

Research Article

Multisensor-Based Heavy Machine Faulty Identification Using Sparse Autoencoder-Based Feature Fusion and Deep Belief Network-Based Ensemble Learning

Yiqing Zhou ¹, Jian Wang,¹ and Zeru Wang²

¹Computer Integrated Manufacturing System (CIMS) Research Center, College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

²CAD Research Center, College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Yiqing Zhou; 1710334@tongji.edu.cn

Received 5 April 2022; Accepted 3 June 2022; Published 21 June 2022

Academic Editor: Mu Zhou

Copyright © 2022 Yiqing Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Faulty identification plays a vital role in the area of prognostic and the health management (PHM) of the industrial equipment which offers great support to the maintenance strategy decision. Owing to the complexity of the machine internal component-system structure, the precise prediction of the heavy machine is hard to be obtained, thus full of uncertainty. Moreover, even for a single component, the feature representation of the acquired conditional monitoring signal can be different due to the different deployment of the sensor location and environmental inference, causing difficulty in feature selection and uncertainty in faulty identification. In order to improve the model identification reliability, a novel hybrid machine faulty identification approach based on sparse autoencoder- (SAE-) and deep belief network- (DBN-) based ensemble learning is proposed in this paper. First, six kinds of statistical features are extracted and normalized from multiple sensors monitoring the same target component. Second, the six extracted features are fused by the two-stage SAE proposed in this paper from the sensor dimension and feature dimension, respectively. The composite feature fused in the feature dimension is regarded as the comprehensive representation of the corresponding component. Finally, the fused features containing comprehensive representation of different components are utilized to predict the machine health condition by the ensemble of multiple deep belief classifiers. The effectiveness of the proposed method is validated by the two case studies of wind turbine gearbox and industrial port crane. The experimental result shows that the proposed ensemble learning approach outperforms other traditional deep learning approaches in terms of the prediction accuracy and the prediction stability when dealing with multisensor feature fusion and the precise faulty identification of the industrial heavy machine.

1. Introduction

With the development of the modern industrial manufacturing scale, the accurate faulty identification of the heavy industrial machine has been becoming increasingly important. The contemporary heavy machine has the characteristic of having highly complex internal component-system structure. Moreover, even for a certain component, multiple sensors are mounted on different locations to acquire complementary information. How to effectively utilize these multisensor information and raise the reliability of the machine faulty identification result remains a great challenge.

The conventional faulty identification approach is usually based on the historical conditional monitoring data which can be very useful for making the appropriate maintenance strategy to avoid catastrophic failure and save excessive maintenance costs. The traditional intelligent faulty identification procedure usually consists of three steps: feature extraction; feature fusion; and faulty identification.

1.1. Feature Extraction. During the feature extraction period, suitable statistical features are extracted and processed from the obtained sensor monitoring signal on the data space based on the expertise knowledge of signal processing to construct a suitable health indicator (HI) which can well

represent the machine health status and provide useful information for faulty identification. The construction approach of the HI can be categorized into two categories: single sensor-based approach and multisensor-based approach. The single sensor-based approach is totally based on the understanding of the extracted single source monitoring signal by selecting the appropriate statistical and signal processing analysis method. Gebrael et al. [1] extracted the average amplitude and harmonic wave of the vibration signal as the health indicator to represent the bearing condition status. Malhi et al. [2] extract the root mean squared error and the peak value of the vibration signal to construct the health indicator of bearing status by using continuous wavelet transform. Kilundu et al. [3] construct the bearing HI indicator by using the singular spectrum analysis of the vibration signal. The single signal-based faulty identification approach can capture the faulty symptom of the target component with specific physical meaning; however, it can only capture the faulty symptom from a single data dimension with lower confidence level.

In order to raise the reliability of the faulty identification result, the multisignal-based approach is proposed which reflects the potential faulty symptom of the target component from multiple data dimensions [4–6]. In the multisignal-based approach, signals including forces, vibrations, temperatures, and acoustics are fused for the faulty identification task [7]. Compared to the single signal-based approach, the multisignal-based faulty identification approach can make the identification result more reliable [8–9]. Hao et al. [10] proposed a multisensor-based approach for the degradation identification of the mechanical component by evaluating the composite index which is combined with multiple sensor signals collected under multiple operational conditions. A motor faulty identification model based on sensor data fusion by using support vector machine and the short-term Fourier transform is proposed by Banerjee and Das [11].

1.2. Feature Fusion. The feature fusion is usually performed on the feature space where different kinds of statistical features extracted during the feature extraction phase are integrated into a composite feature by using technologies such as principal component analysis (PCA), kernel-based methods, and manifold learning. The feature fusion procedure is aimed at achieving considerable information compression and facilitating more effective feature. Xu et al. [12] fuse the time-domain and frequency-domain features extracted from the flame oscillation signal by using PCA. Wang et al. [13] propose the feature fusion and feature selection method from forces and vibration signals by using kernel PCA. Sun et al. [14] proposes a feature fusion method for handwritten recognition by using locally linear embedding.

1.3. Faulty Identification. During the faulty identification phase, the composite feature obtained from the feature fusion step is sent to a classifier for faulty identification [15–16]. Currently, various kinds of machine learning classifiers have been widely explored in the faulty classification stage; however, it is extremely hard for shallow-structured classifier to achieve precise faulty prediction for industrial

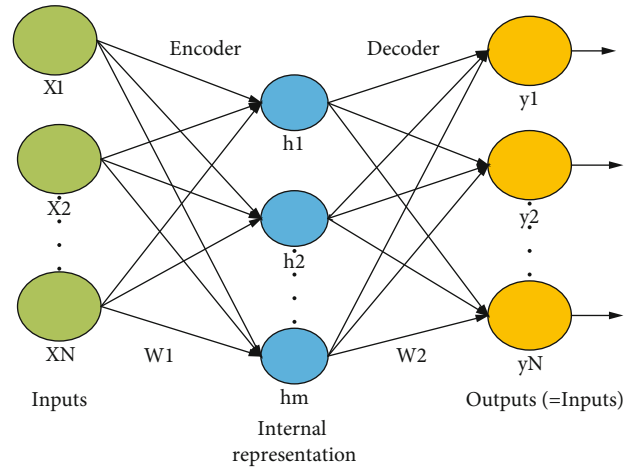


FIGURE 1: The detailed structure of autoencoder.

