IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 27, 2019, accepted June 4, 2019, date of publication June 14, 2019, date of current version July 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923242

Research on ELM Soft Fault Diagnosis of Analog Circuit Based on KSLPP Feature Extraction

GAN XU-SHENG[1], QU HONG[1], MENG XIANG-WEI[1], WANG CHUN-LAN[2], AND ZHU JIE[1]
[1]Air Traffic Control and Navigation College, Air Force Engineering University, Xi'an 710051, China
[2]Xijing University, Xi'an 710123, China

Corresponding author: Gan Xu-Sheng (gxsh15934896556@qq.com)

ABSTRACT In order to improve the capability of soft fault diagnosis in an analog circuit, an integrated diagnosis method based on KSLPP feature extraction and ELM is proposed. The KSLPP feature extraction ability is firstly used to construct the principal component feature set from the fault sample set. Then, the advantage of ELM on solving the complicated nonlinearity problem is applied to build the fault identification model from the principal component feature. Finally, the sample sets of soft fault diagnosis for the analog circuit are respectively established by waveform parameter and wavelet packet transform to conduct the diagnostic test for the built model. The simulation of Elliptic Filter shows that, based on the fault sample set gotten by waveform parameter, the total correct rate of the integrated method is 98.2%; on the basis of the fault sample set built by wavelet packet transform, the total correct rate is 100%. The feasibility and effectiveness of this method are also validated.

INDEX TERMS Extreme learning machine, Kernel supervised locality preserving projection, feature extraction, analog circuit, fault diagnosis.

I. INTRODUCTION

With the development of computer and semiconductor technology, the internal structure of analog circuit and the manufacturing process becomes more complex and more intensive. Naturally, this also requires higher and higher reliability. However, the problem of soft fault caused by element tolerance etc. is usually paid more attention. Such fault may make the circuit deteriorate or even the system collapse, and the fuzziness and uncertainty of soft fault often make it difficult to detect and diagnose. This is also the technical bottleneck of the high cost of circuit chips and the development of integrated circuits, and become the hot spot in analog circuit research.

In 1962, Berkowitz first began to study problems related to fault diagnosis for analog circuit and gave the concept of element parameter solvability [1]. Subsequently, Navid et al studied the solvability of linear resistance value and discussed its sufficient conditions [2], providing the theoretical basis for actual solution of elements in analog circuit, which is of landmark significance. At this stage, the fault dictionary method and the parameter identification method are the main diagnosis methods. These methods are more suitable for single fault, but the amount of calculation is too large when dealing with multiple faults. After 1980, the focus of the research shifted from solving all element values to solving some element values, and the fault diagnosis was divided into two parts: one was fault location, other was fault element solution based on fault location [3]–[5] This can reduce the amount of calculation and weaken the relevant conditions. During this period, the failure element delimitation method and k-fault diagnosis method are more typical fault diagnosis methods. In 1984, Salama *et al.* introduced the idea of network decomposition in the study of fault diagnosis for large-scale analog circuit, and located the fault to the smallest subnetwork [6], but the tearing characteristic of its relation nodes makes it unfeasible.

In 1997, Spina proposed the black-box modeling idea using neural network in the study of analog circuit fault diagnosis, that is, use the self-learning and self-adaptive ability of neural network to automatically realize fault diagnosis [7], which not only has a high recognition rate, but also a good generalization ability. Great attention has been paid in this field and many achievements have been made. However, such defects as slow speed, over-fitting and local convergence in its training still cannot be effectively solved [8]. In 2006, Huang et al proposed a new Single Hidden Layer Feed Forward Neural Networks (SLFNs), namely Extreme

The associate editor coordinating the review of this manuscript and approving it for publication was Valentina E. Balas.

Learning Machine (ELM), which can theoretically provide good generalization ability and extremely fast training speed [9], [10]. However, although ELM performs well in most cases, the features cannot be extracted before modeling. However, the soft fault samples of analog circuits have multidimensional input variables, which not only contain different levels of noise, but also may be correlated with each other. This may have a negative impact on improving the diagnostic accuracy. Therefore, it is not only necessary but also important to study feature extraction before soft fault ELM modeling of analog circuits.

Based on above analysis, a soft fault diagnosis method integrating Kernel Supervised Locality Preserving Projection (KSLPP) [11] with ELM is proposed for analog circuits, and its diagnosis performance for different sample set construction is studied.

The rest of the paper is organized as follows. Section 1 gives the introduction to KSLPP feature extraction technology. The basic principle and learning algorithm on ELM are described in Section 2. The realization steps of integrating KSLPP with ELM are given in Section 3. Section 4 elaborates the waveform parameter method and wavelet packet transform method for sample set construction. The experiments and their result analysis are presented in Section 5. The last section gives some concluding remarks.

II. KSLPP FEATURE EXTRACTION

A. BASIC PRINCIPLE OF KERNEL METHOD

Definition of kernel function: suppose that x_i, x_j are the vectors in the space $\boldsymbol{X}$, and $\Phi(\cdot)$ is the nonlinear function mapping the variables in input space $\boldsymbol{X}$ into the feature space. If the function $\boldsymbol{K}_{i,j} = \langle \Phi(x_i), \Phi(x_j) \rangle$ satisfies Mercer's condition, then $\boldsymbol{K}_{i,j}$ can be called kernel function (the symbol $\langle\ \rangle$ denotes inner product) [12], [13]. In generally, the kernel function method is a kind of method using kernel function technology. The basic principle of kernel function method: the input space is firstly mapped into the high-dimensional feature space (it can be infinite dimension) through the function $\Phi(\cdot)$, and then the processing such as classification or regression etc. is carried out in the feature space. The key is to transform the inner product calculation in high-dimensional feature space into the kernel function calculation in original input space after the nonlinear transformation through the introduction of kernel function, namely, it is achieved by $\boldsymbol{K}_{i,j} = \langle \Phi(x_i), \Phi(x_j) \rangle$. Thus, the linear calculation in the feature space is the nonlinear calculation corresponding to the original input space [14], [15]. The basic principle of kernel function method is shown in Figure 1.

The kernel function is essentially an inner product (dot product), and is also the basis of kernel technique. In the kernel function method, the calculation does not involve the dimension of feature space without specifying the form and parameters of the mapping function $\Phi(\cdot)$. The different kernel functions correspond to the different nonlinear mapping functions. The combination of kernel function method and different algorithms can form the kernel function algorithm which can solve different problems more effectively.

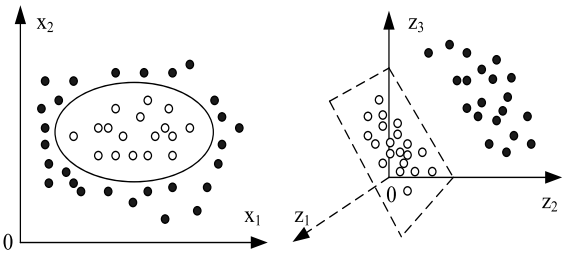


FIGURE 1. Schematic diagram of kernel function method.

