Nonlinear Features of Bark Wavelet Sub-band Filtering for Pathological Voice Recognition

Xiao-Jun Zhang, Xin-Cheng Zhu, Di Wu, Zhong-Zhe Xiao, Zhi Tao and He-Ming Zhao

***Abstract*: This paper proposes a recognition method for extracting nonlinear characteristic parameters of pathological voices using Bark wavelet sub-band filtering. First, the speech signal was processed through 24 Bark filter banks. According to the signal obtained from each channel, the feature multi-frequency band nonlinear coefficient was extracted according to the frequency division factor, α. We used 53 normal voices and 117 pathological voices from the MEEI's pathological voice experimental database, and 14 machine learning methods were used to perform the recognition experiments. The experimental results showed that the proposed method effectively improved the recognition rate. Moreover, the proposed feature was optimal when α = 18, and the highest recognition rate was achieved when the support vector machine learning algorithm was used.**

***Index Terms*—frequency division, nonlinear, pathological voice speech recognition, frequency division factor**

I. INTRODUCTION

Speech is the most common form of communication in human society, and the quality of one's voice directly affects the ability to communicate [1]. However, due the excessive use of vocal cords, smoking, and environmental factors, the number of patients with voice diseases has been increasing. Currently, 6%–9% of people have some form of voice problem, such as pharyngitis or vocal cord inflammation. In severe cases, vocal nodules, vocal polyps, vocal cysts, and vocal cord paralysis can affect health and quality of life [2]. At present, common voice diagnosis and treatment methods, such as laryngoscopy, dynamic endoscopy, electromyography, etc., are invasive examination methods, causing pain and even damage to patients. Furthermore, they cannot make early pre-diagnoses of laryngeal and vocal cord lesions, which can delay treatment [3].

Acoustic analysis methods noninvasively identify voice diseases through the in-depth examination of various parameters of a patient's voice, and by constructing a pathological voice recognition system. Lieberman [4] discovered that pathological voices have abnormal periodic fluctuations. He proposed fundamental frequency perturbation as a characteristic parameter related to pathological voice recognition and studied other acoustic characteristics related to the fundamental frequency [5]. Later, scholars formed a complete acoustic parameter system by studying amplitude perturbation [6], with additional characteristic parameters related to the amplitude of the voice, such as the harmonic noise ratio [7], voice turbulence index [8], soft pronunciation index [9], and other energy-related parameters. In the field of spectral analysis, mel-frequency cepstrum coefficient (MFCC) parameters are widely used for pathological voice detection [10-11]. Fu and Yi [12] proposed a speech signal using the Bark wavelet and its transformation based on the Bark scale obtained from speech perception experiments and the construction of sub-waves. The new feature was superior to MFCC features at distinguishing consonants.

During the process of speech generation, as airflow passes through the vocal cord cavity, eddy currents are generated and the vortex region modulates the speech signal, generating nonlinear phenomena [13-16]. This can be used to extract features of voice signals. Jiang et al. [17] verified the largest Lyapunov exponent and correlation dimension parameters of normal voices and the voices of laryngectomy patients, and found that there is a distinction between the two. Others, including Carlos et al. [18], used nonlinearity. They analyzed the dynamic characteristics of pathological voices, vocal disturbances caused by larynx lesions and Parkinson's disease, and nasal sounds caused by a cleft lip and palate, verifying the validity of the largest Lyapunov exponent and the correlation dimension in distinguishing between different pathological voices.

This paper proposes a method of voice extraction based on the Bark sub-band. First, a Bark wavelet filter that accurately matches the human ear's auditory perception characteristics was used to filter and divide a voice signal into 24 frequency bands. Then, by setting the frequency division factor, α, in the frequency bands 0 to α, the voice logarithm of the voice signal was calculated and the discrete cosine transform (DCT) was used to extract the cepstrum coefficient feature. In frequency bands $\alpha + 1$ to 24, the largest Lyapunov exponent and correlation dimension were more suitable to high frequency signal feature processing, and were extracted to

Manuscript received December 09, 2019; revised September 17, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant No.61372146, No.61271359 and No.61906128.

X.J. Zhang, School of Electronic and Information Engineering & School of Optoelectronic Science and Engineering, Soochow University, Suzhou 215006, China (e-mail:zhangxj@suda.edu.cn).

X.C. Zhu, School of Optoelectronic Science and Engineering, Soochow University, Suzhou 215006, China (e-mail:1713402006 @stu.suda.edu.cn).

D. Wu, School of Optoelectronic Science and Engineering, Soochow University, Suzhou 215006, China (e-mail:wudi@suda.edu.cn).

Z.Z. Xiao, College of Physics, Optoelectronics and Energy, Soochow University, Suzhou 215006, China (e-mail:xiaozhongzhe @suda.edu.cn).

Z. Tao, College of Physics, Optoelectronics and Energy, Soochow University, Suzhou 215006, China (tel.:+86051267870162; fax:+86051267870271;e-mail:taoz@suda.edu.cn).

H.M. Zhao, School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China (e-mail:hmzhao @suda.edu.cn).

ensure that the characteristics of pathological voices could be described in detail in each frequency band. Finally, current machine learning algorithms, such as Bayesian networks [19], deep neural networks [20], support vector machines [21], and so on, were used to identify normal voices and pathological voices.

II. Pathological Voice Analysis

The human voice contains a lot of information about the speaker. Analyzing the voice signal is an important way to understand and process the voice signal to interpret the information contained in it. However, when a person's voice is affected by disease, it will cause changes in the time domain and frequency domain of voice signals. Figures 1 and 2 respectively show waveforms and spectrograms of normal and pathological voices (vocal nodules, vocal cord edema, vocal cord polyps) when vocalizing the vowel /a:/. From the waveform chart in Fig. 1, it can be seen that voices with vocal nodules, vocal cord edema, and vocal cord polyps destroyed the regularity of voice production. Moreover, the amplitude range of the voice changed and there were high frequency oscillations similar to noise. As seen in the spectrogram of Fig. 2, pathological voices introduced high frequency noise, and energy diffused from the low frequency region to the high frequency region. For pathological voice recognition, therefore, it was necessary to extract information that shows the voice characteristics from each frequency band using frequency band processing. Differences between a normal voice and a pathological voice can also be found according to energy distribution characteristics.

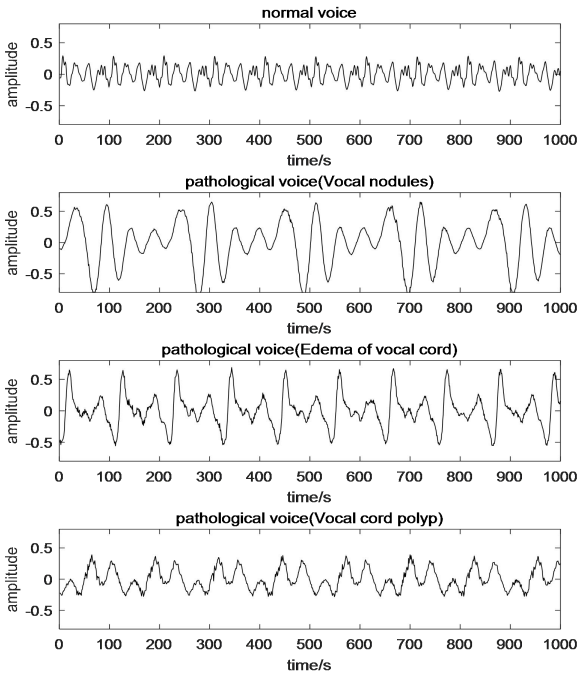

Fig. 1. Time domain diagrams of normal and pathological voices.

