

Different Mother Wavelets and Pathological Voice

Pegah Tayaranian Hosseini¹ Farshad Almasganj²

¹ *Biomedical Eng. Department, Amirkabir University of Technology, Tehran, Iran, p_hosseini@aut.ac.ir*¹

² *Biomedical Eng. Department, Amirkabir University of Technology, Tehran, Iran, almas@yahoo.com*²

Abstract— Early diagnosis of different maladies and pathologies of human vocal system using noninvasive methods and diverse signal processing technics is a problem that is particularly considered by biomedical engineering and signal processing researchers, recently. Automatic detection of voice pathology from speech signal is a new topic and has not been progressed enough. An algorithm able to classify two pathological voice signals based on Wavelet Packets (WP) and Fisher's Linear Discriminant (FLD) is presented in this research. We use WP and different mother wavelets (Daubechies, Coiflet, and Symmlet) for time-frequency analysis giving quantitative evaluation of signal characteristics to identify pathologies in voice signals of subjects with different ages for both male and female. Choosing Coiflet mother wavelet, we use FLD to find the best tree among Coiflet Wavelet Packet trees. After selecting best features from terminal nodes of the best tree with contribution to Genetic Algorithm, we apply Support Vector Machines to separate voice pathologies. Applying our algorithm to separate polyp from some other pathologies we come to much higher conclusions in contrast to previous works that use Daubechies mother wavelet instead of Coiflet mother wavelet (e.g. 92.5% in comparison to 82.5% for separating polyp from adductor spasmodic dysphonia).

Keywords— Wavelet Packet, Genetic Algorithm, Fisher's Linear Discriminant, Daubechies mother wavelet, Coiflet mother wavelet

I. INTRODUCTION

There are too many factors that have negative effects on human vocal system. Early diagnosis of pathological voice is a significant topic that has been recently received considerable attention. Nowadays, the available methods for speech pathology detection are either invasive or need the special analyses of wide range of voice signal parameters. Therefore, attaining a reliable, precise, and noninvasive procedure for voice pathology diagnosis is one of the most necessary proficiencies in speech pathological evaluation. Voice automatic analysis for pathological voice detection has many advantages. The most important ones are as follow: 1) It has quantitative and noninvasive nature; 2) It has the ability of recognizing the extent of pathology progression; 3) It reduces the time and expense of analysis [1].

Although there are some researches on noninvasive separation of normal voice and pathological one, separating two or more pathological voices from each other in not considered seriously. We have considered this matter in this research more attentively.

Diagnosing voice pathology via voice signal is based upon extracting the features of the signal. In this article, we have extracted the signal features with the use of wavelet transform. We wavelet packet of different mother wavelets as our feature extraction method, select the best mother wavelet and find its features with high capability of separation for two specific pathologies with Fisher's Linear Discriminant (FLD). Then, using Genetic Algorithm (GA), we complete our optimization by omitting the less efficient features. These features will then feed into a Support Vector Machine (SVM) classifier to get the final result-the accuracy of separation [2]. The database employed in this study has been developed by Massachusetts Eye and Ear Infirmary Voice and Speech Laboratory (MEEI4337) [3]. We exerted our approach on some common pathologies such as vocal fold polyp, adductor spasmodic dysphonia, keraosis leukoplakia, and vocal nodules and the results were much higher than previous works (replacing Coiflet with Daubechies).

Polyps are growths arising from the vocal fold mucosa. They may be solid or fluid filled, and can become quite large. Their effects on vibration depend on their size and their location on the vocal folds [4]. Adductor spasmodic dysphonia is the most common type of spasmodic dysphonia. The thyroarytenoid muscle (the muscle that lies within each vocal fold) contracts strongly and suddenly as in a muscle spasm. This causes the vocal folds to suddenly squeeze together very tightly. The result is a sudden breaking, stopping, or strangling of the voice [4]. Keratosis is a lesion seen most often on the mucosa of the vocal folds. It typically appears as a white patch, called leukoplakia [4]. Nodules are blister-like or callous-like swellings that form just below the epithelial surface of the vocal folds. They occur on both vocal folds and are symmetrical. The nodules appear as small bumps along the mid portion of the vocal folds, where the vocal folds come into contact with each other. The nodules may create a gap between the two vocal folds allowing air to escape and prevent normal vibration. They may also stiffen the mucosal tissue, causing irregular vibration and a rougher sound [4].