B. KERNEL SLPP ANALYSIS

Supervised Locality Preserving Projection (SLPP) is a linear feature extraction method based on modified Locality Preserving Projection (LPP) [16], [17]. KSLPP is extended version of SLPP based on kernel function in formula (1), which can deal with nonlinear problem through kernel mapping idea.

$$\mathbf{K}_{i,j} = \langle \Phi(x_i),\ \Phi(x_j) \rangle = \exp(-\frac{||x_i - x_j||^2}{2q^2}) \tag{1}$$

where $\boldsymbol{\Phi}(\cdot)$ is the nonlinear mapping function, q is the kernel width parameter, x_i or $x_j \in \boldsymbol{X}(\boldsymbol{X} = \{x_i | i = 1, 2, \cdots, n, x_i \in R^d\})$.

The objective function defined by KSLPP is as follows

$$\max \frac{\sum\limits_{ij} \left(y_i - y_j\right)^2 B_{ij}}{\sum\limits_{ij} \left(y_i - y_j\right)^2 W_{ij}} \tag{2}$$

where y_i or $y_j \in \boldsymbol{Y}\left(\boldsymbol{Y} = \left\{y_i | i = 1, 2, \cdots, n, y_i \in R^l\right\}\right)$, and $\boldsymbol{Y} = \alpha^{\mathbf{T}} \Phi(\boldsymbol{X})$, α is the projection vector. Formula (3) is another transform of Formula (2)

$$\begin{aligned}
&\max \frac{\frac{1}{2}\sum\limits_{ij} \left(y_i - y_j\right)^2 B_{ij}}{\frac{1}{2}\sum\limits_{ij} \left(y_i - y_j\right)^2 W_{ij}} \\
&= \frac{\frac{1}{2}\sum\limits_{ij} \left(\alpha^{\mathrm{T}}\Phi\left(x_i\right) - \alpha^{\mathrm{T}}\Phi\left(x_j\right)\right)^2 B_{ij}}{\frac{1}{2}\sum\limits_{ij} \left(\alpha^{\mathrm{T}}\Phi\left(x_i\right) - \alpha^{\mathrm{T}}\Phi\left(x_j\right)\right)^2 W_{ij}} \\
&= \frac{\sum\limits_{ij} \alpha^{\mathrm{T}}\Phi\left(x_i\right) D_{ii}^B \Phi\left(x_i\right)^{\mathrm{T}} \alpha - \sum\limits_{ij} \alpha^{\mathrm{T}}\Phi\left(x_i\right) B_{ij} \Phi\left(x_j\right)^{\mathrm{T}} \alpha}{\sum\limits_{ij} \alpha^{\mathrm{T}}\Phi\left(x_i\right) D_{ii}^W \Phi\left(x_i\right)^{\mathrm{T}} \alpha - \sum\limits_{y} \alpha^{\mathrm{T}}\Phi\left(x_i\right) W_{ij} \Phi\left(x_j\right)^{\mathrm{T}} \alpha} \\
&= \frac{\alpha^{\mathrm{T}}\Phi(\boldsymbol{X})\left(\boldsymbol{D}^B - \boldsymbol{B}\right)\Phi(\boldsymbol{X})^{\mathrm{T}}\alpha}{\alpha^{\mathrm{T}}\Phi(\boldsymbol{X})\left(\boldsymbol{D}^W - \boldsymbol{W}\right)\Phi(\boldsymbol{X})^{\mathrm{T}}\alpha} \\
&= \frac{\alpha^{\mathrm{T}}\Phi(\boldsymbol{X})\boldsymbol{L}^B\Phi(\boldsymbol{X})^{\mathrm{T}}\alpha}{\alpha^{\mathrm{T}}\Phi(\boldsymbol{X})\boldsymbol{L}^W\Phi(\boldsymbol{X})^{\mathrm{T}}\alpha}
\end{aligned} \tag{3}$$

where, $\boldsymbol{L}^B = \boldsymbol{D}^B - \boldsymbol{B}, \boldsymbol{L}^W = \boldsymbol{D}^W - \boldsymbol{W}, \boldsymbol{D}^B$ and $\boldsymbol{D}^W$ are diagonal matrices, $\boldsymbol{D}^B$ is the sum of the column or row elements

corresponding to similarity matrix $\boldsymbol{B}$ between classes, $\boldsymbol{D}^{\boldsymbol{W}}$ is the sum of the column or row elements corresponding to the similarity matrix $\boldsymbol{W}$ within classes. Formula (3) can also be expressed as the following generalized eigenvalue problem

$$\boldsymbol{K}\boldsymbol{L}^{B}\boldsymbol{K}^{\mathrm{T}}\alpha = \lambda\boldsymbol{K}\boldsymbol{L}^{W}\boldsymbol{K}^{\mathrm{T}}\alpha \tag{4}$$

The realization steps of KSLPP algorithm is:

1. The observation matrix $\boldsymbol{X}$ is normalized and expressed as $\boldsymbol{X}^{b}$

2. $\boldsymbol{X}^{b}$ is mapped into a certain feature space by kernel function $\boldsymbol{K}(\cdot)$, getting $\boldsymbol{\Phi}(\boldsymbol{X}^{b})$.

3. On the basis of the data set $\boldsymbol{\Phi}(\boldsymbol{X}^{b})$, the adjacency graphs within classes and between classes are constructed by k-nearest neighbor method. Suppose the adjacency graph G contains n nodes, the i-th node corresponds to $\boldsymbol{\Phi}(x_i)$, k nearest neighbors $\boldsymbol{\Phi}(x_j)$ of the i-th node within class and between classes are searched respectively, and the node i and j are connected.

4. The between-class weight matrix $\boldsymbol{B}$ and the within-class weight matrix $\boldsymbol{W}$ can respectively be constructed by the following formula

$$B_{ij} = \begin{cases} S_{ij}, & \text{connection between } \Phi(x_i) \text{ and } \Phi(x_j) \\ 0, & \text{other} \end{cases}$$

$$W_{ij} = \begin{cases} S_{ij}, & \text{connection between } \Phi(x_i) \text{ and } \Phi(x_j) \\ 0, & \text{other} \end{cases}$$

where S_{ij} is the element of similarity matrix $\boldsymbol{S}$, and can be obtained by the following formula

$$S_{ij} = \exp\left(-\frac{\left\|\Phi\left(x_i\right) - \Phi\left(x_j\right)\right\|^2}{t}\right) \tag{5}$$

where t is the constant.

5. The generalized eigenvalues and their eigenvectors of Formula (4) are calculated, getting the descending eigenvalue $\lambda_1, \lambda_2, \cdots, \lambda_H$ and their eigenvector $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_H$. The eigenvector corresponding to the first l eigen-values is taken to construct $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_l\}$. Then the projection of $\boldsymbol{\Phi}(\boldsymbol{X}^{b})$ onto α is the low-dimensional feature set to be solved

$$\boldsymbol{Y} = \alpha^{\mathrm{T}}\Phi\left(\boldsymbol{X}^{b}\right) \tag{6}$$

III. EXTREME LEARNING MACHINE

A. BASIC PRINCIPLES OF ELM

ELM is a new neural network with unique form and excellent performance. It can generate all parameters of hidden layer in a random way, and considers the balance between recognition accuracy and algorithm extensibility [9]. At present, it has been widely used in various research fields.