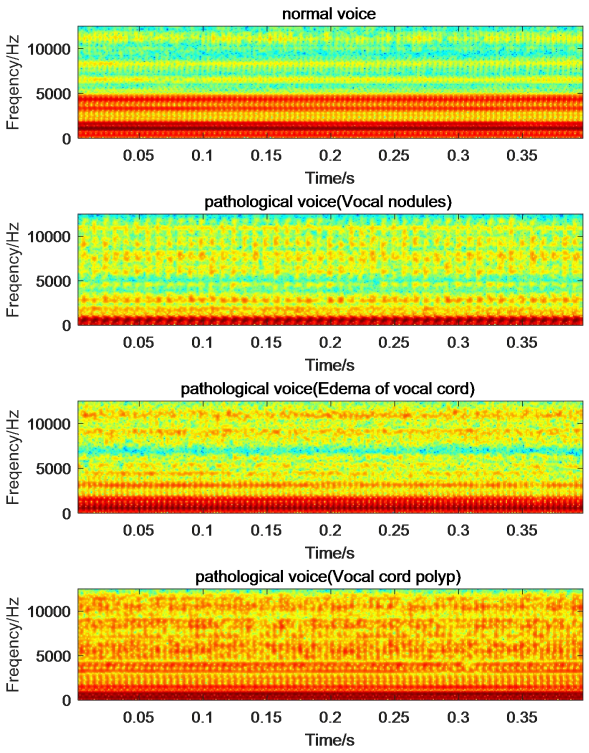

Fig. 2. Spectrograms of a normal voice and pathological voices.

III. Feature Extraction

A. Bark wavelet sub-band filters

A Bark wavelet transform is a wavelet transform based on the idea of a multi-resolution and non-uniform division of time and frequency space. Based on the auditory physiological characteristics of the human ear, it divides the discernible frequencies of the human ear into 24 frequency groups, or 24 Bark domains, and introduces the functional relationship between linear frequencies and auditory perception frequencies given by Traunmulla [22]:

$$b = 6.7\text{asinh}[(f-20)/600] \quad (1)$$

where b denotes the Bark frequency, f is the linear frequency, and "$asinh$" is the inverse hyperbolic sine function. We construct the Bark domain wavelet mother function as follows:

$$\phi(b) = e^{-4\ln 2 \cdot b^2} = 2^{-4b^2} \quad (2)$$

The function expression in auditory perception domain is obtained as follows:

$$\phi_k(b) = \phi(b - b_1 - k\Delta b) = 2^{-4(b-b_1-k\Delta b)^2} \quad (3)$$

Among these, $\Delta b = (b2, b1)/(K-1)$, K is the scale parameter of the auditory perception frequency bandwidth for $(b1, b2)$.

The auditory perception wavelet function at a linear frequency is expressed as follows:

$$\phi_k(f) = 2^{-4[6.7asinh[(f-20)/600]-b_1-k\Delta b]^2} \quad (4)$$

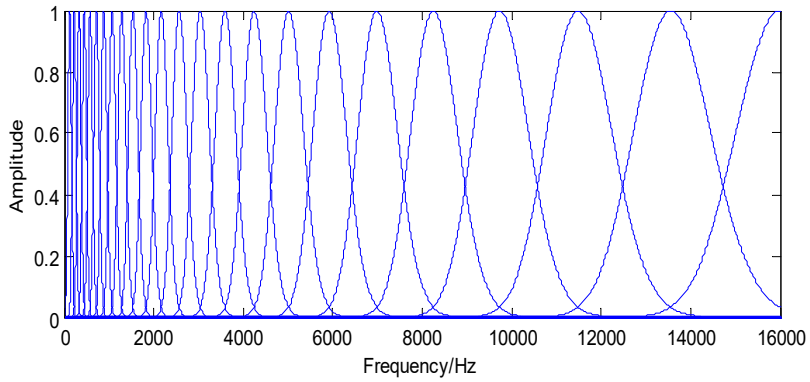

Fig. 3. Sketch of the bark wavelet filter bank.

It can be seen from Fig. 3 that Bark filters with different center frequencies have different bandwidths. Both sides of the center frequency of the band-pass filter have steep edges, indicating that the Bark filter has sharp frequency selection characteristics and focuses on voice features near the center frequency. At the same time, however, the transition between the two sides is smooth, and the leakage of the spectrum is reduced. The amplitude-frequency response characteristics are consistent with the filtering characteristics of the basement membrane, and they completely conform to the auditory characteristics of the human ear.

B. Extracting Nonlinear Features of Voices

The theory of acoustics and aerodynamics indicates that the vibration and vocalization of vocal cords is a complicated nonlinear process. This nonlinear system can produce simple periodic motion and complex chaotic behavior. The study of nonlinear dynamics shows that, unlike pure random signals, chaotic signals are generated by deterministic nonlinear systems and can be described by nonlinear dynamic parameters. As a classical nonlinear dynamic parameter, the correlation and maximum Lyapunov effectively describe chaotic, disorderly, irregular, and aperiodic behaviors.

1. Phase Space Reconstruction

Phase space reconstruction is an effective method for analyzing nonlinear signals. According to the embedded dimension, the data point set under a chaotic system is reconstituted into a visible track (i.e., the chaotic attractor), reflecting the law of the nonlinear system in phase space through a one-dimensional time series.

First, we set the time series embedded to (Δ t for a time sampling interval)

$$x_1,x_2,x_3,\cdots,x_N, x_m = x(t_0 + m\Delta t)\ m = 1,2,\cdots N \quad (5)$$

Then, we construct an $n = N - (m-1)\tau$ m-dimensional phase space vector:

$$\boldsymbol{X}_j = (x_j,x_{j-\tau},\cdots,x_{j-(m-1)\tau})\ \ j = 1,2,\cdots n \quad (6)$$

where m is the embedding dimension, and τ is the time delay.

