KSME

Springer

Journal of Mechanical Science and Technology 34 (11) 2020

Original Article
DOI 10.1007/s12206-020-1002-x

Keywords:
· Rolling element bearings
· Improve intelligent diagnostics
· Deep functional auto-encoder
· Unsupervised learning enhancement
· Massive raw data analysis

Correspondence to:
Jianping Xuan
jpxuan@hust.edu.cn

Citation:
Aljemely, A. H., Xuan, J., Jawad, F. K. J., Al-Azzawi, O., Alhumaima, A. S. (2020). A novel unsupervised learning method for intelligent fault diagnosis of rolling element bearings based on deep functional auto-encoder. Journal of Mechanical Science and Technology 34 (11) (2020) 4367~4381.
http://doi.org/10.1007/s12206-020-1002-x

Received December 3rd, 2019
Revised February 12th, 2020
Accepted August 4th, 2020

† Recommended by Editor
No-cheol Park

A novel unsupervised learning method for intelligent fault diagnosis of rolling element bearings based on deep functional auto-encoder

Anas H. Aljemely[1], Jianping Xuan[1], Farqad K. J. Jawad[2], Osama Al-Azzawi[2] and Ali S. Alhumaima[3]

[1]School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China, [2]School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan 430074, China, [3]School of Electronic Engineering and Computer Science, South Ural State University, Chelyabinsk 454016, Russia

Abstract Recently, several studies tried to develop fault identification models for rolling element bearing based on unsupervised learning techniques. However, an accurate intelligent fault diagnosis system is still a big challenge. In this study, a deep functional auto-encoders (DFAEs) model with SoftMax classifier was designed for valuable feature extraction from massive raw vibration signals. To maximize the unsupervised feature learning ability of the proposed model, various activation functions were applied in an effective methodology, these hidden activation functions enhance significantly the sparsity of the training data-set. The proposed method was validated using the raw vibration signals measured from the machine with different bearing conditions. The achieved results showed that the high-superiority of the proposed model comparing to standard deep learning and other traditional fault diagnosis methods in terms of classification accuracy even with massive input data sets.

1. Introduction

Rotating machinery is extensively used in the newly industry. The key components of the rotating machinery can develop diverse faults under rigorous working situations such as large load, strong impact, high speed, and high background noise [1, 2]. Rolling element bearings are vital components of most rotating machinery and electrical apparatuses. The latter's failures may cause significant losses and serious economic casualties. Therefore, it is worth to diagnose the different bearing faults precisely and automatically that may occur in rotating machinery. Intelligent fault diagnosis of rolling element bearings is an exemplary pattern recognition problem. It designed to analyse the measured vibration data excellently and automatically deliver the diagnosis performance that indicated in classification results. Intelligent fault diagnosis became a new concern directional in the health monitoring (HM) applications [3]. Traditional diagnosis models usually show shallow feature learning in big data. Artificial neural networks (ANNs) and support vector machines (SVMs) are two common traditional methods that used to solve rolling bearings fault diagnosis classification problems and to extract the sensitive and meaningful characterized features from collected vibration signals [4-8]. The developed models of the ANNs used for observing and detecting the health circumstances of machinery tools, such as electric motors, rolling element bearings, gearboxes, etc. The ellipsoid-ARTMAP model and hybrid FAM model improved the network structures which widely utilized to raise the classification performance. These methods supposed that the feature parameters exhibited to FAM had significant point to the categories. However, a big workload of expense and low diagnosis accuracy may occur when fed whole learning features together to models, and the likelihood of some features are irrelevant or extra to the classification architecture [9]. Kernel marginal

Fisher analysis (KMFA) model was utilized in the data learning field, which can extract low-dimensional features rooted in the high-dimensional sphere. Nevertheless, it is unable to resolve the overlap problem in supervised fault diagnosis methods because the supervised learning methods always need a large intake of termed data for fault recognition and it is not available in most work cases [10]. The MODWPT approach is used for rolling vibration signals, and then the cumulative FER value of each frequency band signal is determined and used to produce the FERgram to be used for envelope analysis [11]. The ODL algorithm was used to enhance the fault features extraction. The sparse coding period misses the power to learn information signals, so a large amount of intermingled redundant hype is still in the signal after the sparse coding period. Thus, the standard ODL algorithm cannot effectively explore weak fault feature consolidation of rolling bearings [12]. In Ref. [13], the singular value decomposition (SVD) technique is used to provide a tool for automatically extracting abrupt information from aero-engine signals to identify typical faults. However, most traditional intelligent diagnosis methods still working on shallow architectures lack powerful representation in fault diagnosis tasks and have two intrinsic limitations [14]: feature extraction and selection. These limitations are the main duty of traditional methods to evaluate the diagnosis performance [15]. The measured vibration signals of various fault types, various fault size and various fault locations are always non-linear, and non-stationary escorted high background noise [16, 17]. It is a complex mission to specify the most delicate attributes from the unique characteristic set for diverse diagnosis executions by depending on only engineering experience. Thus, several advanced signal processing methods must be worked versatility to extract useful features. Therefore, powerful feature learning models are seriously necessary to pick up automatically that useful fault features and ensure fitting results. Consequently, this process leads to depreciated time in analyzing tasks to select the most delicate attributes in different diagnosis methods without plentiful prior acquaintances [18]. According to Refs. [19, 20], traditional intelligent fault diagnosis methods cannot eliminate high intricacy pattern recognition problems by applying shallow architecture learning models [21]. Accordingly, there is an imperative issue needed to develop deep unsupervised feature learning architecture to achieve a better intelligent fault diagnosis system. In order to conquer the substance inadequacies of traditional intelligent diagnosis methods, deep learning is an advanced approach of the artificial intelligence (AI) field and it is a good choice to learn the data [22]. The core objective of most deep learning models is that they can automatically learn decisive features from the raw vibration signals directly [23-25]. A deep belief network (DBN) is one of the deep leaning methods used to learn various data sets to extract the features. Also, the convolutional neural network (CNN) and deep auto-encoder (DAE) have same advantages to eliminate the reliance on a variety of developed signal processing techniques and engineering prior-knowledge [26], which can cause more concern and has been increasingly applied in health monitoring systems for near passed years [27-29]. DBN is used to prevent the intricate layout of a deep neural network and to detect roller bearing faults [30]. DAE is a perfectly unsupervised feature learning method unlike DBN and CNN. It is created from manifold auto-encoders (AEs), which has been a more effective and easier learned model in various applications [31]. However, there still exist some problems and many challenges when using the standard DAE directly for intelligent bearing fault diagnosis. These problems can be defined as a shallow training performance and low generalization ability. Standard DAE usually uses a sigmoid function as the activation function in feature learning stages which probably cannot re-presentative the sufficient information from the raw input signals, and it is incapable of generating a precise mapping association between the many classes and the input signals [32].

In most traditional practical engineering, the effectiveness of a standard DAE run decreases gradually because of the measured vibration signals are invariably non-stationary and non-linear with high background of redundant information. Also, it is influenced by different operating conditions [33, 34]. However, existing deep learning models still rely on the functioning research of the standard individual models and manual signal processing techniques. As well as, the capability of automatically capturing useful information from raw vibration data will be rare or ignored, which makes a complex issue that perform a proper deep learning model for input data-set. Therefore, it is useful to upgrade the standard DAE and create a new deep learning model that can yield full privileges of deep learning functionality.

In this paper, a deep functional auto-encoders (DFAEs) model with SoftMax classifier was designed to extract valuable features from massive raw vibration signals. In order to increase the ability of the unsupervised feature learning, different activation functions were implemented in an effective methodology. The designed method is proposed in this paper to describe three major steps: First, various activation functions were used to design each functional auto-encoder (FAE) of the deep learning model. Second, DFAEs are constructed with multiple FAEs for more reinforced of unsupervised feature learning. Finally, SoftMax is utilized to get high accurately classification for the normal and defective bearing. The proposed method was applicable to analyze the empirical measured raw vibration signals of the testing bearings, it was effectively confirmed to overcome the limitations mentioned above by comparing with existing intelligent diagnosis methods and successfully eliminated the reliance on manual feature extraction.

The remainder of this paper is organized as follows. The basic concepts of standard auto-encoder are concisely described in Sec. 2. The proposed method is detailed in Sec. 3. In Sec. 4, the experimental diagnosis results of a bearing data set are studied and deliberated. Finally, in Sec. 5 the conclusions and additional work are summarized.