For N-dimensional sample $\left\{\boldsymbol{x}_j, \boldsymbol{t}_j\right\} \in R^N \times R^n$, the output of SLFN can be expressed as

$$y_j = \sum_{i=1}^{L} \beta_i g_i\left(w_i x + b_i\right) = \sum_{i=1}^{L} \beta_i h_i(x) \tag{7}$$

where L is the number of hidden nodes. y_j is the output of output node j. w_i is the connection weight vector between input node and hidden node. $g(x)$ denotes the excitation function of hidden layer. β_i is the connection weight vector between hidden node and output node. b_i is the threshold of the i-th hidden node. G($\cdot$) is the output function of hidden node. $h(x)$ is the output vector of hidden layer with respect to x.

SLFNs can approximate the training sample x, t with zero error, that is $\sum_{j=1}^{n} \|y_i - t_i\| = 0$, then the parameter w_i, b_i, β_i should meet $\sum_{i=1}^{L} \beta_i g(w_i x + b_i) = t_j, j = 1, 2, \cdots, n$

Suppose that the output matrix of hidden layer is $\boldsymbol{H}$, and the i-th column of $\boldsymbol{H}$ is $h_i = [g(w_{1i}x_1 + b_i), \cdots, g(w_{Ni}x_N + b_i)]^T$, then the following formula can be obtained

$$\boldsymbol{H}\beta = \boldsymbol{T} \tag{8}$$

where

$$\boldsymbol{H} = \begin{bmatrix} g\left(w_{11}x_1 + b_1\right) & \cdots & g\left(w_{1L}x_1 + b_L\right) \\ \cdots & \cdots & \cdots \\ g\left(w_{N1}x_{N1} + b_1\right) & \cdots & g\left(w_{NL}x_N + b_L\right) \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_1^{\mathrm{T}} \\ \vdots \\ \beta_L^{\mathrm{T}} \end{bmatrix}, \quad \boldsymbol{T} = \begin{bmatrix} t_1^{\mathrm{T}} \\ \vdots \\ t_n^{\mathrm{T}} \end{bmatrix}.$$

It has been proved that, if $g(x)$ is infinitely differentiable, the parameter w_j, b_j don't need to be adjusted repeatedly. When w_j, b_j are arbitrarily given, the unique solution $\boldsymbol{H}^{+}$ is solved by Moore-Penrose generalized inverse theorem, then, $\boldsymbol{\beta} = \boldsymbol{H}^{+}\boldsymbol{T}$, where $\boldsymbol{H}^{+}$ represents the generalized inverse matrix of $\boldsymbol{H}$.

B. ELM CLASSIFICATION ALGORITHM

Theoretically, the output of ELM can infinitely approximate the objective function. When there are enough hidden nodes, ELM can be used as a classifier. In order to make the hyperplanes $h(\boldsymbol{x})\beta$ classify correctly, the interval $2/\|\beta\|$ of binary classification should be maximized, namely, $\|\beta\|^2/2$ is minimized. In addition, if there is no training, the category label can be correctly positioned through the training samples, which can be described as the following optimization problem

$$\begin{aligned} &\min\ L_{\mathrm{ELM}} = \frac{1}{2}\|\beta\| + C\frac{1}{2}\sum_{i=1}^{N}\|\xi_i\|^2 \\ &\text{s.t. } h\left(\boldsymbol{x}_i\right)\beta = t_i^{\mathrm{T}} - \xi_i^{\mathrm{T}}, \quad \boldsymbol{i} = 1, 2, \cdots, N \end{aligned} \tag{9}$$

where ξ_i is the training error. C is the regularize- tion parameter corresponding to ξ_i.

Since KKT condition is satisfied, using Lagrange method, formula (9) can be converted into

$$L_{ELM} = \frac{1}{2}||\beta||^2 + \frac{1}{2}C\sum_{i=1}^{N}\xi_i^2 - \sum_{i=1}^{N}\sum_{j=1}^{n}\alpha_{ij}(h(\boldsymbol{x}_i))\beta_j - t_{ij} + \xi_{ij}) \quad (10)$$

where α_{ij} is Lagrange multiplier factor. In order to determine $\beta_j\alpha_i$ and ξ_i, make L maximize. Let $\partial L_{ELM}/\partial\beta_j = 0$, $\partial L_{ELM}/\partial\xi_i = 0$, $\partial L_{ELM}/\partial\alpha_i = 0$, we can obtain $\beta = \boldsymbol{H}^{\mathrm{T}}\left(\frac{\boldsymbol{I}}{C} + \boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right)^{-1}\boldsymbol{T}$, then the output equation of ELM can be written as

$$f(\boldsymbol{x}) = h(\boldsymbol{x})\beta = h(\boldsymbol{x})\boldsymbol{H}^{\mathrm{T}}\left(\frac{\boldsymbol{I}}{C} + \boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right)^{-1}\boldsymbol{T} \quad (11)$$

For binary classification problem, the decision- making equation of ELM is

$$f(\boldsymbol{x}) = \mathrm{sign}\left(h(\boldsymbol{x})\boldsymbol{H}^{\mathrm{T}}\left(\frac{\boldsymbol{I}}{C} + \boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right)^{-1}\boldsymbol{T}\right) \quad (12)$$

For S classification problem, the decision-making equation of ELM is

$$label(\boldsymbol{x}) = \arg\max_{i\in\{1,2,..,S\}} f_i(\boldsymbol{x}) \quad (13)$$

Compared with traditional classifiers, ELM greatly improves the training efficiency under the premise of higher classification accuracy, which makes it have obvious advantages in classification applications.

IV. INTEGRATED ALGORITHM OF KSLPP AND ELM

The integrated modeling idea of KSLPP and ELM: KSLPP is applied to extract the principal component feature, and then some features are selected as ELM model, establishing the integrated model of KSLPP and ELM as shown in Figure 2. The specific realization steps are as follows:

A. EXTRACTION OF PRINCIPAL COMPONENT FEATURE

a. The kernel function is selected to construct the training sample kernel matrix K and test sample kernel matrix K_t, and the centralized processing for K and K_t are conducted to get $\hat{K}$ and $\hat{K}_t$ respectively

b. KSLPP is used to extract the principal component matrix T from $\hat{K}$, and solve the projection vector α from $\hat{K}$.

c. The projection of $\hat{K}_t$ is calculated through $\boldsymbol{\alpha}$ to get the principal component matrix T_{test} of test sample.

B. MODEL TRAINING

T is taken as the training input to build the ELM model by optimizing network parameters.

C. VALIDATION TEST

T_{test} is taken as the test input to validate the built ELM model.

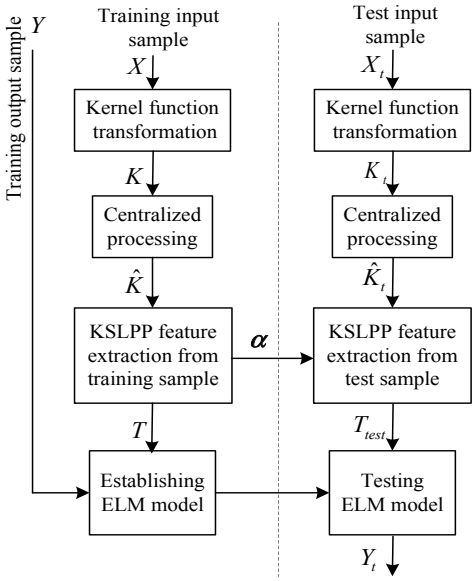


FIGURE 2. Integrated modeling of KSLPP- ELM.