This paper is organized as follows: in the next section the principles of our approach are discussed, in the third section we apply our algorithm on the foresaid pathologies from MEEI and explain our conclusions in the last section.

II. METHODOLOGY

A. Wavelet Analysis

A wavelet is a mathematical function used to divide a given function or continuous-time signal into different frequency components and study each component with a resolution that matches its scale [5]. A wavelet transform is the representation of a function by wavelets. The wavelets are scaled and translated copies (known as "daughter wavelets") of a finite-length or fast-decaying oscillating waveform (known as the "mother wavelet"). Wavelet transforms are classified into discrete wavelet transforms (DWTs) and continuous wavelet transforms (CWTs). Note that both DWT and CWT are continuous-time (analog) transforms. They can be used to represent continuous-time (analog) signals. CWTs operate over every possible scale and translation whereas DWTs use a specific subset of scale and translation values or representation grid.

The frequency bands or subspaces (sub-bands) are scaled versions of a subspace at scale l . This subspace in turn is in most situations generated by the shifts of one generating function $\psi \in L^2(\mathbb{R})$, the mother wavelet. $L^2(\mathbb{R})$ is a continuous family of frequency bands

It is computationally impossible to analyze a signal using all wavelet coefficients, so one may wonder if it is sufficient to pick a discrete subset of the upper half-plane to be able to reconstruct a signal from the corresponding wavelet coefficients. One such system is the affine system for some real parameters $a > 1$, $b > 0$. The corresponding discrete subset of the half-plane consists of all the points $(a^m, na^m b)$ with integers $m, n \in \mathbb{Z}^2$. The corresponding baby wavelets are now given as (1):

$$\psi_{m,n}(t) = a^{-m/2} \psi(a^{-m}t - nb)$$

A sufficient condition for the reconstruction of any signal x of finite energy by (2):

$$x(t) = \sum_{m \in \mathbb{Z}^2} \sum_{n \in \mathbb{Z}^2} \langle x, \psi_{m,n} \rangle \psi_{m,n}(t) \quad (2)$$

is that the functions $\{\psi_{m,n} : m, n \in \mathbb{Z}^2\}$ form a tight frame of $L^2(\mathbb{R})$ [5].

We used different mother wavelets such as Daubechies, Coiflet and Symmlet to extract wavelet packet features till the fifth level.

B. Fisher's Linear Discriminant

After expanding all signals in each class into orthonormal bases by using wavelet packets, this tree is then pruned to

maximize a cost function called Fisher's Linear Discriminant (FLD). The higher FLD means the bigger distance between two classes so it can be known as a dissimilarity measure. For choosing the best tree of the wavelet packet, the algorithm starts from the first level of the wavelet tree. Each node of the level will split if and only if the FLD of the new feature vector is bigger than the previous vector. This will continue till the last level of decomposed wavelet tree [6].

Fisher's linear discriminant is a transform that reduces the dimensionality of the features vector from n into $d=M-1$ (where M is the number of classes involved), while optimally preserving the separability between classes. The idea behind the FLD is the projection of the n dimensional features vectors onto a lower dimensional surface. The surface is to be chosen such that separation between classes is kept as much as possible.

