Diagnostics Method for Analog Circuits Based on Improved KECA and Minimum Variance ELM

Zhijie Yuan, Yigang He and Lifen Yuan

Abstract. Kernel entropy component analysis (KECA) is a new method for data transformation and dimensionality reduction. However, it is sensitive to a single kernel radius. By analysis of the relation of statistics in the kernel feature space, improved KECA introduces two kernel radii and an adjusting factor to make KECA less sensitive to kernel radius. A method for fault diagnosis of analog circuits based on the combination of improved KECA and minimum variance extreme learning machine (ELM) is presented. Through wavelet decomposition of sampled signals, features are extracted. Improved KECA for feature dimension reduction is used. Then the fault patterns are classified by minimum variance ELM. Case studies on two analog circuits demonstrating our diagnostics method are presented.

1. Introduction

Analog circuits are used in industrial systems for conditioning signals, domestic appliance, aerospace vehicle, and more[1]. Circuit failures can affect system functionality, and can cause severe effects[2]. The research for analog circuits fault diagnosis has been deeply carried out and many efficiency methods have sprung up in the past several decades, such as the method neural network with learning by backward error propagation[3-5], support vector machine[6-10], data preprocessing techniques like wavelet decomposition and principal component analysis[11,12] and so on. However, analog circuits fault diagnosis is still full of challenges on account of tolerance of component parameters and noise etc.[13-15]. The key operation of data preprocessing is feature selection and extraction. So an effective data transformation is needed to improve diagnosis efficiency.

Kernel entropy component analysis (KECA) [16,17] as a new data transformation has been used in clustering and image processing[18,19]. Recently the technique was applied to analogue circuit diagnosis by Zhang et al.[20]. Firstly the probability density function (PDF) of data set is estimated by the classical Parzen window, while window function is also as kernel function cleverly to project data onto kernel feature space. The Renyi entropy comes to squared euclidean length of the mean vector of kernel feature space by calculation. So eigendecomposition of kernel matrix is used for kernel PCA axes, which contribute most to Renyi entropy. Then data with reduction dimension is obtained by projection onto those axes.

KECA is actually a principal component analysis method in the kernel feature space, which is sensitive to the kernel parameter[21]. And it has little discussion on the statistics in the space. Against that, our paper gives adequate consideration in three statistics in the kernel feature space and proposes a novel method of data-dimension reduction. This method preserves as much as possible of amount — density minus entropy (DME), which improves KECA in the way that the kernel radius in
the first term of DME differs from the radius in the second term, and moreover there is an adjusting factor in DME.

Classifier is required to verify the performance of the feature selection and extraction method. Extreme learning machine (ELM) is a relatively new learning algorithm with fast network training and low human supervision [22, 23]. Minimum Variance ELM (MVELM) algorithm extends the ELM algorithm [24], which aims at minimizing both the network output weights norm and the dispersion of the training data. Extensive evaluation on two publicly available datasets shows the outperformance of MVELM other ELM-based classification schemes.

A new method based on improved KECA and MVELM to analog circuit soft fault is proposed. Through Wavelet decomposition of the circuit’s response, features are extracted for fault diagnosis. The low-dimensional data is obtained by the proposed data dimension reduction algorithm. Finally, the classifier MVELM is used for fault classification.

This paper is organized as follows. Section 2 outlines the principle and implementing steps of improved KECA. Section 3 briefly describes Minimum Variance ELM (MVELM) algorithm. In section 4, the proposed diagnostics method for analog circuits based on improved KECA and MVELM is put forward. The experimental results are demonstrated in Section 5. Finally, the conclusions are drawn in Section 6.

2. Improved keca

2.1 The principle of improved KECA

KECA is a new method for dimensionality reduction with as much as possible of the Renyi entropy of the original data preserved [16]. The Renyi entropy estimator is given by

$$\hat{V}(P) = \frac{1}{N} \sum_{x_i \in D} p_\sigma(x_i) = \frac{1}{N} \sum_{x_i \in D} \sum_{x_j \in D} k_\sigma(x_i, x_j)$$

where $p_\sigma(x_i)$ is the probability density function generating the data set $D = x_1, \cdots, x_N$ and $k_\sigma(x_i, x_j)$ is the Parzen window, or kernel function to project data onto kernel feature space. Let the map from original data set $D = x_1, \cdots, x_N$ to kernel feature space is $\phi$, that is, $x_i \rightarrow \phi(x_i)$. Under the consideration of the definition of kernel function, we have