V. CONSTRUCTION OF SOFT FAULT SAMPLE SET FOR ANALOG CIRCUIT

A. SAMPLE SET CONSTRUCTION BASED ON WAVE- FORM PARAMETER

As we all know, the greater the difference of fault features contained in the sample set, the better the training and diagnosis effect of the model will be. It is a key step of fault diagnosis for analog circuit to extract the intrinsic characteristics which can reflect the different faults of analog circuit. In actual use, the output response waveform of analog circuit changes with the drift of circuit element parameters. When different elements have soft faults, the output response waveform is also different. Therefore, some parameters of output waveform are extracted, and an appropriate number of parameters that can effectively distinguish the characteristics of different output waveforms are selected to form the input vector of KSLPP-ELM corresponding to each fault state. After repeated compare- sons, 9 waveform parameters are extracted as the feature vectors corresponding to each fault state, namely voltage maximum, low pass cutoff frequency (3dB), high pass cutoff frequency (3dB), bandwidth (3dB), bandwidth (4dB), center frequency (3dB), Q value (3dB), Q value (4dB) and Q value (5dB), as shown in Figure 3. These parameters are enough to reflect the inherent characteristics of the output waveform under each failure state, which is easy to train and test KSLPP-ELM model.

B. SAMPLE SET CONSTRUCTION BASED ON WPT

In fact, for many problems, it is only necessary to extract the information of points at specific time and frequency. At the point in specific frequency domain, the resolution of frequency domain needs to be improved as much as possible; at the point in specific time domain, the resolution of time domain needs to be improved as much as possible. But MRT

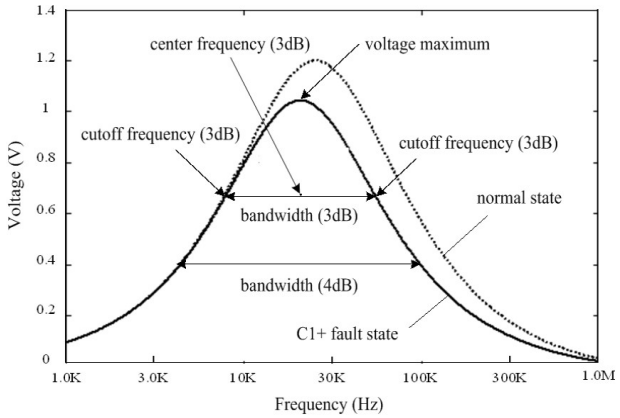


FIGURE 3. Schematic diagram of waveform parameter.

cannot satisfy the requirement. This is because MRT only decomposes V (scale) space, i.e. $V_0 = V_1 \oplus W_1 = V_2 \oplus W_2 \oplus W_1 = \cdots$, and doesn't further decompose W (space). WPT can further decompose W_j, which can satisfy the above requirements with more extensive application.

WPT, derived on the basis of wavelet transform, is a generalization of the wavelet concept [18], [19]. The definition of wavelet packet is given as follows:

For the given orthogonal scaling function $\phi(t)$ and the wavelet function $\psi(t)$, two-scale relationship

$$\phi(t) = \sqrt{2}\sum_k h_{0k}\phi(2t-k) \tag{14}$$

$$\psi(t) = \sqrt{2}\sum_k h_{1k}\psi(2t-k) \tag{15}$$

where h_{0k} and h_{1k} are the filter coefficients in MRT.

Then the recursive relationship can be defined

$$w_{2n}(t) = \sqrt{2}\sum_{k\in Z} h_{0k}w_n(2t-k) \tag{16}$$

$$w_{2n+1}(t) = \sqrt{2}\sum_{k\in Z} h_{1k}w_n(2t-k) \tag{17}$$

when $n = 0$, $w_0(t) = \phi(t)$, $w_1(t) = \phi(t)$. The function set $\{w_n(t)\}_{k\in Z}$ defined above is the wavelet packet determined by $w_0(t) = \phi(t)$.

For the signal $f(t)$, the recursive algorithm of its WPT coefficient is

$$\begin{cases} d_k^{j+1,2n} = \sum_l h_{0(2l-k)}d_k^{j,n} \\ d_k^{j+1,2n+1} = \sum_l h_{1(2l-k)}d_k^{j,n} \end{cases} \tag{18}$$

Then wavelet packet can be reconstructed by the following formula

$$\begin{aligned} d_k^{j,n} &= \sum_k [h_{0(l-2k)}d_k^{j+1,2n} + h_{1(l-2k)}d_k^{j+1,2n+1}] \\ &= \sum_k g_0(l-2k)d_k^{j+1,2n} + \sum_k g_1(l-2k)d_k^{j+1,2n+1} \end{aligned} \tag{19}$$

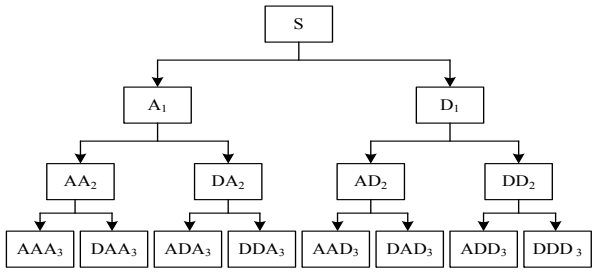


FIGURE 4. Structure of wavelet packet decomposition tree with 3-layers.

A three-layer wavelet packet decomposition tree can be used to help the intuitive understanding of wavelet packet analysis, as shown in Figure 4, where A denotes the low frequency part, D denotes the high frequency part, k denotes the level number of wavelet packet decomposition. Then the decomposition relation is:

$$\begin{aligned} S = {} & AAA_3 + DAA_3 + ADA_3 + DDA_3 \\ & + AAD_3 + DAD_3 + ADD_3 + DDD_3 \end{aligned} \tag{20}$$

As far as WPT is concerned, it can be understood that, at each operation, two decimation one operation is simultaneously performed for the high-frequency part and low-frequency part of previous decompose- tion result (that is, the even or odd part is retained), so that the low frequency band and the high frequency band can have consistent time-frequency resolution. Before the maximum number of the decomposition layers is reached, the spatial decomposition mode is conducted all the way to get higher resolution in frequency and time domain space. According to certain criteria, all frequency bands of last decomposition result are processed, and then the reconstruction is conducted along reverse way.

The output signal waveform of analog circuit contains the rich frequency components. When the circuit fault occurs, these frequency components are affected to varying degrees. Some are weakened, others are enhanced. But the frequency components under fault state are different from that under normal state, that is, its characteristics in amplitude-frequency and phase-frequency have changed, resulting in change of wavelet coefficient with same frequency components. Usually, the energy of enhanced frequency component increase, and the energy of weakened frequency component decrease, and the energy change of different frequency component is significantly different. Analysis shows that, the output signal waveform of analog circuit contains a wealth of fault information, and the energy change of some frequency components can make fault feature of insignificant signals appear with significant change form in wavelet packet subspace under different resolutions. These features extracted can characterize the fault of a component in the circuit. Thus, the wavelet energy of output signal can be used as the fault feature vector to construct the sample set. The method only needs to carry out fault diagnosis according to the change law of wavelet energy of output signal without solving differential

equations [19]. The steps of the sample construction based on WPT are as follows:

1. Carry out $j-$th layer wavelet packet decomposition for the fault output signal $f(t)$, and take out coefficient sequences of each frequency component in the $j-$th layer respectively, obtaining 2^j, namely $\{W_j^0, W_j^1, \cdots, W_j^{2^j-1}\}$.