2. Basic concepts of auto-encoder

An autoencoder (AE) is an asymmetrical purely unsupervised feature learning neural network which is assembled from three key layers. These layers represent input, hidden and output layers in sequential [35, 36]. The AE attempts to diminish the reconstruction error of the learning data between these layers. The typical structure of AE is shown in Fig. 1. The raw input data are unlabeled training samples, described as $x=[x_1,x_2,...,x_m]^T$. The sigmoid function $sigm(.)$ is the general transfer function of the standard AEs that diverts the input data into a new vector of hidden representation, $\mathrm{h}=[h_1,h_2,...,h_p]^T$, as follows:

$$\mathbf{h}=sigm(Wx+b) \tag{1}$$

$$sigm=1/(1+e^{-s}) \tag{2}$$

where W is denoted the weight matrix of NN, b is denoted the bias vector and $\theta=\{\mathrm{W},\mathrm{b}\}$ represented the generated parameter assign between the first layer and second layer. Then, the hidden representation vector, $\mathbf{h}$, is mapping reformatted into a reconstruction data vector, $x^*=[x_1^*,x_2^*,...,x_m^*]^T$, through $sigm$ process as below:

$$\mathbf{x}^*=sigm(W^*\mathbf{h}+b^*) \tag{3}$$

where $\theta^*=\{\mathrm{W}^*,\mathrm{b}^*\}$ represented the generated parameter index between the second (hidden) layer and last (output) layer.

The optimization task of the parameter assign $\theta^*=\{\theta,\theta^*\}=\{\mathrm{W},\mathrm{b},\mathrm{W}^*,\mathrm{b}^*\}$ is performed in the AE training to minimize the reconstruction error. The loss function of the standard AE usually was the mean square error (MSE) used to shape the training reconstructed error. For unlabelled training sample set $\{x^1,x^2,...,x^m,x^M\}$, the training error is defined as:

$$E_{AE}(\theta)=1/M\sum_{i=1}^{m}(1/2\sum_{i=1}^{d}(x_i^{*m}-x_i^m)^2) \tag{4}$$

where $x^m=[x_1^m,x_2^m,...,x_i^m,x_d^M]^T$, $m=(1,2,...,M)$ is *m*th input sample vector, *M* is the number of unlabelled input data of training samples and for each sample assigned d dimensional. x_i^m is the *i*th dimension input data of the *m*th training sample x^m and x_i^{*m} is the *i*th dimension outputs that reconstructed for the *m*th samples.

3. The proposed method

The (DFAEs) with SoftMax classifier combination is proposed based on designing the deep functional auto-encoders to develop the intelligent bearing fault diagnosis. The combination involves the following procedures: DFAE construction, fault pattern recognition-based fine-tuning and SoftMax classifier, and the main strategy of the proposed method.

3.1 DFAE construction

A single standard AE has both simplicity in regulation and ability in unsupervised feature learning for massive measured vibration data. Simple transactions in standard AE cause difficulties in parameter selection for deep learning features. Adopting various activation functions plays a significant role in nonlinear modelling ability, and it has potential to overcome the restrictions of a standard AE. This role is important and effective in the performance of deep neural networks [37, 38]. Thus, most of these structures are robust in feature learning and can produce a new representation of raw vibration data.

During the past years, many activation functions have been applied in the neural network learning. In the present study, nine dissimilar activation functions are used to design AE models. The equations of the nine activation functions are described in Table 1. The exponential functions, such as the sigmoid (Logsig) function and hyperbolic tangent (TanH) function have been widely applied in the diverse structure of NNs in the past years, however, their implementation always has inadequacies of wasting time through extreme computational and pheasants vanishing issue [38]. Training a linear activation function (purelin) always produces only positive numbers over the entire real number range which cannot dodge the vanishing gradient issue. The Elliott (Elliotsig) function is a higher-speed approximation of TanH with a 0-1 output range which is not zero-centred and sometimes may not be stable [39].

In the deep learning process, the rectified linear unit (ReLU), can be used as a non-exponential function to transfer for data transformation [39, 40]. ReLU can overcome the vanishing gradient issue which allowing models to learn faster and can be implemented more accurately. It can be used in default activation when developing multilayer neural networks. Adopting ReLU may easily be considered one of the few sights in the deep learning uprising, e.g. the techniques that now permit the routine expansion of very deep neural networks. A saturating linear (Satlin) function is one of the well-known NN learning functions. Although NNs with the (Satlin) function may take extensive time to train, their learning ability is better than other networks. In addition, these networks can give better pattern recognition and higher average classification accuracy (ACA) than neural networks with other activation functions [41]. Among the various activation functions, Satlin and ReUL are widely used for neural networks due to their advantages with machinery vibration. According to our experimental tasks (test each activation function with the other activation functions), 9×9 = 81 experiments were conducted to produce a better deep learning model. In this paper, the Satlin activation function and ReUL activation function were adopted as the activation functions in FAE in order to allow more efficient network training than the default function. And as a result, using these functions causes increasing of the sparsity to obtain better useful features from raw data [42]. Their waveforms are shown in Fig. 2.

A DAE with different activation functions has a powerful capability for deep feature learning. Thus, a new method of unsu-

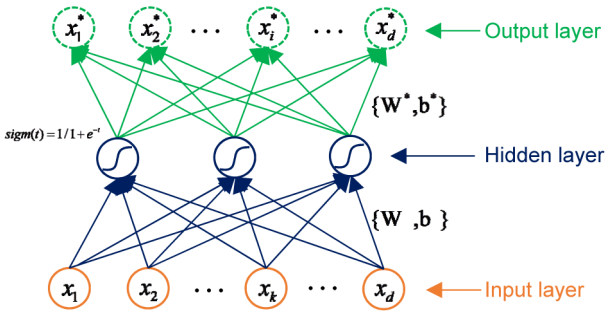


Fig. 1. Standard auto-encoder structure.

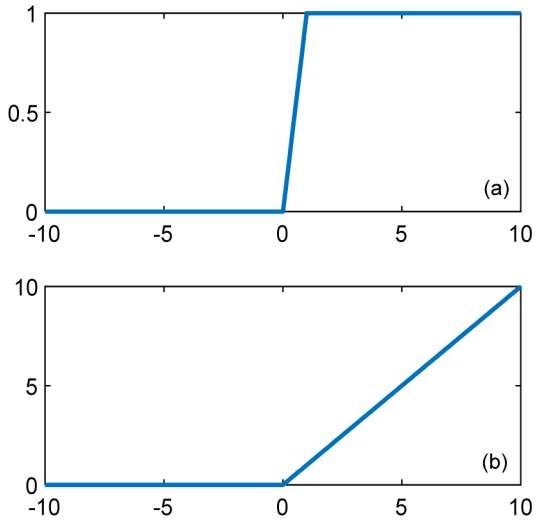


Fig. 2. The two wave forms of activation functions: (a) satlin; (b) ReLU.

Table 1. Equations of the applied activation functions.

Function names	Equations of the functions
Saturating linear (Satlin)	$g(x) = 0$ if $0 \le x$, 1 if $x \ge 1$, x if $0 \le x \le 1$
Rectified linear unit (ReLU)	$g(x) = x/(1+\lvert x \rvert)$
Sigmoid (Logsig)	$g(x) = x/(1+e^{-x})$
Symmetric saturating (Satlins)	$g(x) = -1$ if $-1 \ge x$, 1 if $x \ge 1$, x if $-1 \le x \le -1$
Linear (Purelin)	$g(x) = x$
Radial basis (Radbas)	$g(x) = e^{-x^2}$
Hyperbolic tangent sigmoid (Tansig)	$g(x) = 2/(1+e^{(-2*x)})-1$
Elliot symmetric sigmoid (Elliotsig)	$g(x) = (x/2)/(1+\lvert x \rvert)+0.5$
Sinusoid (sin)	$g(x) = \sin(x)$

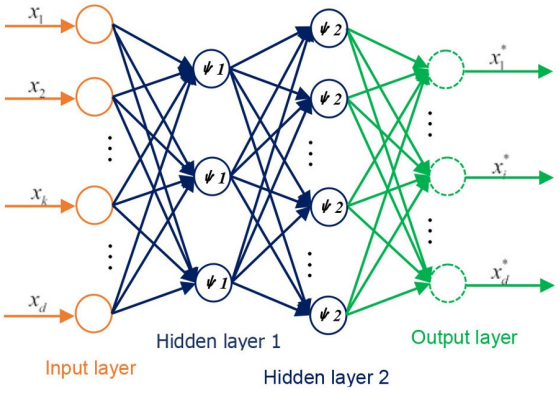


Fig. 3. The structure of DFAEs.

pervised learning, named FAE, is developed in this study. Two of the DAEs with dissimilar activation functions are combined to establish a DFAE model which has a strong potential to capture the representative information of the useful features from raw input data. The structure of the DFAE is shown in Fig. 3. Instead of the traditional sigmoid function, the DFAE model used Satlin and ReLU as the activation functions in the first and second of the AE, respectively. The DFAE model can record different signal properties with different resolutions. FAE is trained to reconstruct the unlabeled sample, into through different activation functions. h = $[h_1, h_2, \ldots, h_m]^T$ is the reconstructed data of the unlabeled sample $x = [x_1, x_2, \ldots, x_m]^T$ as follows:

$$\mathbf{h}_j = \psi_{Sat}\sum_{k=1}^{m}(W_{jk}x_k) \tag{5}$$

where ψ_{Sat} denotes the Satlin activation functions of the first DAE, m is the unit number of the input layer and output layer (*i, k = 1, 2, . . . , m*), x_k is the kth dimension input of the training sample and W_{jk} is the connection weight between input unit k and hidden unit *j*.