$$k_\sigma(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$$

Using matrix notation, (1) is given by

$$\hat{V}(P) = \frac{1}{N^2} \mathbf{1}^T \mathbf{K}_\sigma \mathbf{1}$$

where element $(i, j)$ of the $N \times N$ kernel matrix $\mathbf{K}_\sigma$ equals $k_\sigma(x_i, x_j)$ and $\mathbf{1}$ is an $N \times 1$ vector where each element equals one. Obviously, $\mathbf{K}_\sigma$ is a positive semidefinite matrix and can be eigendecomposed as $\mathbf{K}_\sigma = \mathbf{EDE}^T$, where $\mathbf{D}$ is a diagonal matrix storing the eigenvalues $\lambda_1, \cdots, \lambda_N$ and $\mathbf{E}$ is a matrix with the corresponding eigenvectors $\mathbf{e}_1, \cdots, \mathbf{e}_N$ as columns. Rewriting (2), we have

$$\hat{V}(P) = \frac{1}{N^2} \sum_{i=1}^{N} (\sqrt{\lambda_i} \mathbf{e}_i^T \mathbf{1})^2$$

By selecting eigenvalues and eigenvectors which contribute most to the Renyi entropy estimate, KECA performs dimensionality reduction.

The paper turns attention to the statistics in the kernel feature space. First, we notice that the
relationship between statistics \( \mathbf{mm}^T \) and estimated entropy \( \hat{V}(p) \) is

\[
\hat{V}(p) = \int \phi^T(x) \mathbf{mm}^T \phi(x) dx
\]

(4)

where \( \mathbf{m} = \frac{1}{N} \sum_{x \in D} \phi(x) \) is the mean vector vector of the kernel feature space data set[25]. Concerning the common statistics, the equation illustrates the relations between them:

\[
\mathbf{mm}^T = R_{\phi(x)} - \mathbf{C}_{\phi(x)}
\]

(5)

where \( R_{\phi(x)} = \frac{1}{N} \left[ \phi(x_1), \cdots, \phi(x_N) \right] \left[ \phi(x_1), \cdots, \phi(x_N) \right]^T \) is autocorrelation matrix and

\[
\mathbf{C}_{\phi(x)} = \frac{1}{N} \sum_{x \in D} \left( \phi(x) - \mathbf{m} \right) \left( \phi(x) - \mathbf{m} \right)^T
\]

is covariance matrix of data set \( \phi(x_1), \cdots, \phi(x_N) \) respectively.

Taking (5) to (4), we have

\[
\int \phi^T(x) R_{\phi(x)} \phi(x) dx = \frac{1}{N} \int \left( \sum_{x \in D} k^2(x, x) \right) dx - \hat{V}(p)
\]

(6)

Gaussian kernel function is chosen for ease of calculation, then the first term of the right side of equation(6) is given by

\[
\frac{1}{N} \int \left( \sum_{x \in D} k^2(x, x) \right) dx = \frac{1}{N} \int \left( \sum_{x \in D} k_{\sigma/\sqrt{2}}(x, x) \right) dx \approx \frac{1}{N} \sum_{x \in D} \sum_{x \in D} k_{\sigma/\sqrt{2}}(x, x) = \frac{1}{N} \mathbf{1}^T \mathbf{K}_{\sigma/\sqrt{2}} \mathbf{1}
\]

(7)

Improved KECA concentrates on the following amount, that is

\[
\int \phi^T(x) R_{\phi(x)} \phi(x) dx = \int p_{\sigma/\sqrt{2}}(x) dx - \hat{V}_\sigma(p)
\]

(8)

\[
\approx \frac{1}{N} \left( \mathbf{1}^T \mathbf{K}_{\sigma/\sqrt{2}} \mathbf{1} - \frac{1}{N} \mathbf{1}^T \mathbf{K}_{\sigma} \mathbf{1} \right)
\]

(9)

The first term of (8) is the sum of the probability density, while the second term is estimate entropy. Since the kernel parameter in (9) is determined by Sigma, adjusting factor \( \mu(0 \leq \mu \leq 1) \) is used to make the method be more robust. That is the method preserves as much as possible of amount

\[
\frac{1}{N} \left( \mu \mathbf{1}^T \mathbf{K}_{\sigma/\sqrt{2}} \mathbf{1} - (1 - \mu) \frac{1}{N} \mathbf{1}^T \mathbf{K}_{\sigma} \mathbf{1} \right)
\]

which is called by density minus entropy(DME). Improved KECA is a data transform to preserve as much as possible of DME.

