Research on Feature Extraction and Classification of Ultrasonic Flaw

Hongjia Chen, Hui Liu, Xiaoyan Wang, Yanping Bai

***Abstract*—**By the laser ultrasonic surface wave surface defect detection experiments, we got reflected ultrasonic and the transmitted wave signals in different width and depth cases. Our purpose is to classify the ultrasonic defect signals and search for a way can identify 5 kinds of defect signals. According to non-linear, non-stationary characteristics of ultrasonic detection signal, we use the Mel Frequency Cepstral Coefficient method (MFCC) to extract characteristic coefficients of two waveform signals, to achieve effective detection of different defect categories. We respectively dispose the features with BP neural network and NNRS model. In our experiment, we compare the training and classification of BP neural network to the improved BP neural network with additional momentum. The experiments show that the neural network can recognize defect signals categories. But BP neural network only suit to small simples test. Then we use NNRS model because it is fit for a large number of samples. The result is famous, the recognition rate of two waves over 90%, and the diagnostic accuracies are both 100%. NNRS is more precise and practical.

***Keywords*—**ultrasonic flaw, MFCC, BP neural network, additional momentum , NNRS.

I INTRODUCTION

LASER ultrasonic technique is an important non-destructive testing technology [1,2], which stimulate SAW due to the high sensitivity of surface and sub-surface tiny cracks, very suitable for the detection of tiny cracks. In industrial tests, the feature extraction of ultrasonic echo identify defect primarily extracts temporal characteristics, time - frequency domain characteristics [3] and frequency domain features [4], by extracting the defect features information of relevant field to achieve the purpose of identifying defects. There are many feature extraction methods, , in general, the method based on time-domain feature extraction is time-series model (AR model, ARMA model, etc.); the method based on frequency domain feature extraction is Fast Fourier Transform (FFT); the methods based on time – frequency domain feature extraction are short-time Fourier transform (STFT), time-frequency distribution (Wigner-Ville distribution, Choi-William distribution), wavelet transform, Hilbert-Huang Transform.

Hongjia Chen is with North University of China, School of Science (e-mail:c1224hj@126.com).

Hui Liu is with North University of China, School of Information and Communication Engineering (e-mail:1159894600@qq.com).

Xiaoyan Wang is with North University of China, School of Science (e-mail:xywang00@126.com).

Yanping Bai is with North University of China, School of Science (corresponding author to provide phone: 0351-3942729; e-mail: baiyp666@163.com).

Because of ultrasonic signals of defects recognition extraction, selection of evaluation methods and the eigenvalues of the law are still in the exploratory stage, with uncertainty.

Mel frequency is based on the human auditory characteristic features put forward, and a nonlinear corresponding relationship with it in Hz frequency. Mel Frequency Cepstral Coefficient (MFCC) [5] uses the relationship between them to calculate spectral characteristics of Hz. We will use MFCC to extract feature information of the defect, to achieve the data dimension reduction and easily deal with. Recently, artificial intelligence becomes a popular subject. Artificial neural network [6-10] is an intelligent information processing system built to mimic the human brain, and has a highly nonlinear global mapping. It has a very strong adaptive, self-learning ability and high fault tolerance and robustness on the environment. Then we use BP neural network to dispose the laser ultrasonic wave signal features, owing to the network having powerful massively parallel, distributed processing, self-organizing, self-learning ability. Then we improve BP network with additional momentum. However, the BP algorithm is classical and obsolete. Deep learning [11-15] is an important area of artificial intelligence. It has been popular in the research community, and has become a huge wave of technology trend for big data and artificial intelligence. Deep learning simulates the hierarchical structure of human brain, processing data from lower level to higher level, and gradually composing more and more semantic concepts. In this paper, we use neural network regime switching(NNRS) model [16-18] to dispose the processing. NNRS is based on deep learning with jump connection, and turn weights updating to logistic regression. These are according with human brain complex nonlinear processing, and make the result more scientific and credible.

In this paper, we use the Mel Frequency Cepstral Coefficient method to extract the ultrasonic spectrum characteristics. Then we research its characteristics with BP Neural Network and NNRS to observe classification and recognition results.

The paper is organized as follows. After the statement and formulation of the problem in Section Ⅱ, the information theoretic criteria for feature extraction are introduced in Section Ⅲ. In Section Ⅳ, we draw into the theoretic of BP neural network and improve BP neural network with additional momentum, and give the results of experiments. The experiments comparison between BP neural network and the improved BP neural network are discussed in this part. In

SectionⅤ, we discover the BP infeasible for a larger number of samples. Then we conceive to dispose it with NNRS. NNRS principle concluding deep learning and jump connection is listed here. And we have a good result. In Section Ⅵ, we give comparison the results between BP neural network and NNRS. Finally, some concluding remarks are presented in SectionⅦ.

II THE EXPERIMENTAL PRINCIPLE AND THE ULTRASONIC FLAW SIGNALS TO BE IDENTIFIED

Laser ultrasonic flaw detection technique is based on the theory of sound and light effects. It generates thermal elastic effect when pulsed laser irradiates at the sample surface, and generates an ultrasonic signal containing information of the measured surface. By detecting the defects of the ultrasonic signal after modulate to extract information about the defect for defect detection. If the specimen has discontinuous region such as defects, it will occur scattering, reflection and transmission phenomena when the ultrasonic wave propagates to the region, and then leading to the characteristics of ultrasonic signal will change significantly.

In our experiment, the sample to be tested is a 270*70*40mm aluminum plate, we use 2M ultrasound probe to detect ultrasonic reflected wave and transmitted wave, and the bandwidth of probe is 2M, i.e., the detection frequency range of 1-3M, and the five defect specifications respectively are as follows: (1) width 0.1mm depth 0.3mm; (2) width 0.1mm depth 0.5mm; (3) width 0.1mm depth 0.7mm; (4) wide and 0.1mm deep 0.9mm; (5) non-destructive; each experiment repeatedly measured 5 times. We repeat each experiment measurements for five times, and the sampling rate is 200MHz, the sampling points are 44,000, trigger position 10%. We put probe at 20mm from the laser spot to detect reflected waves, and transmitted wave at 40mm. We show the schematic diagrams for the experiment of reflected wave and transmitted wave measurement respectively in Figure 1 and Figure 2. And in Figure 3 we show a group of signal of two kinds of waves that every wave has five specifications in our experiments.

