A Novel Intelligent Fault Diagnosis Approach for Early Cracks of Turbine Blades via Improved Deep Belief Network Using Three-Dimensional Blade Tip Clearance

XIN HUANG [1], **XIAODONG ZHANG** [1,2], **YIWEI XIONG** [1], **HONGCHENG LIU** [1], **AND YINGJIE ZHANG** [1]

[1]College of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[2]Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an 710049, China

Corresponding authors: Xin Huang (1559196486@qq.com) and Xiaodong Zhang (xdzhang@mail.xjtu.edu.cn)

ABSTRACT Aero-turbines usually work in various non-stationary and harsh operating environments. Its blades easily appear early crack faults in long-time operation, and it is difficult to dig out discriminant features of early crack. To solve these problems, in the paper, a novel intelligent approach for early crack diagnosis of turbine blades using three-dimensional blade tip clearance is presented. In order to improve feature learning ability to obtain better generalization ability, the paper firstly develops a novel deep learning method based on deep belief networks (DBNs). Considering the fact that the feature degradation easily occurs in deeper layers because of the change of the distribution in each layer's outputs with the increase of the layers, it is hard to decide which layer to learn features is useful for fault diagnosis. Accordingly, in the pre-training process, the global back-reconstruction (GBR) mechanism is introduced into DBNs to optimize the feature learning ability. The GBR mechanism can be realized between the input layer and hidden layers by "shortcut connection", and the layer to learn more discriminant features can be determined automatically without prior knowledge. Moreover, due to three-dimensional blade tip clearance (3-DBTC) acquired from three different directions of turbine blades contains much more useful crack failure multiscale information, it is suitable to be used as an input from which to extract multiscale discriminant features. Eventually, in the supervised training, the softmax regression model is employed to classify the health conditions of turbine blades using these sensitive features learned from 3-DBTC. The experimental results show that the proposed method can effectively identify the crack of turbine blades with fairly high diagnostic accuracies and significantly outperform other methods considered in the paper.

INDEX TERMS Turbine blade, three-dimensional blade tip clearance, deep belief networks, feature extraction, fault diagnosis.

I. INTRODUCTION

Turbines are widely used in aero-engines as the key energy transformation component, and their health conditions are highly concerned with the regular operation of the aero-engine. Since turbines usually work in the harsh operating environment caused by high temperature, high stress, and high speeds during operation [1], they are easily prone to

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain.

machinery failures, like blade fatigue cracks, in long-time operation. As cited by Meher-Homji and Cyrus [2], 42% of the total failures in turbines caused by blade cracks. A crack defect, if not be detected in time, may cause catastrophic damage with huge economic losses and human casualties. Therefore, fault diagnosis for turbine blades is essential for operational reliability and timely-maintenance of aero-engine.

The essential parts of fault diagnosis include feature extraction and pattern recognition. Conventional intelligent

fault diagnosis methods mainly depend on manual feature extraction based on signal processing (e.g., spectral analysis or envelop analysis) and shallow learning algorithm, such as support vector machine (SVM) [3], decision tree [4], artificial neural networks (ANNs) [5]. Manual feature extraction based on statistical features (e.g., root-mean-square (RMS), kurtosis, and energy), which often needs enough domain knowledge and human intensive labors, is key step. Then, these manual features are fed into shallow learning algorithm. Whether the representative ability of those statistical features is good or not directly affects the performance of the fault diagnosis models. However, it is difficult and time-consuming to select as much discriminant features as possible for fault diagnosis, because a good set of features varies from case to case in practical application. Therefore, the feature extraction and selection have been an obstacle to further improve diagnostic accuracies.

To address aforementioned those problems, deep learning methods that can automatically extract features from the input data for fault diagnosis has been the effective way in recent years. To be specific, deep learning, which refers to representative learning that has multiple layers of nonlinear transformation, can enable a hierarchical nonlinear learning of high-level features by layer-wise training to discriminate different health conditions (e.g., healthy and faulty). The high-level features usually are better than those statistical features in both representative and generalization ability, and can be further applied to fault diagnosis. By now, deep learning has shown its great potential power in various fields including speech recognition, computer vision, natural language processing, etc. In the field of fault diagnosis and monitoring, convolution neural networks (CNNs) [6], deep belief networks (DBNs) [7], staked auto-encoders (SAEs) [8] are usually popular deep learning models used for fault diagnosis and prognostic application in recent years. Feng *et al.* [9] used deep auto-encoder network to fault diagnosis, which achieved higher accuracy than traditional BPNN. Ince *et al.* [10] applied 1-D CNN to motor fault detection, and yielded higher accuracy than other traditional shallow algorithms. Chen *et al.* [11] utilized SAEs and DBNs to fuse features, and effectively identified the bearing fault. Shen *et al.* [12] presented a new model based on DBNs, which was validated by bearing fault diagnosis. These methods outperform the traditional shallow intelligent algorithm-based fault diagnosis.

But many Deep learning methods, like CNN, often need large-scale data to fit with the increasing complexity of the model and its performance on small-scale data tasks is not excellent. Meanwhile, DBNs has found successful applications applied in fault diagnosis by virtue of two stages: greedy layer-wise unsupervised training and fine-tuning supervised training. An unsupervised way is utilized to effectively pre-train the model in regardless of the amount of training data. Compared with SAEs, DBNs is a generative model that can generate samples based on the features learned during training, so it's suitable for small-scale data task. However,

in unsupervised training, the feature degradation usually occurs in deeper layers and becomes severe with the depth of DBNs increasing due to change of the distribution of each layer's outputs during training. As a consequence, it will make hard to decide which layer to learn features in DBNs is useful for fault diagnosis by determining the depth of DBNs manually. Once the insensitive features in certain layer are selected to fault diagnosis, it is difficult to guarantee to obtain desired diagnostical accuracies. Currently, some researches have been conducted on the problem, such as Shao *et al.* [13] proposed a method based on particle swarm algorithm to optimize the structure of the trained DBNs, and was applied to classify the ten health states of motor bearings with high accuracy. In [14], ACO was introduced to determine the structure of DBNs automatically and achieved good results on fault diagnosis of bearing life cycle prediction. In paper, to solve the problem to an extent, the GBR mechanism is introduced to directly build "short connection" between the layer inputs and each hidden layer outputs, and these parameters in "short connection" block can be optimized to improve the correlation of the layer inputs and the layer outputs, reducing reconstruction error. Thus, the depth of the DBNs can be determined based on the corresponding correlation-based criterion and the suitable representative features from signal can be extracted automatically. In supervised training, the softmax classifier is used to classify the health conditions (i.e., class) based on the learned discriminant features. The study explores a novel diagnosis approach based on the improved DBNs to early crack failure of turbine blades.

Vibration signals collected from turbine blades using non-contact measurement technologies, such as tip clearan-ce measurement, tip-timing measurement, contain importa-nt health condition information of turbine blades. Tip clearance is an important monitoring parameter to on-line operational conditions monitoring for turbines, which has been studied by [15]–[17]. Tip timing is also an important parameter, which has been studied from 1970 [18]. Blade Tip Timing (BTT) is widely applied to acquire blades vibration signal by non-contact ways. Some examples of these works can be found in References [19]–[21]. However, considering the blades usually show three-dimensional vibration behaviors under various working loads, the two methods mentioned above are difficult to reflect the complex dynamical behaviors of blades. Thus, a new non-contact measurement technique has been studied in recent years, called three-dimensional blade tip clearance (3-DBTC) [22]–[26], which are essentially multi-dimensional vibration signals. Teng *et al.* [25] studied the change regulation of 3-DBTC under several typical types of loads by simulation experiments and proved that 3-DBTC is effective, offering a theoretical foreshadow to monitor conditions feasible. Zhang *et al.* [26] has developed a 3-DBTC measuring system, which can acquire the 3-DBTC signals in real-time. In the paper, 3-DBTC signals will be used from which to extract discriminant early crack features.