2.2 The condition using improved KECA

In order to reduce the data dimension while trying to preserve DME, the matrix \( \mathbf{K} = \mu \mathbf{K}_{\sigma/\sqrt{2}} - (1 - \mu) \frac{1}{N} \mathbf{K}_{\sigma} \) is required to be positive semidefinite. This paper will give a sufficient condition about it. First, two lemmas on eigenvalue in the theory of matrices are invoked.

Let \( \lambda_i(\cdot) \) denote the eigenvalue of a \( N \times N \) matrix \( \mathbf{A} \), and \( \lambda_1(\cdot) \leq \cdots \leq \lambda_N(\cdot) \). The following conclusion illustrates a lower bound for the minimum eigenvalue of the sum of two matrix.

**Lemma1[26]:**

\[
\lambda_N(\mathbf{A} + \mathbf{B}) \geq \lambda_N(\mathbf{A}) + \lambda_N(\mathbf{B})
\]

(10)

Bounds for the maximum eigenvalue and minimum eigenvalue of a symmetric matrix is given by Lemma2. We denote by \( S_N \) the sets of symmetric matrices, the trace and Frobenius norm by
respectively, then

\[
\text{Lemma 2[27]: } \quad \lambda_1(A) \leq \frac{tr(A)}{N} + (N-1)\frac{1}{2}s(\lambda) \tag{11}
\]

\[
\lambda_n(A) \leq \frac{tr(A)}{N} - (N-1)\frac{1}{2}s(\lambda) \tag{12}
\]

where \( s(\lambda) = \left[ \frac{\|A\|_F^2 - tr^2(A)/N}{N} \right]^{\frac{1}{2}} \) is the standard deviation of the eigenvalues.

We obtain a sufficient condition for which matrix \( K \) is positive semidefinite based on the two lemmas.

\textbf{Theorem: }\quad \text{If } \sigma \leq \sigma^* = \left( \frac{m^*}{\ln N} \right)^{\frac{1}{2}}, \text{ then } K = \mu K_{\sigma/\sqrt{2}} - (1-\mu)\frac{1}{N} K_{\sigma} \text{ is positive semidefinite, where}

\[
m^* = \min_{x_i, x_j \in B} \|x_i - x_j\|^2
\]

\textbf{Proof:} We denote \( \|x_i - x_j\|^2 = z_{ij} \). Then recall the definition of kernel function, we have

\[
K_{\sigma}(i,j) = \exp(-2z_{ij}/\sigma^2), K_{\sigma/\sqrt{2}}(i,j) = \exp(-z_{ij}/\sigma^2) \tag{13}
\]

Obviously, the following inequality is a consequence of Lemma 1

\[
\lambda_n\left(\mu K_{\sigma/\sqrt{2}} - \frac{1-\mu}{N} K_{\sigma} \right) \geq \mu \lambda_n\left(K_{\sigma/\sqrt{2}}\right) - \frac{1-\mu}{N} \lambda_n\left(K_{\sigma} \right) \tag{14}
\]

Using the fact \( tr(K_{\sigma}) = tr(K_{\sigma/\sqrt{2}}) = N \) and lemma 2, we have

\[
\lambda_n(K_{\sigma}) \geq 1 - (N-1)^{\frac{1}{2}} \left[ \left( \|K_{\sigma}\|_F^2 - N \right)/N \right]^{\frac{1}{2}} = 1 - (N-1)^{\frac{1}{2}} \left[ \sum_{i,j} \exp(-4z_{ij}/\sigma^2) \right]/N \tag{15}
\]

\[
\lambda_1(K_{\sigma/\sqrt{2}}) \leq 1 + (N-1)^{\frac{1}{2}} \left[ \left( \|K_{\sigma/\sqrt{2}}\|_F^2 - N \right)/N \right]^{\frac{1}{2}} = 1 + (N-1)^{\frac{1}{2}} \left[ \sum_{i,j} \exp(-2z_{ij}/\sigma^2) \right]/N \tag{16}
\]

