

Received October 1, 2019, accepted October 24, 2019, date of publication October 29, 2019, date of current version November 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950240

# A Fault Diagnostic Method for Induction Motors Based on Feature Incremental Broad Learning and Singular Value Decomposition

SAI BIAO JIANG<sup>1,2</sup>, PAK KIN WONG<sup>2</sup>, AND YAN CHUN LIANG<sup>1</sup>

<sup>1</sup>Zhuhai College of Jilin University, Zhuhai 519041, China

<sup>2</sup>Department of Electromechanical Engineering, University of Macau, Taipa 999078, Macau

Corresponding author: Pak Kin Wong (fstpkw@um.edu.mo)

This work was supported in part by the Guangdong Premier Key-Discipline Enhancement Scheme under Grant 2016GDYSZDXK036, in part by the Guangdong Key-Project for Applied Fundamental Research under Grant 2018KZDXM076, in part by the Key Platforms and Major Projects of the Education Department of Guangdong Province under Grant 2017KQNCX252 and Grant 2018KQNCX352, in part by the National Natural Science Foundation of China under Grant 61572228 and Grant 61972174, and in part by the 2019 Innovation Strong School Construction Subsidy Project under Grant CR1902.

**ABSTRACT** The occurrence of fault in induction motors is dangerous in our daily life. It is significant to diagnose motor component faults accurately and quickly. In this paper, we propose an efficient and responsive motor fault diagnostic method based on Feature Incremental Broad Learning (FIBL) and Singular Value Decomposition (SVD). Firstly, we extract fault features from raw signals with Particle Swarm Optimization-Variation Model Decomposition, Sample Entropy and Time Domain Statistical Features. Secondly, these features are input into a broad learning system to train a network. Then we use FIBL to retrain the network if the diagnosis accuracy is unsatisfactory. Finally, SVD is used to further simplify the system structure to reduce diagnostic errors. In order to evaluate the performance of the diagnostic system, experiments are conducted. Experimental results show that with the proposed diagnostic method, motor component faults detection is quicker and more accurate.


**INDEX TERMS** Fault diagnosis, feature extraction, incremental broad learning, singular value decomposition, induction motor.

## I. INTRODUCTION

Induction motors are the principal purveyor of motive force in our daily lives. They usually carry out heavy duty tasks and run for a long time under high power and high load conditions. This will potentially affect either production or personal safety when they suffer from failure [1]. Therefore, it is necessary to monitor and diagnose their working conditions in order to avoid serious accidents or personal injuries. Recent studies showed that induction motors usually have problems such as unbalanced windings, an unbalanced stator or rotor, broken rotor bars, eccentricity, and bearing defects [2], [3].

With the advent of machine learning, there is an increasing interest in studying the fault diagnosis of machine learning applying to traditional rotating machinery [4]. At present, Deep Belief Networks (DBN) [5], Deep Boltzmann Machines (DBM) [6], Support Vector Machines (SVM) [7], Extreme

Learning Machines (ELM) [8], and Convolutional Neural Networks (CNN) [9]–[11] are all widely used in the diagnostics of DC motors and AC motors. These methods are particularly suitable for AC motors where the relationship between motor current and speed is non-linear [12]. Deep learning such as DBN, DBM and CNN can improve the diagnostic accuracy by a great number of hyper parameters and complicated structures. However, their training process is highly time-consuming and their complications make it so difficult to analyze the deep structure theoretically. So, deep learning cannot be suitable for electric motor fault diagnosis. In the existing literature [7], support vector machines (SVM) is frequently used as classifiers to diagnose the induction motor faults, but SVM has a disadvantage in respect of the non-probabilistic output. Therefore, SVM is not considered in this study. For ELM, the training time is fast due to a simple structure which contains three layers. However, if the diagnostic accuracy is not good, the whole ELM structure should be adjusted and retraining from scratch is necessary.

The associate editor coordinating the review of this manuscript and approving it for publication was Javier Medina .

In recent years, Broad Learning has been proposed to further improve training performance. In contrast to the previously discussed methods, Broad Learning consists of two layers: one input layer, which contains mapped feature nodes and enhancement nodes, and one output layer [13]. Despite a simple structure, it can yield an improved performance by increasing the number of nodes. In our previous work [14], we have increased the number of enhancement nodes. But the number of enhancement nodes is very large. It is hard to find the best enhancement node. In order to solve this problem, Feature Incremental Broad Learning (FIBL) is proposed to retrain a network. The FIBL is designed to easily update weights without the need for an entire training cycle. Furthermore, feature nodes are small than enhancement nodes. So, it is very effective to remodel a system. This makes them a suitable tool for diagnosing induction motors, improving predictive accuracy while reducing training time and avoiding retraining from scratch.

It should be noted that data collection and pre-processing are required before network training. Typically, some kinds of feature extraction methods are used to improve the diagnostic accuracy in open literature. Fast Fourier Transform (FFT) is a commonly used in signal processing [15]. However, it is not suitable for non-stationary signals which require a time-frequency analysis [16]. Short Time Fourier Transform (STFT) may be a viable alternative but it has defects in the resolution of time and frequency, making it impossible to resolve both time and frequency at the same time [17]. Wavelet Transform is also a good alternative, but it suffers from energy leakage [18] when there are some signal features that do not match the shape of the mother wavelet function. Empirical Mode Decomposition (EMD) is a self-adaptive signal processing method that can be applied to nonlinear and nonstationary processes perfectly. However, the major disadvantage of EMD is the issue of mode mixing. Dragomiretskiy and Zosso [19] recently proposed a new method called Variation Model Decomposition (VMD) which assumes that each extracted pattern has a limited bandwidth and is compressed around a matching center frequency. The sparse prior of each submodule is chosen as the center bandwidth of the spectral domain. However, VMD is not model-adaptive in practical applications and its modulation capability is largely dependent on inherent parameter settings [20]. That is, the different configurations of the penalty ball,  $\alpha$ , and the number of sub-components,  $K$ , result in a variable decomposition performance. Therefore, the parameters  $\alpha$  and  $K$  need to be prioritized for optimization. Particle Swarm Optimization (PSO) algorithm is an evolutionary computation [21]. It comes from the predatory behavior of birds and is easy to adjust parameters. At present, it has been widely used in function optimization and neural network training [22]. In this paper, we employ PSO to optimize the parameters of  $\alpha$  and  $K$  of the original VMD.

After the system finishes training, the FIBL system may contain some redundant nodes due to the wide increment of feature nodes. It will lead the system to have poor

accuracy. Therefore, the system should be simplified by using a low rank approximation. There are many low rank approximations. The common methods are Singular Value Decomposition (SVD) and Non-Negative Matrix Factorization (NMF) [23]. The drawback of NMF is that the weight matrix can't allow to be a negative value. SVD is a nonparametric tuning technique. Furthermore, the signal can usually be converted into a matrix where the singular value represents the nature of the fault signal [24], [25]. It indicates that some singular values of the fault signal matrix can be enhanced, while others can be constrained. This paper uses the SVD algorithm to simplify the incremental broad learning system.

The innovations of this research are as follows:

1. A feature extraction with PSO-VMD is first used in induction motor fault diagnosis to improve the accuracy of the diagnostic system.
2. A Feature Incremental Broad Learning (with feature nodes) method is proposed to optimize the system network to improve the diagnostic system both in accuracy and training speed.
3. SVD is successfully used to simplify the FIBL structure to further reduce the test error of this diagnostic system.

This paper is organized as follows: the proposed motor fault diagnostic method is presented in the coming sections, including a description of the methodology, technology and the relevant experimental setup and data pre-processing; the experimental results and comparisons with other methods are discussed in Section V before presenting our conclusions in the last section.