2. Suppose that S_{j0} denotes reconstructed signal of W_{j0}, S_{j1} denotes the reconstructed signal of W_{j1}, and the like, then

$$S = S_{j0} + S_{j1} + S_{j2} + \cdots + S_{j(2^j-1)} \tag{21}$$

3. Suppose that the energy of S_{ji} is E_{ji} $(i = 0, 1, \cdots, 2^j-1)$, then

$$E_{ji} = \int |S_{ji}|^2 dt = \sum_{k=1}^{n} |X_{ik}|^2 \tag{22}$$

where $X_{ik}(k = 0, 1, 2, \cdots, 2^j - 1)$ is the magnitude of discrete points of reconstructed signal S_{ji}

4. The energy is taken as 2^j-dimension vector of the element

$$P = [E_j^0, E_j^1, E_j^2, \cdots, E_j^{2^j-1}] \tag{23}$$

Generally, the energy of each component of P is not in the same dimension. Therefore, P needs to be normalized before describing the fault feature, and is combined with the expected output vector to obtain the sample set of WNN. It should be noted that, the number of layers of wavelet packet decomposition and the wavelet basis function should appropriately be selected according to the characteristics of fault signals.

VI. EXERIMENT SIMULATION

Experiment environment: Pentium IV 2.4GHz CPU, 4GB DDR memory, 80GB + 7200 RPM hard disk; Windows 7 operating system. The experimental samples are obtained by the software PSPICE simulation. The simulation verification for algorithm is all achieved by MATLAB programming language. In addition, the experiment does not adopt any heuristic algorithm.

A. CONSTRUCTION OF FAULT SAMPLE SET

Elliptic Filter (ITC97) is selected as the circuit to be diagnosed [19]. The nominal value of each element is shown in Figure 5. Suppose that the tolerance of each element is 5%, and the testable node is the output of the circuit. Through comparing the sensitivity of each element affecting the output waveform, it can be confirmed that there are 10 soft fault states in the fault set, that is, R2−, R2+, R6−, R6+, R13−, R13+, C1−, C1+, C5− and C5+, where '−' denotes the small soft fault; '+' denotes the large soft fault, and the nominal deviation value of each element is set to 50%. considering the normal state (no fault state), the experiment involves a total of 11 states.

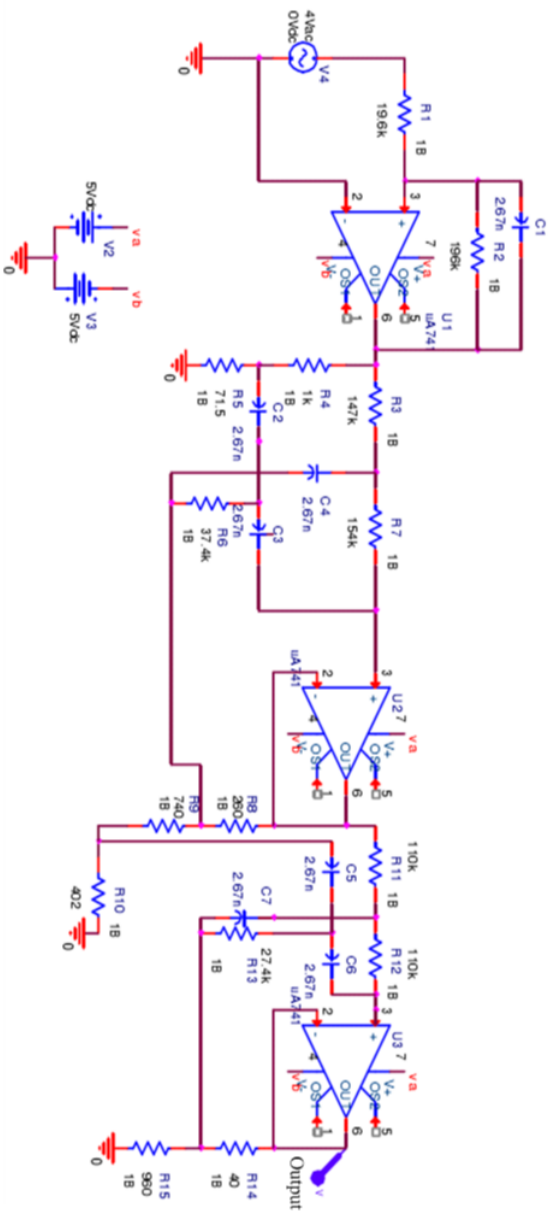


FIGURE 5. Elliptic filter.

1) FAULT SAMPLE SET GOTTEN BY WAVEFORM PARAMETER

PSPICE is adopted to carry out AC Sweep for each failure state of Elliptic Filter. The input V_{in} of the circuit is set as AC signal with voltage amplitude 1V, the settings of other parameters are shown in Figure 5, Sweep frequency is 1.0KHz ∼1MHz, then by AC Sweep the output waveform of the circuit is available under each failure state. The output waveform under R2- fault state is shown in Fig. 6, where the dotted line is the voltage frequency response waveform of the output under normal state. 9 waveform parameters, namely voltage maximum, low pass cutoff frequency (3dB), high pass cutoff frequency (3dB), bandwidth (3dB), bandwidth (4dB), center frequency (3dB), Q value (3dB), Q value (4dB) and Q value (5dB), are represented by solid line. 9 waveform

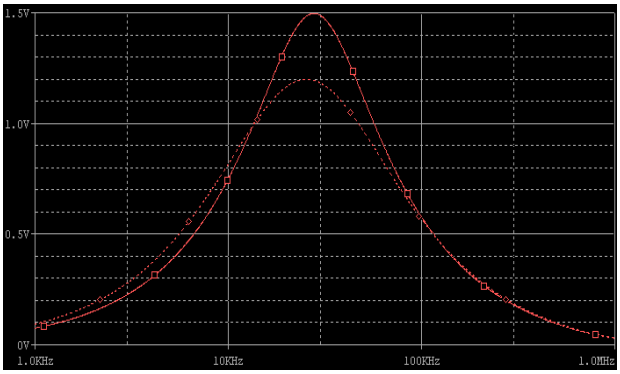


FIGURE 6. Output waveform under R2- failure state based on waveform parameter samples.

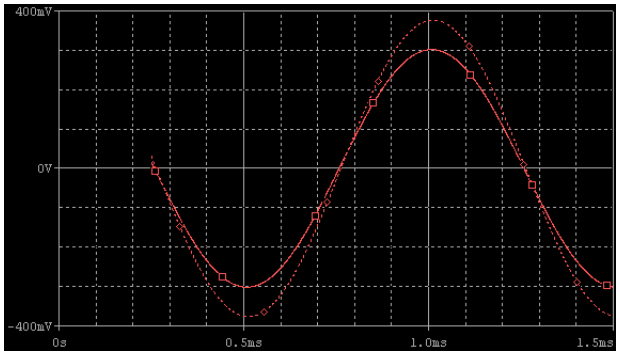


FIGURE 7. Output waveform under R2- failure state based on WPT samples.