From Table 1, $\psi_{\mathrm{Re}} = x_k/(1+|x_k|)$ is the activation function of the second hidden layer. For $h(j=1,\ldots,p)$, the representative data can be calculated similarly to the standard AE formulation. Thus, the FAE2 output can be represented as:

$$x_i^* = \psi_{\mathrm{Re}}\sum_{j=1}^{p}(W_{ij}+b_{ij}) \tag{6}$$

where x_i^* is the ith outputs dimension of the input training samples and W_{ij} is the shared weight between hidden node j and output node *i*.

To improve the sparsity representation, the divergence function of Kullback–Leibler (KL) is adopted to enhance the feature learning in different AEs and is used for measuring the difference between the activation charge of the hidden node and sparse restriction. The definition of the sparsity regularization loss function is explained in Ref. [43]. Therefore, the sparsity loss function can modify the FAE reconstruction error function. For unlabelled sample with m-dimension training examples which set as $\{x^1, x^2, \ldots, x^m, \ldots, x^M\}$, all reconstruction error functions can be rewritten as:

$$E_{AE}(\theta) = 1/M\left[1/2\sum_{i=1}^{d}(x_i^{*m}-x_i^m)^2\right] + \beta\left(\sum_{j=1}^{p}\rho\log\frac{\rho}{\rho_j^*}+(1-\rho)\log\frac{1-\rho}{1-\rho_j^*}\right) \tag{7}$$

where β is the sparse restriction factor, ρ_j^* is the activation rate of hidden node j, ρ is a sparse parameter, x_i^m with ith dimension and the $\mathbf{x}_i^{*m}$ with mth represents the inputs and reconstructed outputs, respectively.

3.2 Fault pattern recognition-based fine-tuning and SoftMax classifier

Basically, N of the hidden layers in the DAEs were trained through stacked DAEs with SoftMax layer training. After entering the vibration data x^m, the mapping of the input layer followed by the first hidden layer can be considered AE1 of the deep network. Through the AE1 training, the reconstruction error in Eq. (7) was minimized based on parameter set θ_1 and the first feature vector ($\mathbf{h}_1^m$) which initialized from the first hidden layer in the DAE network as:

$$\mathbf{h}_1^m = \psi_\theta(x^m). \tag{8}$$

Then, the first feature vector ($\mathbf{h}_1^m$) is considered the input data of the forthcoming hidden layer. Moreover, AE2 is presented the second auto-encoder in the deep network. The second hidden layer was generated through training AE2 to initialize the second feature vector with a different activation function. This process was sequential and conducted until the final hidden layer was initialized for N AEs to train the DAE network. In other words, based on hidden layers, the Nth feature vectors ($\mathbf{h}_N^m$) can be calculated as:

$$\mathbf{h}_N^m = \psi_{\theta_N}(\mathbf{h}_{N-1}^m) \tag{9}$$

where θ_N is denote the W, and b parameters of the series Nth AE. Better classification task can be obtained by pre-training performance effectively.

This performance is demonstrated to produce better local minima than accidental initialization in deep learning algorithms [44, 45]. And then, a fine-tuning operation is conducted in the feature deep learning. The output layer usually works with the output targets for the pattern recognition task. The output of the input signal $\mathbf{x}^m$ for the deep network is described as:

$$\mathbf{y}^m = \psi_{\theta_{N+1}}(\mathbf{h}_N^m) \tag{10}$$

where θ_{N+1} is the output layer parameter.

In order to further output target convergence, backpropagation (BP) is a widespread algorithm which usually used for repeating fed back the data in training the neural network that aims to optimize the parameter set $\theta_{FAEs} = \{\theta_{ij}, \theta_{jk}\} = \{W_{ij}, b_{ij}, W_{jk}, b_{jk}\}$. Therefore, the reconstruction error will be minimized by the BP algorithm during feature training. Based on Eq. (7), further minimizing for the loss function is performed by analyzing the gradient of $E(\theta)$ with respect to W.

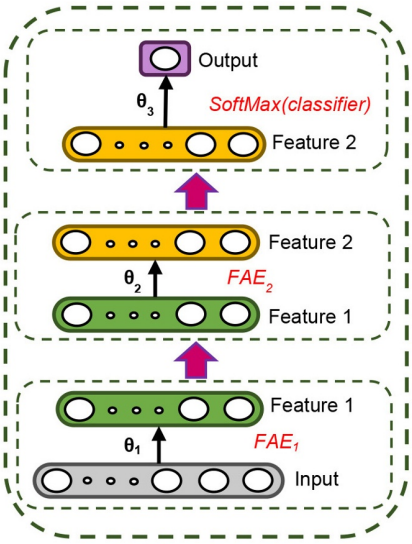


Fig. 4. The structure of DFAEs for two functional auto-encoders.

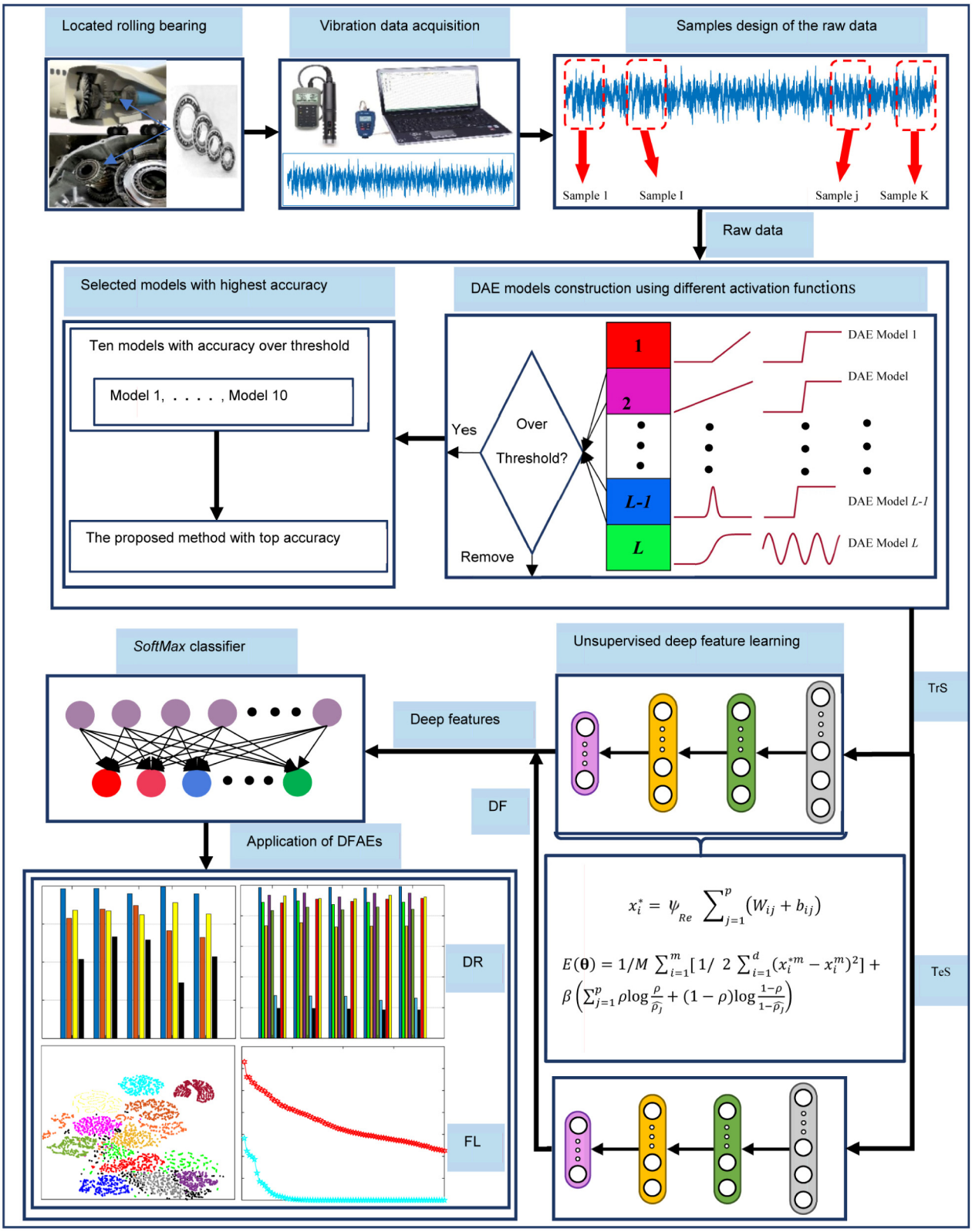
Finally, the updated parameter rules can be stated as:

$$W_{ij}, b_{ij}(I+1) = W_{ij}, b_{ij} - \eta \frac{\partial E(\theta)}{\partial W_{ij}} + \alpha \Delta W_{ij}(I), \Delta b_{ij}(I) \tag{11}$$

$$W_{ik}, b_{ik}(I+1) = W_{ik}, b_{ik} - \eta \frac{\partial E(\theta)}{\partial W_{ik}} + \alpha \Delta W_{ik}(I), \Delta b_{ik}(I) \tag{12}$$

where W and b in Eqs. (11) and (12) are the weight and bias matrix in each AE model, respectively. I is the iterations number, η is a learning rate for the fine-tuning process, α is the momentum factor belong to [0.9, 1] and $E(\theta)$ is the reconstruction error of AE models through each iteration I.