## II. PROPOSED DIAGNOSTIC FRAMEWORK

The proposed motor fault diagnostic method (FIGURE 1) consists of four sub-modules: (a) Data acquisition and data processing, (b) Broad Learning, (c) Feature Incremental Broad Learning, and (d) Simplification by SVD.

### A. DATA ACQUISITION

Since the induction motor is supplied with three-phase symmetrical currents, it just requires two stator currents [26]. The signal acquisition sub-module digitally detects Windings A, B, sound, and the signals are then recorded as  $x_1, x_2, x_3$  respectively. We use a limiting filter to reduce their interference after acquiring these raw signals [27].

### B. DATA PROCESSING

In terms of data pre-processing, the original sound signal  $x_3$  is decomposed by using PSO-VMD. Then, feature extraction is carried out using the effective statistical algorithm of sample entropy (SampEn). The result is saved as  $x_{3-SE}$ . In addition to the signal features, Time Domain Statistical Features (TDSF) are appended to the signals of  $x_1, x_2, x_3$  [28]. To ensure that all features have a uniform contribution, each feature is normalized to [0,1]. Then, each feature is divided into three independent groups namely,  $x_{k-Proc-Train}$ ,  $x_{k-Proc-Vali}$ ,  $x_{k-Proc-Test}$ .

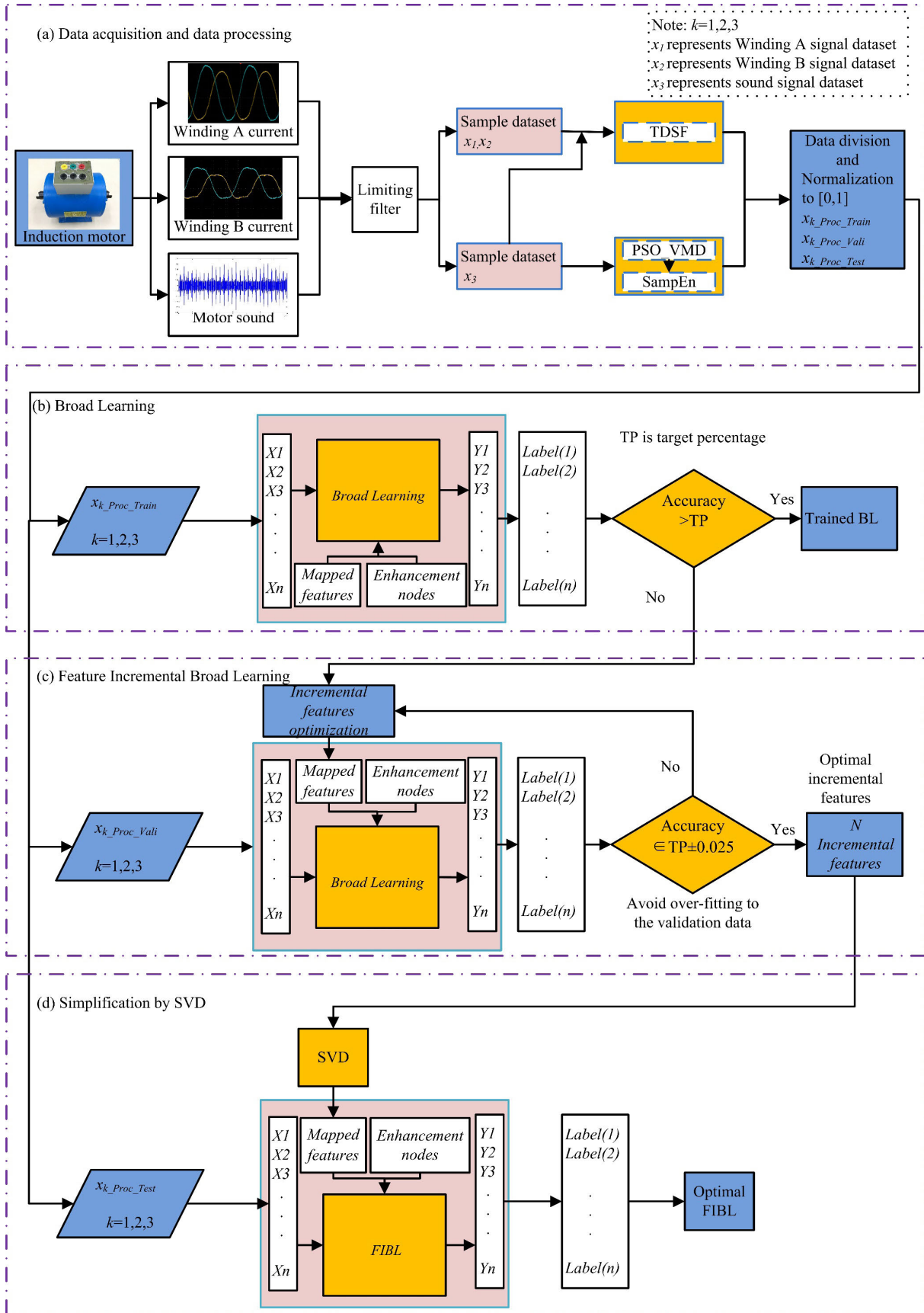


FIGURE 1. Proposed motor diagnostic method.

### C. BROAD LEARNING

In Broad Learning module, the processed datasets,  $\mathbf{x}_{k-Proc-Train}$ , are first trained by using the Broad Learning (BL) network. Secondly, the trained BL network outputs a training accuracy based on training datasets. The network stops training if the training accuracy reaches the target percentage (TP). Otherwise, the Broad Learning is entered to Feature Incremental Broad Learning by increasing the number of feature nodes.

### D. FEATURE INCREMENTAL BROAD LEARNING

The validation dataset is applied to the Feature Incremental Broad Learning sub-module. The accuracy will increase as the number of feature nodes is increased within a range. However, the FIBL network will have over-fitting if the number of feature nodes is increased too much. Therefore, the number of feature nodes should be optimized. This dynamic process will continue until the accuracy is satisfactory. Finally, the optimized feature node  $N$  is obtained. To avoid the system from trapping into infinite loops, we set the maximum iteration with 1000.

### E. STRUCTURE SIMPLIFICATION

It should be noted that redundant nodes may be included due to the wide increment of feature nodes. Firstly, we simplify feature nodes by SVD algorithm. Secondly, we test the accuracy 10 times based on the processed test datasets. Then, we calculate minimum test errors (MTEs) and average test errors (ATEs) to evaluate its effectiveness.

## III. METHODOLOGY AND TECHNOLOGY

### A. VARIATION MODEL DECOMPOSITION AND PARTICLE SWARM OPTIMIZATION

VMD translates the raw signal into a series of dependent subcomponents  $\{u_k\}$ , which have specific sparsity properties [24]. The sparsity prior of each sub-mode is selected to be a bandwidth in frequency domain. That is, VMD requires sub-modes to be close around a center frequency  $\{\omega_k\}$ . To calculate the bandwidth of each sub-component, the VMD method applies  $H^1$  Gaussian smoothness to the demodulated signal. Thus, the constrained variation problem of VMD is then concluded to calculate the minimum of the  $L^2$  norm of the gradient shown in Eq. (1). Where  $\{u_k\} = \{u_1, u_2, \dots, u_K\}$ , and  $\{\omega_k\} = \{\omega_1, \omega_2, \dots, \omega_K\}$  are the modes and the center frequencies.  $K$  is the number of decomposed components.