The key contribution of this study is the development of an improved DBNs, the so-called deep belief networks with
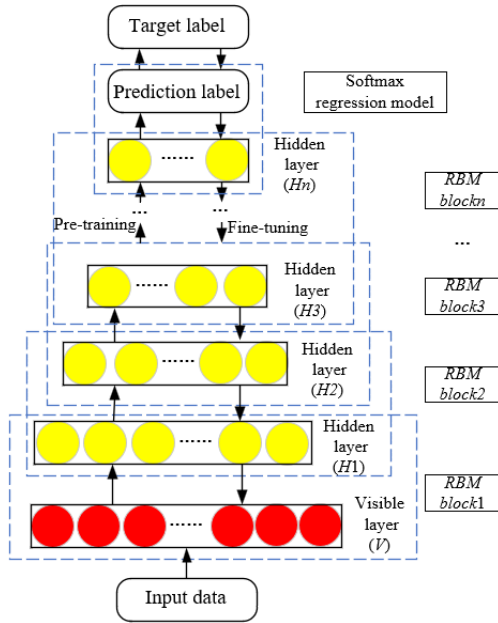
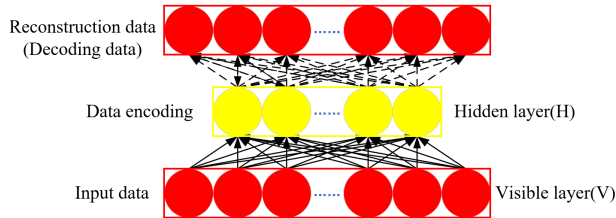**FIGURE 1.** The schematic architecture of DBNs.

FIGURE 2. Architecture of a Restricted Boltzmann Machines.

global back-reconstruction (DBNs+GBR) and applied it for crack fault diagnosis using 3-DBTC signals. The efficacy of the developed DBNs+GBR was verified for turbine blade crack fault diagnosis in this study. the reminder of the paper is organized as follows. Section II introduces the theoretical background of DBNs, and a detailed elaboration on the developed DBNs+GBR based on conventional DBNs. Section III then delineates an intelligent diagnosis approach for turbine blade crack fault using the 3-DBTC signal. Section IV the efficacy of the proposed method is verified by comparing with other considered methods used for fault diagnosis. Some conclusions are addressed in Section V.

II. THEORETICAL CONTEXT OF THE DEVELOPED DBNs+GBR

A. DEEP BELIEF NETWORK(DBNs)

Deep belief networks (DBNs) was first proposed by Hinton *et al.* [27] in 2006, which is essentially comprised of several RBMs stacked with each other and a multi-classifier (softmax regression model used as a classifier in the paper) as the output layer. In DBNs, as depicted in Fig.1, a plurality of RBMs blocks works together to effectively dig out discriminant information of the input data layer by layer through two stages: layer-wise unsupervised pre-training and fine-tuning supervised training.

In DBNs, each RBMs [28], as shown in Fig.2, is composed of two layers: visible layer containing visual neurons denoted

as $v = (v_1, v_2, \cdots, v_n)$ and hidden layer containing hidden neurons denoted as $h = (h_1, h_2, \cdots, h_n)$. It aims to fit the input data with maximum probability and its hidden layer can learn the high-order correlation information with the input data. The energy function [29] of a joint configuration (v, h) can be defined as:

$$E_\theta(v, h) = \sum_i^n \sum_j^m v_i w_{ij} h_j - \sum_i^n b_i v_i - \sum_j^m a_j h_j \quad (1)$$

where $\theta = \{W, b, a\}$ is the parameter of RBMs, W is the symmetric weight matrix, and $w_{ij} \in W$ represents the connection weights between the ith visible neuron and the jth hidden neuron, v_i, h_j represents the state of the ith visible neuron and the jth hidden neuron, respectively. b_i, a_j represents their bias, respectively.

Based on the energy function, the joint probability distribution of both hidden neurons and visible neurons is defined as:

$$P_\theta(v, h) = \frac{1}{Z(\theta)} \exp(-E_\theta(v, h)) \quad (2)$$

$$Z(\theta) = \sum_v \sum_h \exp(-E_\theta(v, h)) \quad (3)$$

where $Z(\theta)$ is a normalization factor, and the sum of the joint probabilities is guaranteed to be 1.

Given the state of the hidden neurons, the conditional probability and edge probability of the visible neurons are as follows:

$$P_\theta(v|h) = \frac{\exp(-E_\theta(v, h))}{\sum_v \exp(-E_\theta(v, h))} \quad (4)$$

$$P_\theta(v) = \frac{\sum_h \exp(-E_\theta(v, h))}{\sum_v \sum_h \exp(-E_\theta(v, h))} \quad (5)$$

Analogously, the conditional probability and edge probability of the hidden neurons are as follows:

$$P_\theta(h|v) = \frac{\exp(-E_\theta(v, h))}{\sum_h \exp(-E_\theta(v, h))} \quad (6)$$

$$P_\theta(h) = \frac{\sum_v \exp(-E_\theta(v, h))}{\sum_v \sum_h \exp(-E_\theta(v, h))} \quad (7)$$

According to the equations (4) and (6), The activation functions [30] of the ith visible neuron and the jth hidden neuron can be obtained, respectively, as follow:

$$P_\theta(v_i = 1|h) = sigmoid(\sum_j^m w_{ij} h_j + a_j) \quad (8)$$

$$P_\theta(h_j = 1|v) = sigmoid(\sum_i^n w_{ij} v_i + b_i) \quad (9)$$

where the sigmoid function is defined as:

$$f_s(x) = sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (10)$$

In the training process of RBMs, the edge probability of the visual neurons in the Gibbs distribution should be as close as
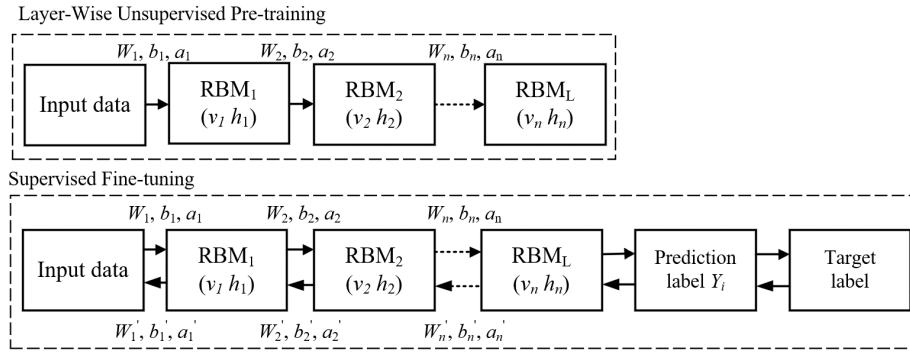
FIGURE 3. Training procedure of DBNs.

possible to the input data distribution. Then, the logarithmic likelihood function $L(\theta)$, objective optimization function, can be defined as follows:

$$L(\theta) = \frac{1}{M} \sum\nolimits_{i=1}^{M} \max_{\theta}(\log(P_\theta(v^{(i)}))) \tag{11}$$

where M is the size of the training sets.

To find optimal parameters $\theta = \{W, b, a\}$, the maximize log-likelihood function $L(\theta)$ can be optimized by the sto-chastic gradient descent (SGD) algorithm, so the partial d-erivative of $L(\theta)$ with respect to the parameters should be calculated as follows:

$$\frac{\partial L(\theta)}{\partial \theta} = \frac{1}{M} \sum\nolimits_{i=1}^{M} \frac{\partial \log P_\theta(v)}{\partial \theta}$$
$$= \frac{1}{M} \sum\nolimits_{i=1}^{M} (\left\langle \frac{\partial E_\theta(v, h)^{(i)}}{\partial \theta} \right\rangle_{data} - \left\langle \frac{\partial E_\theta(v, h)^{(i)}}{\partial \theta} \right\rangle_{model})$$
$$\tag{12}$$

where $\langle \cdot \rangle_{data}$ and $\langle \cdot \rangle_{model}$ model represents the expectations with regard to the distribution of the input data and the distribution learned by the model, respectively.

Then the update for parameter W should be expressed as

$$W = W + \alpha \frac{\partial L(\theta)}{\partial \theta} \tag{13}$$

where α refers to learning rate.

Note that the sample of $\langle \cdot \rangle_{data}$ is easily acquired, whereas the unbiased sample of $\langle \cdot \rangle_{model}$ cannot. In fact, $\langle \cdot \rangle_{model}$ generally is obtained using the Gibbs sampling method [31], but a large number of sampling steps will intractably cause low training efficiency, particularly for a great number of samples. Thus, Hinton *et al.* [27] proposed a contrastive divergence (CD) algorithm to fast approximate the gradient.