2. Correlation dimension

The correlation dimension is an important characteristic quantity for measuring chaotic systems. It reflects the degree of correlation between points in high-dimensional space. The fractal dimension is a key parameter for quantitatively characterizing chaotic attractors in chaos theory. Grassberger and Procaccia proposed the G–P algorithm [23] in 1984. The specific method is as follows:

Step 1. Reconstruct the phase space: For a given one-dimensional time series, $x_1,x_2,x_3,\cdots,x_N$, select the appropriate embedding dimension m_0 and time delay τ to construct an m-dimensional phase space:

$$X_j^0 = (x_j,x_{j-\tau},\cdots,x_{j-(m_0-1)\tau})\ \ j = 1,2,\cdots n_{m_0} \quad (7)$$

Step 2. Calculate the associated integral function:

$$C(r) = \lim_{n_{m_0}\to\infty} \frac{1}{n_{m_0}^2} \sum_{i,j=1}^{n_{m_0}} \theta(r - \left|X_i^0 - X_j^0\right|) \quad (8)$$

where $\left|X_i^0 - X_j^0\right|$ is the distance between the state vector Xi and Xj in Euclidean space, and $\theta(u)$ is a step function, defined as:

$$\theta(u) = \begin{cases} 1 & u > 0 \\ 0 & u \le 0 \end{cases} \quad (9)$$

$C(r)$ is the ratio of the point logarithm with a distance less than r in the phase space attractor to all point logarithms, which reflects the degree of dispersion of the phase points.

Step 3. Estimate the correlation dimension $D: N \to \infty$, which is related to the length of r when it is small, and $r \to 0$ if the correlation integral function $C(r)$ is subject to the index law, $\lim_{r\to 0} C(r) \propto r^D$. At this point, the attractor has fractal characteristics, and the correlation dimension D and the correlation function $C(r)$ approximately satisfy a log-linear relationship:

$$D(m) = ln\, C\,(r)/ln\, r \quad (10)$$

The estimated value corresponding to m_0 can be calculated by fitting.

Step 4. Estimate the embedding dimension: Continue to increase the embedding dimension m_0, and substitute it into Steps 2 and 3. Calculate repeatedly until m_0 gradually converges to a saturation value. At this point, $D(m)$ will not change with an increase of m_0.

3. Largest Lyapunov exponent

The largest Lyapunov exponent (LLE) is an important parameter for describing the nonlinear dynamic characteristics of a system. It measures the exponential mean rate of the divergence and convergence of adjacent trajectories in phase space, with robustness to the embedding dimension, time delay, and data size variation. LLE intuitively judges whether there is chaos in the phase space of a nonlinear dynamic system.

The definition of the LLE is as follows:

$$\lambda = \lim_{t\to\infty} \frac{1}{t} ln \frac{|\Delta R(t)|}{|\Delta R_0|} \quad (11)$$

where $|\Delta R_0|$ is the initial distance between two orbits in phase space, and $|\Delta R_t|$ denotes this distance after time t. There are many methods for calculating the LLE. We adopted the classic Wolf algorithm [24]. The main calculation steps are as follows:

Step 1. For discrete time series $x_1,x_2,x_3,\cdots,x_N$, the reconstruction dimension m is determined by the G–P algorithm. The average mutual information method determines the delay time interval τ, and reconstructs the phase space $X(t)=(x_t,x_{t-\tau},\cdots,x_{t-(m-1)\tau})$. The number of phase points is $n=N-(m-1)\tau$.

Step 2. At the $(N-(m-1)\tau)$ phase point, the initial phase point x_0 is taken as the base point, and a point x_1 closest to x_0 is selected as the end point to form the initial vector. The Euclidean distance between the end points of the base point is recorded as $L(t_0)$.

Step 3. Time step or evolution time t: The initial vector evolves along the trajectory to obtain a new vector. The Euclidean distance between the corresponding point and the end point can be recorded as $L(t_1)$, and the system linear index exponential growth rate during the corresponding period is recorded as :

$$\lambda=\frac{1}{t}ln\frac{L(t_1)}{L(t_0)} \tag{12}$$

Step 4. After iteratively traversing until all phase points are reached, the average value of each exponential growth rate is taken as the estimated LLE value:

$$\lambda=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\tau}ln\frac{L(t_i)}{L(t_{i-1})} \tag{13}$$

C. Nonlinear Feature Extraction of the Frequency Division Band

The sub-band feature extraction scheme adopted in this paper is shown in Fig. 4. First, a Bark wavelet filter bank, which accurately simulates the perception characteristics of the human ear, was used to filter and divide the voice signals into 24 frequency bands. Based on the different noise characteristics of pathological voices in different frequency bands, the frequency division factor α was set. In the frequency band from 0 to α, the Bark frequency cepstrum coefficient (BFCC) feature was extracted by DCT after the speech logarithm operation of the voice signal. In frequency bands $\alpha+1$ to 24, the correlation dimension and the largest Lyapunov characteristic were extracted after embedding the signal into the phase space. To ensure that the voice characteristics were better reflected in different frequency bands, a multi-frequency band nonlinear characteristic (MFNLC) was extracted. For the frequency division factor α, machine learning was needed to recognize pathological voices in different frequency bands, and the optimal frequency division factor α was selected according to the recognition rate.

IV. Experiments and Results

We used the MEEI database [10], which contains 1,384 pathological voices and expert diagnostic results. The vowel /a:/ was selected as the test set. On one hand, the /a:/ sound

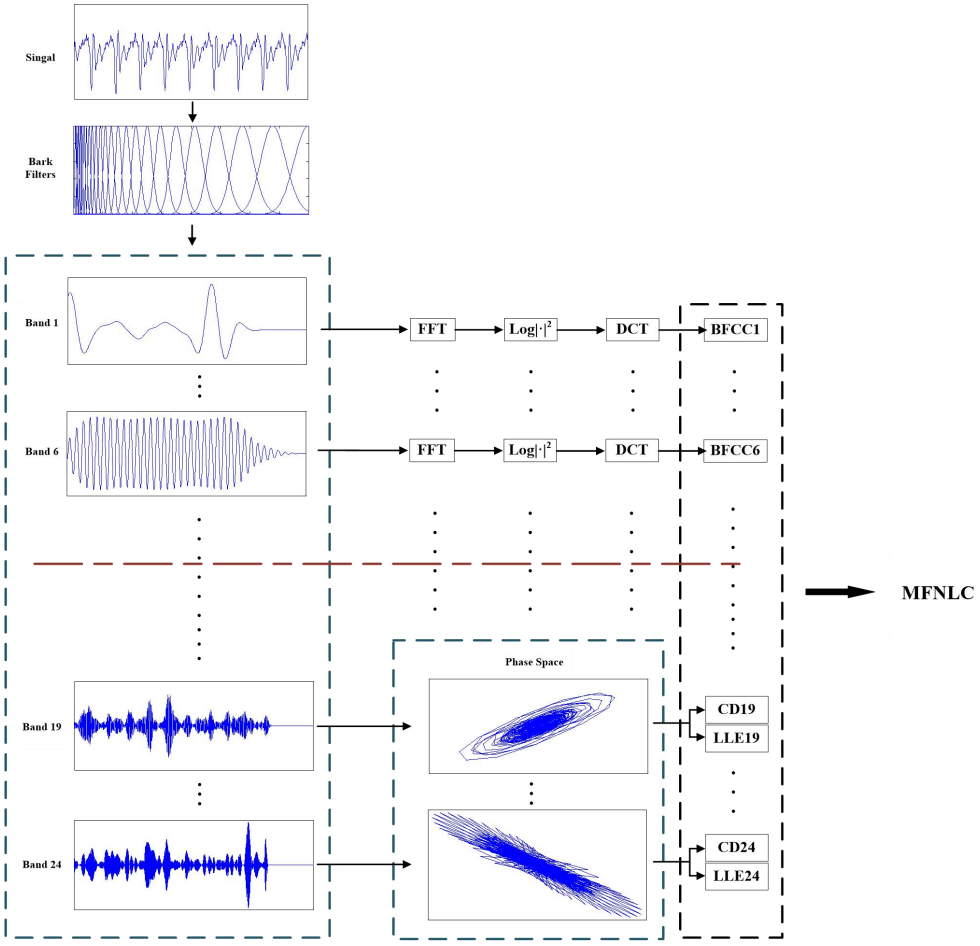

Fig. 4. Schematic diagram of sub-band nonlinear feature extraction.