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (1)$$

Assuming that the sum of sub-modes  $\sum u_k = x$ , the reconstruction constraint given by Eq. (1) can be addressed with a quadratic penalty term  $\alpha$  and Lagrange multiplier  $\lambda(t)$ . The augmented Lagrange  $L(\{u_k\}, \{\omega_k\}, \lambda)$  is shown in Eq. (2).

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \cdot L_1(\{u_k\}, \{\omega_k\}, \lambda) + L_2(\{u_k\}, \{\omega_k\}, \lambda) \quad (2)$$

where the sub-component

$$L_1 = \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2, \text{ and } L_2 = \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle$$

The quadratic penalty term,  $\alpha$ , and the number of sub-modes,  $K$ , are the critical parameters in the process of decomposition. It is impossible to find the optimal pair of  $\alpha$  and  $K$  manually.

The PSO algorithm, introduced by Kennedy and Eberhart [29], [30], is a population based stochastic approach for solving continuous and discrete optimization problems. The position of a particle in PSO represents a candidate solution to the optimization problem. Each particle searches for a better position in the search space by changing its migration velocity according to the rules inspired by behavioral models of bird flocking. The key to run the PSO algorithm is to set up a reasonable objective fitness function [20]

$$\langle \hat{\alpha}, \hat{K} \rangle = \arg \min_{(\alpha, K)} \left\{ \frac{-1}{\hat{K}} \sum_{i=1}^{\hat{K}} \sum_1^K p_i \log_2(p_i) \right\} \quad (3)$$

where  $\hat{\alpha}$  and  $\hat{K}$  indicate an optimal group of parameters and  $p_i$  is the normalized envelope of sub-mode  $u_k$ .

A population of  $n$  particles  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  represents a vector of  $D$ -dimensions.  $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{iD})^T$  represents the position of the  $i$ th particle in the  $D$ -dimensional search space. The fitness value of each particle position  $\mathbf{X}_i$  can be calculated according to the fitness function. The velocity of the  $i$ th particle is  $\mathbf{V}_i = (V_{i1}, V_{i2}, \dots, V_{iD})^T$ . The particle-best is represented as  $\mathbf{P}_i = (P_{i1}, P_{i2}, \dots, P_{iD})^T$ . The global-best is represented as  $\mathbf{P}_g = (P_{g1}, P_{g2}, \dots, P_{gD})^T$ .

The particle updates its speed and position through the particle-best and the global-best in each iteration. That is:

$$V_{id}^{k+1} = \omega V_{id}^k + c_1 r_1 (P_{id}^k - P_{id}^k) + c_2 r_2 (P_{gd}^k - P_{id}^k) \quad (4)$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \quad (5)$$

where  $\omega$  is the weight of inertia;  $d=1, 2, \dots, D$ ;  $i=1, 2, \dots, n$ ;  $k$  is the number of current iterations;  $V_{id}$  is the velocity of the particle;  $c_1$  and  $c_2$  are non-negative constants called acceleration factors;  $r_1$  and  $r_2$  are random numbers from the interval  $[0, 1]$ .

### B. FEATURE INCREMENTAL BROAD LEARNING

In a deep learning network, the number of layers can be increased if it produces a poor training result. This training process is ineffective because the deep learning needs to change the whole weights during a retraining process. However, with the proposed Feature Incremental Broad Learning, a mapped feature can be incremented to create a new structure of the Broad Learning network. The incremental broad learning doesn't need to change the entire structure network. It only adjusts the weights of incremental feature nodes. It is the first study on applying Feature Incremental Broad Learning to the induction motor fault diagnosis.

We suppose the Broad Learning structure consists of  $n$  groups of mapped feature nodes and  $m$  groups enhancement nodes. Then we increase  $(n+1)$ th group mapped feature nodes to this structure. The equation of the  $(n+1)$ th group mapped feature is as follows:

$$\mathbf{Z}_{n+1} = \phi(\mathbf{X}\mathbf{W}_{e_{n+1}} + \boldsymbol{\beta}_{e_{n+1}}) \quad (6)$$

$\mathbf{W}_{e_{n+1}}$  is the weight of the  $(n+1)$ th feature node, and  $\boldsymbol{\beta}_{e_{n+1}}$  is the bias value of  $(n+1)$ th feature node.

The enhancement nodes are generated as follows:

$$\mathbf{H}_{ex_m} = [\xi(\mathbf{Z}_{n+1}\mathbf{W}_{ex_1} + \boldsymbol{\beta}_{ex_1}), \dots, \xi(\mathbf{Z}_{n+1}\mathbf{W}_{ex_m} + \boldsymbol{\beta}_{ex_m})] \quad (7)$$

$\mathbf{W}_{ex_m}$  is the weight of the  $m$ th enhancement node,  $\boldsymbol{\beta}_{ex_m}$  is the bias value of  $m$ th enhancement node. They are randomly generated. Denote  $\mathbf{A}_{n+1}^m = [\mathbf{A}_n^m | \mathbf{Z}_{n+1} | \mathbf{H}_{ex_m}]$ , which is the upgrade of new mapped features and the corresponding enhancement nodes. The relatively upgraded pseudoinverse matrix is given as follows:

$$(\mathbf{A}_{n+1}^m)^+ = \begin{bmatrix} (\mathbf{A}_n^m)^+ & -\mathbf{D}\mathbf{B}^T \\ \mathbf{B}^T & \end{bmatrix} \quad (8)$$

where  $\mathbf{D} = (\mathbf{A}_n^m)^+ [\mathbf{Z}_{n+1} | \mathbf{H}_{ex_m}]$ ,

$$\mathbf{B}^T = \begin{cases} (\mathbf{C})^+ & \text{if } \mathbf{C} \neq 0 \\ (1 + \mathbf{D}^T\mathbf{D})^{-1} \mathbf{D}^T ((\mathbf{A}_n^m)^+)^T & \text{if } \mathbf{C} = 0 \end{cases}, \text{ and}$$

$\mathbf{C} = [\mathbf{Z}_{n+1} | \mathbf{H}_{ex_m}] - \mathbf{A}_n^m \mathbf{D}$ . They are intermediate matrix for calculation.

The new weight is:

$$\mathbf{W}_{n+1}^m = \begin{bmatrix} \mathbf{W}_n^m - \mathbf{D}\mathbf{B}^T \mathbf{Y} \\ \mathbf{B}^T \mathbf{Y} \end{bmatrix} \quad (9)$$

In fact, the new weight of the FIBL can be obtained by calculating the matrix of  $\mathbf{D}$  and  $\mathbf{B}^T \mathbf{Y}$ . The  $\mathbf{W}_n^m$  does not need to be recalculated. FIGURE 2 shows the incremental network of  $(n+1)$  mapped feature nodes.

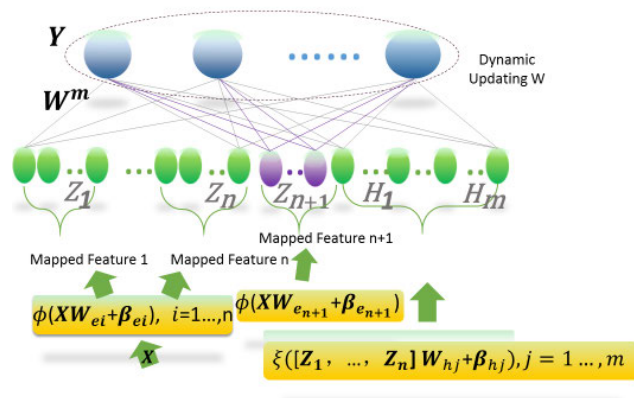


FIGURE 2. Feature incremental broad learning.