The inequalities yield the lower bound for the minimum eigenvalue of \( K \), that is

\[
\lambda_n\left(\mu K_{\sigma/\sqrt{2}} - \frac{1-\mu}{N} K_{\sigma} \right) \geq \left( \mu - \frac{1-\mu}{N} \right) N \exp\left(-\frac{2m^*}{\sigma^2}\right) - \frac{1-\mu}{N} (N-1) \exp\left(-\frac{m^*}{\sigma^2}\right) \tag{17}
\]

where \( m^* = \min_{x_i, x_j \in B} \|x_i - x_j\|^2 \)

Suppose \( \mu \geq \frac{2}{N+1}, \) (17) can be minimized by

\[
\lambda_n\left(\mu K_{\sigma/\sqrt{2}} - \frac{1-\mu}{N} K_{\sigma} \right) \geq 1 - \mu N \exp\left(-\frac{2m^*}{\sigma^2}\right) - (1-\mu) \exp\left(-\frac{m^*}{\sigma^2}\right) \tag{18}
\]

To ensure \( K \) be positive semidefinite, we have

\[
\frac{1}{N} - \mu N \exp\left(-\frac{2m^*}{\sigma^2}\right) - (1-\mu) \exp\left(-\frac{m^*}{\sigma^2}\right) \geq 0
\]
The solution of the inequality is $\sigma \leq \sigma^* = \left(\frac{m^*}{\ln N}\right)^{\frac{1}{2}}$.

When Sigma satisfies the inequality above, $K$ can be eigendecomposed to seek project axes, followed by the dimension reduction of data set.

2.3 The Algorithm of Improved KECA
The procedure for improved KECA transforming data set $\Phi$ to $k$ dimensional data is summarized as:

**step1:** Select kernel function $k_\sigma(x,x_i)$, we use a Gaussian kernel $\sigma$. Calculate kernel matrix $K_\sigma$ and $K_{\sigma/\sqrt{2}}$ based on the data set. Then calculate $K = \mu K_{\sigma/\sqrt{2}} - \frac{1-\mu}{N} K_{\sigma}$.

**step2:** Perform eigenvalue decomposition of $K$, $K = Q^T diag(\zeta_1, \cdots, \zeta_n) Q$, where $\zeta_1, \cdots, \zeta_n$ is the eigenvalue and $Q$ is an orthogonal transform matrix with the corresponding eigenvector $q_i$ as column, here $\zeta_i^{-\frac{1}{2}}Qq_i$ is kernel PCA axes.

**step3:** Calculate the contribution of every axes to the Renyi entropy $\left(\sqrt{\zeta_i}q_i^T, 1\right)^2$, select those $k$ axes contributing most to the entropy estimate which span to $U_k$, the corresponding diagonal matrix is $D_k$ and orthogonal transform matrix is $Q_k$.

**step4:** Project $\Phi$ onto $U_k$, $\Phi_{\text{MECA}} = P_{U_k} \Phi = D_k^{-1} Q_k^T$, $\Phi_{\text{MECA}}$ is the low-dimensional data set.

3. Mvelm
MVELM theory aims to reach the smallest training error and the smallest norm of output weight as a single-hidden layer feedforward neural (SLFN) networks training algorithm.

Let us denote by $\left(x_i, o_i\right)_{i=1, \cdots, N}$ a training set consisting of input vector $x_i = [x_{i1}, \cdots, x_{iD}]^T \in \mathbb{R}^{D\times1}$ and output vectors $o_i = [o_{i1}, \cdots, o_{iC}]^T \in \mathbb{R}^{C\times1}$. The network input weights matrix is $W = [w_1, \cdots, w_L] \in \mathbb{R}^{D\times L}$, where $w_i \in \mathbb{R}^{D\times1}$ is the weight vector connecting the $i$th hidden node and the input nodes. While the network output weights $\beta_i = [\beta_{i1}, \cdots, \beta_{iC}] \in \mathbb{R}^{C\times1}$ connects the $i$th output node and the hidden nodes, the hidden layer bias values is $b \in \mathbb{R}^{L\times1}$.

For a given activation function for the network hidden layer $h(x)$ and a linear activation function for the network output layer, the output $o_i$ of SLFN corresponding to $x_i$ is calculated by:

$$o_k = \sum_{j=1}^{L} \beta_{jk} h(w_j \cdot x_i + b_j), k = 1, \cdots, C$$

(19)

here $w_j \cdot x_i$ denotes the inner product of $w_j$ and $x_i$.