was less affected by the oral cavity, and, on the other hand, it was more affected by auditory feedback. We selected 53 normal voices and 117 pathological voices as data subsets from the database. The subset was selected by taking into account the type of the pathological voice and the sex and age distribution of the normal and pathological voice samples. Specific statistics are shown in Table 1. Ten-fold cross-validation experiments were carried out. All samples were evenly divided into 10 sub-samples. Nine of these were trained, and the remaining subset was used for testing. The experiment was repeated 10 times and each subsample was tested once.

The frequency division factor of 0 to 24 was selected, and 14 machine learning classifiers were used: BayesNet, NaiveBayes, Deep Neural Network (DNN), Linear Discriminant Analysis (LDA), Logistic, Multi-layer Perceptron (MLP), Back Propagation Neural Network (BP), Radial Basis Function Neural Network (RBF), Simple Logistic (SL), Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Kstar, Locally Weighted Learning (LWL), Decision tree (DT), and Random Forest (RF). The data obtained are shown in Table 2.

TABLE 1 VOICE CONDITION STATISTICS

Voice category	Male		Female	
	Sample size	Average age (year), standard deviation	Sample size	Average age (year), standard deviation
Normal	21	38.8, 8.5	32	34.2, 7.9
Pathological	70	41.7, 9.4	103	37.6, 8.2

TABLE 2 STATISTICAL TABLE OF NON-LINEAR CHARACTERISTIC RECOGNITION RATE OF SUB-BAND (%)

Database	α	K-fold	Bayes Net	Naive Bayes	DNN	LDA	Logistic	MLP	BP	RBF
MEEI	0	10	92.04	91.15	87.61	91.15	85.40	88.94	88.94	92.04
MEEI	1	10	91.59	90.71	87.17	90.27	82.74	88.94	89.82	92.48
MEEI	2	10	91.59	90.71	84.96	88.94	85.40	88.50	88.50	92.04
MEEI	3	10	92.48	90.27	85.40	90.71	81.42	88.94	87.61	91.15
MEEI	4	10	92.48	90.27	87.17	90.71	82.30	90.71	89.82	89.82
MEEI	5	10	90.27	89.38	89.38	92.04	87.61	91.15	90.27	90.71
MEEI	6	10	91.59	90.27	85.84	92.04	87.61	90.71	92.48	93.81
MEEI	7	10	93.36	92.04	90.27	96.02	92.48	96.02	94.69	94.25
MEEI	8	10	93.81	92.04	88.50	96.90	88.94	96.46	96.90	94.69
MEEI	9	10	96.46	94.25	89.38	96.02	91.59	97.35	97.79	94.25
MEEI	10	10	95.58	96.02	89.82	96.02	94.69	96.90	96.90	96.02
MEEI	11	10	96.46	95.58	91.59	96.46	91.59	95.58	96.02	93.3
MEEI	12	10	95.13	95.58	91.15	96.46	88.94	95.58	96.02	92.92
MEEI	13	10	96.46	96.02	93.36	96.02	91.15	95.58	96.90	93.81
MEEI	14	10	96.02	96.90	89.82	96.90	91.15	94.25	94.25	93.36
MEEI	15	10	96.46	98.23	92.92	96.46	88.50	94.69	95.13	95.13
MEEI	16	10	96.02	98.23	92.04	96.46	92.04	96.90	96.46	96.02
MEEI	17	10	95.58	97.79	93.81	97.35	92.48	97.79	97.35	96.46
MEEI	18	10	95.58	97.35	92.48	97.35	92.48	99.12	97.35	97.79
MEEI	19	10	96.02	96.46	92.92	96.90	92.92	98.67	96.46	95.13
MEEI	20	10	95.58	96.90	92.48	97.79	93.81	98.23	97.35	96.90
MEEI	21	10	95.58	96.46	93.36	97.79	93.81	97.79	97.35	97.79
MEEI	22	10	96.02	96.90	92.92	97.79	88.94	96.90	96.90	96.90
MEEI	23	10	95.13	96.46	92.04	97.35	92.48	97.35	97.79	97.35
MEEI	24	10	94.69	95.13	93.81	96.46	92.92	96.90	96.02	96.46

Database	α	K-fold	SL	SVM	KNN	Kstar	LWL	DT	RF	Average
MEEI	0	10	90.71	92.04	89.82	88.05	86.73	84.96	91.59	89.41
MEEI	1	10	90.27	92.04	88.05	85.40	86.73	84.96	91.15	88.82
MEEI	2	10	89.38	92.04	88.50	87.17	87.17	84.51	90.27	88.64
MEEI	3	10	88.94	90.27	87.17	85.84	87.17	83.63	90.71	88.11
MEEI	4	10	90.27	90.71	86.28	85.84	86.28	81.86	90.27	88.32
MEEI	5	10	91.59	91.59	86.28	86.28	86.73	85.40	91.15	89.32
MEEI	6	10	91.15	94.69	88.50	86.28	86.28	88.50	91.59	90.09
MEEI	7	10	97.79	96.46	90.71	88.05	86.73	89.82	93.81	92.83
MEEI	8	10	96.46	96.90	91.59	88.94	86.73	90.27	95.13	92.95
MEEI	9	10	96.46	97.35	94.25	90.71	86.73	95.13	95.58	94.22
MEEI	10	10	97.35	96.90	94.69	89.82	86.73	96.02	96.02	94.63
MEEI	11	10	95.58	96.02	95.13	88.94	85.84	92.48	90.27	93.39
MEEI	12	10	94.25	96.90	92.48	89.38	87.61	91.59	95.58	93.30
MEEI	13	10	95.58	97.35	95.58	91.59	87.61	93.36	94.69	94.33
MEEI	14	10	95.13	96.90	94.69	92.04	84.51	88.05	93.81	93.19
MEEI	15	10	95.13	98.23	94.69	91.15	83.19	89.38	96.46	93.72
MEEI	16	10	97.35	97.79	94.69	92.48	84.07	90.27	96.02	94.45
MEEI	17	10	97.35	98.23	95.13	88.50	84.07	91.15	97.79	94.72
MEEI	18	10	97.35	99.56	94.69	88.94	83.63	91.59	96.90	**94.81**
MEEI	19	10	96.46	99.12	94.69	91.15	82.30	92.92	97.35	94.63
MEEI	20	10	97.35	99.12	92.92	86.73	82.74	92.04	97.35	94.48
MEEI	21	10	95.58	98.67	96.46	87.17	82.74	92.04	96.90	94.63
MEEI	22	10	95.58	99.12	97.3	87.17	83.19	92.04	97.79	94.37
MEEI	23	10	96.46	99.12	97.35	88.50	84.96	91.59	96.90	94.72
MEEI	24	10	96.90	98.67	95.58	83.19	83.19	91.59	96.46	93.86

Fig. 5 shows the average recognition rate of the 14 classifiers. When α = 0, the average recognition rate was 89.41%. Between α = 1 and α = 3, the recognition rate decreased. When α = 4, the recognition rate started to increase steadily. When α = 18, the average recognition rate reached 94.81%, before beginning to decrease. Thus, when α = 18, the performance of the extracted sub-band characteristic parameters was the highest. At this time, the low frequency band and the high frequency band best depicted the rich details of the voice signal.