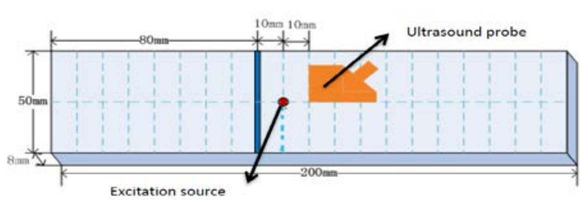


Fig. 1 The schematic diagram for the experiment of reflected wave measurement

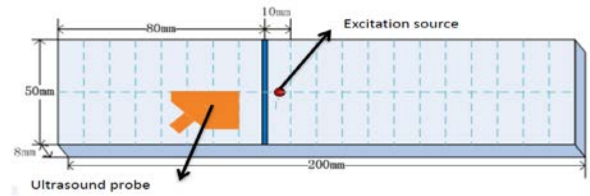
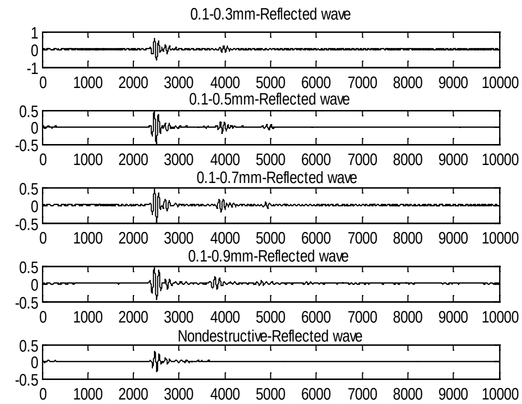
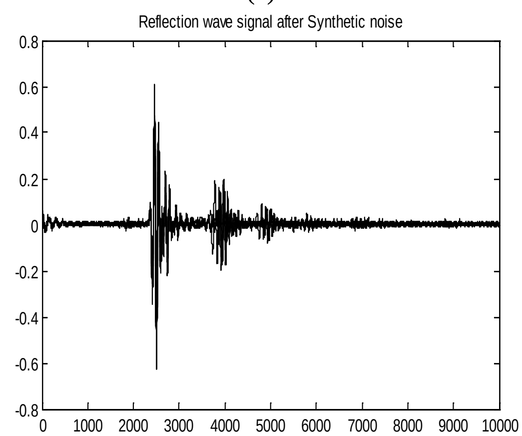


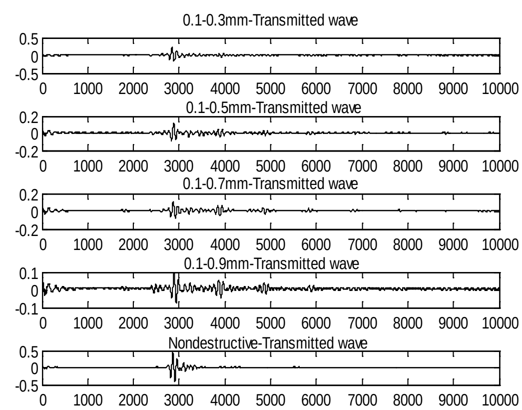
Fig. 2 The schematic diagram for the experiment of transmitted wave measurement



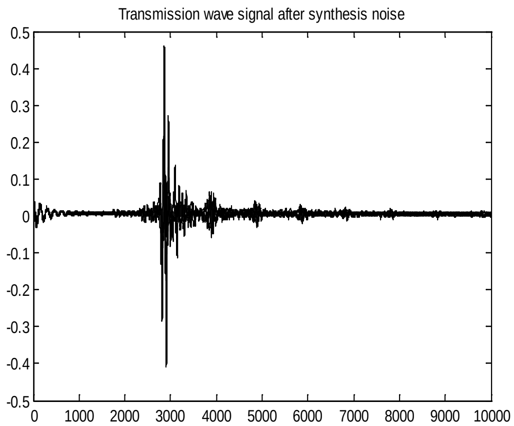
(a)



(b)



(c)



(d)

Fig. 3 Five groups of signal of two kinds of waves (a) Five types of signals of reflected wave. (b) Synthetic reflected wave signals of reflected wave. (c) Five types of signals of transmitted wave. (d) Synthetic reflected wave signals of transmitted wave.

III MFCC PRINCIPLE

Cochlea is substantially equivalent to a filter set, the filtering effect of the cochlea is used on a logarithmic frequency scale .Below 1,000HZ, there is a linear relationship between the human ear's perception and the frequency; In more than 1,000HZ, the human ear's perception does not constitute a linear relationship with frequency, but more inclined to logarithmic relationship, which makes the human ear to low frequency signal is more sensitive than the high-frequency signal. Mel frequency and frequency conversion formula is:

$$F_{mel} = 2595 * \lg(1 + f_{HZ} / 700) \quad (1)$$

1) Pre-emphasis Processing

Pre-emphasis is actually a high-pass filter, the transfer function of the high pass filter is:

$$H(Z) = 1 - \alpha Z^{-1} \quad (2)$$

Where the value of α is 0.97, the role of high-pass filter is filtering low frequency, high frequency characteristics of the ultrasonic signal is more emergent.

2) Frame and Window Treatments

The ultrasonic signal is constantly changing, so to simplify things we assume that on short time scales the signal doesn't change much (when we say it doesn't change, we mean statistically i.e. statistically stationary, obviously the samples are constantly changing on even short time scales). This is why we frame the signal into 20-40ms frames(25ms is standard). If the frame is much shorter we don't have enough samples to get a reliable spectral estimate, if it is longer the signal changes too much throughout the frame.

3) Each Frame Signal FFT Transformation

We make the FFT transform of Sub-frame for each windowed frame signal to get the spectrum of each frame. And get the square norm of the frequency spectrum of ultrasonic signal to get the power of the ultrasonic signals.

4) Calculate Triangular Filter Coefficients

Now we create our filterbanks. The first filterbank will start at the first point, and reach its peak at the second point, then return to zero at the 3rd point. The second filterbank will start at the 2nd point, reach its max at the 3rd, then be zero at the 4th etc. A formula for calculating these is as follows:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \le k \le f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \le k \le f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (3)$$

And it satisfies

$$Mel(f(m)) - Mel(f(m-1)) = Mel(f(m+1)) - Mel(f(m))\text{'}$$

where M is the number of filters we want, and $f(m)$ is the list of M+2 Mel-spaced frequencies, generally considered 10.

Calculated filter coefficients is $m(i), i = 1, \ldots. p$, p is the filter order.

5) Triangular Filtering and Discrete Cosine Transform DCT

$$C_i = \sum_{k=1}^{p} \log(m_k) \cos[l(k - \frac{1}{2}) \frac{\pi}{p}] \quad (4)$$

C_i is the desired extracted feature parameters.