### C. SINGULAR VALUE DECOMPOSITION

There is a risk of redundancy if the mapped features of the BL are increased too much. In general, the structure can be simplified by a series of low-rank approximation algorithms. SVD is one of the low-rank approximation algorithms to provide structural simplifications for BL. The SVD can truncate smaller singular value components [31].

SVD is a nonparametric tuning technique. The feature nodes can be converted into a matrix where the singular value represents the nature of feature nodes. It indicates that some singular values of the feature node matrix can be enhanced, while the others can be constrained.

SVD is applied to the  $i$ th feature  $\mathbf{Z}_i$  as follows

$$\mathbf{Z}_i = \mathbf{U}_{Z_i} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_{Z_i}^P & | & \boldsymbol{\Sigma}_{Z_i}^Q \end{bmatrix} \cdot \begin{bmatrix} \mathbf{V}_{Z_i}^P & | & \mathbf{V}_{Z_i}^Q \end{bmatrix}^T = \mathbf{Z}_i^P + \mathbf{Z}_i^Q \quad (10)$$

where  $\mathbf{U}_{Z_i}$  is an  $k \times k$  orthogonal matrix whose columns are the eigenvectors of  $\mathbf{U}_{Z_i} \mathbf{U}_{Z_i}^T$ ,  $\begin{bmatrix} \mathbf{V}_{Z_i}^P & | & \mathbf{V}_{Z_i}^Q \end{bmatrix}$  is an  $l \times l$  orthogonal matrix whose columns are the eigenvectors of  $\mathbf{U}_{Z_i}^T \mathbf{U}_{Z_i}$ , and  $\begin{bmatrix} \boldsymbol{\Sigma}_{Z_i}^P & | & \boldsymbol{\Sigma}_{Z_i}^Q \end{bmatrix}$  is an  $k \times l$  diagonal matrix.

By compressing  $\mathbf{Z}_i$  by the principal portion  $\mathbf{Z}_i^P$ , the equation between  $\mathbf{Z}_i$  and  $\mathbf{Z}_i^P$  is derived as follows

$$\mathbf{Z}_i^P \mathbf{V}_{Z_i}^P = \mathbf{U}_{Z_i} \boldsymbol{\Sigma}_{Z_i}^P \mathbf{V}_{Z_i}^{P^T} \mathbf{V}_{Z_i}^P + \mathbf{U}_{Z_i} \boldsymbol{\Sigma}_{Z_i}^P \mathbf{V}_{Z_i}^{Q^T} \mathbf{V}_{Z_i}^P = \mathbf{Z}_i \mathbf{V}_{Z_i}^P \quad (11)$$

Assume that the random initial network with  $n$  groups of feature nodes can be represented as the equation of the following form:  $\mathbf{Y} = [\mathbf{Z}_1, \dots, \mathbf{Z}_n] \mathbf{W}_n^0$ . We denote all the feature nodes of  $n$  groups  $[\mathbf{Z}_1, \dots, \mathbf{Z}_n]$  as  $\mathbf{A}_n^0$ .

So

$$\mathbf{Y} = \mathbf{A}_n^0 \mathbf{W}_n^0 \quad (12)$$

For original network, we define:  $\mathbf{W}_n^0 \triangleq [\mathbf{W}_{Z_1}^{(0,n)} | \dots | \mathbf{W}_{Z_n}^{(0,n)}]^T$ .

$$\mathbf{Y} = [\mathbf{Z}_1 \mathbf{V}_{Z_1}^P, \dots, \mathbf{Z}_n \mathbf{V}_{Z_n}^P] \begin{bmatrix} \mathbf{V}_{Z_1}^{P^T} \mathbf{W}_{Z_1}^{(0,n)} \\ \dots \\ \mathbf{V}_{Z_n}^{P^T} \mathbf{W}_{Z_n}^{(0,n)} \end{bmatrix} \quad (13)$$

$$\text{Assume that } \mathbf{W}_F^{(0,n)} = \begin{bmatrix} \mathbf{V}_{Z_1}^{P^T} \mathbf{W}_{Z_1}^{(0,n)} \\ \dots \\ \mathbf{V}_{Z_n}^{P^T} \mathbf{W}_{Z_n}^{(0,n)} \end{bmatrix},$$

$$\mathbf{A}_F^{(0,n)} = [\mathbf{Z}_1 \mathbf{V}_{Z_1}^P, \dots, \mathbf{Z}_n \mathbf{V}_{Z_n}^P]$$

So

$$\mathbf{Y} = \mathbf{A}_F^{(0,n)} \mathbf{W}_F^{(0,n)} \quad (14)$$

Then

$$\mathbf{W}_F^{(0,n)} = (\mathbf{A}_F^{(0,n)})^+ \mathbf{Y} \quad (15)$$

In this way, the original  $\mathbf{A}_n^0$  is simplified to  $\mathbf{A}_F^{(0,n)}$ .

#### IV. EXPERIMENTAL SETUP AND DATA PREPROCESSING

To obtain representative sample data for machine learning and verify the effectiveness of the proposed framework, experiments were carried out. All the proposed methods mentioned were implemented by using MATLAB R2016a and executed on a personal computer with 4 Core i5-4590 @3.30GHz and 12GB RAM onboard.

##### A. TEST RIG AND TEST SCHEME

The test rig used in this example is called TCDJ-03A which was set up in [14]. We test nine different faults of three-phase induction motors according to [32], [33]. TABLE 1 shows details. (Note: In order to collect data from all faults, all the faults are broken slightly. So, the motor can be test under full load or over load)

TABLE 1. Nine different faults.

| Label | Fault   |
|-------|---|
| D1    | Normal  |
| D2    | Phase imbalance                               |
| D3    | Short circuit in the stator of Winding A [34] |
| D4    | Mechanical imbalance                          |
| D5    | Bent rotor                                    |
| D6    | Broken rotor bar                              |
| D7    | Bearing outer raceway defect                  |
| D8    | Bearing inner raceway defect                  |
| D9    | Broken bearing ball                           |

##### B. SAMPLE DATA ACQUISITION

There are three load conditions (underload, rated load and overload). Each case is tested for 100 seconds and generates 800,000 data points. There are 1000 data points in each sample. As mentioned above, each fault under a load condition has a total of 800,000 data points. These data points are divided into 800 sample data (1000 data points in each sample data).

##### C. FEATURE EXTRACTION BY PSO-VMD AND SAMPEN

The parameters of  $K$  and  $\alpha$  should be optimized by PSO before VMD is applied. We select 4 and 50 as the optimized  $K$  and  $\alpha$  respectively according to [20]. FIGURE 3 shows the VMD decomposes the sound signal into 4 intrinsic mode functions (IMFs), but the input dimension of each IMF remains unchanged after decomposition. This leads to poor accuracy of fault diagnosis because the input dimensions of each classifier is very massive. To overcome this problem, an effective feature selection method for dimension reduction is considered. This research uses an effective statistical algorithm, Sample Entropy (SampEn), to compute representative features from each IMF. Therefore, the input cases. It states that there are different SampEn values for each IMF in different cases. It is helpful for classification.

##### D. TIME-DOMAIN STATISTICAL FEATURES

In this paper, ten Time-Domain Statistical Features are used to further analyze the features of current A signal  $x_1$ , current B signal  $x_2$  and sound signal  $x_3$ . TABLE 2 shows the equations of 10 common Time-Domain Statistical Features [35].