In order to find the optimal surface to project onto, a measure of separability is required. Optimality is then understood in the sense of maximizing separability. For this matter, consider the two-class problem. This situation has been considered in this paper. Suppose we have N known samples, β_i , N_i of which belong to w_1 and N_2 of which belong to w_2 . Consider y_i to be linear combination of the features β_i as (3):

$$y_i = \underline{\rho}^T \underline{\beta}_i \quad (3)$$

The n dimensional vector, $\underline{\rho}$, can be considered a line in the n dimensional space; then y_i is the projection of β_i on this line (scaled by $\|\underline{\rho}\|$). Let $\underline{\hat{\mu}}_i$ be the mean of the N_i samples of class w_i in the n dimensional space like (4):

$$\underline{\hat{\mu}}_i = \frac{1}{N_i} \sum_{\beta \in w_i} \beta_i \quad (4)$$

Define the $n \times n$ scatter matrix, W_i , of the i th class as (5):

$$W_i = \sum_{\beta \in w_i} (\beta - \underline{\hat{\mu}}_i)(\beta - \underline{\hat{\mu}}_i)^T, i = 1, 2 \quad (5)$$

W_i is the estimation of the covariance of the i th class in the n -dimensional feature space. It represents a measure of the dispersion of the signals belonging to w_i . The within-class scatter matrix, W , is defined as (6):

$$W = W_1 + W_2 \quad (6)$$

Consider now the variance between the means of the two classes. Denote the matrix, B , the "between class scatter matrix", in the original n -dimensional features space as (7):

$$B = (\underline{\hat{\mu}}_1 - \underline{\hat{\mu}}_2)(\underline{\hat{\mu}}_1 - \underline{\hat{\mu}}_2)^T \quad (7)$$

A criterion of separation can now be formulated in terms of the new scatter matrices. For good separation we require

that the variance of the population of each class be small. Hence, a good separation measure, $J(\underline{\rho})$, is (8):

$$J(\underline{\rho}) = \frac{\underline{\rho}^T B \underline{\rho}}{\underline{\rho}^T W \underline{\rho}} \quad (8)$$

The best $\underline{\rho}$ that gives the maximum ratio of “between” to “within” class scatter is (9) [7]:

$$\underline{\rho} = W^{-1}(\hat{\underline{\mu}}_1 - \hat{\underline{\mu}}_2) \quad (9)$$

C. Genetic Algorithm

Genetic Algorithms (GAs) are robust search and optimization techniques which are finding application in a number of practical problems.

The GA may be viewed as an evolutionary process wherein a population of solutions evolves over a sequence of generations. During each generation, the fitness of each solution is evaluated, and solutions are selected for reproduction based on their fitness. Selection embodies the principle of “Survival of the fittest.” “Good” solutions are selected for reproduction while “bad” solutions are eliminated. The “goodness” of a solution is determined from its fitness value. The selected solutions then undergo recombination under the action of the crossover and mutation operators.

In our approach, each population consists of 300 individuals which have arranged in three subpopulations. Each individual is a binary string which length is the number of features that have been chosen with FLD. In each generation, the feature vector of all the signals in two classes is considered. In each vector, those features which their place in the chromosome is “one” will be picked and the other features will be eliminated. Then, the new feature vectors will feed into a SVM classifier (see subsection II.D). The classifying percent will be the fitness value of that chromosome. This process repeats for each individual of one generation. Then, the next generation will be produced based on fitness values of current generation with the crossover, mutation and reproduction functions.

III. EXPERIMENTAL RESULTS

The database [3] developed by Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Laboratory was used in this article. It contains voice samples of 710 subjects. Included are sustained phonation speech samples of the vowel /a/ from patients with a wide variety of organic, neuralgic, traumatic, and psychogenic voice disorders, as well as 53 normal subjects. The sampling frequency is 44 KHz. The pathologies vocal polyp, vocal nodule, keratosis leukoplakia, and adductor spasmodic dysphonia are considered from the vocal diseases in this database because of their generality. There are both male and female cases in each group of pathologies. The classifying percent of each disease with vocal polyp has been studied using the approach which has been explained and will be presented in the end of this section.

We extracted wavelet packet entropies of each signal in every group of pathologies till the fifth level. We used the “log Energy” entropy (15) and 3 mother wavelets: Deaubechies4, Coiflet4, and symmlet4.

$$E(s) = \sum_i \log(s_i^2) \quad (15)$$

As some features in the terminal nodes of the wavelet packet features are not proper for us and does not have much role in classification, we need to eliminate them from the feature vector. For this reason, we apply GA to the feature vectors as described in section II.C. This step is called “optimization.” You can see the percents of classification with SVM after optimization but before applying FLD for different pathologies and different mother wavelets in Table I. As can be seen in this Table, db4 and sym4 have lower percents in comparison to coif4 after applying GA so we choose coif4 as our best mother wavelet. Knowing that db4 is a common mother wavelet in speech processing, we also choose db4, especially for better resemblance. So, using the 32-feature vector of coiflet and daubechies wavelet packet decompositions, we continue our algorithm.