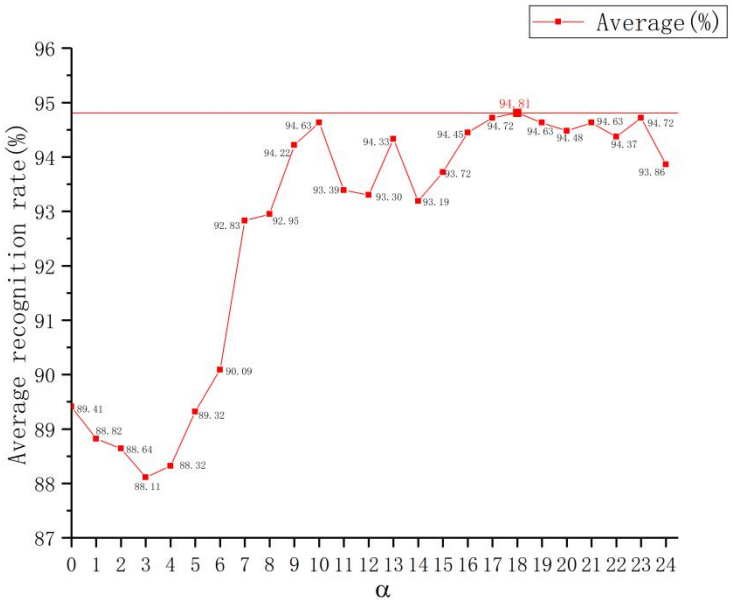

Fig. 5. Average recognition rate of nonlinear features in the frequency division band

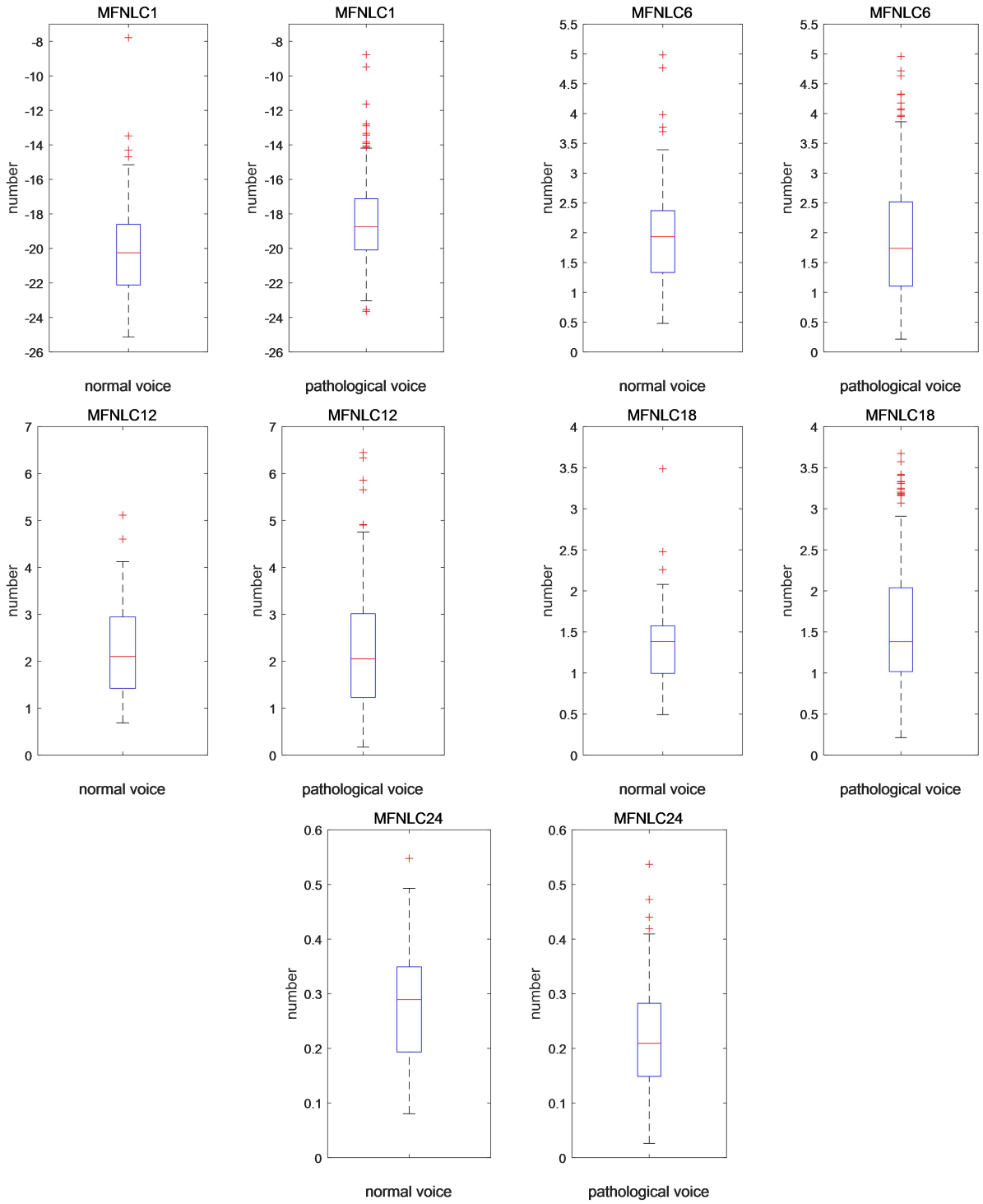

Fig. 6. Box charts of normal and pathological voices (MFNLC1, 6, 12, 18, 24)

Fig. 6 presents the box charts of normal and pathological voices MFNLC1, 6, 12, 18, and 24. The distribution of the characteristics of normal and pathological voices can be seen from the charts. It can be seen that there is little data overlap between normal voices and pathological voices, and there are more discrete points of pathological voice parameters, which also means that these features can better distinguish between normal voices and pathological voices.