By storing the network hidden layer outputs corresponding to the training vectors $x_i, i = 1, \cdots, N$ in a matrix $H = \left[h(w_1 \cdot x_i + b_1), \cdots, h(w_L \cdot x_i + b_L)\right]$:

$$\left[h(x_1), \cdots, h(x_N)\right] = \left[h_1, \cdots, h_N\right] \in \mathbb{R}^{L\times1}$$

hidden layer output corresponding to $x_i$, equation (19) can be expressed in a matrix form as

$$O = \beta^T H$$

(20)
where $O = [o_1, \cdots, o_L] \in R^{C \times N}$ is an output matrix.

MVELM not only tries to minimize the training errors and the norm of the network output weights, but also tries to minimize the variance of the training vectors. The dispersion of the training vectors can be denoted as the total scatter matrix of the training vectors.

$$S = \sum_{i=1}^{N} (h_i - \mu)(h_i - \mu)^T$$  \hspace{1cm} (21)

where $\mu = \frac{1}{N} \sum_{i=1}^{N} h_i$. So $\beta$ can be calculated by solving the following optimization problem:

$$\min L_{MVELM} = \left\| \frac{1}{S} \beta \right\|^2 + \lambda \sum_{i=1}^{N} \| \xi_i \|^2 \quad S.T. \quad \beta^T h_i = t_i - \xi_i, i = 1, \cdots, N$$  \hspace{1cm} (22)

where $\lambda$ is the penalty coefficient on the training errors, $\xi_i$ is the error vector with respect to the ith training pattern, $t_i$ is the target vector corresponding to $x_i$. By calculation, we obtain

$$L_{MVELM} = \sum_{i=1}^{N} \left( o_i - o \right)^2 + \lambda \| o_i - t_i \|^2$$  \hspace{1cm} (23)

where $o = \frac{1}{N} \sum_{i=1}^{N} \beta^T h_i$ denotes the mean network output. (23) shows that MVELM reaches a compromise between the training vectors dispersion in the network output and the network training error.

Based on the Karush-Kuhn_Tucker theorem and Lagrange duality, the output weights can be calculated by

$$\beta = \left( HH^T + \frac{1}{\lambda} S \right)^{-1} HT^T$$  \hspace{1cm} (24)

where $T = [t_1, \cdots, t_N] \in R^{C \times N}$ is a matrix containing the network target vectors.

After the calculation of the network output weights $\beta$, the network response for a given vector $x_m$ is given by $o_m = \beta^T h_m$ where $h_m$ is the network hidden layer output for $x_m$.

4. The new method based on improved keca and mvelm

The proposed diagnostics method for analog electronic circuits can be described as follows:

1) Data preprocessing.

Firstly we use Haar wavelet transform as a preprocessor. Output voltage signal from the circuit under test (CUT) under various fault conditions are gained by Pspice simulation. Then the sampled signal are decomposed into 3 levels using Haar wavelet transform. Energies contained in the coefficients at third level are normalized to get 8-dimensional fault feature vectors.

2) Feature selection.

For the purpose of cutting down training time of the classifier, it is crucial to reduce the dimension of fault features. Thus the novel method-improved KECA is applied to obtain the low-dimension feature vectors.

3) Fault classification.

The features are used to MVELM classifier and the fault patterns are classified.
Obtain lower dimension feature vectors using improved KECA

Fault diagnosis based on MVE LM classifier

Obtain lower dimension feature vectors using improved KECA

FIGURE 1. The flow chart of the proposed method

5. Fault diagnosis results

5.1 Example circuits and faults

In this section, we present two circuits in order to evaluate the performance of the proposed method. In all circuits, a single 10 V pulse with 10 us duration is used as the input. Responses are acquired by sampling the CUTs’ outputs voltage. The resistors and capacitors are assumed to have tolerances of 10% and 5% respectively. We assume that the components are 50% higher or lower than their nominal values. For each fault class, 40 cases are generated. 20 of the cases are used to train the classifier, and the other 20 cases are used to test the method.

FIGURE 2.a. Sallen-Key bandpass filter
FIGURE 2.b. Four opamp biquad highpass filter

The first circuit is a Sallen-Key band-pass filter with 25 kHz central frequency in Fig2a. Fig2a also shows the nominal values of the components. The fault classes include R2↓, R2↑, R3↓, R3↑, C1↓, C1↑, C2↓, C2↑, and normal state, where ↑ and ↓ refer to values higher and lower than the nominal one, respectively.