TABLE 1. State, coding, and icon of fault.

Fault code	F1	F2	F3	F4	F5
Fault state	R2−	R2+	R6−	R6+	R13−
Binary coding	1000000000	0100000000	0010000000	0001000000	0000100000
Fault icon	△	▲	◇	◆	○
Fault code	F6	F7	F8	F9	F10
Fault state	R13+	C1−	C1+	C5−	C5+
Binary coding	0000010000	0000001000	0000000100	0000000010	0000000001
Fault icon	●	☆	★	□	■

parameters constitute a 9-dimensional input vector for each fault state.

50 Monte Carlo analysis are conducted for each fault state to get 50 fault data. The above 9 waveform parameters can be obtained using software PSPICE to form 50 original fault samples, getting 550 (50×11) input vectors. The expected output vector of fault state can be characterized by '0-l' method, as shown in Table 1. It is easy to understand, the expected output vector of normal state F0 (10 numbers) are all '0'. After the normalization of the sample data, 220 samples were randomly selected to form the training sample set, and the remaining 330 samples constitute the test sample set. KSLPP is used to extract the features of the sample set which are input ELM for training and test. If the absolute value of the difference between actual output value and expected output value for any node of output layer is less than the determination value 0.3 (different values are taken as the case may be), the diagnosis is considered to be correct. Otherwise, the diagnosis is wrong.

2) FAULT SAMPLE SET GOTTEN BY WPT

In Figure 5, the input power V_{in} of the circuit is set as sine wave input excitation with voltage amplitude 4V and voltage frequency 1KHz. The scan starts at 0.25ms and ends at 1.5ms with a step length of 0. 5 μs. According to the characteristics of each fault state, the fault element is set. The trouble-free element is within the normal tolerance range, and the characteristic analysis of circuit is set as transient analysis. The software PSPICE is used for Monte Carlo analysis, getting the output waveform under each fault state. The output waveform under R2- fault state is shown in Figure 7, where the dotted line denotes the voltage frequency response waveform of the output under normal state. 50 transient response signal waveforms can be obtained for each mode by simulation, which are sampled and stored in output text file of PSPICE.

Read the sample data of each waveform from 0.25ms to 1.5ms in the ∗.out file, and get 550 signals. db2 wavelet is firstly selected as wavelet basis function to decompose the collected data through 3 layers wavelet packet, and then according to steps of sample set construction based on wavelet packet transform in section 4.2 calculate the energy of 8 frequency bands in the 3rd layer, and constitute 8-dimensional input vector after normalization, finally combining with the expected output vector characterized by '0-1' method, 550 groups of samples can be obtained, 220 of which are randomly selected for training KSLPP-ELM model, and the remaining for testing.

B. EXPERIMENT ANALYSIS

For the convenience of verification, ELM, LPP- ELM, SLPP-ELM and KSLPP-ELM are respectively adopted to build the fault model and test for Elliptic Filter. In the experiment, LPP, SLPP and KSLPP method are respectively used to reduce the dimension of the fault sample set gotten by waveform parameter or WPT. Figure 8(a) ∼ (c) show the dimension reduction effect of test sample on two-dimensional scale. It is not difficult to find that, the low-dimensional features extracted by LPP have the worst discrimination ability. Except for C5-, other 9 kinds of fault states are confused and cannot be cleared up. For SLPP, the distance within classes becomes smaller, but the distance between classes is still unclear with edge overlap. After KSLPP processing, the distance between classes is obvious, and the distance within classes is also very small.

In order to quantitatively present the above dimension reduction effect, on the basis of waveform parameter samples, low dimensional feature set extracted by LPP, SLPP and KSLPP is respectively input into ELM for modeling and diagnosis. The calculation results are shown in Table 2, where 'C' represents the correct diagnosis number, 'R' represents the correct diagnosis rate.

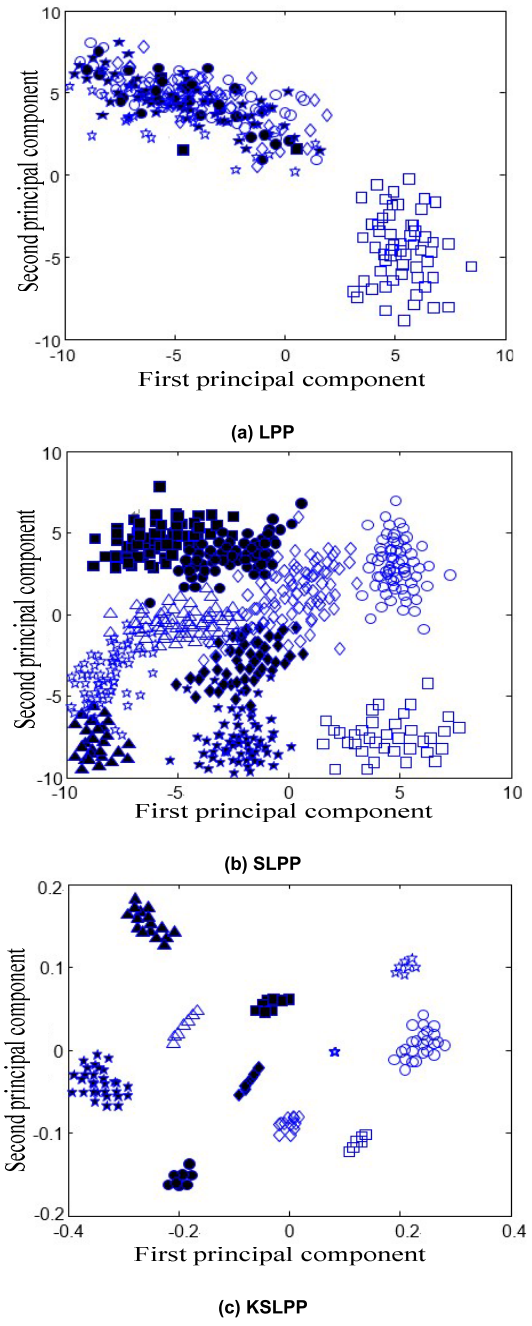


FIGURE 8. Dimension reduction effect of different methods in 2-dimensional scale.

In Table 2, comparing the diagnosis results of ELM and LPP-ELM, LPP extracts 5 principal components as the input of ELM model. The ELM model requires fewer hidden nodes. When 45 hidden nodes are selected, the diagnosis accuracy of LPP-ELM is 88.5%, and that of ELM with the same number of hidden nodes is 79.7%. Only when the number of hidden nodes of the single ELM is selected as 75 can the diagnosis accuracy similar to that of LPP-ELM be achieved. When the number of hidden nodes is fewer, LPP-ELM can achieve higher diagnosis accuracy. It is proved that, LPP-ELM has higher precision, better stability and simpler

TABLE 2. Comparison of diagnosis results of four methods.