Due to the simplicity of unsupervised feature learning for the AE, it was requisite to build a deep construction based on a string of trained FAEs and to follow Hinton's training procedures. The DFAEs was established with multiple AEs that have different characteristics, and it attained via the sequential training efforts of each single AE. Fig. 4 shows the layer-by-layer structure progression of a DFAE with two FAEs. First, the samples of measured raw vibration signal $\mathbf{x}$ are used as input data to train the first functional autoencoder (FAE1), and then, low-level features $\mathbf{h}_1^m$ are learned. Second, the feature vector $\mathbf{h}_1^m$ is fed into the second tracking functional autoencoder (FAE2) to learn $\mathbf{h}_2^m$ with the highest-dimensional features. The DFAEs training stage is over. Then, the learned features of FAE2 are used to train the SoftMax for much accurate fault identification. This process enables the trained construction to extract the inherent characteristics from raw vibration signals and setting up an intricate non-linear mapping between the learned features and bearing classes. Hence, the designated architecture can effectively achieve better fault characteristic extraction and robust fault classification for rolling element bearings.



Note: DR, FL, DF, TrS and TeS represents the diagnosis result of feature learning, deep features, training samples and testing samples, respectively.

Fig. 5. The illustration of the proposed method.

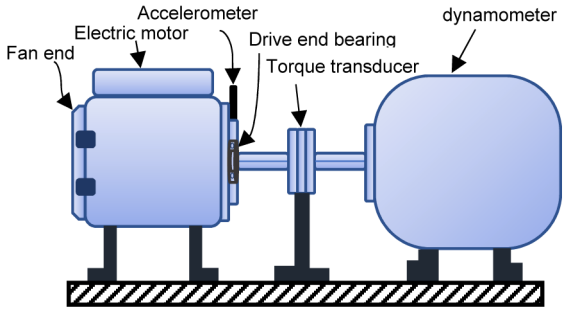


Fig. 6. The rolling bearing experimental setup.

3.3 Main strategy of the proposed method

The proposed method is detailed as a diagram illustrated in Fig. 5 and the fundamental execution steps are defined as follows:

- Step 1: The rolling bearing vibration signals are collected by acceleration sensors and acquisition instruments.
- Step 2: The raw vibration signals are divided into relative training samples and testing samples without any signal pre-processing or manual feature selection and extraction.
- Step 3: Different FAE models are constructed with different pairs of activation functions to build a better DAE.
- Step 4: The model DFAEs is established from two FAEs with Satlin and ReLU in FAE1 and FAE2, respectively.
- Step 5: The model DFAEs is completely trained for unsupervised feature learning using the training samples.
- Step 6: The learned deep features are utilized to train the SoftMax classification layer.
- Step 7: A fine-tuning and BP training algorithm are used to reduce the training error in feature deep learning.
- Step 8: Visualizing the high-dimensional deep learned features using the t-SNE algorithm.
- Step 9: The trained DFAEs is verified using the test sample to prove that the strongly proposed method.

4. Experimental rolling element bearing fault diagnosis

4.1 Experimental rolling element bearing data set

In this study, the bearing vibration data is given by experiments at a Case Western Reserve University (CWRU) testbed [46]. The experimental test-rig consists of four main parts: induction motor, sensor, different defective bearings, and a changeable load motor as shown in Fig. 6. The experiments examined 6205-2RS JEM SKF bearings under diverse loads (0, 1, 2 and 3 hp) and the measured signals corresponding to bearing with fault sizes of 0.1778 mm, 0.3556 mm, 0.5334 mm and 0.7112 mm. One sensor was located near the drive end bearing to collect the vibration signals accurately.

The collected vibration signals at 1772 rpm motor speed (1 hp load) with a sampling frequency of 12 kHz were applied

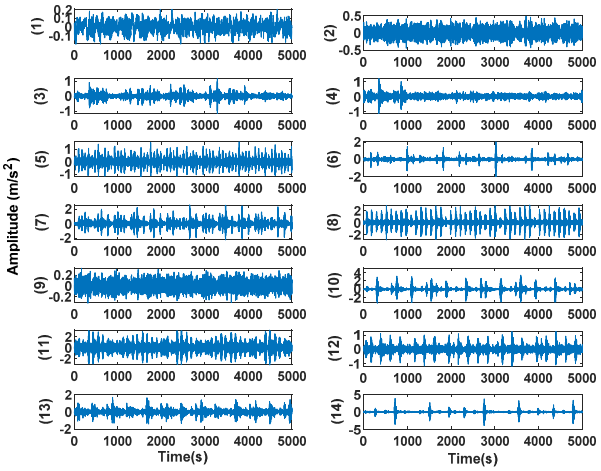


Fig. 7. Vibration signals (5000 data points) of the fourteen rolling bearing operating conditions ((1) Healthy condition, (2) Ball defect condition (0.007), (3) Ball defect condition (0.014), (4) Ball defect condition (0.021), (5) Inner race defect condition (0.007), (6) Inner race defect condition (0.021), (7) Inner race defect condition (0.028), (8) Outer race defect condition (0.007Centered@6:00), (9) Outer race defect condition (0.014Centered @6:00), (10) Outer race defect condition (0.021Centered@6:00), (11) Outer race defect condition (0.007Orthogonal@3:00), (12) Outer race defect condition (0.007Opposite@12:00), (13) Outer race defect condition (0.021Orthogonal@3:00), (14) Outer race defect condition (0.021Opposite @12:00).

for identification of the fault in the DFAEs structure and simulate the real-life applications. The fourteen signals of various bearing working conditions were chosen including different faulty components, fault sizes, and fault trends. 300 samples are created for each signal and each sample contains 400 sampling data points.

For each condition, 210 randomly selected samples were used for training and the residual 90 for testing. Further details of the fourteen bearing signals were described in Table 2. In Fig. 7, the raw time-signals of the first 5000 sampling points for each bearing operation are plotted.

4.2 The creation of the proposed method and diagnosis results

The advantages and superiority of the proposed method can be demonstrated by comparison with three deep learning methods. First, it is compared with different DAE models of two kinds of activation functions. Second, the proposed method is compared with three standard intelligent deep learning methods. A 400-dimensional for each input data set of vibration signals is used to train the networks. In this paper, unsupervised learning is used by the new model of deep learning. The feature extraction is enhanced by the following points:

- The proposed method is wholly dissimilar to the traditional intelligent methods. It is performed without any signal preprocessing or feature extraction for improving the intelligent diagnosis systems.

Table 2. Descriptions of the training and testing samples for fourteen bearing conditions.

Bearing operating conditions	Fault diameter (mm)	Outer race fault orientation	Training samples & testing samples	Condition label
Healthy	0	–	210/90	1
Ball defect	0.1778	–	210/90	2
Ball defect	0.3556	–	210/90	3
Ball defect	0.5334	–	210/90	4
Inner race defect	0.1778	–	210/90	5
Inner race defect	0.3556	–	210/90	6
Inner race defect	0.5334	–	210/90	7
Outer race defect	0.1778	Centered@6:00	210/90	8
Outer race defect	0.3556	Centered@6:00	210/90	9
Outer race defect	0.5334	Centered@6:00	210/90	10
Outer race defect	0.1778	Orthogonal@3:00	210/90	11
Outer race defect	0.1778	Opposite@12:00	210/90	12
Outer race defect	0.021	Orthogonal@3:00	210/90	13
Outer race defect	0.021	Opposite@12:00	210/90	14

- The saturated linear (Satlin) and rectified linear unit (ReLU) functions are selected as activation functions to learn the proposed method while two other different activation functions are used in the DAE models.
- The recent studies always compared only with shallow learning methods. Whereas, the proposed method is compared with various DAE models, standard deep learning methods and traditional fault diagnosis methods which is quite different from previous studies.
- The proposed method architecture is constructed based on two different activation functions while previous studies are used only one activation function in deep learning features, i.e. [47].

As mentioned above, this procedure is validated in two experiments (1, 2). Based on our experimental trials and according to Fig. 5, each activation function is tested with eight different activation functions to construct DAE models from model 1 to model L. Only ten DAE models are over the threshold value. The proposed method is selected according to the highest accuracy reported from these models and it is constructed from the Satlin function with the ReLU function as a hidden activation function in DAE1 and DAE2, respectively. The 14 bearing data sets with different fault severities and orientations are used to train the deep networks. The experiments are described as follow:

Experiment 1: Ten trials are implemented to illustrate the robustness steady of the proposed method compared with other deep AE learning models. The results of the proposed method and those other DAE models for each trial are shown in Fig. 8. Table 3 described the types of the used activation functions of DAE models and their results. In diagnosis results, the average testing accuracy is the criteria to determine the fault diagnosis performance which is 99.44 % for the proposed method and it is slightly higher than other models numbered 1-9 which are 95.39 %, 96.1 %, 95.97 %, 97.33 %, 96.98 %, 97.05 %, 96.56 %, 94.22 %, and 93.33 %, respectively. In performance comparison, the other DAE models still have limitations in dealing with massive data while the present method has the potential to overcome this constraint. The standard deviation of the

Table 3. Descriptions of the classification results for the various deep AE learning models.