The second circuit is more complicated, as shown in Fig2b. This is a 2-Four-Opamp Biquad High-pass Filter. The nominal values of the components are shown in the same figure. 11 fault classes are considered, including C1↑, C1↓, C2↑, C2↓, R1↑, R1↓, R2↑, R2↓, R3↑, R3↓, and the normal state.

TABLE 1.a. Fault classes for Sallen-Key highpass filter

| Fault code | Model | Standard value | Fault value |
|------------|-------|----------------|-------------|
| F1         | Normal | ——             | ——          |
| F2         | R2↓   | 3k Ω           | 1.5k Ω      |
| F3         | R2↑   | 3k Ω           | 6k Ω        |
| F4         | R3↓   | 2k Ω           | 1k Ω        |
| F5         | R3↑   | 2k Ω           | 4k Ω        |
| F6         | C1↓   | 5nF            | 2.5nF       |
| F7         | C1↑   | 5nF            | 10nF        |
| F8         | C2↓   | 5nF            | 2.5nF       |
| F9         | C2↑   | 5nF            | 10nF        |
TABLE 1.b. Fault classes for Four opamp biquad highpass filter

| Fault code | Model   | Standard value | Fault value |
|------------|---------|----------------|-------------|
| F1         | Normal  | ——             | ——          |
| F2         | C1↑     | 5nF            | 10nF        |
| F3         | C1↓     | 5nF            | 2.5nF       |
| F4         | C2↑     | 5nF            | 15nF        |
| F5         | C2↓     | 5nF            | 2.5nF       |
| F6         | R1↑     | 6.2kΩ          | 15kΩ        |
| F7         | R1↓     | 6.2kΩ          | 3kΩ         |
| F8         | R2↑     | 6.2kΩ          | 18kΩ        |
| F9         | R2↓     | 6.2kΩ          | 2kΩ         |
| F10        | R3↑     | 6.2kΩ          | 12kΩ        |
| F11        | R3↓     | 6.2kΩ          | 2.7kΩ       |

5.2 Diagnosis results

5.2.1 Features extracted by improved KECA and KECA methods. In order to verify the effectiveness of the proposed method, a comparison is conducted with the two CUTs for improved KECA and KECA[20], that is the two methods are applied to reduce the dimension of the energy features.

Using the proposed improved KECA we deprive the first two most significant component vectors which contribute most to DME for the first circuit. Two-dimensional features are displayed in Fig.3.a. It is obviously that all fault classes are fallen into different 9 ambiguity groups. And there is none of asunder distributed fault classes. A 2-dimension feature set extracted by KECA for the same circuit, is shown in Fig.3.b. Although all fault classes are distinct ambiguity groups, there are several flaws. Obviously F2 fault classes is divided into two parts, as well as F8. Moreover dots from F9 fault classes distributes very separately. The results reveal that our method is an improvement of KECA and it can generate better extraction performance than KECA in this experiment.
FIGURE 3.a. The two-dimensional feature of Sallen-Key filter by improved KECA

FIGURE 3.b. The two-dimensional feature of Sallen-Key filter by KECA

Three-dimensional features obtained by improved KECA and KECA is displayed in Fig4.a. and Fig4.b respectively for the second circuit. The figure Fig4.a shows that all fault classes can be distinct. It is obviously that F2, F3, F4, F5, F6, F7, F9, F10 and F11 fault classes are different ambiguity groups, only F1 and F8 fault classes are partially overlapping. Figure4.b shows that there is completely overlapping for F4 and F11 fault classes, while F2, F4 and F11 fault classes are partially overlapping ambiguity groups. Overlap between different fault classes means bad extraction performance. Besides, F3, F4 and F5 are not compact fault classes. Fig4 shows improved KECA is still superior to KECA for more complicated circuit in extraction performance.
5.2.2 Diagnosis Accuracies. The followed table reveals the classing accuracy for two circuits based on MVELM. Table 2 is the comparison of the test accuracy between improved KECA method and KECA using MVELM. From table2, it can be seen that the proposed approach is accurate (100%) in the first circuit and the accuracy is 99.09% in the second circuit in the fault diagnosis. While the method KECA have accuracies of 98.89% and a more lower 97.72% in the two circuits. From table 2, it can be concluded that the proposed method has a higher accuracy than the method KECA in the case of the two CUTs.

| Method     | Circuit 1     | Circuit 2     |
|------------|---------------|---------------|
| KECA       | 98.33%        | 97.72%        |
| Our Method | 100%          | 99.09%        |
5.2.3 Sensitivity of Kernel Radius. In the section, we discuss the sensitivity of kernel radius. Cluster validity of reduced dimension of energy feature and classification accuracy of faults will be used.