Index	ELM	ELM	LPP-ELM	SLPP-ELM	KSLPP-ELM
Hidden node number	75	45	45	45	45
C	293	263	292	313	324
R	88.7%	79.7%	88.5%	94.8%	98.2%

network structure. Similarly, KSLPP-ELM modeling and diagnosis performance can be proved to be advanced.

As can be seen from Table 3 and Table 4, ① The diagnosis accuracy of LPP-ELM, SLPP-ELM and KSLPP-ELM is higher than that of single ELM. It indicates that the diagnosis effect of ELM with feature extraction of LPP, SLPP and KSLPP is indeed better than that without feature extraction. ② Compared with ELM with unsupervised LPP, ELM with supervised SLPP and KSLPP has obvious advantages in C and R. This is because the supervised dimension reduction method with category label is used to extract the low-dimensional features that are conducive to improving the discrimination ability, rather than the descriptive features that characterize circuit faults. ③ After KSLPP feature extraction, the diagnosis effect of ELM is satisfactory. For waveform parameter samples, the total R of KSLPP-ELM is 98.2%, and R of 6 states in 10 fault states plus normal state is 100%. For WPT samples, R of KSLPP-ELM are all 100% for 10 fault states plus normal state. The reason of KSLPP-ELM high accuracy is just because, in KSLPP, the kernel mapping can transform the nonlinear problem in feature set into the linear problem and SLPP based on the local structure be used, which can extract the most discriminating low-dimensional sensitive features in the feature space. ④ In contrast to the waveform parameter samples, the WPT samples can make 4 methods to achieve better accuracy of fault diagnosis. The reason is that, through WPT, the indistinct features can be displayed in the form of significant energy changes in wavelet packet subspace under different resolutions, which is more advantageous than the waveform parameter which directly describe the fault characteristics.

To test the robustness of ELM, firstly, the random disturbance with the noise level $\sigma = 0.1, 0.2, 0.3$ is respectively added in the sample set from waveform parameter or WPT, and then through KSLPP feature extraction the low-dimensional features gotten are respectively input BP, RBF, WNN, SVM and ELM to conduct the fault pattern recognition. The diagnosis results are shown in Table 5 and Table 6, where the parameters of BP, RBF and WNN to be optimized can be initialized as the random number in [0, 1]. The maximum training times are set as 1000. The number of hidden nodes can be determined by 10-fold cross- validation method. SVM was originally designed for binary classification problems. When dealing with multiple classification, a suitable type of classifier needs to be constructed. In the paper, a certain category of samples is successively classified into one category during training, the remaining samples are classified into another category. In this way, for k categories

TABLE 3. Comparison of diagnosis results based on waveform parameter samples.

Fault code	Sample number	ELM		LPP-ELM		SLPP-ELM		KSLPP-ELM	
		N	R	N	R	N	R	N	R
F0	32	26	81.3%	29	90.6%	32	100%	32	100%
F1	29	24	82.8%	27	93.1%	28	96.6%	28	96.6%
F2	30	22	73.3%	24	80.0%	26	86.7%	28	93.3%
F3	32	27	84.3%	29	90.6%	31	96.9%	31	96.9%
F4	27	21	77.8%	24	88.9%	26	96.3%	27	100%
F5	31	25	80.6%	26	83.9%	29	93.5%	31	100%
F6	30	23	76.7%	26	86.7%	29	96.7%	30	100%
F7	28	23	82.1%	25	89.3%	27	96.4%	28	100%
F8	33	26	78.8%	28	84.8%	29	87.9%	32	97.0%
F9	28	22	78.6%	28	100%	28	100%	28	100%
F10	30	24	80.0%	26	86.7%	28	93.3%	29	96.7%
总计	330	263	79.7%	292	88.5%	313	94.8%	324	98.2%

TABLE 4. Comparison of diagnosis results based on WPT samples.

Fault code	Sample number	ELM		LPP-ELM		SLPP-ELM		KSLPP-ELM	
		N	R	N	R	N	R	N	R
F0	30	25	83.3%	27	90.0%	29	96.7%	30	100%
F1	28	24	85.7%	26	92.9%	28	100%	28	100%
F2	29	25	86.2%	27	93.1%	28	96.5%	29	100%
F3	33	28	84.8%	29	87.9%	32	96.9%	33	100%
F4	29	24	82.7%	26	89.7%	27	93.9%	29	100%
F5	32	27	84.4%	28	87.5%	30	93.8%	32	100%
F6	31	26	83.9%	28	90.3%	30	96.8%	31	100%
F7	29	25	86.2%	27	93.1%	29	100%	29	100%
F8	34	28	82.4%	30	88.2%	33	97.1%	34	100%
F9	27	22	81.5%	27	100%	27	100%	27	100%
F10	28	21	75.0%	24	85.7%	28	100%	28	100%
总计	330	275	83.3%	299	90.6%	321	97.3%	330	100%

of samples, k SVMs need to be constructed, and unknown samples are classified into the category with the maximum classification function value during classification. It should be noted that, TL-CDCN is a deep Coupled Dense Convolutional Network (CDCN) proposed in [20] with Transfer Learning (TL) [21]. CDCN can integrate information fusion, feature extraction and fault classification together for intelligent diagnosis (Therefore, this method does not require KSLPP for feature extraction.). TL is dedicated to transfer the

TABLE 5. Comparison of anti-interference ability of classification algorithm based on waveform parameter samples.

Classification algorithm	Average R under different noise levels			
	σ = 0.0	σ = 0.1	σ = 0.2	σ = 0.3
BP	87.2%	83.8%	72.8%	45.0%
RBF	96.5%	84.4%	71.1%	48.2%
WNN	98.0%	87.3%	73.6%	57.4%
SVM	100%	98.7%	87.1%	62.5%
TL-CDCN	100%	99.4%	97.5%	89.7%
ELM	100%	99.2%	96.0%	89.2%

TABLE 6. Comparison of anti-interference ability of classification algorithm based on WPT samples.

Classification algorithm	Average R under different noise levels			
	σ = 0.0	σ = 0.1	σ = 0.2	σ = 0.3
BP	89.6%	87.5%	82.6%	55.9%
RBF	93.2%	86.0%	84.1%	59.3%
WNN	99.4%	90.7%	79.8%	68.6%
SVM	100%	100%	91.5%	77.2%
TL-CDCN	100%	100%	98.3%	91.6%
ELM	100%	100%	98.8%	91.2%

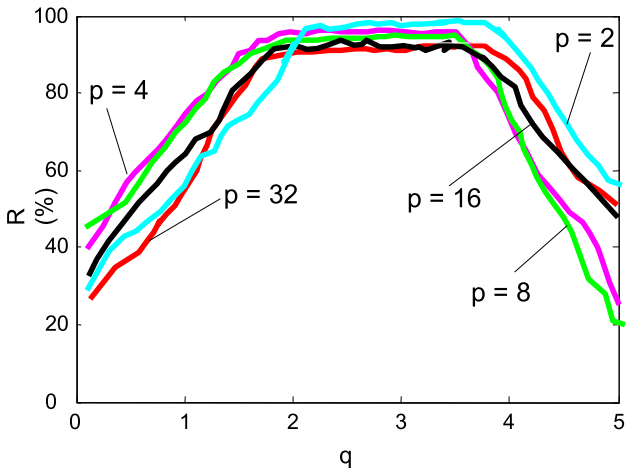


FIGURE 9. Influence of p and q on diagnosis result.

knowledge from previous tasks to target tasks when the target task lacks high quality training data. It can be seen from the table that: ① R of SVM, TL-CDCN and ELM is significantly better than that of BP, RBF and WNN, and TL-CDCN and ELM has higher accuracy compared with SVM, the reason is that, SVM is greatly affected by kernel function. In the absence of optimization for kernel function, the recognition rate of SVM cannot reach the optimization. ② Even if the sample set contains some interference, TL-CDCN and ELM based on KSLPP can still get a satisfactory diagnosis result. It is proven that TL-CDCN and ELM has advantages in

anti-interference, recognition and other. In contrast, ELM is roughly equivalent to TL-CDCN, but the calculation complexity of the latter is much higher than that of the former. From the perspective of modeling efficiency, ELM has certain advantages. ③ Compared with waveform parameter samples, WPT samples can get the stronger anti-interference for each classification algorithm.