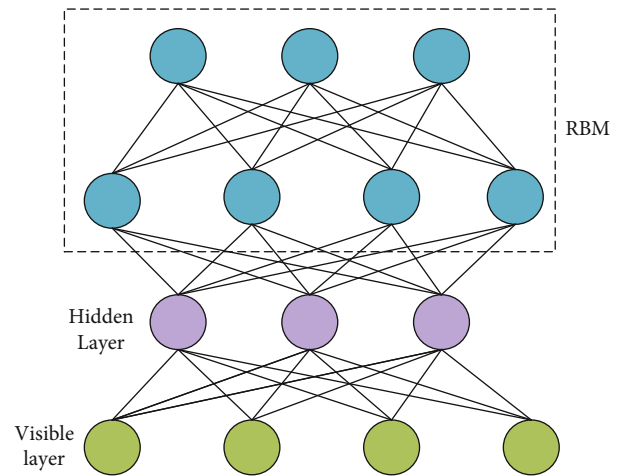


FIGURE 2: The structure of deep belief network.

heavy machine with complex internal component-system structure [17–19].

Recently, with the development of the deep learning technology, it has been widely used in feature extraction: feature fusion and faulty identification phase due to its powerful automatic feature learning and input-output mapping ability with its multi-layer structure. The deep learning-based faulty identification approach can release the dependence on the expertise knowledge of faulty identification, and it has been proved to be effective in improving the identification result [20–23]. However, there are three issues needed to be further considered.

- (1) During the feature extraction period, most existing methods directly extract statistical features from multiple sensor signals. However, these sensors are mounted on different locations to acquire a variety of condition monitoring signals with different characteristics due to sensors' arrangement and environment inference. Therefore, the features extracted from these sensors are disordered and correlated

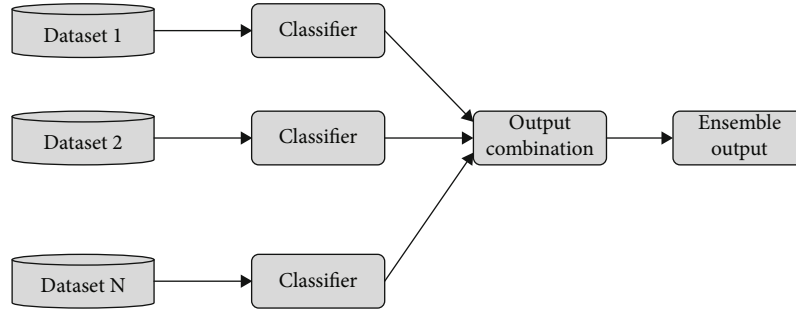


FIGURE 3: The scheme of homogeneous approach.

with different sources, increasing the difficulty of representative feature extraction. How to effectively merge these multisensor features to learn an over complete representation of these multisource information remains a challenge

- (2) During the feature fusion period, conventional feature fusion methods such as PCA, NPE (neighborhood preserving embedding), and LLE (locally linear embedding) fail to learn high-level features due to the nonstationary characteristics of fault vibration signals. How to jointly utilize different kinds of extracted features to construct a more robust feature remains a challenge
- (3) During the faulty identification period, most existing methods only use a single classifier as the final decision maker to output the final prediction result. Since the machine fault reason can be related to multiple components due to its highly complex internal component-system structure, the prediction result of a single classifier can be full of uncertainly. Since different components can reflect machine faulty conditions from different aspects, how to construct a hierarchical faulty prediction approach reflecting the complex component-system relationship network remains a challenge

Dealing with the three issues listed above, this paper proposes a multisensor-based hierarchical machine faulty identification approach by using a two-stage sparse autoencoder-based feature fusion and the deep belief network-based ensemble learning. The innovation point of the proposed paper is listed as follows: Considering the first issue listed above, six kinds of statistical features are extracted and normalized from sensors monitoring the same component, and these features are sent to six different three-layer SAEs for feature extraction. The middle layers of six SAEs are extracted as the complete representation of the multisensor input signals. Considering the second issue listed above, the six extracted middle-layer features are concatenated and sent to a specially designed three-layer SAE for feature fusion to obtain a more robust feature. The six extracted middle-layer features are merged into a six-node compressed feature which is regarded as the composite health indicator of the target component. Considering the

third issue listed above, the composite health indicators obtained during the feature fusion period which reflects the comprehensive health statuses of different components are sent to multiple DBN classifiers to independently classify the machine faulty status, and the outputs of these classifiers are aggregated by using the Bayesian weighting strategy which represent the affiliated degree between the target component and the certain machine conditions. The rest of this paper is organized as follows: Section 2 briefly reviews the related research literature including the sparse autoencoder, the deep belief network, and the ensemble learning; Section 3 presents the framework of the proposed hybrid faulty prognostic approach; Section 4 describes the case study and the competition with other deep learning models; and finally, Section 5 summarizes the main contribution and future work of the proposed paper.

2. Methodology

2.1. Autoencoder and Sparse Autoencoder. The autoencoder is a kind of unsupervised feature learning approach which can learn the high-level representation of the raw input signal. The output layer of the autoencoder has the same dimension as the input layer which minimizes the reconstruction error between the input and the output so that the high-level features contained in the hidden layer can be obtained. The learning process of the autoencoder consists of two procedures: encoding and decoding. The encoding process acts as a feature extractor realizing the transformation from the raw input into the high-level feature, and the decoding process acts as a feature detector which reconstruct the input in the output layer based on the obtained high-level feature. The detailed structure of the autoencoder can be illustrated in Figure 1.