Because our filterbanks are all overlapping, the filterbank energies are quite correlated with each other. The DCT decorrelates the energies which means diagonal covariance matrices can be used to model the features in e.g. a HMM classifier. But notice that only 12 of the 26 DCT coefficients are kept. This is because the higher DCT coefficients represent fast changes in the filterbank energies and it turns out that these fast changes actually degrade ASR performance, so we get a small improvement by dropping them. We have 12 MFCC coefficients, we would also get 12 delta coefficients, which would combine to give a feature vector of length 24

MFCC has been widely used in speech recognition. Its extraction process is as Figure 4:

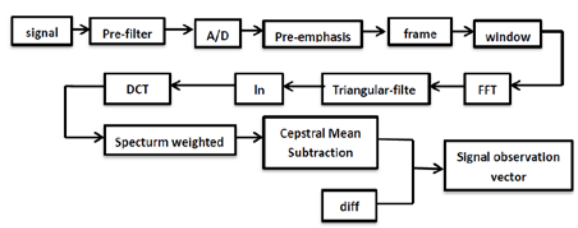


Fig. 4 The processing of MFCC extracts signal characteristics

We sample five groups of defect signals, and get data of 10,000*25. Then we extract feature with Mel Cepstral method. Due to our sampling rate is 200MHz, $f_{HZ} = 20000$. Each defect signal extract 24 features, sample length is 75, then the five kinds of defect signals turn into the sample characteristics 1875 * 24.

IV The Principle Of Bp Network

A. BP Network Structure

Back-propagation (BP) neural network idea was first proposed in 1969 by Bryson etc., it was not until 1986 that Rumelhart and his team published their findings in the journal Nature, the BP network to get the attention of people. BP network is actually a multi-layer perception and a supervised learning algorithm, The network consists of a large number of processing units constructed through an extensive interconnected network system, with massively parallel, distributed processing, self-organizing, self-learning, etc. advantage, is widely used in function approximation, pattern recognition, classification, data compression, and many other fields.

BP neural network is a multilayer feedforward neural network, its main features is to transmit before the signal, the error back-propagation. In the forward pass, an input signal from the input layer through the hidden layer processing layer by layer, until the output layer. The states of neurons of each layer only affect the next layer neuron state. If the output layer is not expected, then transferred back propagation, adjust the network weights and thresholds based on prediction error, allowing BP neural network to predict the output constantly approaching the desired. And BP neural network topology is shown in Figure 5.

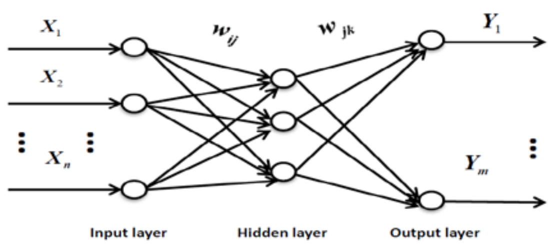


Fig. 5 BP neural network topology

Where $X_1, X_2, \cdots, X_n$ are the input values of BP neural network, $Y_1, Y_2, \cdots, Y_m$ are the predicted values of BP neural network, w_{ij} and w_{jk} is the weights of BP neural network. As we can be seen from the figure, BP neural network can be viewed as a nonlinear function, network input is independent variable of the function, and predicted values is the dependent variable of the function. When the input nodes are n, the output nodes is m, BP neural network expresses a function mapping relationship of n independent variables with m dependent variables. BP algorithm's learning rule is based on the gradient descent method, due to the gradient descent method is an effective nonlinear data fitting method, and it's the direct and effective method needs to calculate the derivative of unconstrained optimization algorithm, this is good maneuverability and overall convergence. Here, the gradient is a vector that we call it derivative.

B. The Working Principle of BP Network

Before BP neural network forecasting, first to train network, through training the network has associative memory and forecasting capabilities. The algorithm's basic idea that is turning the input and output problems of a set of sample into a nonlinear optimization problem, using gradient descent method that commonly used in optimization to realize mean square error (mse) minimum of the network between actual output and desired output, and finishing BP network training task. Training process of BP network is consist of forward transmission of work signals and back propagation of error signals:

1) Forward transmission of work signals. In the process of the spreading, the input samples income from input layer, finally to output layer after hidden layer processing step by step. We compare the actual output value to the desired output value of output layer, if there is a deviation, immediately go to the back propagation of error signals process;
2) Back propagation of error signals. The deviation between network's actual output and desired output is the error signal. This process is that the output error propagates along the original connection path step by step, and according to the way of minimizing error to adjust the weight matrix of the network. We adjust the network weights constantly and make the real output value of the network gradually approaching the designer's expectations.

Each layer weights adjustment process of forward transmission of work signals and back propagation of error signals is iterative, and the constantly adjust process of network weights is the training process of the network.

BP neural network training process includes the following steps.

1) Network initialization. According the system input and output sequence (X, Y) to determine the network input layer nodes n, hidden layer nodes l, output layer nodes m. Initialize the connection weights of neurons w_{ij} , w_{jk} between input layer and hidden layer and output layer, initialize the hidden layer threshold a, the output layer threshold b , given the learning rate and neuronal excitation functions.
2) Calculate the hidden layer output. According to the input vector X , the connection weights w_{ij} between the input layer and the hidden layer and the hidden layer threshold a, calculate the output of the hidden layer H.

$$H_j = f(\sum_{i=1}^{n} w_{ij} x_i - a_j) \qquad j = 1, 2, ..., l \tag{5}$$

Where l is the hidden layer node; f representatives hidden layer excitation function, which has a variety of forms, our paper selected the function is:

$$f(x) = \frac{1}{1+e^{x}} \tag{6}$$

3) Output layer output calculation. According to the output of the hidden layer H, the connection weights w_{jk} and thresholds b, calculate predicted output T of BP neural network.

$$T_k = \sum_{j=1}^{i} H_j w_{jk} - b_k \qquad k = 1, 2, ..., m \tag{7}$$

4) Error calculation. According to the network predicted output T and the desired output Y, computing network prediction error e.

$$e_k = Y_k - T_k \qquad k = 1, 2, ..., m \tag{8}$$

5) Update weights. According to the predicted error e update the network weights w_{ij}, w_{jk}.

$$w_{ij} = w_{ij} + \eta H_j (1 - H_j) x(i) \sum_{k=1}^{m} w_{jk} e_k \qquad i = 1, 2, ..., n;\ j = 1, 2, ..., l \tag{9}$$

$$w_{jk} = w_{jk} + \eta H_j e_k \qquad j = 1, 2, ..., l; k = 1, 2, ..., m \tag{10}$$

Where η is the learning rate.

6) Threshold update. According to the network predicted error e update the network node threshold a, b.

$$a_j = a_j + \eta H_j (1 - H_j) \sum_{k=1}^{m} w_{jk} e_k \qquad j = 1, 2, ..., l \tag{11}$$

$$b_k = b_k + e_k \qquad k = 1, 2, ..., m \tag{12}$$

7) Judge iterative algorithm to determine whether the end, if not end, return (2).