The mentioned above is the pre-training process in DBNs, and this greedy learning process is unsupervised without label information, as demonstrated in Fig.3. In the pre-training procedure, several RBMs blocks are trained bottom-up individually. Each RBMs block (consists of two adjacent layers) in DBNs is used to fit its input data by updating the connection weights W. Once an RBMs block is trained well, another RBMs block is stacked atop it. Thus, the abstract high-level

features can be obtained by a series of nonlinear transformation through different-level RBMs blocks. The number of neurons in visible layer are determined by the size of the input data.

When applied to classification tasks, the fine-tuning procedure is implemented to fine-tune the parameter space. As shown in Fig.3, to get good discriminative performance in classification, the labeled data need to be utilized to refine the parameter space W. Likewise, a multi-classifier, which can be used to predict the desired label samples, is carried out with this discriminative fine-tune in the training data set. More specifically, the weights of all layers, initialized by the unsupervised pre-training, are further adjusted top-down by error back-propagation using the labeled data. The high-level features learned by the final RBMs block are further fed into a softmax regression model to get probability prediction results. Then, the prediction error can be estimated by comparison with the target labels. According to the maximum likelihood function, the error back-propagation algorithm based on SGD is applied to fine-tune the weights and biases of each layer simultaneously.

B. GLOBAL BACK-RECONSTRUCTION (GBR)

In the developed DBNs+GBR, the global back-reconstruction (GBR) mechanism is employed in each RBMs block in DBNs, as illustrated in Fig.4. The role of the GBR mechanism is to restrict the accumulation of the reconstruction error to a range. More specifically, the GBR is achieved by building a "short connection" between the input layer and each hidden layer output. Then, the output data of each hidden layer is directly used to reversely reconstruct the input data to the most extent, which can be expressed by:

$$p_{\theta_n}(v) = \frac{\sum_{h^1, h^2, \cdots, h^n} \exp(-E_{\theta_n}(v, h^1, h^2, \cdots, h^n))}{\sum_v \sum_{h^1, h^2, \cdots, h^n} \exp(-E_{\theta_n}(v, h^1, h^2, \cdots, h^n))} \tag{14}$$

where θ_n represents parameters in the nth hidden layer, v represents the reconstruction status of the visible neurons in input layer. It should be noticed that the parameters of the current RBMs block are only updated, other RBMs blocks
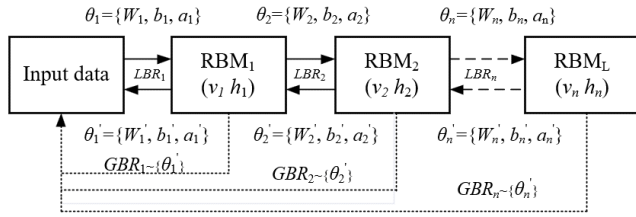
FIGURE 4. The GBR mechanism of the developed DBNs+GBR.

ahead are fixed. To be specific, the $GBR_n \sim \{\theta_n'\}$ shown in Fig.4 indicates the parameters θ_n of the nth RBMs block are required to be optimized in the GBR phase, but θ_1, θ_2, $\cdots$, θ_{n-1} all remain unchanged.

Aiming to reconstruct the input data to the maximum extent, the parameters θ_1, θ_2, $\cdots$, θ_n should be further optimized by maximizing the edge probability $P_{\theta_n}(v)$. The related logarithmic likelihood function is shown as:

$$L(\theta_n) = \frac{1}{M} \sum_{i=1}^{M} \max_{\theta_n} \sum_{v} \log(P_{\theta_n}(v)) \qquad (15)$$

where M is the number of the training dataset.

Then the reconstruction error is further estimated based on the two-order norm, taking the origin input data as the benchmark, mathematically expressed as follows:

$$Rerr = \frac{\sum_{i=1}^{M} \sum_{j=1}^{P} (v_{i,j} - y_{i,j})^2}{MPG} \qquad (16)$$

where P is the data point of single sample, Y is the input data, and G is the number of sampling points.

In addition, for the sake of effectively restricting the accumulation of the reconstruction error, it's necessary to preset threshold, so the following criterion is defined as follows:

$$\begin{cases} N_{rbms} = N_{rbms} + 1, & Rerr < \varepsilon \\ L = N_{rbms}, & Rerr \geq \varepsilon \end{cases} \qquad (17)$$

where ε is the preset threshold of the accumulation of the reconstruction error, and L is the number of stacked RBMs blocks in DBNs.

C. THE DEVELOPED DBNs+GBR
In this subsection, the motivations for developing DBNs+GBR are introduced, and the theory background of the developed DBNs+GBR is elaborated in detail.

1) THE REASONS FOR DEVELOPING DBNs+GBR
As above mentioned, the multiple RBMs blocks are bottom-up trained one by one during the pre-training phase in DBNs, which means each RBMs block only probabilistically reconstruct its own inputs (the process is called local back-reconstruction (LBR) mechanism in this paper). In the LBR mechanism, the distribution of each layer's outputs will change with the increase of layers during pre-training because of reconstruction error existence, leading to feature degradation in deeper layers. As a consequence, it is difficult to decide which layer to learn features in DBNs is useful for fault diagnosis by constructing the depth of DBNs artificially.

Hence, it is necessary to explore effective way to adaptively determine the depth of layers, avoiding the reconstruction error accumulation reaches a certain level and improving classification accuracy.

On the basis of CD algorithm and Gibbs sampling theory, the minimum reconstruction error between the input data distribution and the distribution in visual neurons learned by the first RBMs block can be quantified given by:

$$\Delta E_1 = \left\langle v^1 h^1 \right\rangle_{data} - \left\langle v^1 h^1 \right\rangle_{recon} \qquad (18)$$

where $\langle \cdot \rangle_{data}$ stands for the input data distribution, and $\langle \cdot \rangle_{recon}$ stands for the distribution learned by the first RBMs block after Gibbs sampling.

Then, based on the pre-training rule, the minimum reconstruction error between different-level RBMs blocks can be quantitatively estimated by:

$$\begin{cases} \Delta E_2 = \left\langle h^1 h^2 \right\rangle_{data} - \left\langle h^1 h^2 \right\rangle_{recon} \\ \Delta E_3 = \left\langle h^2 h^3 \right\rangle_{data} - \left\langle h^2 h^3 \right\rangle_{recon} \\ \vdots \\ \Delta E_n = \left\langle h^{n-1} h^n \right\rangle_{data} - \left\langle h^{n-1} h^n \right\rangle_{recon} \end{cases} \qquad (19)$$

where ΔE_n indicates the minimum reconstruction error of the nth RBMs block, h^n can be regarded as the high-level features learned by the nth RBMs block.

It is obvious that the probabilistic distribution of each hidden layer's output will gradually change to some extent as the depth of the DBNs increases during the unsupervised pre-training phase owing to the accumulation of reconstruction error. When reaching a certain threshold, the accumulation will pose an effect on the probabilistic distribution of the final hidden layer's outputs that will significantly deviate from the distribution of the input data. Thus, the representative ability of the high-level features also becomes inferior which can't reflect essential information of the input data, or in other words, the correlation between high-level features and the input data is weak. It is notable that the weight optimization in the fine-tuning training process will be affected unavoidably, and the performance of DBNs (e.g., generalization ability and robust ability) will begin to degrade eventually.

To solve the problem above mentioned, in the developed DBNs + GBR, the GBR mechanism is employed in each RBMs block to construct the novel DBNs architecture [see Fig.4]. and the GBR mechanism has many characteristics, which will be beneficial to improve the expression ability of high-level features extracted from 3-DBTC, such as:

(1) The weights to be further optimized by the GBR in each RBMs block in the unsupervised process and accelerate the speed of the fine-tuning training.

(2) The GBR has the ability to determine the depth of DBNs automatically by way of limiting the reconstruction error, avoiding the feature degradation in deep layers.