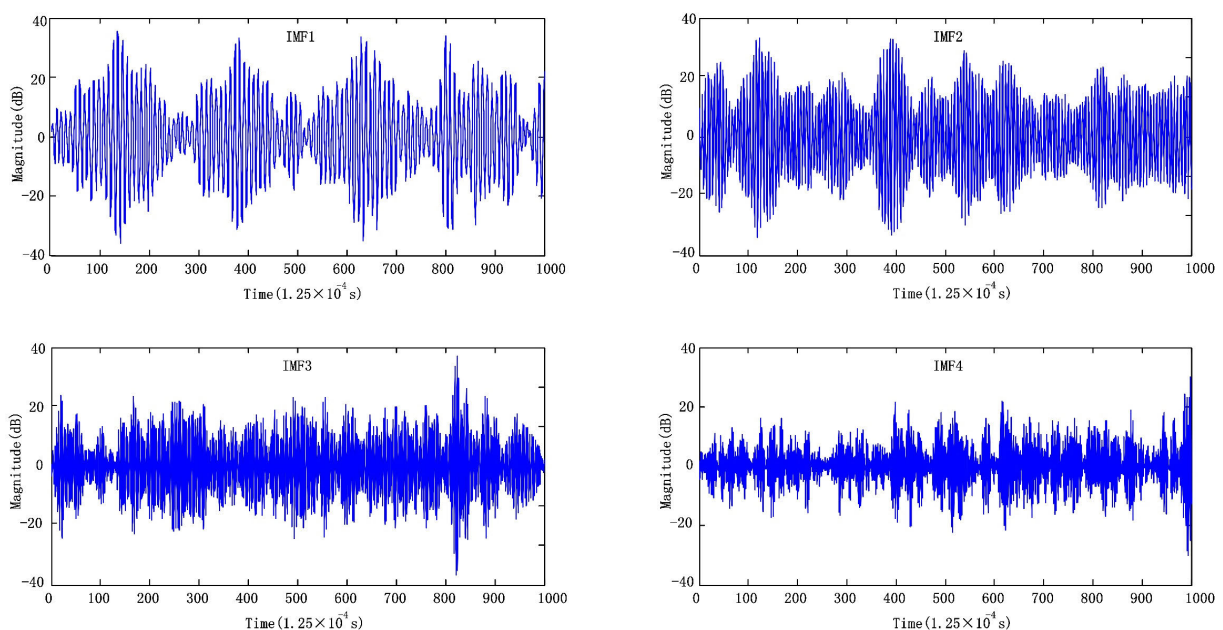


FIGURE 3. VMD of sound signals in normal condition (one sample with 1000 data points).

**TABLE 2. Definitions of common statistics in time-domain for motor current and sound signals.**

| Time-domain feature   | Equation  | Time-domain feature | Equation  |
|-----------------------|---|---------------------|---|
| 1. Mean               | $\frac{1}{N} \sum_{i=1}^N x_i$                        | 6. Kurtosis         | $\frac{\sum_{i=1}^N (x_i - x_m)^4}{(N - 1)x_{std}^4}$ |
| 2. Standard deviation | $\sqrt{\frac{\sum_{i=1}^N (x_i - x_m)^2}{N - 1}}$     | 7. Crest factor     | $\frac{x_{pk}}{x_{rms}}$                              |
| 3. Root mean square   | $\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$               | 8. Clearance factor | $\frac{x_{pk}}{(\frac{1}{N} \sum_{i=1}^N  x_i )}$     |
| 4. Peak               | $\max x_i $   | 9. Shape factor     | $\frac{x_{rms}}{\frac{1}{N} \sum_{i=1}^N  x_i }$      |
| 5. Skewness           | $\frac{\sum_{i=1}^N (x_i - x_m)^3}{(N - 1)x_{std}^3}$ | 10. Impulse factor  | $\frac{x_{pk}}{\frac{1}{N} \sum_{i=1}^N  x_i }$       |

**E. NUMBER OF EXTRACTED FEATURES**

TABLE 3 indicates the number of extracted features after using PSO-VMD+SampEn and TDSF. As shown in TABLE 3, the sound signal contains 14 features which consist of 4 PSO-VMD+SampEn features (i.e. IMF1, IMF2, IMF3, IMF4) and 10 TDSF features which is calculated by the equations in TABLE 2. Winding A current contains 10 TDSF features, whereas number of feature of Winding B current is the same as that of Winding A current. All of these 34 features are used altogether to import to Broad Learning algorithm. To ensure that all these features have a uniform contribution, all of them are normalized to [0, 1].

**TABLE 3. Number of extracted features from different signal types.**

| Type of induction motor signal | PSO-VMD+SampEn | TDSF | Total features |
|--------------------------------|----------------|------|----------------|
| Winding A current signal       | 0              | 10   | 10             |
| Winding B current signal       | 0              | 10   | 10             |
| Sound signal                   | 4              | 10   | 14             |

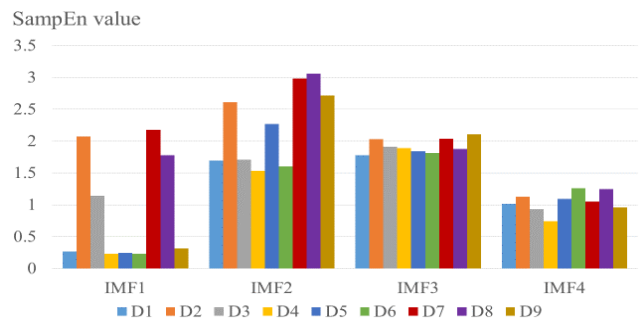
**V. EXPERIMENTAL RESULTS AND DISCUSSIONS**

Since there are many combinations of the techniques for feature extraction and machine learning, a set of experiments was carried out on a computer to determine the best combination for the proposed framework.

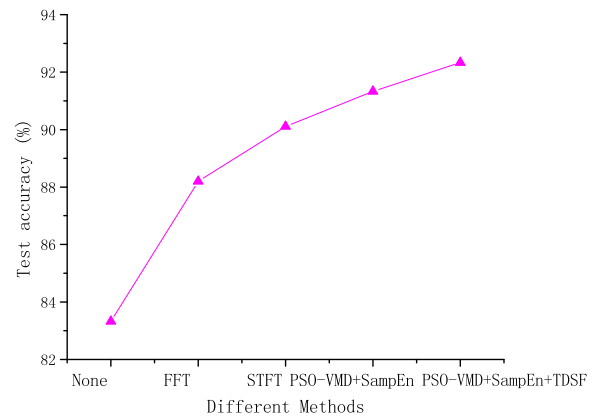
**A. RESULT AND DISCUSSION OF DIFFERENT FEATURE EXTRACTION METHODS**

There are many feature extraction methods for data classification, e.g. FFT, STFT, PSO-VMD+SampEn, PSO-VMD+SampEn+TDSF [36]. Their parameters must

be set before use. The number of TDSF of Winding A, Winding B and sound is set to 10 [35]. For FFT, the sample frequency is set to 8000, the sample points are the same as the time domain which is set as 1000 [16]. For STFT, the window function is set as a Hamming window [17]. For PSO-VMD,  $K$  and  $\alpha$  are optimized to 4 and 50 [20]. For SampEn,  $m$  and  $r$  are set to 2 and 0.2, respectively according to [37]. FIGURE 5 shows that the test accuracy is the highest for the PSO-VMD+SampEn+TDSF method, while the other feature extraction methods are lower, especially for methods that do not use any feature extraction. The main reason is that PSO-VMD is a self-adaptive time-frequency technique. It can decompose the signal into several IMFs and vary with the signal itself. SampEn can reduce dimension by calculating representative features for each IMF. TDSF can enhance the diagnostic accuracy through time-domain statistical features. So, the diagnostic accuracy of PSO-VMD+SampEn+TDSF is better than the other approaches. (Note: The experiment was performed in BL with 70 feature nodes and 240 enhancement nodes).