TABLE 3 P VALUES OF MFNLC (MANN WHITNEY U TEST)

	MFNLC1	MFNLC2	MFNLC3	MFNLC4	MFNLC5	MFNLC6	MFNLC7	MFNLC8	MFNLC9	MFNLC10	MFNLC11	MFNLC12
P	**0.001**	**0.000**	0.058	**0.002**	**0.015**	0.339	**0.003**	0.613	**0.000**	0.562	**0.000**	0.791

	MFNLC13	MFNLC14	MFNLC15	MFNLC16	MFNLC17	MFNLC18	MFNLC19	MFNLC20	MFNLC21	MFNLC22	MFNLC23	MFNLC24
P	**0.000**	**0.029**	**0.000**	0.938	**0.000**	0.229	**0.000**	0.062	**0.000**	0.371	**0.003**	**0.000**

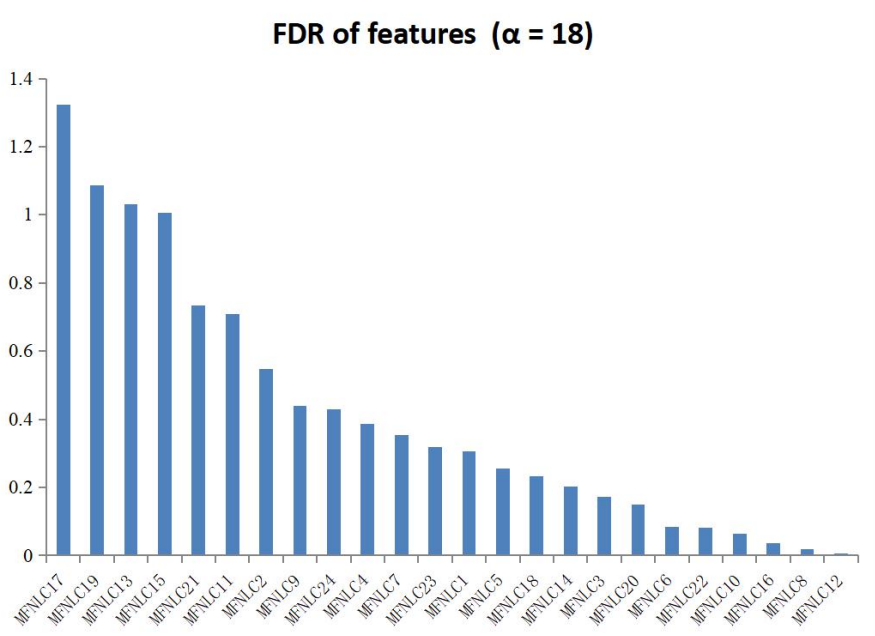

Fig. 7. FDR of MFNLC features

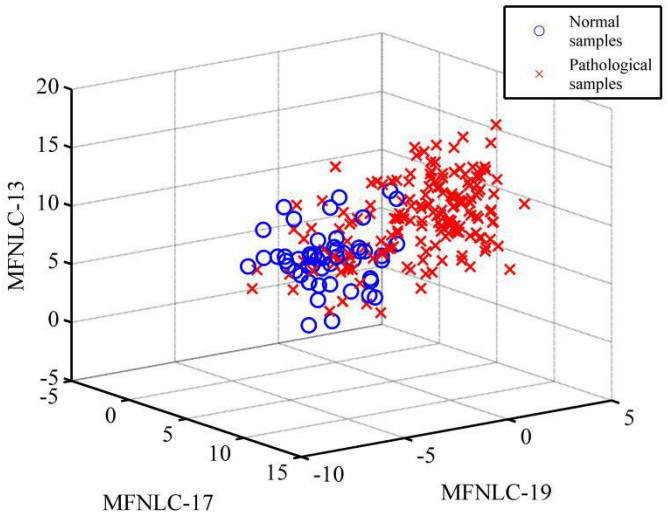

Fig. 8. Three-dimensional scatter plot (MFNLC17, 19, 13)

Table 3 shows the P values of MFNLC obtained by the Mann Whitney U test. The values marked in bold indicate statistical differences (P<0.05). It was found that the MFNLC proposed in this paper can increase the statistical difference in the overall mean values of normal voices and pathological voices.

Fig. 7 shows the false discovery rate (FDR) of MFNLC features; MFNLC17, 19, 13 have the biggest FDR. Fig. 8 shows a three-dimensional scatter plot of the top three distinguishing features (MFNLC17, 19, 13) obtained using the FDR and P-test. It can be seen from the figure that although there are a few normal (blue circle) and pathological (red cross) samples with slight overlap, the high differentiation of the overall distribution of the two categories demonstrates that the multi-band nonlinear features can effectively separate normal voices from pathological voices.

Acoustic parameters (APs) (specific acoustic parameters that we used, as listed in Table 4), the linear predictive cepstrum coefficient (LPCC), the MFCC, and our proposed sub-band nonlinear feature (α = 18) were used to extract the parameters of the voice data subsets, identified under the 14 classifiers. According to the data in Table 5, the recognition rate of the nonlinear feature of the frequency band with α = 18 under the SVM classifier reached 99.56%, considerably better than that of APs, the LPCC, and the MFCC. The data in the table show that the average pathological voice recognition rate with APs, the LPCC, and the MFCC under the 14 machine learning classifiers was 86.07%, 92.60%, and 94.13%, respectively, or 8.7%, 2.2%, and 0.7% lower than that of the proposed method. Thus, the method of extracting α = 18 sub-band nonlinear features better reflected the characteristics of pathological voices. The specific identification rate is shown in Table 4.

TABLE 4 ACOUSTIC PARAMETERS

Feature	Description
Fo	Average fundamental frequency for the vocalization.
STD	Standard deviation of the fundamental frequency in the vocalization.
Jita	The absolute jitter gives an evaluation in microseconds of the period-to-period variability of the pitch period within the analyzed voice sample.
RAP	The relative average perturbation gives an evaluation of the variability of the pitch period within the analyzed voice sample at a smoothing factor of three periods.
PPQ	The pitch period perturbation quotient gives an evaluation in percent of the variability of the pitch period within the analyzed voice sample at a smoothing factor of five periods.
sPPQ	The smoothed pitch period perturbation quotient gives an evaluation in percent of the long-term variability of the pitch period within the analyzed voice sample at a smoothing factor of 55 periods.
vFo	The fundamental frequency variation represents the relative standard deviation of the period-to-period calculated fundamental frequency.
ShdB	The shimmer percent gives an evaluation in percent of the variability of the peak-to-peak amplitude within the analyzed voice sample.
APQ	The amplitude perturbation quotient gives an evaluation in percent of the variability of the peak-to-peak amplitude within the analyzed voice sample at a smoothing factor of 11 periods.
sAPQ	The smoothed amplitude perturbation quotient gives an evaluation in percent of the long-term variability of the peak-to-peak amplitude within the analyzed voice sample with a smoothing factor of 55 periods.
vAm	The peak amplitude variation represents the relative standard deviation of the period-to-period calculated peak-to-peak amplitude.
NHR	The noise-to-harmonic ratio is the average ratio energy of the in-harmonic components in the range 1500–4500 Hz to the harmonic components energy in the range 70–4500 Hz.
VTI	The voice turbulence index is the average ratio of the spectral in-harmonic high frequency energy to the spectral harmonic energy in stable phonation areas.
SPI	The soft phonation index is the average ratio of the lower frequency to the higher frequency harmonic energy.