2) THE THEORY OF DBNs+GBR
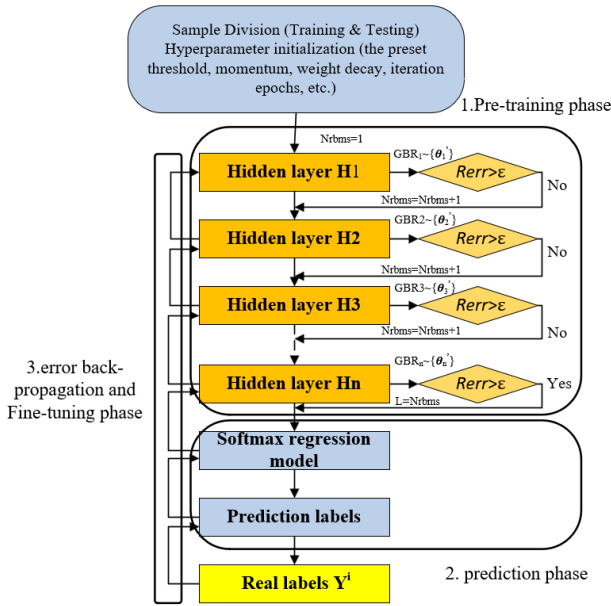As shown in Fig.5, the developed DBNs+GBR is a variant of DBNs that uses GBR mechanism to improve the ability

FIGURE 5. Computation schematic of the developed DBNs+GBR.

FIGURE 6. The definition of 3-DBTC and its parameters.

of feature learning. The GBR mechanism is inserted into the architecture of the conventional DBNs to improve the representative ability for input data. Moreover, the GBR mechanism can be beneficial to optimize the weights in DBNs and adaptively determine the depth of the DBNs, making less dependent on prior human labor and knowledge.

In Fig.5, the developed DBNs+GBR will probabilistically reconstruct the input data to a maximum extent using each hidden layer' output through GBR mechanism, optimizing weights in several RBMs blocks to learn high-quality features. In the GBR mechanism, the reconstruction error will be further calculated in each hidden layer according to equation (16). If the accumulation of reconstruction error exceeds the preset threshold ε, the developed DBNs+GBR will globally optimizes these trainable parameters in the supervised training. Otherwise, the RBMs block will be continuously stacked atop the RBMs block located ahead until the condition is satisfied.

Note that those parameters of all hidden layer will get optimized in the LBR and GBR phase, respectively, and it is useful to extract high-quality high-level features in the unsupervised pre-training. After the supervised fine-tuning, the high-level features can be employed to classify different conditions.

III. INTELLIGENT DIAGNOSIS APPROACH FOR THE CRACK FAULT OF TURBINE BLADES USING 3-DBTC SIGNALS

A. DESCRIPTION OF 3-DBTC SIGNALS

In the operating environment, turbine blades have to bear multiple dynamic loads, such as high temperature, mechanical stress, strong centrifugal force, or strong aerodynamic loading, etc. Thus, the turbine blades are prone to fatigue cracks. When the crack faults occur, the blades appear
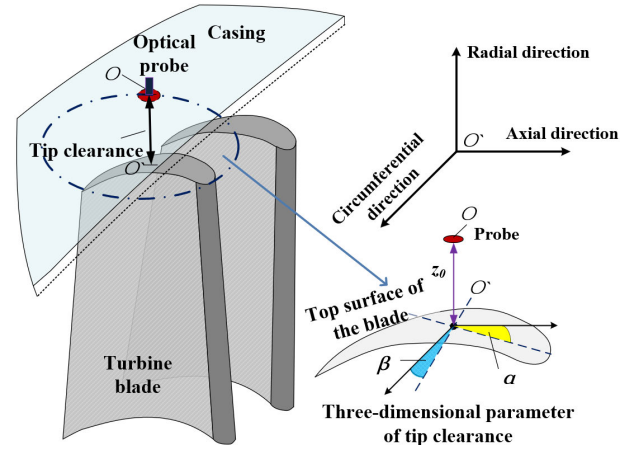
complex dynamic behaviors resulting in the change of three-dimension parameters. Hence, in the paper, the three-dimension parameters are simply called three-dimensional blade tip-clearance [22], abbreviated to 3-DBTC. In fact, the 3-DBTC refers to three geometric parameters (i.e., radial direction, axial direction, and circumferential direction.) of the whole space between the surface of the cross-section of the optical probe installed on the casing and the surface of the turbine blades, which can effectively reveal the evolution process of dynamical behaviors of turbine blades. It then follows that we can diagnose the crack fault of turbine blades by detecting the changes of 3-DBTC. Its diagrammatic description as illustrated in Fig.6.

In order to facilitate monitoring three geometric parameters, space coordinate system [see Fig.6] is built through setting axial direction, circumferential direction, and radial direction as the x-axis, y-axis, and z-axis, respectively, and the original point is set as the projective point O' of the fixed point O in which installs the optical probe on the casing. when the fatigue cracks occur on turbine blades, the deformation of the surface of the blades will appear, resulting in the projective point O' displaces along three-dimensional directions. To be specific, the description of the three-dimensional parameters as follows: 1) The radial tip clearance (RTC) Z_0, which refers to the radial distance between the midpoint O of the optical probe and its corresponding projective point O' on the surface of the blades; 2) The axial deviation angle (ADA) α, which refers to the intersection angle along axial direction; 3) The circumferential deviation angle (CDA) β, which refers to the angle slippage along the circumferential direction.

As shown in Fig.7 and Fig.8, the 3-DBTC change tendency of turbine blades in different health conditions (i.e., healthy and crack) under loads was simulated using Ansys software. For comparison between the 3-DBTC signals acquired in different locations, we chose nine sets of measurement points shown in Fig.7. As depicted in Fig.8, we can observe that the tendency change is consistent with each other among different health conditions, but the values are different in different measurement points, especially for the ninth point located in
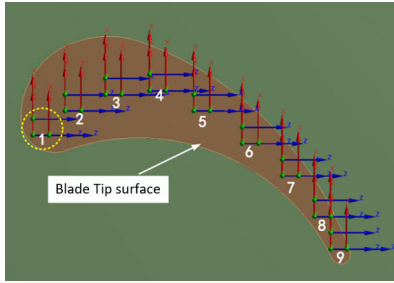
FIGURE 7. Distribution of measurement points.

the blade trailing edge, which indicates that the difference among different health conditions. In 8(c), we can observe that the distinguished difference in CDA between healthy and crack blades. Thus, we can dig out valuable information to identify health conditions of turbine blades from 3-DBTC. It is notable that the 3-DBTC signals can provide more rich information about health conditions from several aspects than single-dimensional monitoring signals like vibration signals.

B. SOFTMAX REGRESSION MODEL FOR CLASSIFICATION
In neural network, the softmax regression model is usually used at the final layer to establish the perfect mapping between the learned features and the health conditions to complete the model building. The softmax regression model can provide probabilistic classification and is often applied to process the multi-classification tasks. It is computed fast and is easy to be implemented [33]. Suppose that there is a training set $\{x^i\}_{i=1}^R$ with its labels $\{y^i\}_{i=1}^R$ where $x^i \in R^{P \times 1}$ an-d $y^i \in \{1, 2, \cdots, S\}$. For each input sample, the model attempts to estimate the probability for each label belonging to $\{y^i\}_{i=1}^R$. Then, the model will output a probability vector that includes S estimated probabilities of the input sample x^i to each label, in which the label having maximum probability is the real label for the input sample x^i. To be specific, the model can be defined as follows:

$$O(x^i) = \begin{bmatrix} P(y^i = 1|x^i; \theta_1^T) \\ P(y^i = 2|x^i; \theta_2^T) \\ \vdots \\ P(y^i = K|x^i; \theta_K^T) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^i}} \begin{bmatrix} e^{\theta_1^T x^i} \\ e^{\theta_2^T x^i} \\ \vdots \\ e^{\theta_K^T x^i} \end{bmatrix}$$
(20)

where $\theta = [\theta_1, \theta_2, \cdots, \theta_k]^T$ is the parameter vector of softmax regression model. the term $\sum_{j=1}^k e^{\theta_j^T x^i}$ is normalization f-actor, and the sum of all estimated results is guaranteed to equal 1.

Based on the estimated output, the model can be trained by error back-propagation (BP) algorithm to minimize the loss function $J(\theta)$.

$$J(\theta) = -\frac{1}{R} \left[\sum_{j=1}^R \sum_{s=1}^S 1\{y^i = s\} \log \frac{e^{\theta_k^T x^j}}{\sum_{j=1}^k e^{\theta_j^T x^j}} \right] + \frac{\lambda}{2} \sum_{s=1}^S \sum_{p=1}^P \theta_{sp}^2$$
(21)

where $1\{\cdot\}$ is an indicator function returning 1 when the estimated result is correct, and 0 otherwise. λ is the weight decay coefficient that force some parameters to take small absolute values, while other parameters to retain relatively large values. Thus, the generalization ability of the softmax regression model can be improved obviously [34].