From the scatter plots of fault classes by improved KECA and KECA in figure 3 and figure 4, it is obviously that the reduced-dimensional data has been clustered. In this part, cluster indexes are used to analyze sensitivity of kernel parameter Sigma. Dunn Validity Index (DVI) is based on geometrical considerations as a cluster validity indexes, which can provide the validation results for evaluating clustering. Suppose we have partitioned dataset X into n clusters C, the Dunn Index for the set is defined as:

\[
DVI = \min_{1 \leq i \neq j \leq n} \frac{\min_{x_u \in C_i, x_v \in C_j} \| x_u - x_v \|}{\max_{C_i} \text{diam}(C_i)}
\]

where \( \text{diam}(C_i) = \max_{x_u, x_v \in C_i} \| x_u - x_v \| \) and \( \text{dist}(C_i, C_j) = \min_{x_u \in C_i, x_v \in C_j} \| x_u - x_v \| \). \( n = 9 \)

DVI can evaluate the effectiveness of techniques improved KECA and KECA. Fault classes are characterized by 2-dimensional features of low-pass filter by improved KECA and KECA. For this circuit, \( \sigma^2 = 0.178 \) and here \( n = 9 \), \( C_i \) contains 40 feature vectors of 2-dimension. We adjust the value of \( \mu \) and select the best one corresponding to every Sigma. Figure 5 shows relation between DVI and kernel parameter Sigma, using data-dimension reduction techniques—improved KECA and KECA. It can be seen that when Sigma is less than 0.145, DVI increases slowly and fluctuated narrowly in the range 0.145 \( \leq \sigma \leq 0.155 \), then keeps a value around DVI = 0.264 using our method. The combinatorial kernel matrix and different kernel radius in MDE can be responsible to it. In the contrast, there are several suddenly smaller or larger value of DVI because of fixed kernel radius in KECA. On the other hand, the red curve lines always above the blue curve, which partially can be explained by figure 3.b. Parted fault classes can cause larger distance within class. Figure 5 shows that DVI in the dataset using KECA is more sensitive than our method. We can find that improved KECA has ascendancy over KECA in robustness.

FIGURE 5. DVI of clustering as a function of Sigma

6. Conclusion
In this paper, improved KECA is presented, which is a stable method for data transformation and dimensionality reduction. The method is different from KECA in the consideration of three statistics in the kernel feature space. As an application of improved KECA, a fault diagnosis method based on improved KECA and MVELM is proposed. The reduced data and fault diagnosis efficiency is demonstrated. It is showed that the proposed method can achieve better effect.
Acknowledgements
This work was supported by the National Natural Science Foundation of China under Grant No. 51577046, the State Key Program of National Natural Science Foundation of China under Grant No. 51637004, the national key research and development plan "important scientific instruments and equipment development" Grant No.2016YFF010220, Anhui Provincial Science and Technology Foundation of China under Grant No.1301022036.