Assuming the n -dimensional raw input vector is $\{X_1, X_2, \dots, X_N\}$, $X_i \in R^n$. During the encoding process, the raw input vector was transformed into the high-level feature represented in the following equations:

$$h_j = f \left(\sum_{i=1}^N w_{ij}^1 \times X_i + b_j^1 \right), \quad (1)$$

$$f(z) = 1/(1 + \exp(-z)), \quad (2)$$

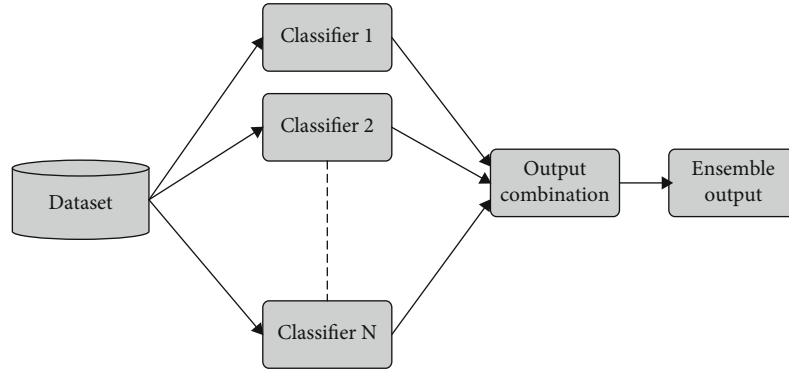


FIGURE 4: The scheme of heterogeneous approach.

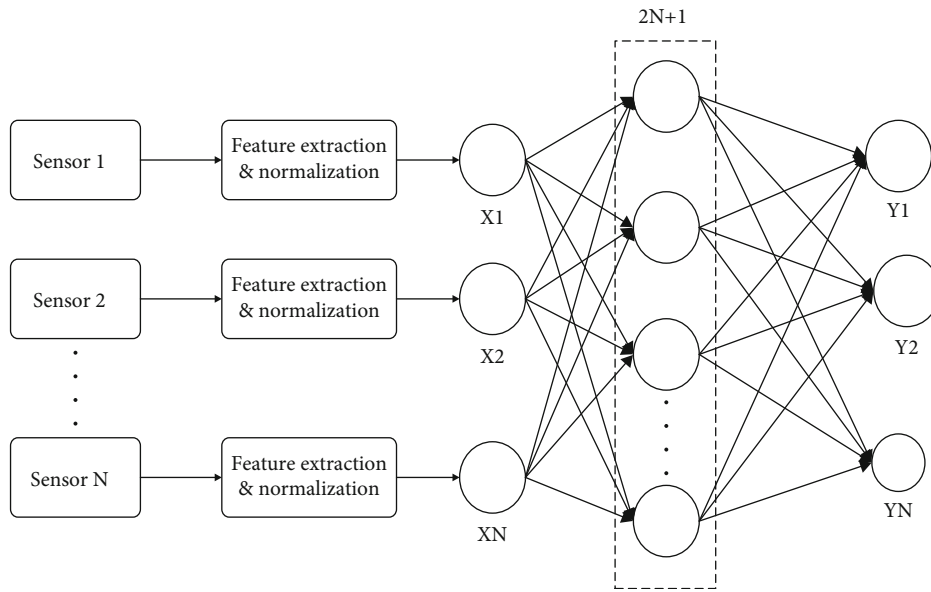


FIGURE 5: SAE fusion used for comprehensive feature representation.

where w_{ij}^1 denotes the weight between the i th dimension in the input layer and the j th dimension in the hidden layer. The parameter b_j^1 denotes the bias of the j th node in the hidden layer. The function $f(*)$ denotes the activation function which transforms the raw input vector into the high-level feature.

During the decoding process, the decoder transforms the hidden layer h_j into the output layer y , and the transform function can be illustrated in the following equation:

$$y_{k,b}(x) = f\left(\sum_{j=1}^m w_{jk}^2 \times h_j + b_k^2\right), \quad (3)$$

where w_{jk}^2 denotes the j th dimension in the hidden layer and the k th dimension in the output layer. The parameter b_k^2 denotes the bias of the k th node in the output layer. In order to obtain the compressed high-level information of the raw

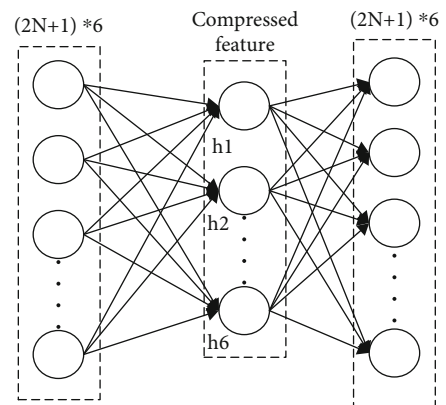


FIGURE 6: SAE fusion used for compressed feature representation.

input signal, the “sparse” restriction is applied on the hidden layer. The “sparse” restriction is aimed at making the statuses of the majority nodes at an inactive stage whose output

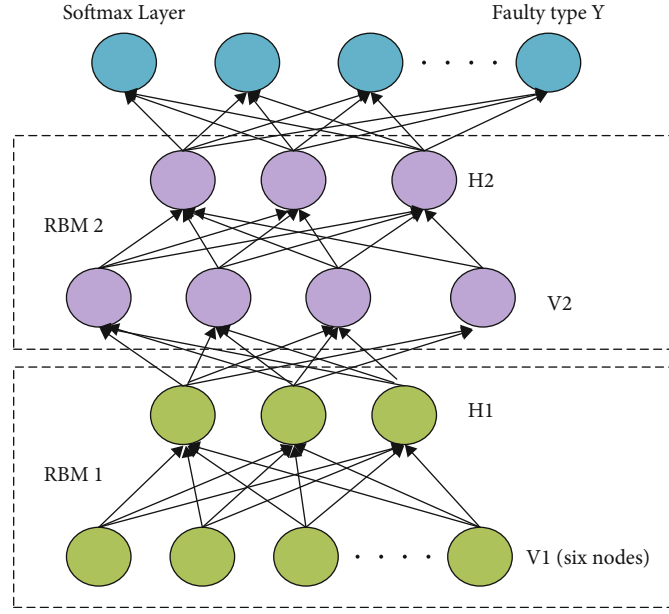


FIGURE 7: The detailed structure of the proposed DBNs.

is close to zero. The “sparse” restriction can be illustrated as shown in the following:

$$\tilde{p}_j = \frac{1}{n} \times \sum_{i=1}^n [\alpha_j(x_i)]. \quad (4)$$

Here, the parameter \tilde{p}_j denotes the average activation value of the j th node of the hidden layer, and n denotes the number of the training samples. $\alpha_j(x_i)$ denotes the activation value of the j th hidden node of the i th sample.