**FIGURE 4. SampEn value of each IMF in 9 faults.**



**FIGURE 5. Comparison of different feature extraction methods.**

**B. RESULT AND DISCUSSION OF DIFFERENT CLASSIFICATION**

We compare different recent classification methods, including Deep Belief Network (DBN), Convolutional Neural Network (CNN), Extreme Learning Machine (ELM), Broad Learning (BL), and Feature Incremental Broad Learning (FIBL). For DBN, we use a 5400-100-240-9 structure with

100 iterations and set a learning rate to 0.1 [38]. For CNN, we use the original deep CNN structure (LeNet-5) [39], which contains 7 layers. The kernel size of the convolutional layer 1 is  $5 \times 5$ . The combined size of the sampling layer 2 is  $2 \times 2$ . The kernel size of the convolutional layer 3 is  $5 \times 5$ . The combined size of the sampling layer 4 is  $2 \times 2$ . The kernel size of the convolutional layer 5 is  $5 \times 5$ . The number of nodes in the fully connected layer is 84. The number of nodes in the output layer is 9. The learning rate is set to 0.1, and the activation function is sigmoid. For Extreme Learning Machine [40], the number of hidden nodes is set to 310. The activation function is configured as a sigmoid function. For Broad Learning [13], the regularization parameter  $\lambda$  of the ridge regression is set to  $1 \times 10^{(-8)}$ . The sigmoid function is selected to build the enhancement features. For Feature Incremental Broad Learning, we add 10 additional feature nodes in each incremental broad learning. In order to play fair for experimental comparisons, the inputs for DBN and CNN are the same as that for BL which are extracted features listed in TABLE 4.

**TABLE 4. Comparison of diagnostic accuracies and training time for different methods.**

| Method   | Test accuracy (%) | Total training time (s) |
|--|-------------------|-------------------------|
| DBN  | 92.71             | 378.7479                |
| CNN (1 epoch)  | 87.25             | 63.6417                 |
| CNN (10 epochs)  | 92.27             | 642.3905                |
| ELM (310 hidden nodes)   | 70.73             | 0.3362                  |
| ELM (3000 hidden nodes)  | 92.52             | 65.5183                 |
| BL (300 nodes including 60 feature nodes and 240 enhancement nodes)  | 91.83             | 1.2329                  |
| FIBL(310 nodes including 70 feature nodes and 240 enhancement nodes) | <b>92.73</b>      | <b>1.2329+0.0528</b>    |

The experimental result is shown in TABLE 5. The highest test accuracy is FIBL, which is up to 92.73%. The second highest is DBN. However, the training time is the longest with 378.7479 seconds. The training time of ELM, which contains 310 hidden nodes, is the fastest, but the test accuracy is very low. This test accuracy can be increased to 92.52% by adding 3000 hidden nodes to the system, but the training time increases at the same time. In contrast to the above methods, the test accuracy of FIBL is 92.73% and the training time is only 1.2857 seconds. Therefore, FIBL is more efficient.

The reason is that DBN or CNN can improve the test accuracy by a deep structure. However, this takes more time to train the network. ELM runs fast. However, its test accuracy is very low. BL has only two layers. It can improve a test accuracy by feature increment. As shown in TABLE 4, Feature Incremental Broad Learning can improve test accuracy by increasing the feature nodes. Since a mapped feature can

**TABLE 5. Network structure compression by SVD.**

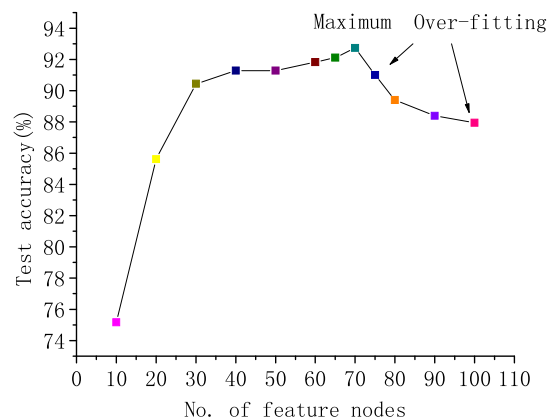
| FIBL         |          |          | SVD-FIBL                      |                 |                 |
|--------------|----------|----------|-------------------------------|-----------------|-----------------|
| ( $\Omega$ ) | MTE(%)   | ATE(%)   | ( $\Omega_1$ )-( $\Omega_2$ ) | MTE(%)          | ATE(%)          |
| (30,240)     | 0.157883 | 1.163821 | (30,240)-<br>(17,240)         | 0.095301        | 0.354399        |
| (40,240)     | 0.093911 | 0.745168 | (40,240)-<br>(26,240)         | 0.048731        | 0.315043        |
| (50,240)     | 0.062781 | 0.531963 | (50,240)-<br>(36,240)         | 0.047178        | 0.244777        |
| (60,240)     | 0.057769 | 0.314961 | (60,240)-<br>(45,240)         | 0.046956        | 0.243691        |
| (70,240)     | 0.054991 | 0.298432 | (70,240)-<br>(55,240)         | <b>0.037882</b> | <b>0.224981</b> |

**Remarks:** ( $\Omega$ ) denotes (Number of feature nodes, Number of enhancement nodes); ( $\Omega_1$ )-( $\Omega_2$ ) denotes the structure is simplified from ( $\Omega_1$ ) to ( $\Omega_2$ ).

be added to create a new structure of the Broad Learning network to improve its accuracy. However, the FIBL doesn't need to change the entire network structure. It only adjusts the weights of the incremental feature nodes. Therefore, by adding 10 feature nodes, the FIBL can reach 92.73% and the retraining time is only 0.0528 seconds. That is very effective.

### C. RESULT AND DISCUSSION OF FEATURE INCREMENTAL BROAD LEARNING

Although Feature Incremental Broad Learning can effectively improve test accuracy by adding feature nodes, when the number of feature nodes is increased too much, the FIBL network will suffer from over-fitting. Feature nodes should, therefore, be optimized to an appropriate number. In this experiment, the number of enhancement node is set to 240. The number of feature nodes is initially specified as 10 and increased by 10 nodes per step. The number increased 10 times to 100. FIGURE 6 indicates that the original BL network has the lowest test accuracy. The maximum accuracy is 92.73% when the number of feature nodes reaches 70.



**FIGURE 6. Test accuracy of feature incremental broad learning.**



It drops as the number of feature nodes increases to 80. This is due to overfitting. Therefore, the optimal number of feature nodes is selected to be 70. The maximum accuracy is 92.73%, which requires initial training and six times of incremental retraining. (Note: The network needs more available training data during every incremental training process. We generate these training data from raw signal data by using the random distribution function which is defined in MATLAB.)

In order to ensure that the best feature node number is close to 70, more additional numbers of feature nodes between 60 and 80 are also tested. Starting with 60, the model is further expanded with additional 5 feature nodes at every iteration until reaching 80. The interval between 60 and 80 in Figure 6 shows that the extra test result and the optimal number of feature nodes is exactly 70.