C. Improved BP network

At present, in the field of the application of neural network, BP algorithm is the most widely used. Although BP network can approximate any nonlinear function in theory, because there are many parameters in network training learning choice without theoretical basis, in practice, the algorithm itself has some limitations and shortcomings, mainly includes the following aspects:

1) The learning rate of BP network is fixed, its slow convergence speed and long-time training, it often requires thousands of times even more iterative training.
2) The gradient descent method of BP network makes the network easy to fall into local minimum and can't get the global optimal, and when the training patterns learn one by one, the network connection weights will be readjusted, this makes training process time longer.
3) The number of nodes in the hidden layer of network is usually determined by experiences, there is no exact theory instruction, which makes the design of the network model become more complex, the network training time can be longer.

Aiming at the limitations and disadvantages of BP network, the researchers spent a lot of energy in improving the performance of its research work, and put forward many improvements.

In the ultrasonic wave signal recognition and classification, using the steepest descent algorithm of BP network can make weights and threshold vector to get a stable solution, but the learning process is slow convergence, network easily trapped in a local minimum. At the same time due to the BP network is sensitive to learning rate, simply increase the learning rate to accelerate the convergence, the algorithm may be unstable and oscillating. Therefore, to solve these problems is very important. Thus we adopt additional momentum to solve.

On the basis of the back-propagation, each weight change pluses a value proportional to the previous weight change value, and in accordance with back propagation method to generate a new weight value conversion. With additional momentum factor weights value adjustment formula is:

$$w_{ij}(k+1) = w_{ij}(k) + \Delta w_{ij}(k+1) + \delta[w_{ij}(k) - w_{ij}(k-1)] \tag{13}$$

$$b_j(k+1) = b_j(k) + \Delta b_j(k+1) + \delta[b_j(k) - b_j(k-1)] \tag{14}$$

Where δ is the momentum factor, generally take around 0.95.

D. Feature disposing

We randomly selected 300 samples as the training set of the network, each type of defect signals selected 15 samples for testing set, that the number of samples Tested set was 75.

Now, in order to cancel the order of magnitude difference between the various dimensions of data, we convert sample data into [0,1] by the maximum and minimum normalization method, to avoid big network test errors caused by the magnitude of the difference between the input and output data. Maximum and minimum normalized form follows function:

$$x_n = \frac{x_n - x_{\min}}{x_{\max} - x_{\min}} \tag{15}$$

Where $x_{\min}$ represents the minimum value of the input data sequence, $x_{\max}$ is the maximum value of the sequence. According to the desired type of the signal identifies, we encode each category as the goal desired output vector, and the defect types code as shown in Table 1:

TABLE 1: DEFECT CATEGORY CODE DESIGN.

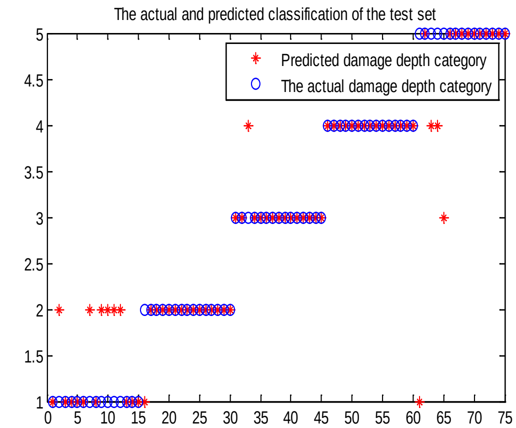
Parameters	Fault type（width(mm)*depth(mm)）				
	0.1*0.3	0.1*0.5	0.1*0.7	0.1*0.9	Non-destructive
Defect code	1 0 0 0 0	0 1 0 0 0	0 0 1 0 0	0 0 0 1 0	0 0 0 0 1

In the experiments, as the result of the existence of error, we specified the following rules, if the output is ‘ 0.9801 ,0.0050 ,0.0129 ,0.0020 ,4.141e-07 ’, it is equal to ‘1

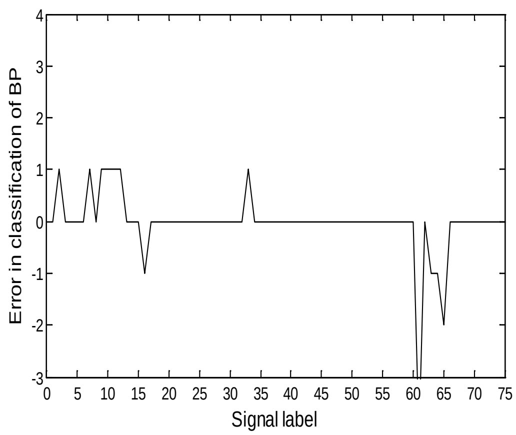
0 0 0 0', then we order it the first category, and so on.

E. Laser Ultrasonic Surface Wave Flaw Detection Experiments Contrast

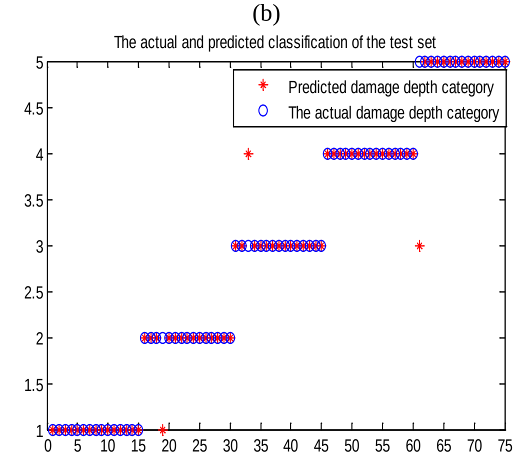
The number of neurons in the input layer is 24, one hidden layer and the number of the hidden layer neurons is 9, number of neurons in the output layer is 5. After the network is fully trained using a test set for testing, BP network and improved BP network results in the Figure 6 and Figure 7.



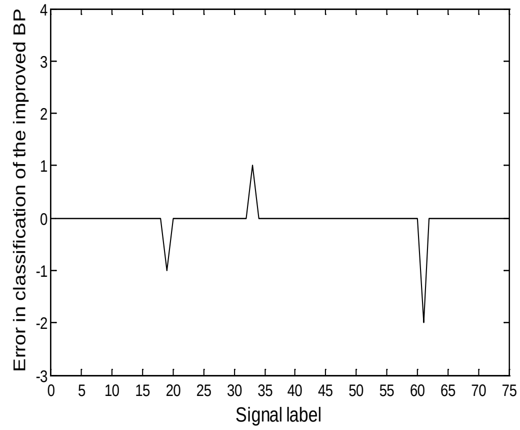
(a)



(b)



(c)



(d)

Fig. 6 Reflected wave experiment contrast. (a) BP network classification results of reflected wave. (b) Error of BP network classification. (c) The improved BP network classification results of reflected wave. (d) Error of the improved BP network classification.