C. INTELLIGENT FAULT DIAGNOSIS APPROACH FOR TURBINE BLADES'CRACK
As mentioned above, 3-DBTC signals can represent the dynamical behaviors of turbine blades from several aspects, which contains abundant information of turbine blades'-rack. Hence, the sensitive features extracted from the 3-DBTC signals can reflect the health conditions of turbine blades. Moreover, it is difficult to construct a set of suitable statistical features for blades' crack, and if the distributions of the features extracted are not separable enough for differentiating the health conditions (i.e., healthy and crack considered in paper), it can be challenging to achieve high diagnostic accuracy. Thus, in the paper, A novel intelligent approach based on improved DBNs for early crack of turbine blades using 3-DBTC signals is proposed to achieve high diagnostic accuracy for blade cracks, as illustrated in Fig.9.

The main procedures of the proposed method as follows: (1) the 3-DBTC signals of turbine blade are acquired by sensor and collected by data acquisition system under different running speeds. Further, the signals are split into two groups directly, constructing training dataset and testing dataset both involving different health states; (2) according to our experiments and experiences shared in other papers, the hyperparameters (e.g., weight decay, learning rate, mini-batch, momentum coefficient, preset threshold), which are used to define the detailed structure of the neural network, are determined. It is elaborated in detail in the next section; (3) the training samples are fed into the developed DBNs+GBR, and the high-level features are gradually extracted by several hidden layers in pre-training process. Then, those features are further input into softmax regression model concatenated in the final layer to get prediction results; (4) the error back-propagation algorithm is used to update all weights of hidden layers by top-down using the prediction errors, reducing the prediction mistakes. Finally, the fault diagnosis model trained is completed; (5) the testing samples are used to validate the efficacy of the proposed method in diagnosing crack fault of turbine blade.

IV. EXPERIMENTAL STUDY
A. DESCRIPTION OF EXPERIMENTAL PLATFORM AND DATA COLLECTION
The developed DBNs+GBR was used to pinpoint the health conditions (i.e., healthy and crack fault) in turbine blades. The experiments in the study used a turbine blade simulation test platform, which mainly consisted of a two-circle coaxial optical fiber probe, simulated rotor, driving motor, 3 DOF calibration device, measuring system etc., as shown
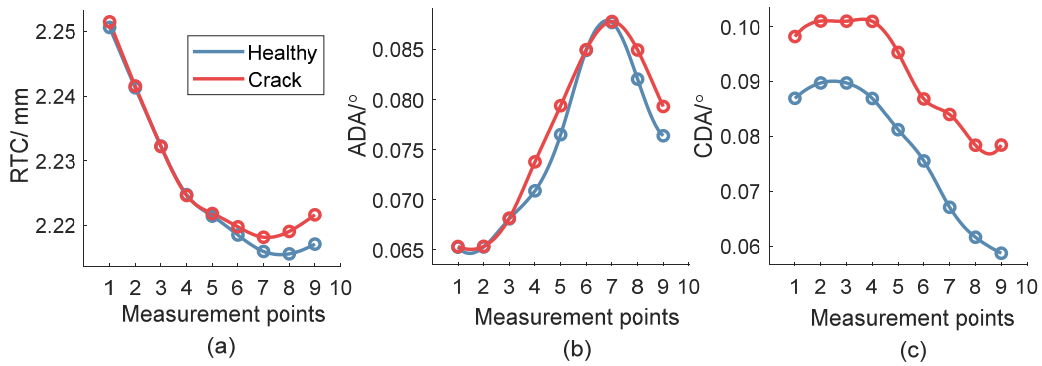
FIGURE 8. The 3-DBTC change tendency of turbine blades in different measurement points.
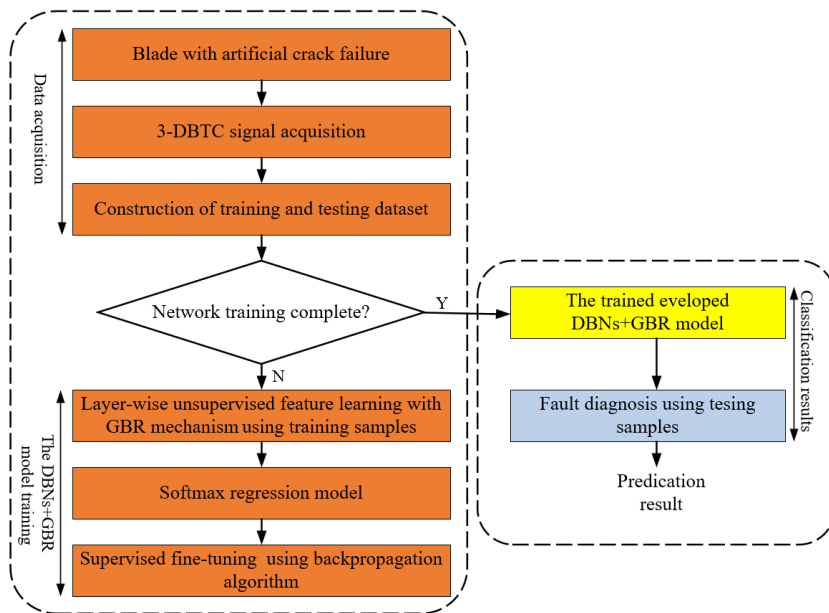

FIGURE 9. Approach for intelligent fault diagnosis for turbine blades' crack.
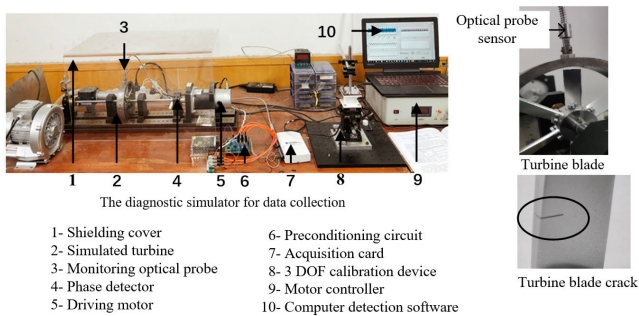

FIGURE 10. Turbine blade simulation test platform used for experiments.

1- Shielding cover
2- Simulated turbine
3- Monitoring optical probe
4- Phase detector
5- Driving motor
6- Preconditioning circuit
7- Acquisition card
8- 3 DOF calibration device
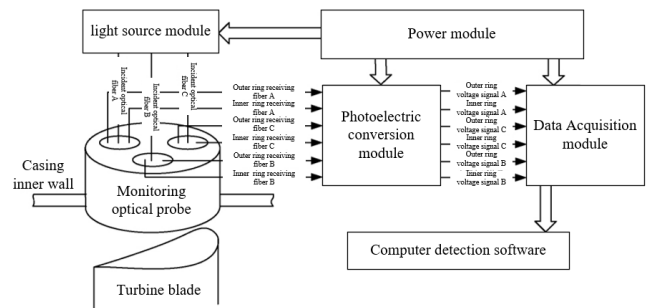9- Motor controller
10- Computer detection software

FIGURE 11. Diagrammatic description of the measuring system used for the simulation test platform.

in Fig.10 and Fig.11. The optical fiber probe mounted on the casing of the simulation turbine was used to acquire the 3-DBTC signals, then the feature extraction and fault diagnosis can be performed using the proposed method.