During the training stage, the activated output of the hidden layer is restricted within the predefined sparse value as is illustrated in the following equation:

$$\sum_{j=1}^n KL(p|\tilde{p}_j) = \sum_{j=1}^n p \times \log\left(\frac{p}{\tilde{p}_j}\right) + (1-p) \times \log\left(\frac{1-p}{1-\tilde{p}_j}\right). \quad (5)$$

Here, the parameter p denotes the “sparse” parameter. The KL (Kullback-Leibler) divergence is used for the similarity evaluation between the “sparse” parameter p and the actual average activation \tilde{p} . Once the $\tilde{p}_j = p$, the KL divergence equals zero, and the loss function of SAE can be represented as shown in the following equations:

$$J_{\text{Sparse}}(w, b) = J(w, b) + \beta \times \sum_{j=1}^{S_2} KL(p|\tilde{p}_j), \quad (6)$$

$$J(w, b) = \left[\frac{1}{n} \times \sum_{i=1}^n J(w, b; x^{(i)}, y^{(i)}) \right] + \frac{\lambda}{2} \times \sum_{l=1}^{n_{l-1}} \sum_{i=1}^{S_l} \sum_{j=i}^{S_{l+1}} \left(w_{ji}^{(l)} \right)^2. \quad (7)$$

Equation (6) denotes the loss function of the autoencoder, and Equation (7) denotes the “sparse” restriction. The sparse autoencoder is aimed at making the output as equivalent as possible to the input, obtaining the high-quality internal representation of the input vector. In this paper, the sparse autoencoder is used for the construction of the health indicator of individual component, containing the information of different sensor locations and feature characteristics.

2.2. Deep Belief Network. The deep belief network is a kind of unsupervised greedy layer-by-layer training algorithm which was proposed by Hinton in 2006 for solving the problem of structural optimization existed in the deep learning algorithm. The DBN is constructed by stacking multiple layers of restricted Boltzmann machines which is illustrated in Figure 2.

Each RBM consists of two layers with visible units and hidden units, respectively, and there is no connection between the same layers which only exists between the visible layer and the hidden layer. The training process of the DBN consists of two stages: pretraining the stacked RBM one by one in a greedy way and fine-tuning the whole network for obtaining an ideal performance. Since the DBN is composed of multiple stacked RBMs, it can be trained through contrastive divergence [24] in an unsupervised manner.

In recent years, the DBNs have been successfully used in faulty identification areas [25–27]. In this paper, multiple DBN classifiers are used independently for machine faulty classification from the aspect of different components.

2.3. Ensemble Learning. The ensemble learning is based on the notion that “two heads are better than one head.” The performance of aggregating multiple classifiers has been

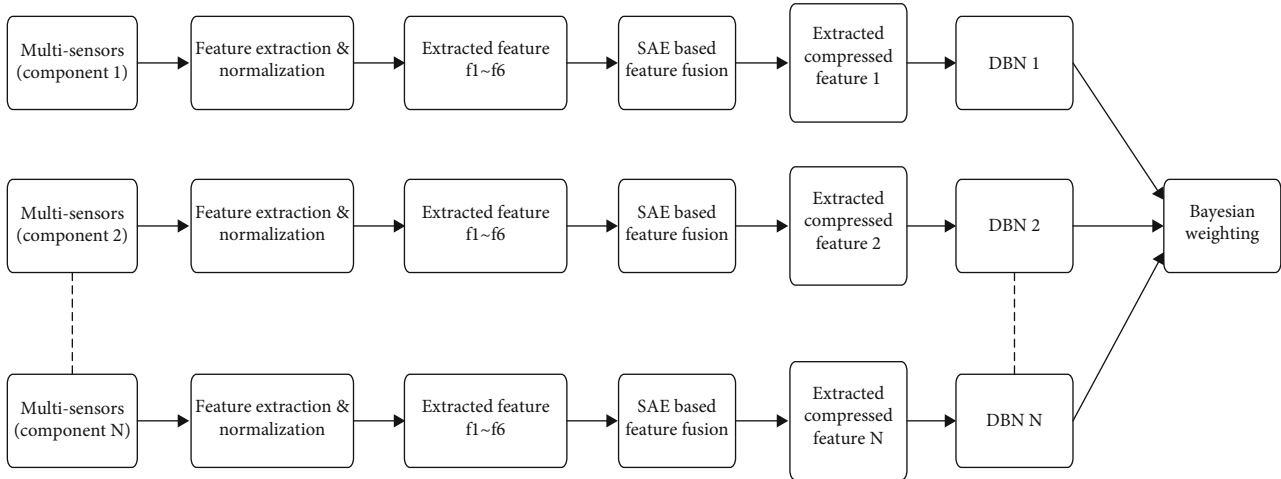


FIGURE 8: The overall framework of the proposed approach.