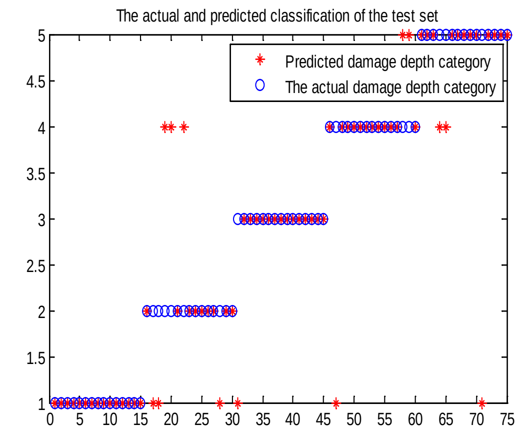
In the Figure 6, (a) represents BP network classification results of reflected wave, and the red '*' is the predicted damage depth category, the blue 'O' is the actual damage depth category; (b) represents error in classification of BP network; (c) represents the improved BP network classification results of reflected wave, and the red '*' is the predicted damage depth category, the blue 'O' is the actual damage depth category; (d) represents error in classification of the improved BP network.

TABLE 2: THE RESULT OF TWO KINDS OF NETWORKS TEST REFLECTED WAVE SIGNAL.

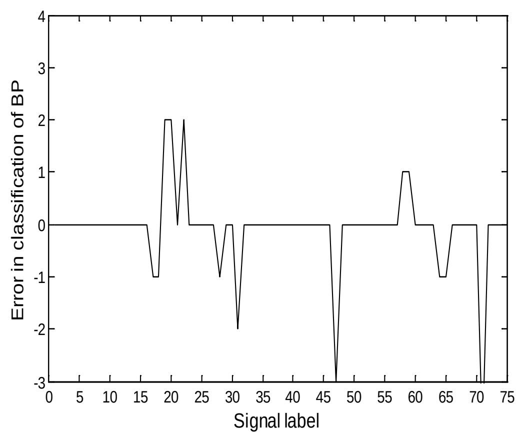
Identification method	Number of training samples	Test samples	Identify the correct number	The correct rate
BP network	300	75	63	0.8400
Improved BP network	300	75	72	0.9600

By the TABLE 2, we can see that for the reflected wave, the number of test set samples is 75, the correct identification of BP network number is 63, the correct identification rate is 84%, while the correct identification of the improved BP neural network number is 72, right recognition rate is 96%.Both methods recognition rate are in excess of 80%.

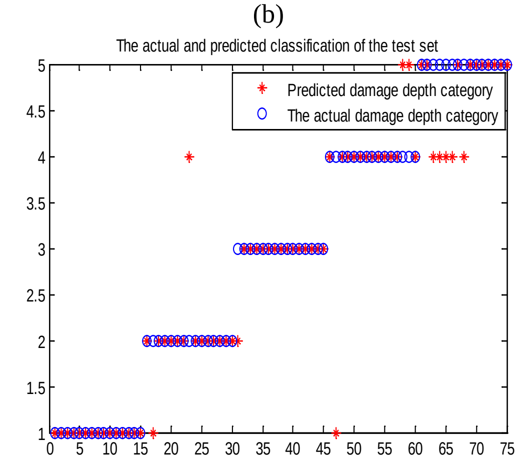
In the Figure 7, (a) represents BP network classification results of transmitted wave, and the red '*' is the predicted damage depth category, the blue 'O' is the actual damage depth category; (b) represents error in classification of BP network; (c) represents the improved BP network classification results of transmitted wave, and the red '*' is the predicted damage depth category, the blue 'O' is the actual damage depth category; (d) represents error in classification of the improved BP network.



(a)



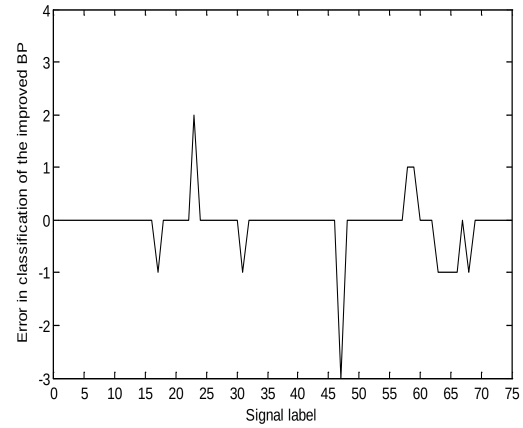
(b)



(c)

TABLE 3: THE RESULT OF TWO KINDS OF NETWORKS TEST TRANSMITTED WAVE SIGNAL.

Identification method	Number of training samples	Test samples	Identify the correct number	The correct rate
BP network	300	75	62	0.8267
Improved BP network	300	75	64	0.8533



(d)

Fig.7 Transmitted wave experiment contrast. (a) BP network classification results of transmitted wave. (b) Error of BP network classification. (c) The improved BP network classification results of transmitted wave. (d) Error of the improved BP network classification

As can be seen from the TABLE 3, for the transmitted wave, the number of test set samples is 75, the correct identification of BP network number is 62, the correct identification rate is 82.67%, while the correct identification of the improved BP neural network number is 64, right recognition rate is 85.33%. The correct recognition rate of two networks for the transmission wave is lower than reflected waves, but the recognition rate of both over 80%.

From the above two kinds of networks for reflected wave and transmission wave flaw detection experimental results, it can be seen that the two methods both can be effectively used for defect detection, and the improved BP network performance is better.

After the network is fully trained, we take one group of samples for forecasting, the prediction results are shown in Table 4. As can be seen, the actual output of the network is very consistent with the expected output, and the diagnostic accuracy is 100%. This proofs that the neural network model is reliable and can accurately defect types of ultrasonic signal for effective identification and classification. Our experiments show, BP and improved BP network can achieve effective detection of different types of defects. However, in general, the improved BP network is much more precise.

V WHY USE NNRS MODEL AND WHAT IS NNRS?

In the above discussion, we concluded that the BP network can accurately classify defect types of ultrasonic signal. But small samples tested, the method can't indicate that it is also fit for a large number of samples. In the further work, it has confirmed that the BP network does not apply to mass samples of defect classification, regardless of which kind of its accuracy is not high.

In fact, BP algorithm has been undesirable training method for several layers as a typical algorithm of traditional training multi-layer network. The main source of training hard is local minimum is common in the nonconvex target cost function of depth structure(involving multiple nonlinear

TABLE 4: SAMPLE PREDICTION RESULTS.