DAE models	Type of activation function		Average testing samples accuracy (%) ±standard deviation	Average testing time (s)
	AE1	AE2		
The proposed model	Satlin	ReLU	99.44±1.81	0.1018
Model 1	Satlin	Satlin	95.39±2.02	0.1423
Model 2	Satlin	Purelin	96.1±2.04	0.1077
Model 3	Satlin	Rabas	95.97±2.08	0.1458
Model 4	Satlin	Tansig	97.33±1.99	0.1130
Model 5	Satlin	Satlins	96.98±2.1	0.1073
Model 6	Satlin	Elliotsig	97.05±2.07	0.1019
Model 7	Satlin	Hardlim	96.56±2.21	0.1255
Model 8	Satlins	Satlin	94.22±2.34	0.1362
Model 9	ReLU	ReLU	93.33±3.01	0.1522

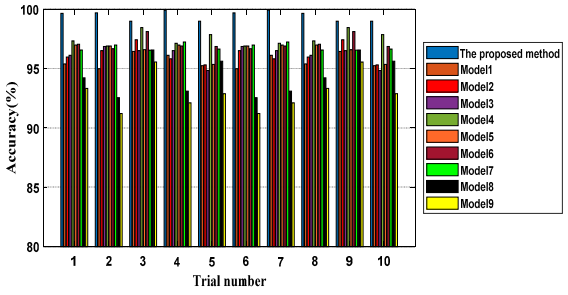


Fig. 8. Detailed classification results of different deep AE models for ten trials in experiment 1.

proposed method is 1.81 which is the least value compared with other DAE models which were 2.02, 2.04, 2.08, 1.99, 2.1, 2.07, 2.21, 2.34 and 3.01, respectively. This comparison proved that the selection reason for the proposed method among different DAE models to distinguish the fault type of rolling-element bearing. In Table 3, the testing time of the proposed method is (0.1018 s) which is less than other DAE models. It's worth mention that this work is conducted by (Core i5, 12-GB memory, MATLAB R2018b (PC)) that contains the fault classification stage while the testing times of the other methods were 0.1423 s, 0.1077 s, 0.1458 s, 0.1130 s, 0.1073 s, 0.1015 s, 0.1255 s, 0.1362 s, and 0.1522 s, respectively.

Experimental 2: For further classification robustness confirmation, the training of the proposed method and three advanced deep learning methods are carried out. These methods are applied to diagnose and identify the same data set including convolution neural network (CNN), long short-term memory (LSTM) deep network and the standard DAE. The main parameters of these methods are described as follows:

- Method 1 (standard CNN): Its construction comprises of an input layer with size 28 / 28. Two convolutional layers with 4 kernels for each one and two pooling layers have scale value 3 for each one. The learning rate is 1×10-4 and the epoch number is 120. SoftMax layer was assumed as a classification layer to identify the bearing fault condition at the end of the structure.
- Method 2 (standard LSTM): The architecture of the LSTM deep learning method consists of sequence input layer with 400-input size. Two LSTM layers are included with two hidden units of 100 neurons for each layer. The minimum batch size is set to 27. The maximum epoch number is 120 and the SoftMax layer is employed after fully connected layer to identify the bearing fault condition.
- Method 3 (standard DAE): The architecture of the standard DAE is 400-200-100-14 which is determined by experimentation. The log-sigmoid function is utilized as a hidden unit in hidden layers. The learning rate is 0.1 and the max-iteration number is 150. SoftMax layer is used to identify and classify the bearing fault condition.

Five trials are carried out to show the superiority and the diagnosis identification accuracy for the proposed method and compared with other deep learning methods. The average testing accuracies, the standard deviation and computed testing time for each method are listed in Table 4. For each trial, all diagnosis results of compared methods are completely presented in Fig. 9. From Table 4, the highest average testing accuracy of the proposed method is 99.52 % while all other methods (standard CNN, standard LSTM and standard DAE) obtain 95.5 %, 96.78 %, and 90.93 %, respectively. Moreover, the standard deviation of the proposed method is 1.79 which is the smallest compared with other methods which are 2.52, 2.34, and 3.78, respectively. The average testing time of the proposed method is 0.1021 s which is the lowest computed time to validate the testing samples of the proposed method. Whereas, the average computed time of the testing samples for other methods are 7.4131 s, 1.5387 s and 0.1179 s as shown in Table 4, respectively.

Table 4. Descriptions of the classification results for the proposed method and other deep learning methods.

Methods description	Average testing samples accuracy (%) ±standard deviation	Average testing time (s)
DFAEs	99.52±1.79	0.1021
Method 1 (standard CNN)	95.51±2.52	7.4131
Method 2 (standard LSTM)	96.78±2.34	1.5387
Method 3 (standard DAE)	90.93±3.78	0.1179

Note: The proposed method and all methods have similar dimension (400-dimensional).

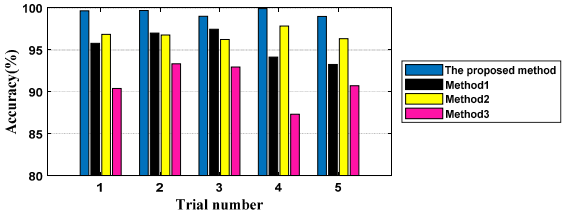


Fig. 9. Detailed classification results of different deep learning methods for five trials in experiment 2.

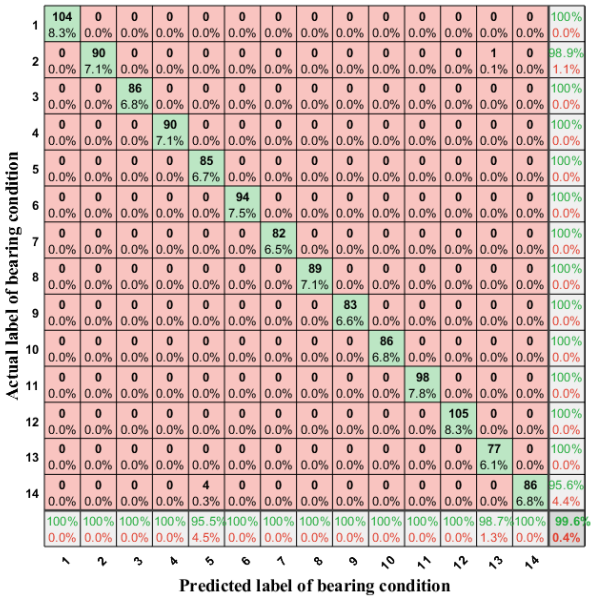


Fig. 10. The confusion matrix of the proposed method in the first trial.

These results are given clearly that the proposed method can classify powerfully several fault types with several fault orientations of element rolling bearings with the lowest testing time. For the first trial, the confusion matrix of the proposed method is shown in Fig. 10. The actual and predicted bearing classification labels are represented in the ordinate and the abscissa of a confusion matrix, respectively. In the confusion

Table 5. Main parameters used in rolling bearing fault diagnosis.

Descriptions	Value
The dimension of each input sample	400
The number of hidden layers	2
The number units of the FAE1	200
The number units of the second FAE2	100
Learning rate of each FAE	0.001
Iteration number of each FAE	120
Sparsity parameter of each FAE	0.1
Sparse penalty factor of each FAE	7
The threshold of the allowed accuracy	90 %
The number of repeated trials	10
The input units of SoftMax classifier	100
The output units of SoftMax classifier	14

matrix, the component on the right column presents the classification accuracy of the classified conditions. The lowest accuracy occurs in condition 14 (95.6 %). The results of the proposed method confirmed that it produced the highest robustness and stability in different numbers of trials and it can effectively learn the vital features using raw vibration data. The main reasons relate to the following notes:

1) The meaningful information can effectively be captured due to the deep feature learning task of the proposed method.

2) The DFAE training accuracy and testing accuracy reported higher results than other methods because it can take full advantages of DAE and two activation functions (satlin, ReLU) with SoftMax classifier which can further encourage deep learning performance and enhance the classification accuracy for non-stable signals. Otherwise, it is very time-consuming and heavy labor to design a master architecture for deep learning with massive raw data. All parameters of the proposed method are offered in Table 5 with full architecture involves DFAEs with SoftMax classifier which used to train the raw bearing vibration signals. From Eq. (7), two substantial constraints of the DFAEs, which are the sparse penalty constraint, β, and the sparse constraint, ρ, were used to optimize the representative in the network. The optimal parameters β and ρ were determined by the cross-validation method, in which the nominee set of β was chosen as [1-9] and ρ was [0.1-0.9]. For the first trial, a better correlation between the accuracy and parameter set (β, ρ) is shown in Fig. 11. It gives a great idea that the accuracy is very delicate to the sparse constraint ρ slighter values of sparse can be induced to provide recovering choices.