Input: Given the condition monitoring data collected from multiple sensors

Output: The machine faulty identification result

Step 1: Categorizing the multisensor data according to component

The conditional monitoring data collected from the sensors monitoring the same component with different sensing locations are categorized into the same group

Step 2: Feature extraction and feature fusion

2.1 The categorized sensor data in step 1 is segmented into two categories, one for training and the other for testing

2.2 The six time-domain features ($f_1 \sim f_6$) illustrated in Equations (8)–(13) are extracted and normalized by using Equation (14)

2.3 The six normalized features are sent to six SAEs, respectively, with the structure of Figure 5 for unsupervised training by using Equations (1)–(7) and obtain the complete expression of the six normalized features, respectively

2.4 The complete representation of the six normalized feature in Section 2.3 are concatenated and sent to the SAE with the structure of Figure 6 for unsupervised training by using Equations (1)–(7) and obtain the six-node compressed feature

2.5 The six-node compressed features are extracted and regarded as the comprehensive representation of a certain target component

Step 3: Machine faulty identification based on the proposed multi-DBN-based ensemble learning

3.1 The constructed compressed comprehensive features with respect to multiple components are sent to multiple individual DBN classifiers for predicting the machine health status by adding the Softmax prediction layer from the component perspective and the individual classifiers are trained by using Equations (15)–(19)

3.2 The Softmax outputs of the multiple DBN classifiers are aggregated based on the dynamically updated correlation weight obtained by using the Bayesian weighting strategy as shown in Equations (20)–(22)

3.3 Output the heavy machine condition result based on the maximum weighted output class

3.4 Calculate the mean value of the correlation weight of each individual classifier during the training process

Step 4: Evaluate the proposed methodology

Validate the proposed ensemble learning methodology in the testing datasets by using the mean value of the correlation weight obtained during the training process

ALGORITHM 1: General procedure of the proposed methodology.

proved to be better than a single classifier in many fields [28–31]. By using the appropriate combination strategy, the ensemble learning can fully take advantage of each individual classifier so as to improve generalization [32]. The ensemble learning can be usually categorized into two types: sequential ensemble learning and parallel ensemble learning. In sequential ensemble learning, different algorithms are sequentially combined in the way that the first algorithm is used to generate a model, and then the second algorithm is used to correct the first model and so on which is also called boosting. In parallel ensemble learning, different algorithms are used for model training independently which is also

called bagging. There are two kinds of parallel ensemble learning namely homogeneous ensemble learning and heterogeneous ensemble learning [33].

2.3.1. Homogeneous Ensemble Learning. In the homogeneous ensemble learning scheme, the same types of classifier are used but with different training datasets. These datasets are collected from multiple data sources, and the data federation of multisources is achieved which is illustrated in Figure 3. This type of ensemble is also named as data variation ensemble, and it is mainly used for the management of industrial big data under large-scale scenarios.

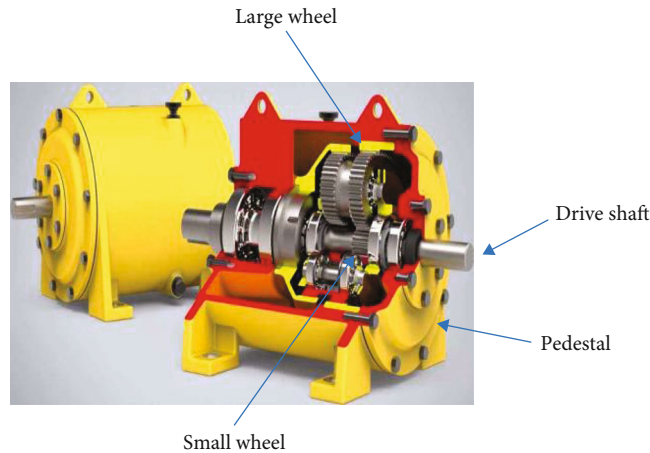


FIGURE 9: The component-system structure of the turbine gearbox.

TABLE 1: Description of datasets of Case Study I.

Datasets	Speed range (rpm)	Sample number in each condition	Condition status
Training/testing	800-1600/800-1600	1080/1080	Healthy
		180/180	Rotor unbalance
		180/180	Rotor misalignment
		180/180	Rotor friction
		180/180	Bearing loosing

2.3.2. Heterogeneous Ensemble Learning. In the heterogeneous ensemble learning scheme, a number of different types of classifiers but over the same training datasets are applied as can be seen in Figure 4, and the model diversity is achieved in this scheme. This type of ensemble is also called function variation ensemble which is mainly used for the generalization enhancement of model outputs.

In this paper, the homogeneous ensemble learning of several DBN classifiers are used for the construction of the hierarchical data processing framework of the component-system network.

3. Proposed Methodology

3.1. Data Preprocessing and Feature Fusion with SAE

3.1.1. Data Segmentation. The conditional monitoring data are obtained from multiple sensors monitoring the same target component but with the different installed locations. The obtained datasets were segmented into two categories: one for training and the other for testing.

3.1.2. Feature Extraction and Normalization. During the feature extraction and normalization period, the six time-domain features including the impulse $X_{Impulse}$, kurtosis $X_{Kurtosis}$, skewness $X_{Skewness}$, shape factor $X_{Shape\ factor}$, clearance factor $X_{Clearance\ Factor}$, and crest factor X_{Crest} are used

in this paper. Formulas for the six used features are presented as follows:

$$X_{Impulse} = \frac{\max(x(i))}{1/N \sum_{i=1}^N |x(i)|}, \quad (8)$$

$$X_{Kurtosis} = \frac{\sum_{i=1}^N (x(i) - \bar{x})^4}{N \times \left(\sqrt{\frac{\sum_{i=1}^N (x(i) - \bar{x})^2}{n}} \right)^4}, \quad (9)$$