Experiment Type	Defect types	Target	Prediction	Correct classify-cation rate
BP Reflected wave	0.1*0.3	1 0 0 0 0	0.9801 0.0050 0.0129 0.0020 4.141e-07	100%
	0.1*0.5	0 1 0 0 0	0.0101 0.8871 0.0573 3.336e-04 0.0451	
	0.1*0.7	0 0 1 0 0	0.1444 0.2436 0.4685 0.0014 0.1421	
	0.1*0.9	0 0 0 1 0	0.0283 0.0318 0.0554 0.8603 0.0242	
	Non-destructive	0 0 0 0 1	0.0931 0.0767 0.0082 0.0191 0.8029	
Improved BP Reflected wave	0.1*0.3	1 0 0 0 0	0.8025 0.0131 0.0699 0.0728 0.0417	100%
	0.1*0.5	0 1 0 0 0	0.2539 0.7109 0.0026 0.0269 0.0057	
	0.1*0.7	0 0 1 0 0	0.0244 0.0401 0.8348 0.0881 0.0126	
	0.1*0.9	0 0 0 1 0	0.0571 0.1036 0.0162 0.7494 0.0737	
	Non-destructive	0 0 0 0 1	0.0115 0.1114 0.1263 0.0709 0.6799	
BP Transmitted wave	0.1*0.3	1 0 0 0 0	0.9801 0.0050 0.0129 0.0020 4.141e-07	100%
	0.1*0.5	0 1 0 0 0	0.0023 0.9915 0.0016 0.0015 0.0031	
	0.1*0.7	0 0 1 0 0	0.0074 0.0121 0.8427 0.1379 2.941e-10	
	0.1*0.9	0 0 0 1 0	5.999e-14 0.0002 3.406e-09 0.9885 0.0114	
	Non- destructive	0 0 0 0 1	2.419e-08 0.0003 4.073e-12 0.0001 0.9995	
Improved BP Transmitted wave	0.1*0.3	1 0 0 0 0	0.4033 0.0478 0.2090 0.0598 0.2801	100%
	0.1*0.5	0 1 0 0 0	0.1734 0.4095 0.1947 0.1759 0.0465	
	0.1*0.7	0 0 1 0 0	0.0263 0.1533 0.4429 0.1209 0.2566	
	0.1*0.9	0 0 0 1 0	9.490e-04 0.0306 0.0961 0.5959 0.2764	
	Non- destructive	0 0 0 0 1	0.1714 0.0047 6.074e-04 0.0974 0.8973	

processing unit)

The problems existing in the BP algorithm:

1) Gradient sparser: from the top down, the error correction signal is smaller and smaller;
2) Converge to local minimum: especially from the beginning of time away from the optimal region (random value initialization will lead to this situation).

The study found that the neural network regime switching(NNRS) can have a good performance in large sample defect recognition. Before introduce NNRS, we illustrate two problems, jump connections and deep learning.

A. Jump Connections

One alternative to the pure feedforward network or sieve network is a feedforward network with jump connections, in which the inputs X have direct linear links to output Y, as well as to the output through the hidden layer of squashed functions. Figure 8 pictures a feedforward jump connection network with n inputs, one hidden layer, and two neurons.

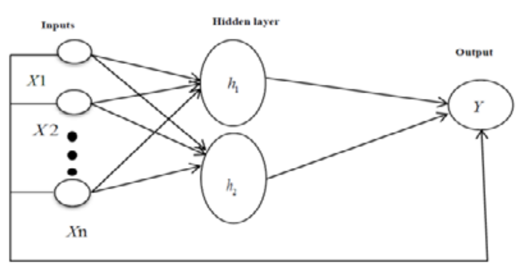


Fig.8 Feedforward neural network with jump connections

The mathematical representation of the feedforward network pictured in Figure 5, for logsigmoid activation functions, is given by the following system:

$$h_j = \sum_{i=1}^{n} w_{ji} x_i + w_{j0} \quad (16)$$

$$H_j = \frac{1}{1+e^{-h_j}} \quad (17)$$

$$y_k = \sum_{i=1}^{n} \beta_i x_i + \sum_{j=1}^{2} \lambda_j H_j + \lambda_0 \quad (18)$$

In this system there are n input variables $x_i, i = 1,2,...,n$, and $j = 1,2.$ neurons, with the coefficient vector or set of input weights w_{ji}, as well as the constant term, w_{j0}, form the variable h_j. This variable is squashed by the logistic function, and becomes a neuron H_j. The set of j neurons are combined in a linear way with the coefficient vector $\{\lambda_j\}$, $j = 1,2.$, and taken with a constant term λ_0, to form the forecast $y_k, k = 1,...,m$.

Note that the feedforward network with the jump connections increases the number of parameters in the network by j, the number of inputs. An appealing advantage of the feedforward network with jump connections is that it nests the pure linear model as well as the feedforward neural network. It allows the possibility that a nonlinear function may have a linear component as well as a nonlinear component. If the underlying relationship between the inputs and the output is a pure linear one, then only the direct jump connectors, given by the coefficient set $\{\beta_i\}, i = 1,...,n$, should be significant. However, if the true relationship is a complex nonlinear one, then one would expect the coefficient sets $\{w\}$ and $\{\lambda\}$ to be highly significant, and the coefficient set $\{\beta\}$ to be relatively insignificant. Finally, the relationship between the input variables $\{x\}$ and the output variable $\{y\}$ can be decomposed into linear and nonlinear components, and then we would expect all three sets of coefficients, $\{\beta\}$, $\{w\}$, and $\{\lambda\}$, to be significant.

A practical use of the jump connection network is as a useful test for neglected nonlinearities in a relationship between the input variables x and the output variable y. In this vein, we can also estimate a partitioned network. We first do linear least squares regression of the dependent variable y on the regression, x, and obtain the residuals, e. We then set up a feedforward network in which the residuals from the linear regression become the dependent variable, while we use the same regression as the input variables for the network. If there are indeed neglected nonlinearities in the linear regression, then the second-stage, partitioned network should have significant explanatory power. Of course, the jump connection network and the partitioned linear and feedforward network should give equivalent results, at least in theory. However, as we discuss in the next section, due to problems of convergence to local rather than global optima, we may find that the results may be different, especially for networks with a large number of regression and neurons in one or more hidden layers.

B. *Deep Learning*

The concept of deep learning is a new way of training multi-layer neural network. The optimization difficulty associated with the deep models can be alleviated. Containing many hidden layers of multilayer perceptron is a kind of deep learning structure. Deep learning represents the attribute category or feature by combining low-level features to form more abstract high-level characteristics to discover distributed characteristic presentation of data. Below is a simple deep learning model, depth = 3, using the hierarchy similar to neural network. The system is multi-layer network including input layer, the hidden layer (multilayer), the output layers; only connection between adjacent layers of nodes, no connection between each other and cross-layer node or in the same layer, each layer can be seen as a logistic regression model.