TABLE 5 STATISTICAL TABLE COMPARING THE IDENTIFICATION RATE (%)

	APs	**LPCC**	**MFCC**	**Proposed method (α=18)**
Bayes Net	85.40	92.04	99.12	95.58
Naive Bayes	70.80	91.59	98.23	97.35
DNN	84.96	88.94	91.15	92.48
LDA	88.39	97.35	95.13	97.35
Logistic	86.73	94.69	93.36	92.48
MLP	91.15	95.13	95.58	99.12
BP	88.50	95.13	94.69	97.35
RBF	86.28	93.81	93.36	97.79
SL	88.05	97.35	96.46	97.35
SVM	83.19	96.46	96.90	**99.56**
KNN	88.50	87.17	94.69	94.69
Kstar	87.61	84.51	85.84	88.94
LWL	84.51	88.94	87.17	83.63
DT	86.73	90.27	92.92	91.59
RF	90.27	95.58	97.35	96.90
Average	86.07	92.60	94.13	**94.81**

TABLE 6 COMPARISON OF RESULTS (%)

	Correctly Classified	Kappa statistic	Mean absolute error	Relative absolute error	Precision	Recall	F-Measure	ROC Area	F1 Score
AP+SVM	83.19	40.76	16.81	46.64	83.60	83.20	80.30	66.10	50.00
LPCC+SVM	96.46	89.88	3.54	9.82	96.40	96.50	96.40	93.80	93.68
MFCC+SVM	96.90	91.43	3.10	8.59	96.90	96.90	96.90	96.00	95.92
Proposed method+SVM	**99.56**	**98.76**	**0.44**	**1.23**	**99.60**	**99.60**	**99.60**	**99.10**	**98.99**

The recognition rate of the proposed $\alpha = 18$ sub-band nonlinear feature under the SVM classifier reached 99.56%, outperforming traditional parametric methods. To highlight the performance of this method, a confusion matrix is shown in Fig. 9, comparing the classifier's prediction results using APs, the LPCC, the MFCC and the MFNLC under the SVM classifier. The confusion matrix shows that the TP and TN values of the proposed method are higher than those of APs, the LPCC and the MFCC. The corresponding FN and FP values are extremely small, indicating the effectiveness of the proposed method.

Table 6 shows the correct classification rate, kappa statistic, mean absolute error, relative absolute error, precision, recall, F-measure, ROC Area and F1 score of the above four methods. The results showed that the proposed method had the highest correct classification rate and kappa statistic and the lowest mean absolute error and relative absolute error. These results demonstrated that the method was the most accurate, with the smallest error and the best algorithmic performance. The ROC curves of the four methods are shown in Figs. 10–13. From the curve, we can see that the ROC area of the proposed method is 0.9910, demonstrating the effectiveness of the nonlinear features.