For the sake of verifying the efficacy of the developed DBNs+GBR, two health states in turbine blade were used

for diagnosis, as described in Table 1. A crack fault of 6mm in depth is firstly set artificially on turbine blade. Then, 3-DBTC signals sampled at 10kHz were collected via the optical fiber probe and measuring system [See Fig.11] under different rotational speeds (1000rpm/1500rpm) generated by driving motor. More specifically, each dimension component

TABLE 1. Summary of health conditions of the turbine blades considered in the study.

| Health Conditions / Three-Dimensional Blade Tip Clearance (3-DBTC)	Healthy		Crack	
	Number of Samples	Rotation Speed/(r·min⁻¹)	Number of Samples	Rotation Speed/(r·min⁻¹)
Racial Tip Clearance Z_0 (RTC)	210	1000/1500	210	1000/1500
Axial Deviation Angle α (ADA)	210	1000/1500	210	1000/1500
Circumferential Deviation Angle β (CDA)	210	1000/1500	210	1000/1500
The total number of datasets	630		630	
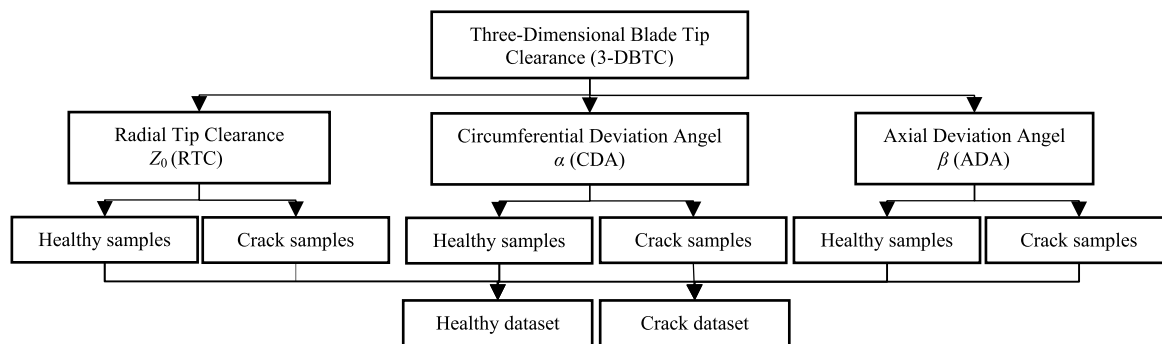
FIGURE 12. The preprocessing scheme for 3-DBTC signal.

in 3-DBTC signals was further divided into 210 samples by overlapped way, containing 1024 data points per sample. Hence, the total number of samples for each health condition considered in the study is 630. As shown in Fig.12, the procedure of sample construction for the DBNs+GBR is introduced.

B. HYPERPARAMETER SETUP

The hyperparameter selection greatly influences the performance of the developed DBNs+GBR, and therefore it is necessary to set these hyperparameters properly, including learning rate, mini-batch, momentum coefficient, weight decay coefficient, iteration epochs, and so forth, as well as preset threshold defined in this paper. Moreover, because no consensus has been reached as to how to set these common hyperparameters, many empirical suggestions for the hyperparameters selection are adopted according to [35], [36], and [37].

To be specific, the training rate is set to 0.001 at 60 epochs, so that the trainable parameters can be updated smoothly in training process. Momentum is a weight-updating strategy to accelerate the training process and avoid local optima using the updates of previous iterations. In general, the momentum coefficient is recommended to be 0.9. L2 regularization technology is adopted in the paper to increase the generalization ability and yield higher test accuracy. In L2 regularization,

a penalty term known as weight decay is applied in the objective function to push the weights towards to small absolute value, and the weight decay coefficient is set to 0.0001. The mini-batch refers to a group of arbitrarily selected samples that are fed into the networks at the same time, and the size of mini-batch is set to 20. In the paper, the initial number of RBMs blocks was to set to 1, and the number of hidden neurons in each hidden layer is set to 70% of that in its former layer accord-ing to [35]. In the training process, 60% of the total of dataset were used to train the developed DBNs+GBR, others were used to test.

Besides, it is important to select a proper preset threshold ε given those hyperparameters above. Hence, it is a need to explore how to set the threshold in order to achieve higher accuracy. In the paper, the setups for setting the preset threshold ε are detailed as follows:

Setup 1: Select randomly a part of training samples to build and train a conventional DBNs and calculates the reconstruction error of each RBMs block in the pre-training process according to equation (19).

Setup 2: Roughly estimate a proper range $\left[\varepsilon_{lower}, \varepsilon_{upper}\right]$ of the preset threshold ε based on the Setup 1.

Setup 3: Optimize the preset threshold ε by step-wise refinement. In general, as the preset threshold increases, the classification accuracy of the developed DBNs+GBR will increase, then decrease. In order to

TABLE 2. The diagnostic results using different methods.

Methods	The average fault diagnostic accuracy (%)	The standard deviation of accuracy (%)	The average F1 score
The proposed method	98.43	0.092	0.997
Improved DBNs in paper [35]	94.09	1.730	0.941
Conventional DBNs with 5 RBMs	88.11	2.231	0.843
Conventional DBNs with 6 RBMs	83.47	2.511	0.778
SAE	88.57	3.866	0.885
BPNN	58.10	1.116	0.501
BPNN+statistical features	98.08	0.336	0.991

narrow the researching range $\left[\varepsilon_{lower}, \varepsilon_{upper}\right]$, it is can be divided into N subintervals $\left[\varepsilon_{loweri}, \varepsilon_{upperj}\right]$, $i, j = 1, \cdots, N$. Then the ε_{loweri} and the ε_{upperj} both with highest accuracies constitute the new searching range $\left[\varepsilon_{loweri\_new}, \varepsilon_{upperj\_new}\right]$. If the $\left|\varepsilon_{loweri\_new} - \varepsilon_{upperj\_new}\right|$ is definitely small, then the preset threshold can be obtained from two endpoints $\varepsilon_{loweri\_new}, \varepsilon_{upperj\_new}$ and the midpoint $\left(\varepsilon_{loweri\_new} + \varepsilon_{upperj\_new}\right)/2$ by comparing accuracy. In the process, the test accuracy is the main selection criterion.

C. EXPERIMENTAL RESULTS AND DISCUSSIONS

In the section, some samples were applied to find an optimal preset threshold ε given other hyperparameters. With the method proposed in section 4.2, the optimal threshold ε was set to $7e^{-4}$ according to the previous experiment. Then, to further demonstrate the superiority of the proposed method, it was compared with three different methods including the improved DBNs in [35], two kinds of standard DBNs with the different numbers of RBMs blocks (i.e., the number of RBMs blocks in standard DBNs are 5 and 6, respectively), SAE used in [38] and back-propagation neural network (BPNN) with different inputs. The experimental results are shown below and further discussions are as follows.

1) PERFORMANCE COMPARISON

To evaluate the efficiency of the proposed method in diagnosing crack faults, different methods with feature learning ability were conducted. Considering the impact of stochastic factors, 14 repeated experiments were performed, and the average accuracies of the results are as displayed in Table 2. As indicated in Table 2, the proposed method obviously outperforms other considered different methods in this paper, yielding an average accuracy of 98.43%.

More specifically, the proposed method yields improvement of 4.34%, 10.32%, 14.96%, 9.86%, 40.33%, and 0.35% in terms of average testing accuracy, compared to other methods, respectively. Besides, the proposed method also
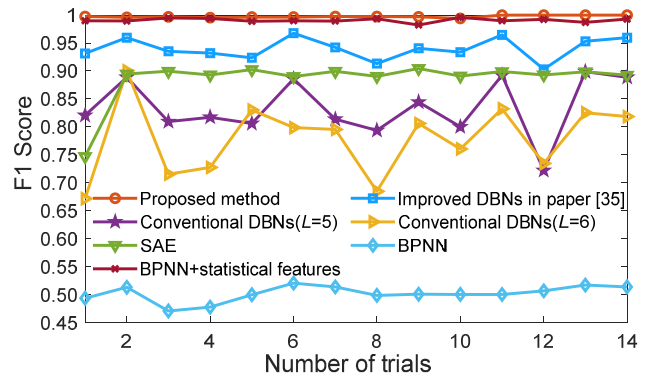

FIGURE 13. *F*1-score using different methods in 14 repeated experiments.

(a) The confusion matrix of the training set
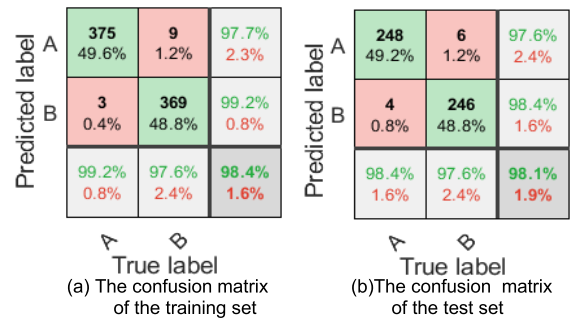
(b) The confusion matrix of the test set

FIGURE 14. Confusion matrix of the training set and test set using the proposed method.

has better performance in generalization and stability capability than those in other methods by reducing 1.638%, 2.139%, 2.419%, 3.774%, 1.024%, and 0.244% in terms of the standard deviation (SD) of accuracies, respectively, in repeated 14 experiments. Moreover, it worth noting that the BPNN+statistical features has as good diagnostical accuracy as the proposed method, but the BPNN and SAE are inferior, particularly for BPNN which is used to learn features from 3-DBTC. It mainly due to those statistical features retain discriminant information of blade crack, but the BPNN is hard to learn valuable features from 3-DBTC resulting in poor diagnostical effects. In fact, it is labor-consuming and hard to find a reliable set of hand-crafted features in real operating environments. Our analysis demonstrated that the proposed method in our paper could learn discriminant information from 3-DBTC to recognize health conditions of turbine blades. Compared with [35] and other two conventional DBNs, the proposed method also has significant advantages in diagnosing blade cracks with higher average accuracy and lower SD. The reason is that the parameter initialization is optimized well in the pre-training phase, avoiding the influence of the reconstruction error in deep layers, so that the high-level features are more discriminant than other methods.