References
[1] A.Vasan,L.Bing and M.Pecht, Diagnostics and prognosis method for analog electronic circuits, IEEE Trans. Ind. Electron. 60(11), pp.5277-5291, 2013.
[2] J. Sikorska, M. Hodkiewicz and L. Ma, Prognostic modeling options for remaining useful life estimation by industry, Mech. Syst. Signal Process. vol. 25, no. 5, pp.1803–1836, Jul. 2011.
[3] Y. Tan, Y. He and C. Cui, G. Qiu, A novel method for analog fault diagnosis based on neural networks and genetic algorithms, IEEE Trans. Instrum. Meas. 57(11), pp.2631-2639, 2008.
[4] M. A. El-Gama, and M. D. A. Mohamed, Ensembles of Neural Networks for Fault Diagnosis in Analog Circuits, Journal of Electronic Testing. 23(4), pp.323-339, 2011.
[5] L. Yuan, Y. He, J. Huang, and Y. Sun, A new neural-network-based fault diagnosis approach for analog circuits by using kurtosis and entropy as a preprocessor, IEEE Trans. Instrum. Meas., vol. 59, no. 3, pp. 586–595, Mar. 2010.
[6] J. Huang, X. Hu, and F. Yang, Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker, Measurement, vol. 44, no. 6, pp. 1018–1027, Jul. 2011.
[7] B. Long, S. L. Tian, Q. Miao, and M. Pecht, Research on features for diagnostics of filtered analog circuits based on LS-SVM, IEEE AUTOTESTCON, pp. 360–366, 2011.
[8] B. Long, S. Tian, and H. Wang (Jun. 2012), Diagnostics of filtered analog circuits with tolerance based on LS-SVM using frequency features, J. Electron. Test.: Theory Appl., vol. 28, no. 3, pp. 291–300.
[9] C.-W. Hsu, C.-J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Netw. 13(2), pp. 415-425, 2002.
[10] A. Zhang, C. Chen and B. Jian, Analog circuit fault diagnosis based UCISVM, Neurocomputing, Volume 173, Part 3, 15 January 2016, pp.1752-1760, 2016.
[11] M. Aminian and F. Aminian, Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor, IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process., vol. 47, no. 2, pp. 151–156, Feb. 2000.
[12] Y. Xiao and Y. He, A novel approach for analog fault diagnosis based on neural networks and improved kernel PCA, Neurocomputing, vol. 74, no. 7, pp. 1102–1115, Mar. 2011.
[13] M. Pecht and R. Jaa, A prognostics and health management roadmap for information and electronics-rich systems, Microelectron. Reliab., vol. 50, no. 3, pp. 317–323, Mar. 2010.
[14] J. Sikorska, M. Hodkiewicz, and L. Ma, Prognostic modeling options for remaining useful life estimation by industry, Mech. Syst. Signal Process., vol. 25, no. 5, pp. 1803–1836, Jul. 2011.
[15] D. Binu, B. S. Kariyappa, A Survey on Fault Diagnosis of Analog Circuits: Taxonomy and State of the Art, International Journal of Electronics and Communications. 73, pp. 68-83, 2017.
[16] JENSSEN R. Kernel entropy component analysis, IEEE Trans. Pattern Analysis and Machine Intelligence, 32(5), pp.847-860, 2010.
[17] J. Shi, Q. Jiang, Q. Zhang and Q. Huang, X. Li, Sparse kernel entropy component analysis for dimensionality reduction of biomedical data, Neurocomputing, vol. 168, pp. 930-940, 2015.
[18] SHEKAR B H, SHARMILA KUMARI M and MESTETSKIY L M, et al. Face recognition using kernel entropy component analysis[J], Neurocomputing, 74(6): pp. 1053-1057, 2011.
[19] XIE Z and GUAN L. Multimodal information fusion of audio emotion recognition based on kernel entropy component analysis[J]. Multimedia, IEEE International Symposium on, 41(11): pp.1-8, 2012.
[20] C. Zhang and Y. He, et.. A novel approach to diagnosis of analog circuit incipient faults based of
[21] Jonas Nordhaug Myhre. Semi-Supervised Classification using Kernel Entropy Component Analysis and the LASSO. FYS-3921 Master’s thesis in Electrical Engineering December 2011. The University of Tromsø.

[22] G.B. Huang, L. Chen, and C.K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes, IEEE Transactions on Neural Networks, vol. 17, no. 4, pp. 879–892, 2006.

[23] Y. Wang, F. Cao, and Y. Yuan. A study on effectiveness of extreme learning machine, Neurocomputing, vol. 74, no. 16, pp. 2483–2490, 2011.

[24] A. Iosifidis, A. Tefas, and I. Pitas, Minimum class variance extreme learning machine for human action recognition, IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 11, pp. 1968–1979, Nov. 2013.

[25] X. Yan, S. Wang, H. Guo, Feature reduction of high-order statistics based on Parzen window, CAAI Transactions on Intelligent Systems, 8(1): pp. 1-10, 2013.

[26] R.A. Horn, C.R. Johnson. Matrix Analysis [M]. Cambridge: Cambridge University Press, 1985.

[27] P. Tarazag, Eigenvalue Estimate for Symmetric Matrices. Linear Algebra and its Application, 135(1), pp. 171-179, 1990.