In addition, when KSLPP is used to extract the features from sample set, the kernel width parameter q and principal component number p have different effects on diagnosis. Figure 9 shows the influence curve of q and p on the diagnosis result of KSLPP-ELM. As can be seen, when p is taken as 2, 4, 8, 16 or 32 and q is in the range from 2 to 4, the established KSLPP-ELM model can get relatively satisfactory diagnosis results, which is very important for the practical promotion of KSLPP-ELM.

VII. CONCLUSION

For the soft fault detection and diagnosis of analog circuits, an integrated diagnosis method based on KSLPP and ELM is proposed. KSLPP is used to extract low-dimensional sensitive features with identification ability from soft fault sample of analog circuit. ELM is introduced to solve the modeling and diagnosis problems for complex nonlinear and uncertain problems. Combining KSLPP and ELM can improve the accuracy and efficiency of soft fault diagnosis in analog circuit. The simulation for Elliptic Filter shows that, compared with other classification algorithm, KSLPP-ELM has advantages in model establishment, calculation complexity, anti-interference and diagnosis accuracy. In addition, compared with the waveform parameter samples, the WPT samples can make KSLPP-ELM better in fault diagnosis, so as to provide an effective method to solve the problem of soft fault diagnosis in analog circuit. The next step is to develop a practical circuit detection device by using the proposed method.

REFERENCES

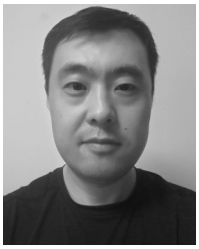
[1] R. S. Berkowitz, "Conditions for network-element-value solvability," *IRE Trans. Circuit Theory*, vol. 15, no. 9, pp. 24–29, 1962.
[2] N. Navid and A. Willson, "A theory and an algorithm for analog circuit fault diagnosis," *IEEE Trans. Circuits Syst.*, vol. 26, no. 7, pp. 440–457, Jul. 1979.
[3] Z. Huang, C. Lin, and R. Liu, "Node-fault diagnosis and a design of testability," *IEEE Trans. Circuits Syst.*, vol. 30, no. 5, pp. 257–265, May 1983.
[4] C.-C. Wu, K. Nakajima, C.-L. Wey, and R. Saeks, "Analog fault diagnosis with failure bounds," *IEEE Trans. Circuits Syst.*, vol. 29, no. 5, pp. 277–284, May 1982.
[5] A. Salama, J. Starzyk, and J. Bandler, "A unified decomposition approach for fault location in large analog circuits," *IEEE Trans. Circuits Syst.*, vol. 31, no. 7, pp. 609–622, Jul. 1984.
[6] T. Ozawa, J. Bandler, and A. Salama, "Diagnosability in the decomposition approach for fault location in large analog networks," *IEEE Trans. Circuits Syst.*, vol. CAS-32, no. 4, pp. 415–426, Apr. 1985.
[7] R. Spina and S. Upadhyaya, "Linear circuit fault diagnosis using neuromorphic analyzers," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 44, no. 3, pp. 188–196, Mar. 1997.
[8] D. Bisen and Y. Sharma, "Empirical investigation of genetic algorithm parameters on neural network based fault diagnosis in analog circuits," *Artif. Intell. Syst. Mach. Learn.*, vol. 9, no. 7, pp. 140–144, 2017.
[9] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and application," *Neurocomputing*, vol. 70, nos. 1–3, pp. 277–284, 2006.
[10] H. C. Leung, C. S. Leung, E. W. M. Wong, and S. Li, "Extreme learning machine for estimating blocking probability of bufferless OBS/OPS networks," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 9, no. 8, pp. 682–692, Aug. 2017.
[11] X.-D. Wang, R.-Z. Zhao, and L.-F. Deng, "Rotating machinery fault diagnosis based on KSLPP and RWKNN," *J. Vib. Shock*, vol. 35, no. 8, pp. 219–223, 2016.
[12] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1999.
[13] L. Oneto, F. Bisio, E. Cambria, and D. Anguita, "Statistical learning theory and ELM for big social data analysis," *IEEE Comput. Intell. Mag.*, vol. 11, no. 3, pp. 45–55, Aug. 2016.
[14] K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–202, Mar. 2001.
[15] J. K. A. Suykens, T. Van Gestel, and J. De Brabanter, *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.
[16] Z.-W. Zhang, F. Yang, K.-W. Xia, and R.-X. Yang, "A supervised LPP algorithm and its application to face recognition," *J. Electron. Inf. Technol.*, vol. 30, no. 3, pp. 539–541, 2008.
[17] Z.-H. Shen, Y.-H. Pan, and S.-T. Wang, "A Supervised locality preserving projection algorithm for dimensionality reduction," *Pattern Recognit. Artif. Intell.*, vol. 23, no. 2, pp. 233–239, 2012.
[18] N. Jamia, P. Rajendran, S. El-Borgi, and M. I. Friswell, "Mistuning identification in a bladed disk using wavelet packet transform," *Acta Mech.*, vol. 229, no. 3, pp. 1275–1295, 2018.
[19] G. H. Man, "The research of analog circuit soft fault diagnosis based on wavelet analysis and neural network," M.S. thesis, Guangxi Univ., Nanning, China, 2009.
[20] J. Jiao, M. Zhao, J. Lin, and C. Ding, "Deep coupled dense convolutional network with complementary data for intelligent fault diagnosis," *IEEE Trans. Ind. Electron.*, to be published.
[21] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowl.-Based Syst.*, vol. 80, pp. 14–23, May 2015.

GAN XU-SHENG was born in 1972. He is currently pursuing the Ph.D. degree in philosophy. His research activities mainly focus on pattern recognition, analog circuit, and nonlinear system modeling.

QU HONG was born in 1985. She is currently pursuing the Ph.D. degree in military science. Her research interests are in the area of command, and control theory and ground navigation.

MENG XIANG-WEI was born in 1968. He is currently pursuing the Ph.D. degree in philosophy. His research activity mainly focuses on nonlinear system modeling.

WANG CHUN-LAN was born in 1988. She is currently an Associate Professor. Her research activity mainly focuses on pattern recognition and analog circuit.



ZHU JIE was born in 1980. He is currently an Associate Professor. His research activity mainly focuses on pattern recognition and analog circuit.