C. *NNRS Model*

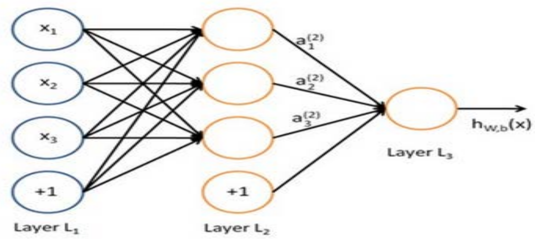


Fig.9A simple deep learning model, depth = 3

And then we use NNRS for defect classification. NNRS is a kind of regime switching model based on depth of learning with jump connections, and use logistic regression model to replace weights adjustment of two hidden layer. NNRS as shown in the Figure 10:

One way to model a regime switching framework with

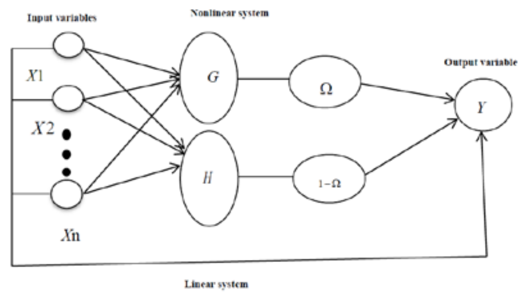


Fig.10NNRS model

neural networks is to adapt the feedforward network with jump connections. In addition to the direct linear links from the inputs or regressors x to the dependent variable y, holding in all states, we can model the regime switching as a jump-connection neural network with one hidden layer and two neurons, one for each regime. These two regimes are weighted by a logistic connector which determines the relative influence of each regime or neuron in the hidden layer. This system appears in the following equations:

$$y_j = \alpha x_i + \beta\{[\Omega(y_{j-1};\theta,p)]G(x_i;\alpha_1) + [1-\Omega(y_{j-1};\theta,p)]H(x_i;\alpha_2)\} + \eta_i \quad (19)$$

where x_i is the vector of independent variables, and α represents the set of coefficients for the direct link.

$$\eta_i = \varepsilon_i + \lambda(L)\varepsilon_{i-1},\ \varepsilon_i \sim N(0,\sigma^2) \quad (20)$$

Where η_i is a disturbance term, $\lambda(L)$ are lag operators.

The functions $G(x_i;\alpha_1)$ and $H(x_i;\alpha_2)$, which capture the two regimes, are logsigmoid and have the following representations:

$$G(x_i;\alpha_1) = \frac{1}{1+\exp(-\alpha_1 x_i)} \quad (21)$$

$$H(x_i;\alpha_2) = \frac{1}{1+\exp(-\alpha_2 x_i)} \quad (22)$$

The transition function Ω which determines the influence of each regime or state depends on the value of y_{j-1} as well as a smoothness parameter vector θ and a threshold parameter p, with $p=0$. We use a logistic or logsigmoid specification for $\Omega(y_{j-1};\theta,p)$.

$$\Omega_j = \Omega(\theta \cdot y_{j-1} - p) = \frac{1}{1+\exp(\theta \cdot y_{j-1} - p)} \tag{23}$$

Of course, we can also use a cumulative Gaussian function instead of the logistic function. Measures of Ω are highly useful, since they indicate the likelihood of continuing in a given state. This model, of course, can be extended to multiple states or regimes.

D. Experiment and Result

In experiment of BP, we only test the first group sample, without repetition. Now, we test all data collected. As previously described, each defect signal extracts 24 features, sample length is 75, the five groups of defect signals will turn into the sample characteristics 1875 * 24. Then, the input $X=[x_1,x_2,...,x_{24}]$ is 1875 * 24, the output Y is 1875*1, is the category label of X. The category label rules show in table 5. We choose the last set of test data for forecast and are very excited to get the following results.

As the table 6 showing, NNRS can well predict two waves.

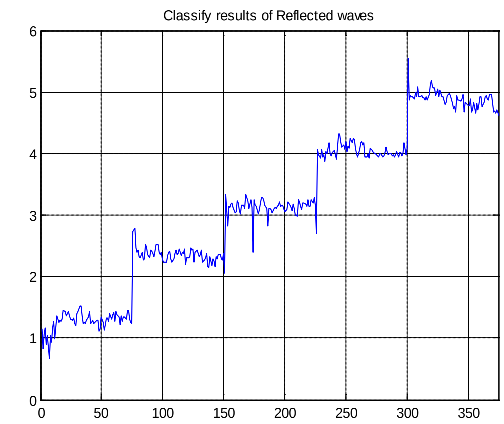
TABLE 5: THE CATEGORY LABEL RULES

Para-meters	Fault type（width(mm)*depth(mm)）				
	0.1*0.3	0.1*0.5	0.1*0.7	0.1*0.9	Non-destructive
label	1	2	3	4	5

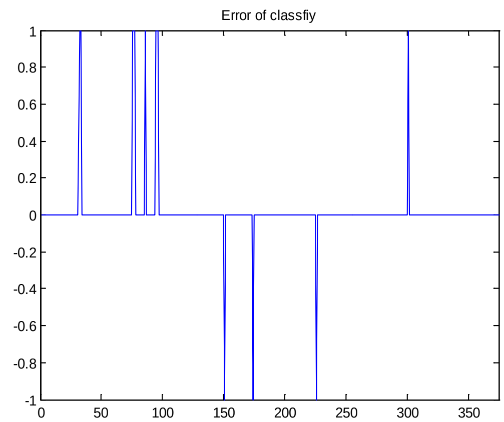
TABLE 6: CORRECT IDENTIFICATION RATE OF NNRS FOR TWO KINDS OF WAVE

WAVE	CORRECT RATE
REFLECTED	97.07%
TRANSMITTED	94.93%

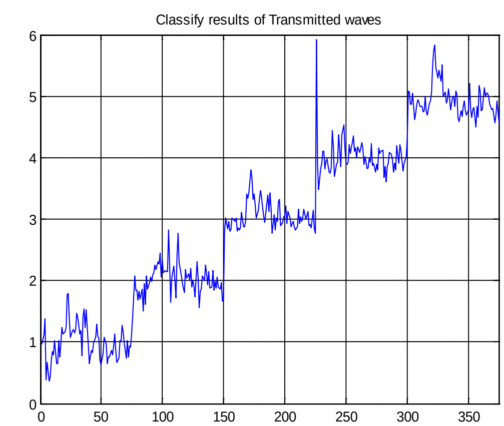
For the reflected waves, the correct identification rate is 97.07%; for the transmitted waves, the correct identification rate is 94.93%. Both the recognition rate of two waves over 90%, this means NNRS can be used to classify the ultrasonic defect signals and identify 5 kinds of defect signals. We also can say the diagnostic accuracy is 100%. In order to confirm our results, we show them in Figure 11. It is easy to see the results; five kinds of signals are obviously separated. And the errors of two experiments also are demonstrated, they are acceptable. Figure 11 is experiment results of NNRS model, where (a) represents classification results of reflected waves, Y-axis is the category label; (b) represents error of reflected wave classification. We define the output rule as below: the round value of real output y is the result of the final desired; (c) represents classification results of transmitted waves; (d) represents error of transmitted wave classification.