The training of each FAE attempts to reduce the reconstruction error with the optimized parameters. In Fig. 12, the proposed method and the standard DAE method reconstruction error curves are shown from the first only eighty trial iterations are selected from 120 iterations to present the gradient curves. The proposed method training shows high convergence and faster gradient for reconstruction error compared with standard DAE.

For further evaluation, 100 is the dimension number of the

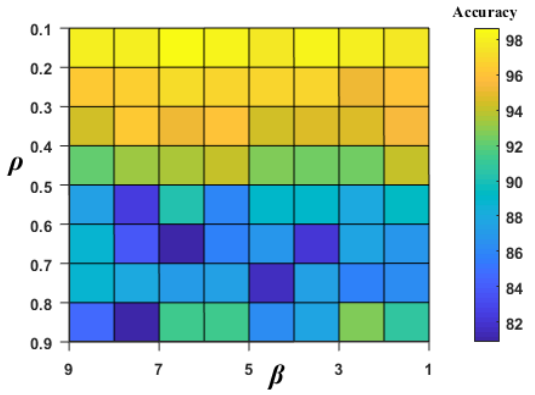


Fig. 11. The relationship between accuracy and parameter set (*β, ρ*) for the first trial.

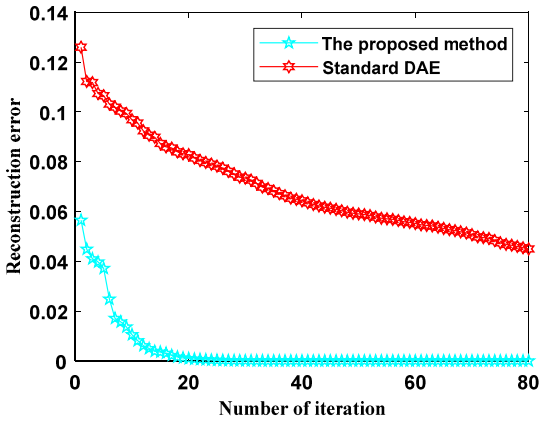


Fig. 12. The reconstruction error curves of the proposed method and standard DAE method.

extracted features learned by the proposed method and the standard DAE. All are high-dimensional data (second-layer feature). The t-distributed stochastic neighbor embedding (t-SNE) method is used for feature visualization by reducing the dimensionality of the high-dimensional features. The main learned deep features are visualized to illustrate the effectiveness of each learning method. Fig. 13 shows that the two-dimensional and three-dimensional visualizations of the features in the first trial which t-SNE1, t-SNE2, and t-SNE3 denote that the first three principal components of 14 corresponding bearing condition labels. From Fig. 13, the deep features learned by DFAEs have more ability to represent the input vibration data recognizably compared with learned deep features of the standard DAE. These results confirm that a DFAEs model was more vigorous for feature learning than the standard DAE with massive input raw data.

4.3 The impact of data set dimension and number of hidden units on the proposed architecture

In this section, the influence of the sample dimension and the first AE hidden size on the presentation of the proposed

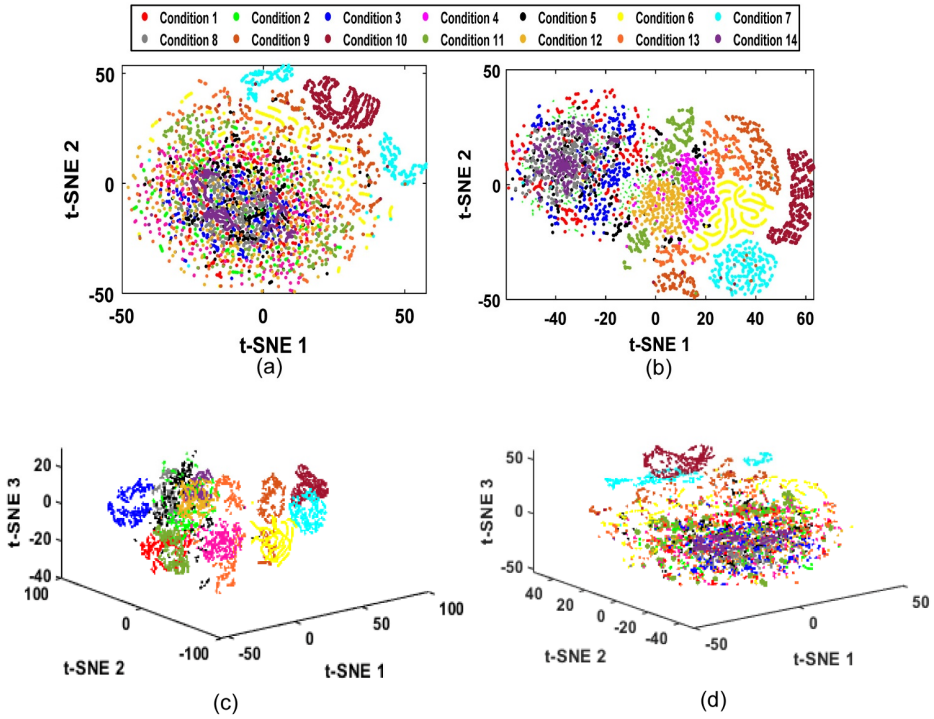


Fig. 13. Dimensional visualizations of different features using t-SNE: (a) two dimensional deep features learned by DFAEs; (b) two dimensional deep features learned by standard DAE; (c) three dimensional deep features learned by DFAEs; (d) three dimensional deep features learned by standard DAE.

method is investigated. The testing accuracy is calculated with different selected dimensions of each sample, 150, 200, 250, 300, 400, 500, 500, 600 and 800. The curve of the testing accuracy shows that the highest accuracy is 98 % with the 400-dimensional sample as shown in Fig. 14. In the results, better neural network construction can be trained with 400-dimensional raw vibration data. After the sample dimension is determined, the next step is finding the accurate number of the hidden units in each layer of the proposed method architecture. The number of hidden units in each layer of the proposed deep architecture is very important for diagnostic accuracy. Therefore, the number of hidden units is increased in each layer of FAE1 and FAE2 from 50 to 500 by 50 units. Fig. 15 shows that the testing accuracy curves for each number of hidden layers in the proposed method of determined raw data dimension.

The evaluated accuracy demonstrates that only two hidden layers are more robust to represent the massive raw data than others. Therefore, the performance of diagnosis results is very affected by the number of hidden units in the first auto-encoder and the optimal architecture of the proposed method is selected as (400-200-100-14). Flexibility in changing the hidden layer number and units can be provided to permit the designer a lot of freedom. From Figs. 14 and 15, we can determine that the optimal structure and optimal setting of deep learning methods rely on a certain diagnosis issue [3].

Different deep learning methods can show non-stable classification performance for bearing element condition in each trial. In this section, quantitative performance is evaluated by F-measure, which is utilized extensively for assessing the bear-

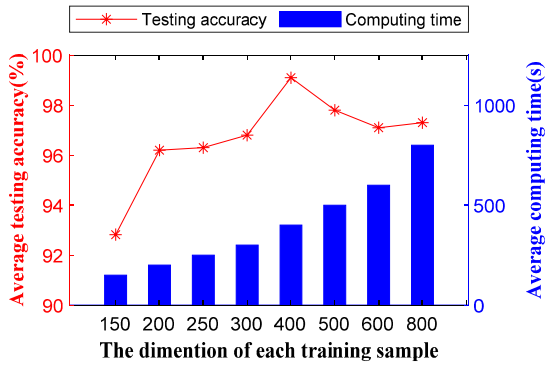


Fig. 14. The classification accuracy corresponding to the dimension of each training sample.

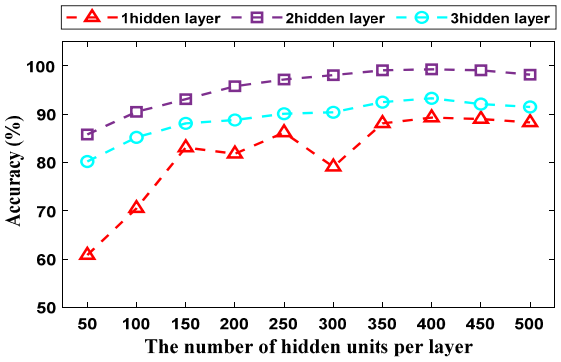


Fig. 15. The classification accuracy of the proposed model corresponding to the number of hidden layer.

Table 6. The precision and recall rate of the proposed method and other deep learning methods in the first trial.