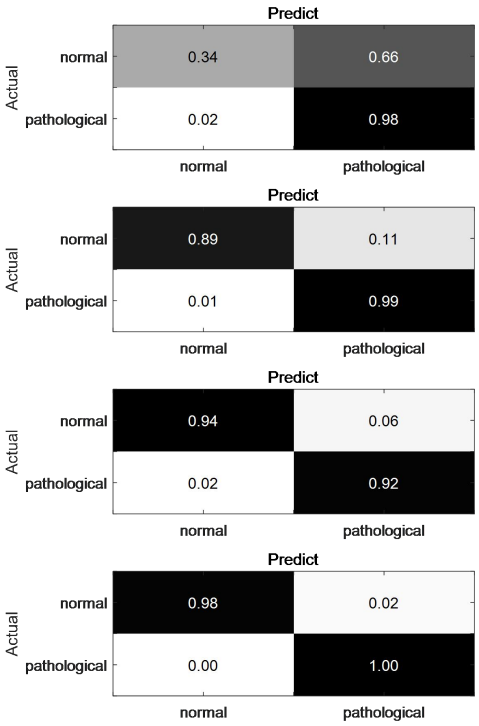

Fig. 9. Confusion matrix comparison chart

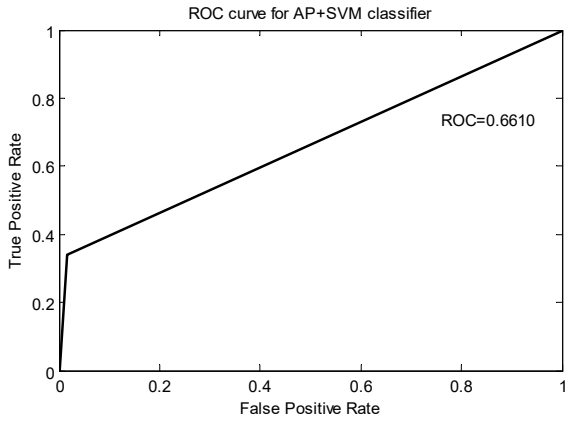

Fig. 10. ROC curve for AP+SVM classifier

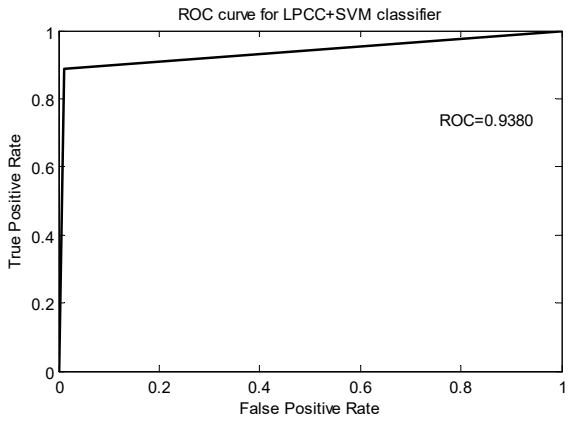

Fig. 11. ROC curve for LPCC+SVM classifier

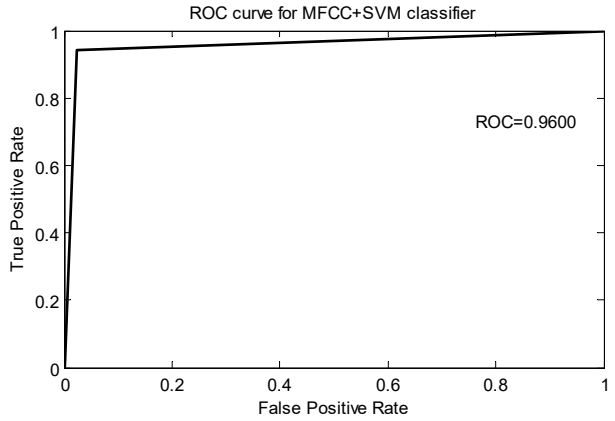

Fig. 12. ROC curve for MFCC+SVM classifier

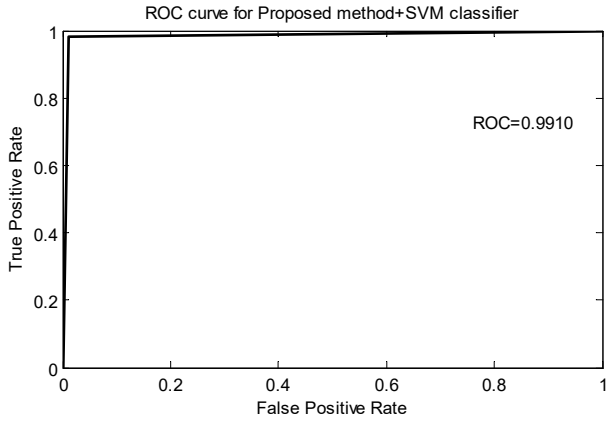

Fig. 13. ROC curve for Proposed method+SVM classifier

V. Conclusion

We used Bark wavelet sub-band filtering to extract characteristic parameters unique to voice pathologies according to the characteristics of energy distribution. According to the simulation results, when the frequency division factor $\alpha = 18$, the average recognition rate of pathological voices was 94.81% (with 14 standard machine learning methods). A recognition rate of 99.56% was achieved with a SVM classifier, and each evaluation index was the best, demonstrating the effectiveness of the proposed method.

There are still some problems to be solved and improvements to be made when identifying pathological voices using Bark wavelet frequency-division nonlinear characteristic parameters. For example, our method of calculating the correlation dimension and largest Lyapunov characteristic used in the $\alpha + 1$ to 24 frequency bands is too complex and time-consuming. In future experiments, we will study optimizing MFNLC features and the algorithm, and we will explore adaptive selection of the frequency division factor α according to different samples. We also plan to optimize the kernel function in the SVM, to improve the recognition rate further.

References

[1] J. F. Restrepo and G. Schlotthauer , "Invariant measures based on the U-correlation integral: An application to the study of human voice," Complexity, pp1-9, 2018

[2] N. Roy, R. M. Merrill, S. Thibeault, R. A. Parsa, S. D. Sary and E. M. Smith, "Prevalence of voice disorders in teachers and the general population, "Journal of Speech, Language & Hearing Research, vol. 47, no. 2, pp281-293, 2004

[3] X. Zhang, T. Zhi, H. Zhao and T. Xu, "Pathological voice recognition by deep neural network," 2017 4th International Conference on Systems and Informatics, pp456-460, 2017

[4] P. Lieberman, "Perturbations in Vocal Pitch," Journal of the Acoustical Society of America, vol. 33, no. 5, pp597-603, 1961

[5] P. Lieberman, "Some Acoustic Measures of the Fundamental Periodicity of Normal and Pathologic Larynges," Journal of the Acoustical Society of America, vol. 35, no. 3, pp344-353, 1963

[6] H. Kasuya, Y. Kobayashi, T. Kobayashi and S. Ebihara, "Characteristics pf pitch period and amplitude perturbations in pathologic voice," Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP, vol.8,no.3,pp1372-1375,1983

[7] E. Yumoto , "Harmonics-to-noise ratio as an index of the degree of hoarseness," Journal of the Acoustical Society of America, vol. 71, no. 6, pp1544-1549, 1982

[8] T. Fukazawa, A. el-Assuooty and I. Honjo, "A new index for evaluation of the turbulent noise in pathological voice," Journal of the Acoustical Society of America, vol. 83, no. 3, pp1189-1193, 1988

[9] N. C. Roussel and M. Lobdell, "The clinical utility of the soft phonation index," Clinical Linguistics & Phonetics, vol. 20, no. 2-3, pp181-186, 2009

[10] J. I. Godino-Llorente, P. Gómez-Vilda and M. Blanco-Velasco, "Dimensionality Reduction of a Pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters," IEEE Transactions on Biomedical Engineering, vol. 53, no. 10, pp1943-1953, 2006

[11] D. Pravena, S. Dhivya and D. A. Durga, "Pathological voice recognition for vocal fold disease," International Journal of Computer Applications, vol. 47, no. 13, pp31-37, 2012

[12] Q. Fu, and K. Yi, "Bark wavelet transform of speech and its application in speech recognition," Acta Electronica Sinica, vol. 28, no. 10, pp102-105, 2000

[13] I. R. Titze, R. Baken, and H. Herzel, "Evidence of chaos in vocal fold vibration physiology: Frontiers in Basic Science," San Diego, CA: Singular Publishing Group, pp143–188, 1993

[14] H. Herzel, D. Berry, I. R. Titze and M. saleh, "Analysis of Vocal Disorders With Methods From Nonlinear Dynamics," Journal of Speech Language and Hearing Research, vol. 37, no. 5, pp1008-1019, 1994

[15] J. K. Maccallum, L. Cai, L. Zhou, Y. Zhang and J. J. Jiang, "Acoustic Analysis of Aperiodic Voice: Perturbation and Nonlinear Dynamic Properties in Esophageal Phonation," Journal of Voice, vol. 23, no.3, pp283-290, 2009

[16] Y. Zhang, J. J. Jiang, S. M. Wallace and L. Zhou, "Comparison of nonlinear dynamic methods and perturbation methods for voice analysis," Journal of the Acoustical Society of America, vol. 118, no. 4, pp2551-2560, 2005

[17] J. J. Jiang, Y. Zhang, C. N. Ford, "Nonlinear dynamics of phonations in excised larynx experiments," Journal of the Acoustical Society of America, vol. 118, no. 4, pp2198-2205, 2003

[18] C. M. Travieso, J. B. Alonso, J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, E. Noeth and A. G. Ravelo-Garcia, "Detection of different voice diseases based on the nonlinear characterization of speech signals," Expert Systems With Applications, vol. 82, pp184-195, 2017

[19] Y. Yang, Y. Li, W. Wang and W. Yang, "Fast and Efficient Evolutionary Algorithms Based on Bayesian Networks," IAENG International Journal of Computer Science, vol. 42, no. 3, pp174-182, 2015

[20] J. Ma, S. Yan and I. McLoughlin, "End-to-end DNN-CNN Classification for Language Identification," Proceedings of The World Congress on Engineering , pp199-203, 2017

[21] U. Iruansi, J. R. Tapamo and I. E. Davidson, "Classification of Power-line Insulator Condition using Local Binary Patterns with Support Vector Machines," IAENG International Journal of Computer Science, vol. 46, no. 2, pp300-310, 2019

[22] H. Traunmuller, "Analytical expressions for the tonotopic sensory scale," Journal of the Acoustical Society of America, vol. 88, no. 1, pp97-100, 1990

[23] P. Grassberger, I. Procaccia, "Measuring the strangeness of strange attractors," Physica D, vol. 9, no. 1-2, pp189-208, 1983

[24] A. Wolf, J. B. Swift, H. L. Swinney and J. A. Vastano, "Determining Lyapunov exponents from a time series," Physica D Nonlinear Phenomena, vol. 16, no. 3, pp285–317, 1985