Now, we apply FLD algorithm to the wavelet trees of coiflet and daubechies mother wavelets as explained in II.B. For example, for two specific pathologies polyp and adductor, the tree in Fig.1 was attained for coif4.

In the next step, the feature vectors of FLD step were used and GA algorithm applied to them to eliminate the non-proper features that have fewer roles in classification. You can see a chromosome example in Fig.2. This specific chromosome refers to the best coif4 features for polyp and adductor. The red parts in this chromosome mean “one” and the white ones mean “zero”. As we explained in section II.C, where we have a “one” in a chromosome, we should pick that specific feature from the feature vector. As we see in Fig.3, we should pick the features which have been denoted in red circles from the best tree of FLD step with respect to the chromosome of Fig.2 and the best tree of Fig.1.

Our GA algorithm runs till generation 30 with 300 chromosomes in each generation. In every generation, 10 chromosomes are transferred to the next generation without any change. Our crossover fraction is 0.2 and our mutation function is Gaussian.

The fitness function of GA contains the SVM classifier supported with regular basis function (RBF). GA will choose the chromosome which has the best classifying percent. We used 75% of each disease samples for training and 25% for testing. Instead of choosing signals randomly, we circled the 25% test signals 4 times and calculated the average percent. In fact, this percent will be considered as a chromosome’s fitness value.

In [8] there is a similar approach for separating paralysis from normal voice with 100% result. It also uses the MEEI database. Its difference with our work is that in this article, we have replaced the 32 feature vector of the fifth level of

TABLE I. CLASSIFYING PERCENTS BETWEEN POLYP AND THREE OTHER DISEASES BEFORE AND AFTER OPTIMIZATION LEVEL WITHOUT APPLYING FLD (WITH DB4, COIF4, AND SYM4)

		Before GA			After GA		
		db	coif	sym	db	coif	sym
Polyp	Adductor	55	57.5	45	75	87.5	72.5
Polyp	Keratosiis	45	59.9	47.7	77	80.3	75
Polyp	Nodule	37	62	57.5	80	92.5	76.5

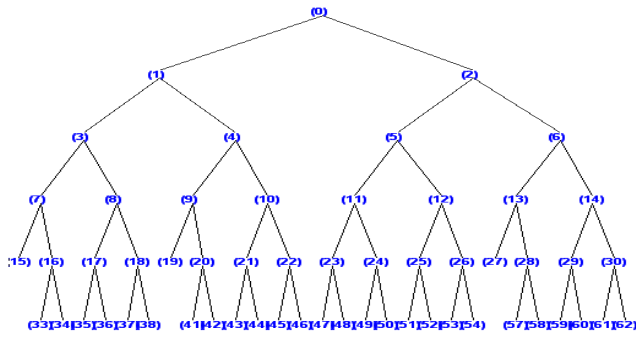


Figure 1. The best tree attained from FLD for polyp and adductor spasmodic dysphonia

15	19	27	33	34	35	36	37	38	41	42	43	44	45	46
47	48	49	50	51	52	53	54	57	58	59	60	61	62	

Figure 2. The best chromosome for selecting features after FLD for best separating polyp and adductor. The numbers 1 to 25 show the number of features. Red and white means “1” and “0”, respectively

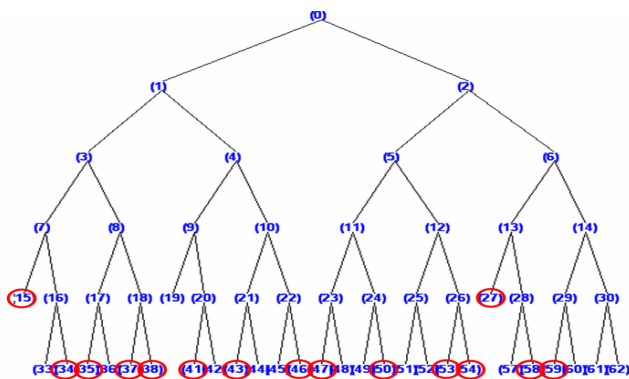


Figure 3. Red circles show which features should be picked from the feature vector with respect to chromosome in Fig.5.