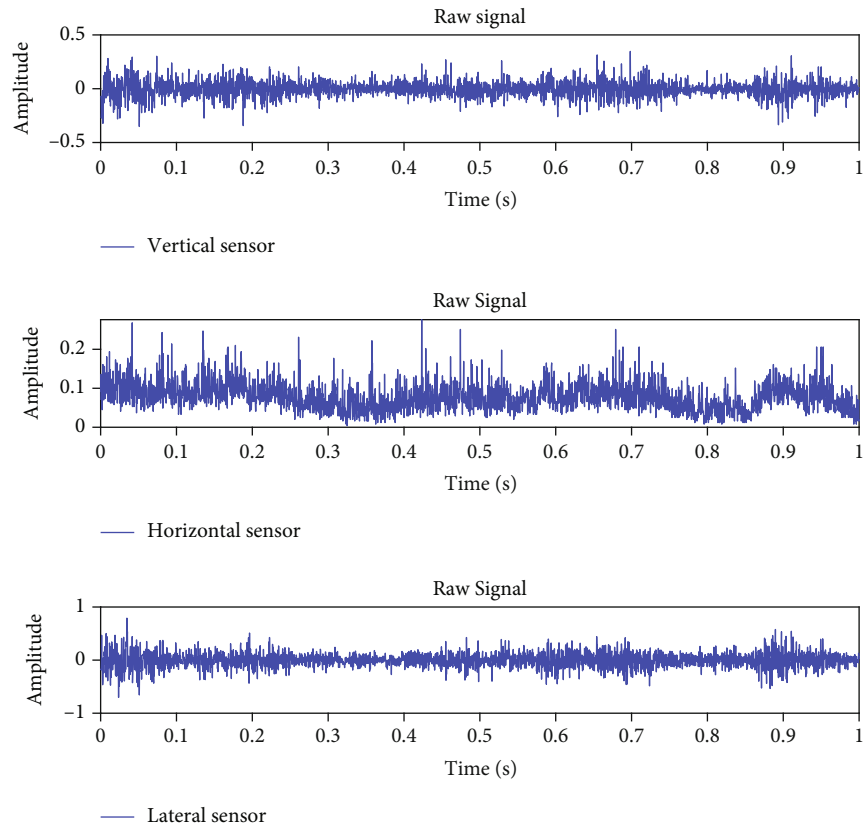
$$X_{Skewness} = \frac{1}{N} \times \frac{\sum_{i=1}^N (x(i) - \bar{x})^3}{\left(\sqrt{1/N \times \sum_{i=1}^N (x(i) - \bar{x})^2} \right)^3}, \quad (10)$$

$$X_{Shape\ factor} = \frac{\sqrt{1/N \times \sum_{i=1}^N (x(i))^2}}{1/N \sum_{i=1}^N |x(i)|}, \quad (11)$$

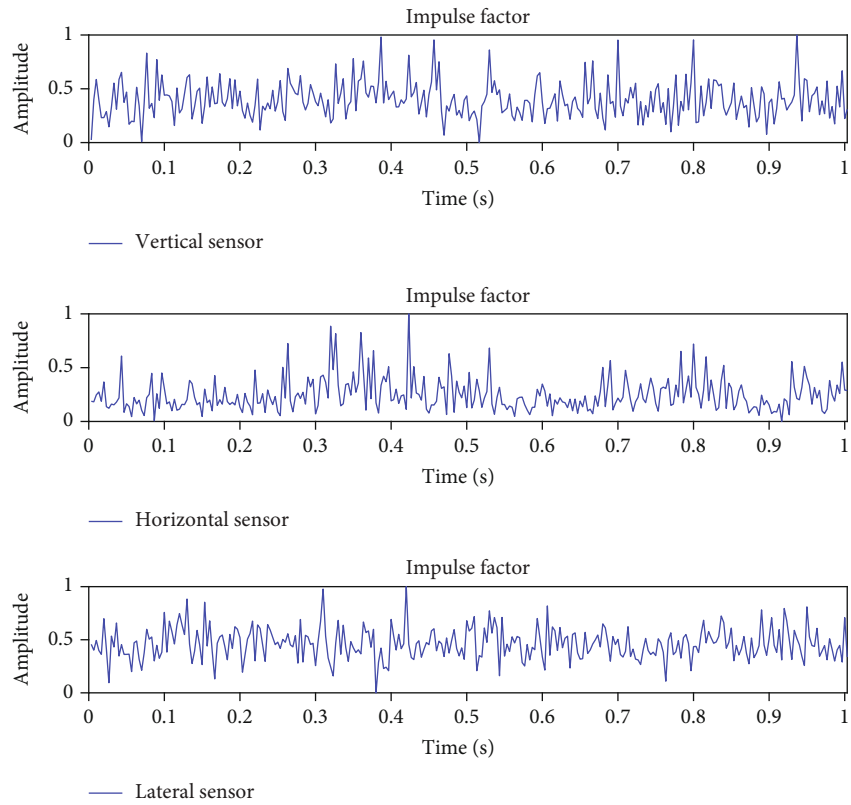
$$X_{Clearance\ Factor} = \frac{\max |x(i)|}{\left(1/N \times \sum_{i=1}^N |x(i)|^{1/2} \right)^2}, \quad (12)$$

$$X_{Crest} = \frac{\max(x(i))}{\sqrt{\left(\sum_{i=1}^N (x(i))^2 \right) / N}}. \quad (13)$$

where $x(i)$ denotes a series of sensor signal and the parameter "N" refers to window length. The six extracted



(a)



(b)

FIGURE 10: Continued.