To further compare the performances of different methods, a comprehensive evaluation criterion, so-called F1-score (also called F1-measure) [39], was adopted. The F-score considers both the precision P and the recall R of the diagnostic results, where the precision P is the function of the positive
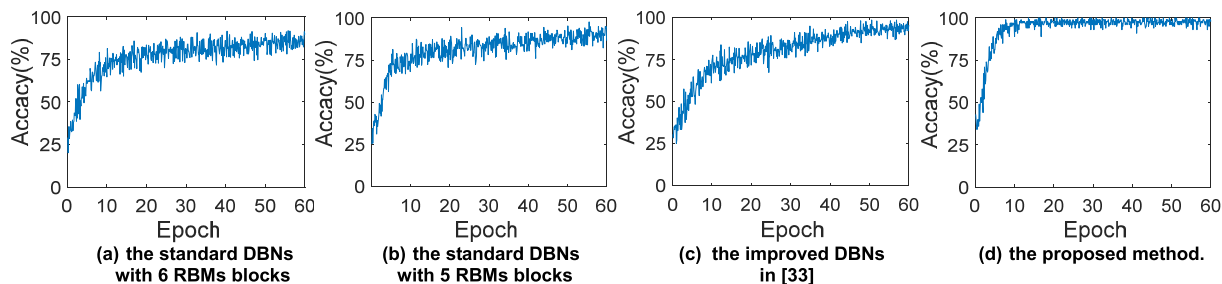
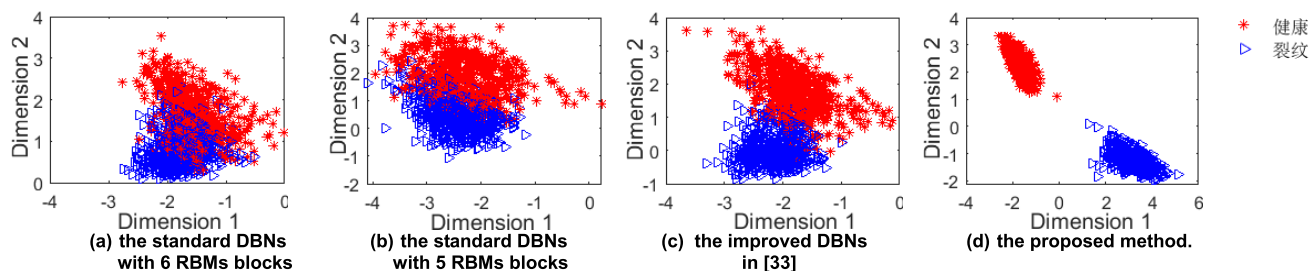**FIGURE 15.** Testing accuracy using different methods.

FIGURE 16. 2D visualizations of high-dimensional features at the final layer.

samples correctly classified (true positive, *TP*) and negative samples misclassified as positive (false positive, *FP*), and the recall *R* is the function of *TP* and its misclassified samples (false negative, *FN*). The *F*1-score reaches its best value at 1 (*P* and *R* both are perfect) and worst at 0.

$$p = \frac{TP}{TP + FP} \ and \ R = \frac{TP}{TP + FN} \quad (22)$$

$$F1\text{-score} = 2 \times \frac{P \times R}{P + R} \quad (23)$$

As depicted in Fig.13, the *F*1-scores using different methods in 14 repeated experiments are shown. Fig.14 presents confusion matrix of classification results in training and testing samples using the proposed method (label A and B present healthy and crack, respectively). In Fig.13, It can be seen that the *F*1-scores using the proposed method range from 0.9940 to 1 with the SD of 0.0019, and its average value is 0.997 in 14 experiments, which are higher than those obtained by other different methods considered in paper. Besides, the *F*1-scores using the improved DBNs in [35] ranges from 0.9025 to 0.9677 with the SD of 0.0194, and its average value is 0.941. Additionally, we can observe that the *F*1-scores using the BPNN+statistical features range from 0.9823 to 0.9950 with the SD of 0.3556 as good as the proposed method. However, other methods considered in paper were inferior, especially for BPNN. Otherwise, in contrast to the proposed method and [35], the *F*1-scores obtained by the two conventional DBNs were bad with more fluctuation than other methods in 14 repeated experiments, because the feature learning ability is easily influenced by feature degradation occurred in deeper layers.

As shown in Fig.14, It can be observed that the proposed method can recognize the crack and healthy samples correctly in training and testing datasets. Particularly for crack, the proposed method yields 97.6% and 97.6% in terms of recall *R* as well as 99.2% and 98.4% in terms of precision *P* in training and testing datasets, respectively. It indicated that the proposed method had the ability to identify the health conditions of turbine blade.

2) USEFULNESS OF FEATURE LEARNING IN FAULT DIAGNOSIS

For the sake of further demonstrating the validation of the proposed method, as shown in Fig.15, the training accuracies change tendency obtained from the proposed method and [35] as well as the two conventional DBNs were presented. It can be observed that all accuracy curves consist of two stages: the first one is a rapid learning phase, and the second one is the slow learning phase with almost constant accuracy. The proposed method has a steep learning curve at the first stage, but other methods have almost the same learning curve with a slow learning process. Besides, the proposed method moves into the second stage at the epochs of around 10, which is earlier than other methods obviously. It is noting that the accuracy curve of the proposed method is less fluctuation than other methods in the entire training process and converges to a final diagnostic accuracy of nearly 98.40% at the end of the training process, which is better than other methods. Our analysis indicated that the use of the GBR mechanism to dynamically adjust the weights and adaptively determine the depth of DBNs can effectively avoid feature degradation and improve the capability of learning

sensitive features for identifying crack faults in the turbine blade.

In addition, the t-distributed stochastic neighbor embedding (*t*-SNE) [40] was utilized to visualize the high-level features at the final layer in two-dimension space. The mapped features of different methods are shown in Fig.16(a)-(b), respectively. In Fig.16(a)-(b), the features of the two health states are mixed together and don't be separable well in the two standard DBNs. In Fig.14(c), the features learned by the improved DBNs [35] cluster better than those by the two standard DBNs, but still have a certain level of overlapping. In contrast, in Fig.14(d), it is seen that the features of the same health state are mostly gathered in the same region and are also basically separable from each other in the proposed method.

V. CONCLUSION

Finding a good set of features has been a long-standing problem in the fault diagnosis of turbine blades of the aero-engine. To solve this problem, a DBNs + GBR method was developed to learn a set of features from 3-DBTC signals that could discriminate diverse health conditions in the turbine blades—one health state and crack state. More specifically, the global-back-reconstruction (GBR) mechan-ism in the developed deep learning architecture was used to optimize weights applied to 3-DBTC signals contributed to discriminating the turbine blades' health states.

The efficacy of feature learning was verified through a set of comparisons between the developed DBNs + GBR method and three kinds of DBNs and SAEs as well as back-propagation neural network (BPNN). The developed method can automatically learn more sensitive features from the training data. Experimental results indicated that this method outperformed other three kinds of DBNs and SAE by yielding 4.34%, 10.32%, 14.96%, and 9.86% performance improvement in terms of average testing accuracy. That is, the inclusion of the GBR mechanism to learn a good set of features is significant for fault diagnosis of the turbine blades. Likewise, compared with BPNN with different input (original signal and the good statistical features), this method yielded 40.33% and 0.35% performance improvement in terms of average testing accuracy. it demonstrated that this method actually learned a good set of discriminant features from 3-DBTC signals.

In this study, the validation of the developed DBNs + GBR that learns a good set of features was verified by fault diagnosis of the turbine blades. However, there are still some shortcomings in this study. This paper only focused on the single crack fault on turbine blades, and in the coming research, the developed method will be further extended on the multiple crack fault diagnosis.