TABLE II. CLASSIFYING RESULTS BEFORE AND AFTER APPLYING GA FOR FEATURES AFTER FLD

		After FLD, Before GA		After FLD, After GA	
		db4	coif4	db4	coif4
Polyp	Adductor	57.5	70	82.5	92.5
Polyp	Keratosiis	59.09	61.17	81.81	84.27
Polyp	Nodule	65	70	85	92.5

wavelet packet with the feature vector of lower length that has attained from FLD; besides, we changed the “Shannon” entropy with the “log energy” one. Still, we attain 100% separation for paralysis and normal voice. TableII represents higher results of applying FLD and GA in comparison with TableI. It also shows that applying FLD gives better results for separating two pathological voices from each other in contrast to the previous work (without FLD and just with GA); besides, we get much higher results for coif4 in comparison with db4. We can see the increasing procedure of classification percent in these two tables.

IV. CONCLUSION

In this paper, the role of Fisher’s Linear Discriminant, applied on fifth level of coiflet and daubechies wavelet-packet decomposition, in clinical diagnosis of vocal fold polyp and some other diseases (keratosiis leukoplakia, adductor spasmodic dysphonia, and vocal nodules) is investigated. The optimal feature sub-set selection procedure is implemented using GA after we applied FLD on the feature vectors. The GA uses the SVMs recognition rates as the fitness value for each individual. The genetic-based optimization procedure is applied to the feature vector pool in order to find the best individual. Each individual’s chromosome string comprises a set of active and inactive binary genes. The demonstrated active genes represent the frequency sub-bands over which the extracted entropy features would participate in the classification task. The results of the simulations show that for the SVM, with a “RBF” kernel, the optimized set of entropy features leads to high recognition. Getting better results for fisher, after applying GA, means there are some features that have been forced to produce. For more explanation we can say that when we split a node and the fisher value of the new feature vector is more than the previous one, that node is obliged to split. But maybe one of these nodes is not really important for the classification and it is the second node that has compensated the fisher value of the first one. When we apply GA on the feature vector, that non-important node or its productions are eliminated.

REFERENCES

- [1] J. Wang, Ch. Jo, “Vocal Folds Disorder Detection using Pattern Recognition Methods”, *Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale*, Lyon, France, August 23-26, 2007
- [2] E. S. Fonseca, R. C. Guido, A. C. Silvestre, J. C. Pereira, “Discrete wavelet Transform and Support Vector Machine Applied to Pathological Voice Signals Identification”, *Proceedings of the Seventh IEEE International Symposium on Multimedia*, 2005
- [3] Disordered Voice Database, Version 1.03, October 1994, Massachusetts Eye and Ear Infirmary, Voice and Speech Lab, Boston, MA, Kay Elemetrics Corp
- [4] Lions Voice Clinic, University of Minnesota, Department of Otolaryngology, P.O. Box 487 420 Delaware St. SE.Minneapolis, MN 55455, USA
- [5] Available at: www.wikipedia.org
- [6] N. F. Ince, A. Tewfik, S. Arica, “Classification of Movement EEG with Local Discriminant Bases”, *ICASSP*, 2005
- [7] A. Cohen, “*Biomedical Signal Processing*”, CRC Press, 1983
- [8] R. Behroozmand, F. Almasganj, “Optimal selection of wavelet-packet-based features using genetic algorithm in pathological assessment of patients’ speech signal with unilateral vocal fold paralysis”, *Computers in Biology and Medicine*, Vol.37, pp.474 – 485, 2007