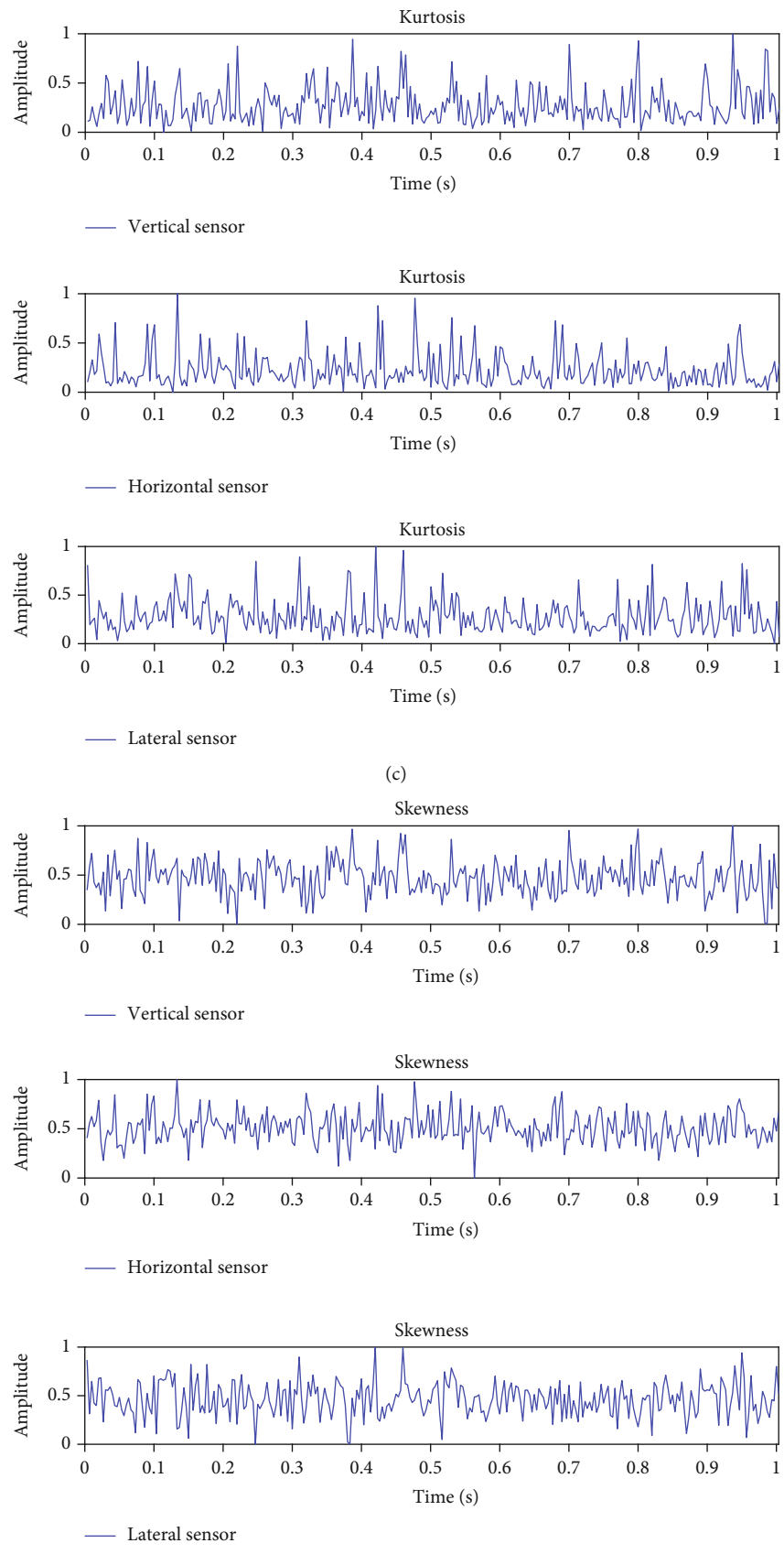
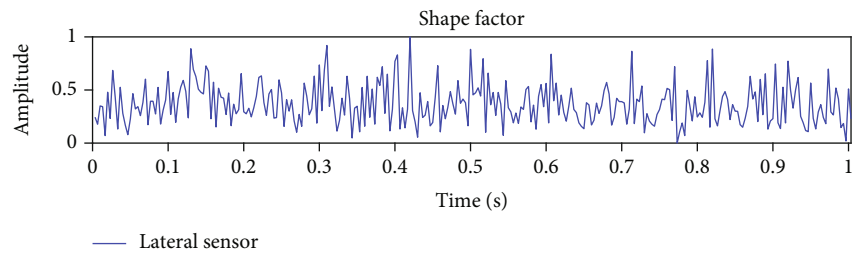
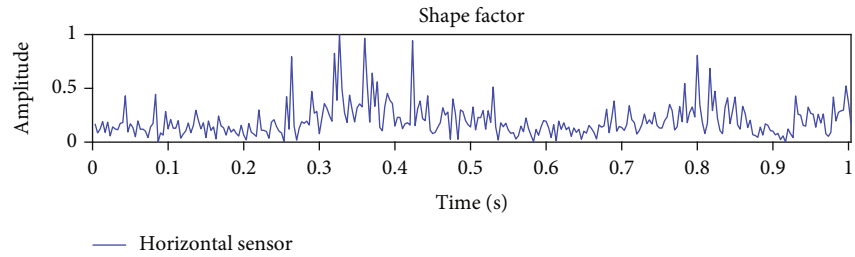
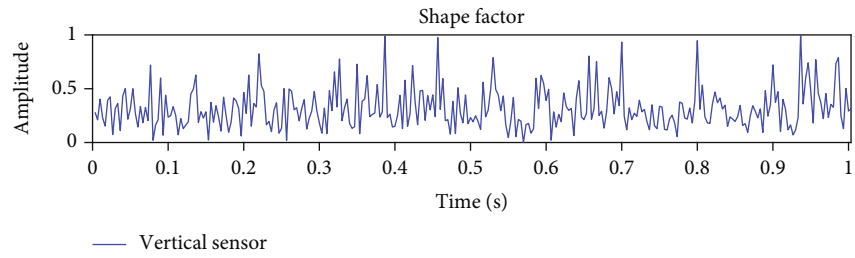
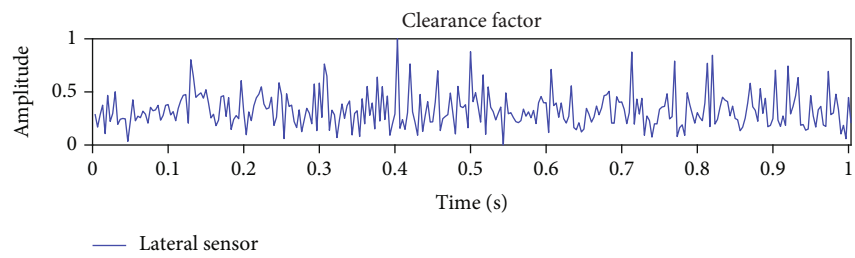
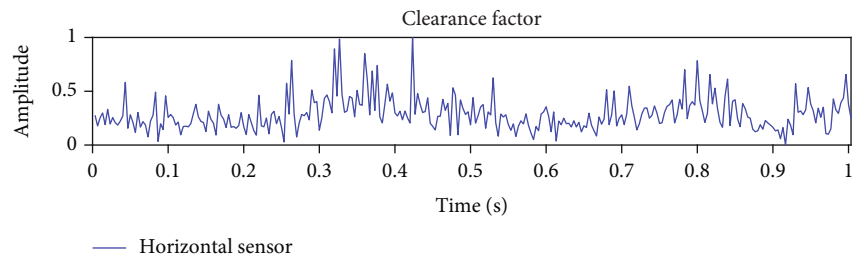
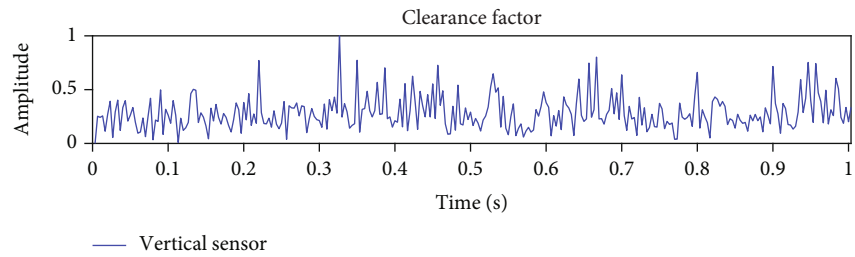