#### D. RESULT AND DISCUSSION OF RETRAINING TIME

To prove the efficiency of the retraining of Feature Incremental Broad Learning, the training time and retraining time are introduced to the same enhancement node for comparison (the number of enhancement nodes is set to 240). FIGURE 7 presents that the initial training time is the longest. It takes 0.9398 seconds for the network training. However, retraining time is shorter in the next nine steps of Feature Incremental Broad Learning. The time required for the nine steps is 0.0465s, 0.0771s, 0.0656s, 0.0423s, 0.0616s, 0.0528s, 0.0533s, 0.0868s, 0.0657s respectively. It is because the FIBL algorithm does not need to change its original weights. It only needs to calculate the weights of the incremental nodes. Therefore, its retraining time is faster. The total training time for weight optimization is 1.2887 seconds, which includes 7 steps (0.9388s, 0.0465s, 0.0771s, 0.0656s, 0.0423s, 0.0616s, 0.0528s).

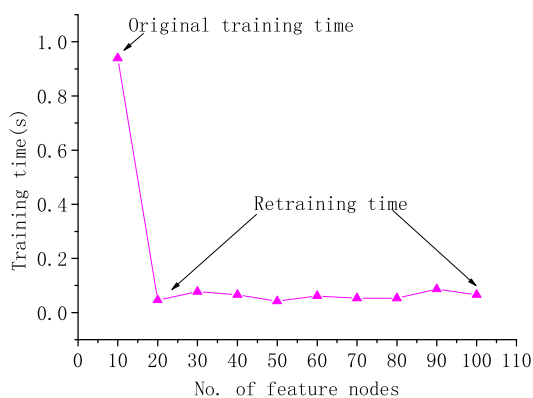


FIGURE 7. Training time of feature increment broad learning.

#### E. RESULT AND DISCUSSION OF STRUCTURE SIMPLIFICATION

This study uses SVD to compress the FIBL structure to reduce its error. TABLE 5 shows the comparison between FIBL (original structure without compression) and SVD-FIBL (new structure with SVD compression).

In TABLE 5, the symbol  $\Omega$  represents the network structure of FIBL in which the first number represents the number of feature nodes and the second number represents the number of enhancement nodes. The column SVD-FIBL in TABLE 5 means the network is compressed by SVD from an original structure  $\Omega_1$  to a new structure  $\Omega_2$ . Both FIBL and SVD FIBL methods are repeated ten times. Then, the minimum test error (MTE) and average test error (ATE) are recorded in TABLE 5.

*Remarks:* ( $\Omega$ ) denotes (Number of feature nodes, Number of enhancement nodes); ( $\Omega_1$ )-( $\Omega_2$ ) denotes the structure is simplified from ( $\Omega_1$ ) to ( $\Omega_2$ ).

TABLE 5 reveals that both MTE and ATE of FIBL and SVD-FIBL are generally reduced as the total number of nodes is increased. The MTE & ATE of SVD-FIBL reach the maximum values of 0.037882% and 0.224981% respectively when the structure of the FIBL network is simplified from (70,240) to (55,240). In general, the more the nodes, the better the accuracy (except for overfitting networks). TABLE 5 also shows that both MTE and ATE in SVD-FIBL are smaller than those in FIBL. It means that the SVD can effectively reduce redundant nodes. Therefore, the simplified structure is more concise and stable.

#### VI. CONCLUSIONS

In this paper, a framework for three-phase induction motor fault diagnosis is proposed that combines feature extraction, Feature Incremental Broad Learning, and SVD-FIBL. Firstly, raw sample data from Winding current A, Winding current B and acoustic signals are collected. Subsequently, the raw sample data are processed by using filters, PSO-VMD, SampEn, TDSF, and Normalization. Secondly, this processed data are import into Broad Learning to train the network. Then, the network is continually trained through Feature Incremental Broad Learning until the test accuracy is satisfactory. Finally, the network structure is simplified by SVD. The experimental results show that the Feature Incremental Broad Learning and SVD can improve the diagnostic accuracy and training speed. The contributions of this research are as follows:

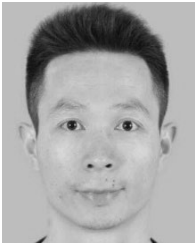
1. A feature extraction with PSO-VMD is first used in induction motor fault diagnosis to improve the accuracy of the diagnostic system.
2. A Feature Incremental Broad Learning (with feature nodes) method is proposed to optimize the system network to improve the diagnostic system both in accuracy and training speed.
3. SVD is successfully used to simplify the FIBL structure to further reduce the test error of this diagnostic system.

The proposed fault diagnostic method is suitable for three-phase induction motors and other similar motors such as DC motors or permanent magnet synchronous motors. However, some points should be improved in the future. Firstly, it takes more attempts to choose a better feature extraction method and the feature extraction method of PSO-VMD+SampEn+TDSF is perhaps not the best.

Secondly, the number of feature nodes is manually tested in the Feature Incremental Broad Learning process. It can be tested automatically by developing a proper optimization algorithm. Thirdly, we can try to increase the number of inputs instead of adding the number of feature nodes to improve system performance. Finally, we need more comparisons with many existing activation functions instead of the sigmoid function in FIBL network.

