphia, pp. 749-752, 1996.

[2] R. J. Baken and R. Orlikoff, *Clinical Measurement of Speech and Voice*, 2nd ed. San Diego, CA: Singular, 2000.

[3] I.R. Titze, *Principles of Voice Production*, Prentice Hall, 2nd Edition, 1998.

[4] M. Greene, L. Mathieson, *The Voice and its Disorders*, John Wily & Sons, 6nd Edition, 2001.

[5] B. Boyanov and S. Hadjitodorov, "Acoustic analysis of pathological voices,", *IEEE Eng. Med. Biol. Mag.*, pp. 74–82, July 1997.

[6] D. Michaelis, H. Gramss, and W. Strube, "Glottal-to noise ratio- a new measure for describing pathological voices," *Acustica/Acta Acustica*, vol. 83, pp. 700–706, 1997.

[7] H. Kasuya, S. Ogawa, K. Mashima, and S. Ebihara, "Normalized noise energy as an acoustic measure to evaluate pathologic voice,", *J. Acoust. Soc. Amer.*, vol. 80, no. 5, pp. 1329–1334, 1986.

[8] G. de Krom, "A cepstrum-based technique for determining harmonic-to-noise ratio in speech signals," *J. Speech Hearing Res.*, vol. 36, pp. 254–266, Apr. 1993.

[9] Y. Qi and R. E. Hillman, "Temporal and spectral estimations of harmonics- to-noise ratio in human voice signals,", *J. Acoust. Soc. Amer.*, vol. 102, no. 1, pp. 537–543, 1997.

[10] W. Winholtz, "Vocal tremor analysis with the vocal demodulator", *J. Speech Hearing Res.* vol. 35, pp. 562–573, 1992.

[11] Disordered Voice Database (CD-ROM), Version 1.03, Massachusetts Eye and Ear Infirmary, Kay Elemetrics Corporation, Boston, MA, Voice and Speech Lab., October 1994.

[12] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern classification*, 2nd edition, John Wiley and Sons, New York, 2001.

[13] A. Webb, *Statistical Pattern Recognition*, John Wiley & Sons, New York, 2002.

[14] R. J. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches*. New York: Wiley, 1991.