(a)



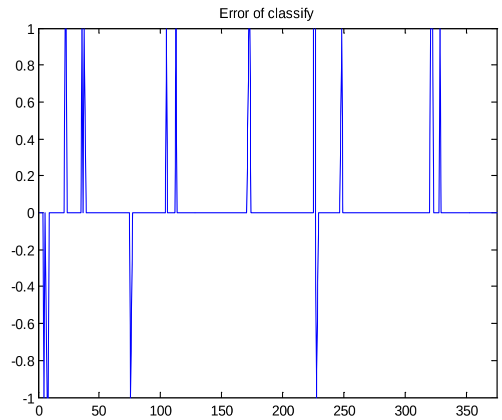
(b)



(c)

VI EXPERIMENTAL COMPARISION

In the part Ⅳ and Ⅴ, we respectively discuss two models. The experiment results of models have showed in figures and tables. The results of BP neural network, the testing accuracies are over 80%, the diagnostic accuracy is over 100%. However, we can't say the way is valid. The experiment data only account for 20% of the total. NNRS use all data, and the correct identification rates are as high as 90%. The diagnostic accuracy is 100%. We summarize NNRS is suitable for laser ultrasonic signal features disposing. NNRS is more suitable for practical application and can availably classify five kinds of signals.



(d)

Fig.11 Experiment results of NNRS model. (a) Classification results of Reflected waves. (b) Error of reflected wave classification. (c) Classification results of transmitted waves. (d) Error of transmitted wave classification

VII CONCLUSION

Our purpose is to classify the ultrasonic defect signals and searching for a way can identify 5 kinds of defect signals. In this paper, we extract the characteristic coefficients of ultrasonic reflected and transmitted wave signals with Mel Cepstral method, and successfully reduce the sample data dimension from 10,000*25 to 1875*24. The method has no need to consider temporal characteristics and frequency domain characteristics. We respectively dispose the features with neural network and NNRS. The experiments show that BP neural network has limitations. Small sample test is not suitable for practical application. NNRS is better than BP neural network in classifying ultrasonic flaw. Because NNRS is based on deep learning with jump connection, and turn weights updating to logistic regression. This makes it more close to the complex structure of human brain. As we know, human brain has complex nonlinear processing, so the model makes the result more scientific and credible. We still can use the method to explore and research more practical problems.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests regarding the publication of this paper.

ACKNOWLEDGEMENTS

This work was financially supported by the National Natural Science Foundation of China (61275120).

REFERENCES

[1] Gang Li. Laser ultrasonic technology and its application in non-destructive testing of metal [D]. Central China Normal University, 2004.

[2] Zhang, Yanhua. Study on feature extraction and classification of ultrasonic flaw signals[J].WSEAS Transactions on Mathematics,2010,7(9):529-538

[3] Astuti, W. Time domain feature extraction technique for earth's electric field signal prior to the earthquake[J].IOP Conference Series: Materials Science and Engineering,2013,1

[4] Yuan, Yu. Signal feature extraction base on factral dimensions of time-frequency domain[J].Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument,2009,10(30):39-42

[5] Yichuan Wang, Zhizhong Li. Ship target classification based on Mel Cepstral and BP neural network[J]. Sensors and Microsystems, 2011,06: 55-57 + 67.

[6] Jiaohong Yi, Weihong Xu, Yuantao Chen. Novel Back Propagation Optimization by Cuckoo Search Algorithm [J].The Scientific World Journal, 2014.

[7] Jianyong Liu, Huaixiao Wang, Yangyang Sun, Chengqun Fu, JieGuo. Real-Coded Quantum-Inspired Genetic Algorithm-Based BP Neural Network Algorithm [J]. Mathematical Problems in Engineering , 2015

[8] Chaoqin Peng, Chun Cao, Jiaoying Huang, Qiusheng Liu. Seismic signal recognition using improved BP neural network and combined feature extraction method[J]. Journal of Central South University,2014,05:1898-1906.

[9] Xiaochuan Wang, Feng Shi, Yang Li. The analysis of 43 cases in MATLAB Neural Network[M]. Beijing: Beijing University of Aeronautics and Astronautics Press,2013.8.

[10] Junqin He, Ting Jiang, Zhihao Xing. A method of target detection and identification based on RPROP and UWB channel characteristic parameters, *Globecom Workshops (GC Wkshps), 2012 IEEE,* On page(s): 1460 – 1463.

[11] Shusen Zhou, Qingcai Chen, Xiaolong Wang. Active deep learning method for semi-supervised sentiment classification[J]. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings,2014:* 1626-1630.

[12] Juan Fombellida, Santiago Torres-Alegre, Juan Antonio Piñuela-Izquierdo, Andina, Diego. Artificial metaplasticity for deep learning: Application to WBCD breast cancer database classification[J]. *Lecture Notes in Computer Science,2015:* 399-408.

[13] Schmidhuber, Jürgen.Deep Learning in neural networks: An overview[J]. *Neural Networks, 2015, 1, 01:* 85-117.

[14] Deng, Li, Platt, John C. Ensemble deep learning for speech recognition[J]. *Proceedings of the Annual Conference of the International Speech Communication Association, 2014*：1915-1919.

[15] Yu, Dong; Deng, Li; Jang, Inseon; Kudumakis, Panos; Sandler, Mark; Kang, Kyeongok. Deep learning and its applications to signal and information processing[J]. *IEEE Signal Processing Magazine,2011,1(1):* 145-149+154.

[16] Hamilton, James D. A New Approach to the Economic Analysis of Nonstationary Time Series Subject to Changes in Regime[J]. Econometrica,1989(57):357-384.

[17] Martinez, Angel Mario Castro ; Moritz, Niko; Meyer, Bernd T. Should Deep Neural nets have ears? The role of auditory features in deep learning approaches[J]. *Proceedings of the Annual Conference of the International Speech Communication Association,2014:* 2435-2439.

[18] Deng, Li ; Yu, Dong.Deep learning: Methods and applications[J]. *Foundations and Trends in Signal Processing*,2013(3): 197-387.