Bearing condition label	The proposed method		Method 1		Method 2		Method 3	
	PR(%)	*RE*(%)	*PR*(%)	*RE*(%)	*PR*(%)	*RE*(%)	*PR*(%)	*RE*(%)
Condition 1	100	100	100	100	94.73	98.9	100	89.33
Condition 2	98.96	100	100	92.38	93.97	98.73	98.93	93.93
Condition 3	100	97.37	100	89.52	98.19	99.09	100	97.84
Condition 4	97	98.98	96.6	96.6	96.29	92.85	93.54	100
Condition 5	97.94	97.94	85.9	99	98.78	96.42	92.39	98.83
Condition 6	98.94	100	93.61	96.7	96.51	100	96.93	95.95
Condition 7	100	100	96.51	88.3	98.94	97.91	94.11	80
Condition 8	100	100	100	87.1	100	100	98.71	91.66
Condition 9	100	100	92.1	95.9	97.61	90.1	85.04	96.8
Condition 10	100	100	84.84	96.55	97.64	100	86.17	94.18
Condition 11	100	96.3	100	96.55	95.69	100	95.14	97.02
Condition 12	97.96	90.38	89	96.42	94.84	92	93.25	97.64
Condition 13	94.05	98.75	86.66	97	91.86	87.77	86.25	95.83
Condition 14	100	97.78	100	95.7	97.87	100	100	90

ing condition classification performance of various learning methods [14] and can be expressed as:

$$\mathrm{PR} = \mathrm{precesion} = \frac{TP}{TP + FP} \times 100 \tag{13}$$

$$\mathrm{RE} = \mathrm{recall} = \frac{TP}{TP + FN} \times 100 \tag{14}$$

$$\mathrm{F-measure} = \frac{2TP}{2TP + FP + FN} \times 100 \tag{15}$$

where *TP* symbolizes the number of correct positive cases, *FP* symbolizes the number of incorrect positive cases and *FN* symbolizes the number of false-negative cases. From Eq. (15), the index of F-measure covers the precision rate and recall rate since its value varieties from the worst (0) to the best (1). From Table 4, method 1, method 2 and method 3 show that varying performances in deep learning features of the raw signals. Therefore, these methods are utilized for evaluating the functioning in each method using the F-measure index for each signal of element bearing condition. In Table 6, the precision and recall rates for the fourteen bearing conditions are described and the corresponding F-measure values are shown in Fig. 16. The mean value of F-measure values for the proposed method, method 1, method 2 and method 3 are 98.9 %, 94.57 %, 96.63 %, and 94.1 %, respectively. In the results, the F-measure of the present method is more superior to the other methods which has advantages in dealing with massive data to represent the effective model in intelligent fault diagnosis.

4.4 The diagnosis results of the proposed method and different traditional intelligent methods

In this section, the power of the proposed method was compared with different eight traditional machine learning methods to show its effectiveness, robustness, and stability. The training of these methods has been done using the same input raw data set (400- dimensional). All data set included the signals of fourteen element bearing conditions with different severities and orientations. The main parameters of these methods are defined as follows:

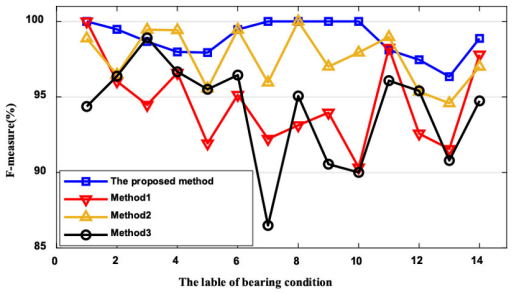


Fig. 16. F-measures of the proposed method, method 1, method 2 and method 3 in Table 4.

- Backpropagation neural network (BPNN): The log-sigmoid was used as an activation function in the hidden layer and the number of neurons in the hidden layer is 80, 80softmax layer is adopted as an output layer. The main parameters of the network are 400-80-14 (input size, hidden size and output size). It is important to note that the mentioned parameters have determined by the prior knowledge and experiences. The number of epochs and learning rate are set to 1000, 0.4, respectively.
- Support vector machine (SVM): In this method, the type of kernel is (RBF), the numbers of penalty factor and epsilon are 1000, 0.001, respectively.
- Random forest (RF): The numbers of trees and minimum leaf size are 1000, 0.001, respectively. The number of

Table 7. Descriptions of the classification results for the proposed method and traditional machine learning methods.

Method	Average testing samples accuracy (%) ±standard deviation	Average testing time (s)
DFAEs	99.67±1.71	0.1034
BPNN	88.79±2.86	10.325
SVM	73.45±3.21	6.8952
RF	94.05±2.01	3.2287
AB	84.6±2.95	2.5821
DT	26.69±5.58	4.8925
DA	19.01±6.07	2.1467
NB	90.21±2.56	2.0648
K-NN	92.23±2.11	3.0748

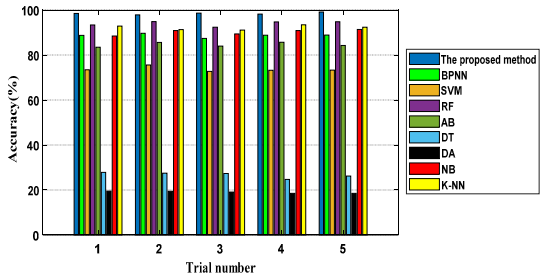


Fig. 17. Detailed classification results of the proposed method and eight traditional learning methods for five trials.

predictors to samples is set to 20.

- Linear discriminant analysis (LDA), the delta and gamma values are set to be 0.00025, 0.997, respectively.
- The minimum leaf size in decision trees (DT) is 1, which is also decided based on the prior knowledge and experiences.
- Adaptive boost (AB): The number of trees and a maximum number of splits are 100, 80, respectively.
- Naive Bayes (NB), the distribution name is set to normal.
- K-nearest neighbors (K-NN): The number of neighbors(k) is 4, and distance is set to cosine, which is decided by the prior knowledge and experiences.

Five trials were carried out for the proposed method and all of these traditional machine learning methods to show the diagnosis result for each method. Table 7 describes the fault diagnosis results in detail. For each trial, Fig. 17 shows that the diagnosis results of all methods. In these trials, the average testing accuracy of the proposed method is 99.67 % which is higher than the highest accuracy of the (RF) method, which is 94.05 %. The other methods BPNN, SVM, AB, DT, DA, NB, and K-NN, have obtained an accuracy of 88.79 %, 73.45 %, 84.6 %, 26.69 %, 19.01 %, 90.21 %, and 92.23 %, respectively.

As in Table 7, the average standard deviation of the proposed method is 1.71 of these trials, which is less than other methods, which have values 2.86, 3.21, 2.01 2.95, 5.58, 6.07, 2.56 and 2.11, respectively. The average testing time of the proposed method is 0.1034 s which is lowest all other compared methods. From this comparison, it clear that the proposed method is more accurate and stable in multiple trials than other intelligent traditional methods. The robustness and reliability of the proposed method stand out due to two main facts. (1) The deep architecture of the proposed model has a greater capacity to learn and extract meaningful features from the raw input data of each bearing condition. Nevertheless, other traditional methods have weakness to extract valuable features from the raw massive data. It has prominent powerful in automatic learning and taking full advantage of the non-missing important information features due to different characteristics of the DAEs. Other traditional methods have many constraints in automatic learning. They could miss important information features during their process. (2) The proper selection of the activation function has awarded the model power and effective learning ability to encourage sparse activation.

5. Conclusions

A novel deep network architecture is proposed. A deep functional auto-encoders (DFAEs) have designed with SoftMax classifier for an accurate deep learning-based intelligent bearing fault diagnosis method. The proposed method is carried out on following main steps:

First, the designed functional auto-encoder (FAE) is performed better in capturing the signal characteristics through the adoption of the saturating linear (Satlin) and rectified linear unit (ReLU) as activation functions for a hidden layer. Second, the DFAE architecture is built with multiple FAEs based on various DAE models performances for further enhancing the unsupervised feature learning method. Third, based on DFAEs training, the learned deep features are utilized as dataset to train the Softmax classifier for an accurate categorization pattern of bearing fault classification. Fourth, the results have shown that the Satlin function always shows a higher accuracy in deep network learning and it gives a new phenomenon that leads to advantages in using this function with dissimilar activation functions to develop the deep neural network learning methods. It is powerful for capturing useful information from raw vibration data and get the best-learned features. Finally, the experimental research has approved that the proposed methodology has robustness in the identification of massive data on various rolling element bearing conditions, which could give us more opportunity to use this model with other datasets calculated from machine components to achieve a highly accurate system in the field of intelligent fault diagnostics.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (Grant No. 5157520, 51675204), the National Science and Technology Major Project of China (Grant No. 2018ZX04035002-002), and the Science Challenge Project of China (Grant No. TZ2018006-0102-01).

References

[1] H. Jiang, C. Li and H. Li, An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis, *Mech. Syst. Signal Process.*, 36 (2) (2013) 225-239.

[2] J. Chen et al., Wavelet transform based on inner product in fault diagnosis of rotating machinery: a review, *Mech. Syst. Signal Process.*, 70-71 (2016) 1-35.

[3] F. Jia, Y. Lei, J. Lin, X. Zhou and N. Lu, Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, *Mech. Syst. Signal Process.*, 72-73 (2016) 303-315.

[4] X. Jin, M. Zhao, T. W. S. Chow and M. Pecht, Motor bearing fault diagnosis using trace ratio linear discriminant analysis, *IEEE Trans. Ind. Electron.*, 61 (5) (2014) 2441-2451.