## REFERENCES

- [1] E. Levi, "Multiphase electric machines for variable-speed applications," *IEEE Trans. Ind. Electron.*, vol. 55, no. 5, pp. 1893–1909, May 2008.
- [2] J. Faiz, B. M. Ebrahimi, and M. B. B. Sharifian, "Different faults and their diagnosis techniques in three-phase squirrel-cage induction motors—A review," *Electromagnetics*, vol. 26, no. 7, pp. 543–569, Oct. 2006.
- [3] J. Faiz and M. Ojaghi, "Different indexes for eccentricity faults diagnosis in three-phase squirrel-cage induction motors: A review," *Mechatronics*, vol. 19, no. 1, pp. 2–13, Feb. 2009.
- [4] P. K. Wong, J. H. Zhong, Z. X. Yang, and C. M. Vong, "A new framework for intelligent simultaneous-fault diagnosis of rotating machinery using pairwise-coupled sparse Bayesian extreme learning committee machine," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 231, no. 6, pp. 1146–1161, 2016.
- [5] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.
- [6] C. L. P. Chen, C.-Y. Zhang, L. Chen, and M. Gan, "Fuzzy restricted Boltzmann machine for the enhancement of deep learning," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 2163–2173, Dec. 2015.
- [7] P. Konar and P. Chattopadhyay, "Bearing fault detection of induction motor using wavelet and support vector machines (SVMs)," *Appl. Soft Comput.*, vol. 11, no. 6, pp. 4203–4211, Sep. 2011.
- [8] J.-H. Zhong, P. K. Wong, and Z.-X. Yang, "Fault diagnosis of rotating machinery based on multiple probabilistic classifiers," *Mech. Syst. Signal Process.*, vol. 108, pp. 99–114, Apr. 2018.
- [9] G. Wang, "Design of damage identification algorithm for mechanical structures based on convolutional neural network," presented at the Int. Conf. Modern Comput. Sci. Appl., Wuhan, China, 2018.
- [10] S.-Y. Shao, W.-J. Sun, R.-Q. Yan, P. Wang, and R. X. Gao, "A deep learning approach for fault diagnosis of induction motors in manufacturing," *Chin. J. Mech. Eng.*, vol. 30, no. 6, pp. 1347–1356, 2017.
- [11] W. Sun, S. Shao, R. Zhaob, R. Yana, X. Zhangc, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, Jul. 2016.
- [12] I. Bravo-Imaz, H. D. Ardakani, Z. Liu, A. Garcia-Arribas, A. Arnaiz, and J. Lee, "Motor current signature analysis for gearbox condition monitoring under transient speeds using wavelet analysis and dual-level time synchronous averaging," *Mech. Syst. Signal Process.*, vol. 94, pp. 73–84, Sep. 2017.
- [13] C. L. P. Chen and Z. L. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.
- [14] S. B. Jiang, P. K. Wong, R. Guan, Y. Liang, and J. Li, "An efficient fault diagnostic method for three-phase induction motors based on incremental broad learning and non-negative matrix factorization," *IEEE Access*, vol. 7, pp. 17780–17790, 2019.
- [15] G. S. Maruthi and V. Hegde, "Application of MEMS accelerometer for detection and diagnosis of multiple faults in the roller element bearings of three phase induction motor," *IEEE Sensors J.*, vol. 16, no. 1, pp. 145–152, Jan. 2016.
- [16] T. Ameid, A. Menacer, H. Talhaoui, and I. Harzelli, "Broken rotor bar fault diagnosis using fast Fourier transform applied to field-oriented control induction machine: Simulation and experimental study," *Int. J. Adv. Manuf. Technol.*, vol. 92, nos. 1–4, pp. 917–928, Sep. 2017.
- [17] L. H. Wang, X. P. Zhao, J. X. Wu, Y. Y. Xie, and Y. H. Zhang, "Motor fault diagnosis based on short-time Fourier transform and convolutional neural network," *Chin. J. Mech. Eng.*, vol. 30, pp. 1357–1368, Nov. 2017.
- [18] J. Cusidó, L. Romeral, J. A. Ortega, J. A. Rosero, and A. G. Espinosa, "Fault detection in induction machines using power spectral density in wavelet decomposition," *IEEE Trans. Ind. Electron.*, vol. 55, no. 2, pp. 633–643, Feb. 2008.
- [19] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
- [20] X.-B. Wang, Z.-X. Yang, and X.-A. Yan, "Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 68–79, Feb. 2018.
- [21] Z.-H. Zhan, J. Zhang, Y. Li, and H. S.-H. Chung, "Adaptive particle swarm optimization," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 6, pp. 1362–1381, Dec. 2009.
- [22] J.-R. Zhang, J. Zhang, T.-M. Lok, and M. R. Lyu, "A hybrid particle swarm optimization-back-propagation algorithm for feedforward neural network training," *Appl. Math. Comput.*, vol. 185, no. 2, pp. 1026–1037, Feb. 2007.
- [23] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, no. 10, pp. 3102–3112, Dec. 2015.
- [24] X. An and H. Zeng, "Fault diagnosis method for spherical roller bearing of wind turbine based on variational mode decomposition and singular value decomposition," *J. Vibroeng.*, vol. 18, no. 6, pp. 3548–3556, Sep. 2016.
- [25] S. W. Fei, "Fault diagnosis of bearing based on wavelet packet transform-phase space reconstruction-singular value decomposition and SVM classifier," *Arabian J. Sci. Eng.* vol. 42, no. 5, pp. 1967–1975, May 2017.
- [26] G. Cablea, P. Granjon, and C. Berenguer, "Three-phase electrical signals analysis for mechanical faults monitoring in rotating machine systems," *Mech. Syst. Signal Process.*, vol. 92, pp. 278–292, Apr. 2017.
- [27] S. Solecki, "Filters and sequences," *Fundamenta Mathematicae*, vol. 163, no. 3, pp. 215–228, Jan. 2000.
- [28] Z.-X. Yang, X.-B. Wang, and P. K. Wong, "Single and simultaneous fault diagnosis with application to a multistage gearbox: A versatile dual-ELM network approach," *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5245–5255, Dec. 2018.
- [29] R. C. Eberhart, D. J. Groves, and J. K. Woodward, "Deep swarm: Nested particle swarm optimization," presented at the IEEE Symp. Ser. Comput. Intell. (SSCI), Honolulu, HI, 2017.
- [30] J. Kennedy and R. Eberhart, "Particle swarm optimization," presented at the IEEE Int. Conf. Neural Netw. (ICNN), Univ. Western Australia, Perth, WA, Australia, 1995.
- [31] Y. C. Liang, H. P. Lee, S. P. Lim, W. Z. Lin, K. H. Lee, and C. G. Wu, "Proper orthogonal decomposition and its applications—Part I: Theory," *J. Sound Vibrat.*, vol. 252, no. 3, pp. 527–544, May 2002.
- [32] A. Bellini, F. Filippetti, C. Tassoni, and G.-A. Capolino, "Advances in diagnostic techniques for induction machines," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4109–4126, Dec. 2008.
- [33] A. Glowacz, "Acoustic based fault diagnosis of three-phase induction motor," *Appl. Acoust.*, vol. 137, pp. 82–89, Aug. 2018.
- [34] V. Hegde and M. G. S. Rao, "Detection of stator winding interturn short circuit fault in induction motor using vibration signals by MEMS accelerometer," *Electr. Power Compon. Syst.*, vol. 45, no. 13, pp. 1463–1473, Nov. 2017.
- [35] A. Khorshidtalab, M. J. E. Salami, and M. Hamed, "Robust classification of motor imagery EEG signals using statistical time-domain features," *Physiol. Meas.*, vol. 34, no. 11, pp. 1563–1579, Nov. 2013.
- [36] J. Liang, Y. Zhang, J.-H. Zhong, and H. Yang, "A novel multi-segment feature fusion based fault classification approach for rotating machinery," *Mech. Syst. Signal Process.*, vol. 122, pp. 19–41, May 2019.
- [37] W. Zhang, Y. Pu, and J. Zhu, "Gear fault diagnosis method using EEMD sample entropy and grey incidence," presented at the 4th Int. Conf. Manuf. Sci. Eng. (ICMSE), Dalian, China, 2013.
- [38] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.
- [39] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.
- [40] Z.-X. Yang, X.-B. Wang, and J.-H. Zhong, "Representational learning for fault diagnosis of wind turbine equipment: A multi-layered extreme learning machines approach," *Energies*, vol. 9, no. 6, pp. 379–396, 2016.



**SAI BIAO JIANG** received the M.Sc. degree from the Guangdong University of Technology, in 2011. He is currently pursuing the Ph.D. degree in electromechanical engineering with the University of Macau. He is also a Lecturer with the Zhuhai College of Jilin University. His research interests include data-driven fault diagnosis, feature extraction, and artificial intelligence.



**PAK KIN WONG** received the Ph.D. degree in mechanical engineering from The Hong Kong Polytechnic University, Hong Kong, in 1997. He is currently a Professor with the Department of Electromechanical Engineering and the Associate Dean (Academic Affairs) of the Faculty of Science and Technology, University of Macau. He has published over 225 scientific articles in refereed journals, book chapters, and conference proceedings. His research interests include automotive engineering, fluid transmission and control, artificial intelligence (AI), mechanical vibration, and AI for medical applications.



**YAN CHUN LIANG** received the Ph.D. degree in applied mathematics from Jilin University, Changchun, China, in 1997. He was a Visiting Scholar with the University of Manchester, U.K., from 1990 to 1991, a Visiting Professor with the National University of Singapore, from 2000 to 2001, a Guest Professor with the Institute of High Performance Computing of Singapore, from 2002 to 2004, a Guest Professor with Trento University, Italy, from 2006 to 2008, and a Visiting Professor with Missouri University, Columbia, MO, USA, from 2010 to 2018. He is currently a Professor with the College of Computer Science and Technology, Jilin University, and also a Professor with the School of Computer, Zhuhai College of Jilin University, China. His research interests include computational intelligence, machine learning methods, and bioinformatics. He has published over 400 articles. His research was featured in the *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, the *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, the *Journal of Micromechanics and Micro Engineering*, the *Physical Review E*, and the *Bioinformatics*.

• • •