REFERENCES

[1] N. S. Vyas and J. S. Rao, "Dynamic stress analysis and a fracture mechanics approach to life prediction of turbine blades," *Mechanism Mach. Theory*, vol. 32, no. 4, pp. 511–527, May 1997.

[2] C. B. Meher-Homji, "Blading vibration and failures in gas turbines: Part A—Blading dynamics and the operating environment," in *Turbo Expo, Power for Land, Sea, and Air*, vol. 78811. New York, NY, USA: American Society of Mechanical Engineers, 1995

[3] M. M. M. Islam and J.-M. Kim, "Reliable multiple combined fault diagnosis of bearings using heterogeneous feature models and multiclass support vector machines," *Rel. Eng. Syst. Saf.*, vol. 184, pp. 55–66, Apr. 2019.

[4] A. Bustillo, G. Urbikain, J. M. Perez, O. M. Pereira, and L. N. L. De Lacalle, "Smart optimization of a friction-drilling process based on boosting ensembles," *J. Manuf. Syst.*, vol. 48, pp. 108–121, Jul. 2018.

[5] Á. Arnaiz-González, A. Fernández-Valdivielso, A. Bustillo, and L. N. L. De Lacalle, "Using artificial neural networks for the prediction of dimensional error on inclined surfaces manufactured by ball-end milling," *Int. J. Adv. Manuf. Technol.*, vol. 83, nos. 5–8, pp. 847–859, Mar. 2016.

[6] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[7] T. A. Ngo, Z. Lu, and G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Med. Image Anal.*, vol. 35, pp. 159–171, Jan. 2017.

[8] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3235–3243, Jul. 2018.

[9] 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.*, vols. 72–73, pp. 303–315, May 2016.

[10] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.

[11] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.

[12] C. Shen, J. Xie, D. Wang, X. Jiang, J. Shi, and Z. Zhu, "Improved hierarchical adaptive deep belief network for bearing fault diagnosis," *Appl. Sci.*, vol. 9, no. 16, p. 3374, Aug. 2019.

[13] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, Nov. 2015, Art. no. 115002.

[14] M. Ma, C. Sun, and X. Chen, "Discriminative deep belief networks with ant colony optimization for health status assessment of machine," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 12, pp. 3115–3125, Dec. 2017.

[15] D. Müller, A. G. Sheard, S. Mozumdar, and E. Johann, "Capacitive measurement of compressor and turbine blade tip to casing running clearance," *J. Eng. Gas Turbines Power*, vol. 119, no. 4, pp. 877–884, Oct. 1997.

[16] J. Chivers, "A technique for the measurement of blade tip clearance in a gas turbine," in *Proc. 25th Joint Propuls. Conf.*, Jul. 1989, p. 2916.

[17] S. J. Gill, M. D. Ingallinera, and A. G. Sheard, "Turbine tip clearance measurement system evaluation on an industrial gas turbine," in *Turbo Expo, Power for Land, Sea, and Air*, vol. 78712. New York, NY, USA: American Society of Mechanical Engineers, 1997.

[18] S. Heath and M. Imregun, "An improved single-parameter tip-timing method for turbomachinery blade vibration measurements using optical laser probes," *Int. J. Mech. Sci.*, vol. 38, no. 10, pp. 1047–1058, Oct. 1996.

[19] S. Heath and M. Imregun, "A survey of blade tip-timing measurement techniques for turbomachinery vibration," *J. Eng. Gas Turbines Power*, vol. 120, no. 4, pp. 784–791, Oct. 1998.

[20] S. Heath and M. Imregun, "A review of analysis techniques for blade tip-timing measurements," in *Turbo Expo: Power for Land, Sea, and Air*, vol. 78712. New York, NY, USA: American Society of Mechanical Engineers, 1997.

[21] E. Kramer, "Optical vibration measuring system for long, freestanding LP rotor blades," *ABB Rev.*, vol. 5, pp. 4–9, 1997.

[22] S. Xie, X. Zhang, B. Wu, and Y. Xiong, "Output characteristics of two-circle coaxial optical fiber bundle with regard to three-dimensional tip clearance," *Opt. Exp.*, vol. 26, no. 19, pp. 25244–25256, 2018.

[23] S. Xie, "Demodulation technique for 3-D tip clearance measurements based on output signals from optical fiber probe with three two-circle coaxial optical fiber bundles," *Opt. Exp.*, vol. 27, no. 9, pp. 12600–12615, 2019.

[24] S. Xie and X. Zhang, "Design and modeling of three-dimensional tip-clearance optical probe based on two-circle reflective coaxial fiber bundle," in *Proc. IEEE SENSORS*, Oct. 2016, pp. 1–3.

[25] F. Teng, X. Zhang, and S. Xie, "Research on variation mechanism of three-dimensional blade tip clearance of aero-engine," in *Proc. 13th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, Aug. 2016, pp. 1–6.

[26] X. D. Zhang, B. Wu, and S. Y. Xie, "Optical fiber measurement system for 3-D variation of turbine blade tip clearance," *Opt. Precis. Eng.*, vol. 26, no. 7, pp. 1578–1587, 2018.

[27] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[28] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," Dept. Comput. Sci., Univ. Colorado Boulder, Boulder, CO, USA, Tech. Rep. CU-CS-321-86, 1986.

[29] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 599–619.

[30] A. Fischer and C. Igel, "An introduction to restricted Boltzmann machines," in *Iberoamerican Congress on Pattern Recognition*. Berlin, Germany: Springer, 2012.

[31] R. Salakhutdinov and I. Murray, "On the quantitative analysis of deep belief networks," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, 2008, pp. 872–879.

[32] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.

[33] J. Behley, V. Steinhage, and A. B. Cremers, "Laser-based segment classification using a mixture of bag-of-words," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 4195–4200.

[34] J. Han, D. Zhang, S. Wen, L. Guo, T. Liu, and X. Li, "Two-stage learning to predict human eye fixations via SDAEs," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 487–498, Feb. 2016.

[35] G. Y. Pan, W. Chai, and J. F. Qiao, "Calculation for depth of deep belief network," *Control Decis.*, vol. 30, no. 2, pp. 256–260, Feb. 2015.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[37] 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.*, vol. 61, no. 5, pp. 2441–2451, May 2014.

[38] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017.

[39] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation," in *Proc. Australasian Joint Conf. Artif. Intell.* Berlin, Germany: Springer, 2006.

[40] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

XIAODONG ZHANG received the B.S. degree in thermal energy and power engineering, the M.S. degree, and the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 1982, 1992, and 1996, respectively.

He is currently a Professor and a Ph.D. Supervisor with the College of Mechanical Engineering, Xi'an Jiaotong University. More than 130 articles have been published in his research career. His main research interests include intelligent robotics, intelligent measurement, diagnosis and control technology, and bio-signals measurement technology and its applications in engineering.

Dr. Zhang was a recipient of the State Technological Invention 3rd Prize of China, in 1999, and the Shanxi University Education Technological Invention 2nd Prize of China, in 2013 and 2019.

YIWEI XIONG is currently pursuing the Ph.D. degree with the College of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. His current research interests include fault diagnosis and health management for aero-engine.

HONGCHENG LIU is currently pursuing the Ph.D. degree with the College of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. His current research interests include bio-signals measurement technology and its applications in engineering.

XIN HUANG was born in Sichuan, China, in May 1993. He received the M.S. degree in mechanical engineering from Chongqing Jiaotong University, Chongqing, China, in 2019. He is currently pursuing the Ph.D. degree under the supervision of Prof. Xiaodong Zhang with the College of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. His research interests include data-driven fault diagnosis, signal processing, and prognostics and health management of aero-engine.

YINGJIE ZHANG received the B.S. and M.S. degrees in mechanical manufacturing, and the Ph.D. degree from the Key Laboratory of Ministry of Education, CAD/CAM Research Center, Northwestern Polytechnical University, Xi'an, China, in 1986, 1989, and 1995, respectively.

He is currently a Professor and a Ph.D. Supervisor with the College of Mechanical Engineering, Xi'an Jiaotong University, Xi'an. More than 120 articles have been published in his research career. His main research interests include CAD/CAPP/CAM and intelligent manufacturing, robot and numerical control systems, and graphics and image processing.

· · ·