[5] H. Wang and P. Chen, Intelligent diagnosis method for rolling element bearing faults using possibility theory and neural network, *Comput. Ind. Eng.*, 60 (4) (2011) 511-518.

[6] S. G. Barad, P. V. Ramaiah, R. K. Giridhar and G. Krishnaiah, Neural network approach for a combined performance and mechanical health monitoring of a gas turbine engine, *Mech. Syst. Signal Process.*, 27 (2012) 729-742.

[7] X. Zhang, B. Wang and X. Chen, Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine, *Knowledge-Based Syst.*, 89 (2015) 56-85.

[8] Y. Lei, Z. Liu, X. Wu, N. Li, W. Chen and J. Lin, Health condition identification of multi-stage planetary gearboxes using a mRVM-based method, *Mech. Syst. Signal Process.*, 60-61 (2015).

[9] Z. Xu, J. Xuan, T. Shi, B. Wu and Y. Hu, Application of a modified fuzzy ARTMAP with feature-weight learning for the fault diagnosis of bearing, *Expert Syst. Appl.*, 36 (6) (2009) 9961-9968.

[10] L. Jiang, J. Xuan and T. Shi, Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis, *Mech. Syst. Signal Process.*, 41 (1) (2013) 113-126.

[11] S. Wan and B. Peng, The FERgram: a rolling bearing compound fault diagnosis based on maximal overlap discrete wavelet packet transform and fault energy ratio, *J. Mech. Sci. Technol.*, 33 (2019) 157-172.

[12] H. Wang, P. Wang, L. Song, B. Ren and L. Cui, A novel feature enhancement method based on improved constraint model of online dictionary learning, *IEEE Access*, PP (2019) 1.

[13] Q. Pan, Y. Liu, R. Zhou, H. Wang, H. Chen and T. He, An automatic abrupt signal extraction method for fault diagnosis of aero-engines, *J. Mech. Sci. Technol.*, 33 (2019) 1633-1640.

[14] Y. Lei, F. Jia, J. Lin, S. Xing and S. X. Ding, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, *IEEE Trans. Ind. Electron.*, 63 (5) (2016) 3137-3147.

[15] Y. Bengio, A. Courville and P. Vincent, Representation learning: a review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.*, 35 (8) (2013) 1798-1828.

[16] C. Shang, F. Yang, D. Huang and W. Lyu, Data-driven soft sensor development based on deep learning technique, *J. Process Control*, 24 (3) (2014) 223-233.

[17] X. Guan and G. Chen, Sharing pattern feature selection using multiple improved genetic algorithms and its application in bearing fault diagnosis, *J. Mech. Sci. Technol.*, 33 (2019) 129-138.

[18] H. Shao, J. Hongkai, Z. Huiwei and W. fuan, An enhancement deep feature fusion method for rotating machinery fault diagnosis, *Knowledge-Based Syst.*, 119 (2016).

[19] Y. Zhang, X. Li, L. Gao and P. Li, A new subset based deep feature learning method for intelligent fault diagnosis of bearing, *Expert Syst. Appl.*, 110 (2018) 125-142.

[20] W. Mao, W. Feng and X. Liang, A novel deep output kernel learning method for bearing fault structural diagnosis, *Mech. Syst. Signal Process.*, 117 (2019) 293-318.

[21] H. Wang, S. Li, L. Song and L. Cui, A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals, *Comput. Ind.*, 105 (2019) 182-190.

[22] G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science (80-.).*, 313 (5786) (2006) 504 LP-507.

[23] J. Leng and P. Jiang, A deep learning approach for relationship extraction from interaction context in social manufacturing paradigm, *Knowledge-Based Syst.*, 100 (2016) 188-199.

[24] J. Schmidhuber, Deep learning in neural networks: an overview, *Neural Networks*, 61 (2015) 85-117.

[25] T. Jiahui, J. Wu, B. Hu, C. Guo and J. Zhang, A fault diagnosis method using Interval coded deep belief network, *J. Mech. Sci. Technol.*, 34 (2020).

[26] Y. Ma et al., A new fault diagnosis method based on convolutional neural network and compressive sensing, *J. Mech. Sci. Technol.*, 33 (2019) 5177-5188.

[27] H. Shao, H. Jiang, X. Zhang and M. Niu, Rolling bearing fault diagnosis using an optimization deep belief network, *Meas. Sci. Technol.*, 26 (2015).

[28] H. Shao, H. Jiang, H. Zhao and F. Wang, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, *Mech. Syst. Signal Process.*, 95 (2017) 187-204.

[29] O. Janssens et al., Convolutional neural network based fault detection for rotating machinery, *J. Sound Vib.*, 377 (2016) 331-345.

[30] Z. Shang, X. Liao, R. Geng, M. Gao and X. Liu, Fault diagnosis method of rolling bearing based on deep belief network, *J. Mech. Sci. Technol.*, 32 (2018) 5139-5145.

[31] S.-Z. Su, Z.-H. Liu, S.-P. Xu, S.-Z. Li and R. Ji, Sparse auto-encoder based feature learning for human body detection in depth image, *Signal Processing*, 112 (2015) 43-52.

[32] Y. Lei, Z. He and Y. Zi, EEMD method and WNN for fault diagnosis of locomotive roller bearings, *Expert Syst. Appl.*, 38 (2011) 7334-7341.

[33] M. Zhao and X. Jia, A novel strategy for signal denoising

using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery, *Mech. Syst. Signal Process.*, 94 (2017) 129-147.
[34] K.-H. Jeong, W. Liu, S. Han, E. Hasanbelliu and J. C. Principe, The correntropy MACE filter, *Pattern Recognit.*, 42 (5) (2009) 871-885.
[35] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés and C. Li, Fault diagnosis in spur gears based on genetic algorithm and random forest, *Mech. Syst. Signal Process.*, 70-71 (2016) 87-103.
[36] J. Zabalza et al., Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging, *Neurocomputing*, 185 (2016) 1-10.
[37] S. S. Liew, M. Khalil-Hani and R. Bakhteri, Bounded activation functions for enhanced training stability of deep neural networks on visual pattern recognition problems, *Neurocomputing*, 216 (2016) 718-734.
[38] X. Jiang, S. Mahadevan and Y. Yuan, Fuzzy stochastic neural network model for structural system identification, *Mech. Syst. Signal Process.*, 82 (2017) 394-411.
[39] S. Ittiyavirah, S. Jones and P. Siddarth, Analysis of different activation functions using backpropagation neural networks, *J. Theor. Appl. Inf. Technol.*, 47 (2013) 1344-1348.
[40] J. Heaton, Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning, *Genet Program Evolvable Mach*, 19 (2018) 305-307.
[41] T. K. K. R. Mediliyegedara, A. K. M. De Silva, D. K. Harrison and J. A. Mcgeough, An intelligent pulse classification system for electro-chemical discharge machining (ECDM)-A preliminary study, *J. Mater. Process. Technol.*, 149 (2004) 499-503.
[42] A. Elsaeidy, K. Munasinghe, D. Sharma and A. Jamalipour, Intrusion detection in smart cities using restricted Boltzmann machines, *J. Netw. Comput. Appl.*, 135 (2019).
[43] W. Liu, T. Ma, D. Tao and J. You, HSAE: a Hessian regularized sparse auto-encoders, *Neurocomputing*, 187 (2016) 59-65.
[44] G. Hinton, S. Osindero and Y.-W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.*, 18 (2006) 1527-1554.
[45] A. C. D. Erhan, Y. Bengio and P. Vincent, Why does unsupervised pre-training help deep learning?, *Journal of Machine Learning Research*, 9 (2010) 201-208.
[46] W. A. Smith and R. B. Randall, Rolling element bearing diagnostics using the case western reserve university data: a benchmark study, *Mech. Syst. Signal Process.*, 64-65 (2015) 100-131.
[47] S. Haidong, J. Hongkai, L. Xingqiu and W. Shuaipeng, Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine, *Knowledge-Based Syst.*, 140 (2018) 1-14.

Anas Hamid Aljemely received the B.S. degree in Mechanical Engineering in University of Technology, Baghdad, Iraq, in 2003. His Master's degree is achieved in Mechanical Engineering in Huazhong University of Science and Technology (HUST) in 2016, Wuhan, China. He is currently a Ph.D. candidate of the School of Mechanical Science and Engineering, HUST. His research interests include signal processing analysis techniques, mechanical fault diagnosis, machine leaning algorithms and deep learning methods.

Jianping Xuan earned his Ph.D. degree in Mechanical Engineering from Huazhong University of Science and Technology (HUST) in 1999. After postdoctoral work ended, he joined the Mechanical Engineering faculty, HUST, in 2002. From February 2013 to February 2014, he was a visiting scientist in Department of Mechanical Engineering, Massachusetts Institute of Technology, USA. Currently, he is a Professor with the Department of Mechanical Engineering, HUST. His research interests include Digital Manufacturing for difficult-to-machine Materials and Structures; Big Data Based Intelligent Maintenance Systems; PHM for Structures, Machinery, CNC systems and Machine Tools. He has published more than 60 journal papers.