FIGURE 10: Continued.



(e)



(f)

FIGURE 10: Continued.

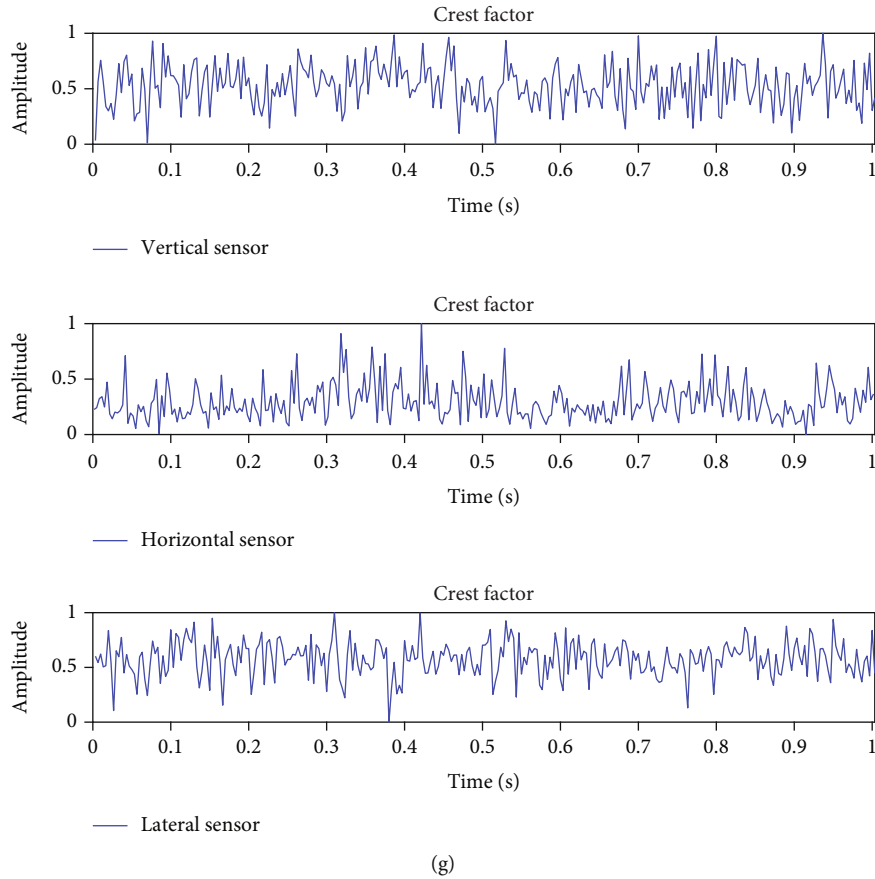


FIGURE 10: Six extracted normalized feature of turbine system of (a) raw signal, (b) impulse factor, (c) kurtosis, (d) skewness, (e) shape factor, (f) clearance factor, and (g) crest factor.

features are normalized by using Equation (14) where $x_{i,j}(t)$ denotes the extracted feature of the j th sensor within the i th signal segmentation. $\max(x_{i,j})$ and $\min(x_{i,j})$ represent the maximum feature value and the minimum feature value within the i th signal segmentation of the j th sensor.

$$\hat{x}_{i,j}(t) = \frac{x_{i,j}(t) - \min(x_{i,j})}{\max(x_{i,j}) - \min(x_{i,j})}. \quad (14)$$

3.1.3. Feature Fusion and Feature Construction Based on SAE. To construct the comprehensive health indicator of the target component, the proposed two-stage SAE-based feature fusion method is executed in this paper. First, six kinds of normalized features extracted from a group of sensors monitoring the same component are sent to six three-layer SAEs, respectively, with the structure as shown in Figure 5. Assuming that there are N sensors monitoring the same target component, the node number of the hidden layer is set as $2N + 1$ so that the layer can be forced to learn an over complete representation of each kind of feature obtained from multisensor signals [34]. Since there are six kinds of features extracted from the sensor signal

which is mentioned in Section 3.1.2, the six comprehensive features are concatenated and sent to the second SAE as shown in Figure 6. The node number of the hidden layer of the second SAE is set to 6 according to the parameter expertise adjustment in literature [34] so that the network is forced to learn a highly compressed representation of the six extracted features. The six-node compressed feature is extracted and regarded as the composite health indicator of the target component.

3.2. Ensemble Learning Based on DBNs

3.2.1. Individual Faulty Classification Based on Single DBN. The extracted six-node compressed features are fed to multiple DBN models which are used as the subclassifier of the ensemble learning network. In this paper, the DBN is constructed by stacking two RBM layers and one softmax layer for faulty classification with the structure as shown in Figure 7, where “ V ” denotes visible vector and “ h ” denotes the hidden vector. The parameter “ W ” denotes the connection weight between the visible layer and the hidden layer.

The input dimension of the proposed DBN is set to six, and the output dimension of the Softmax layer is equivalent to the faulty type of the corresponding faulty

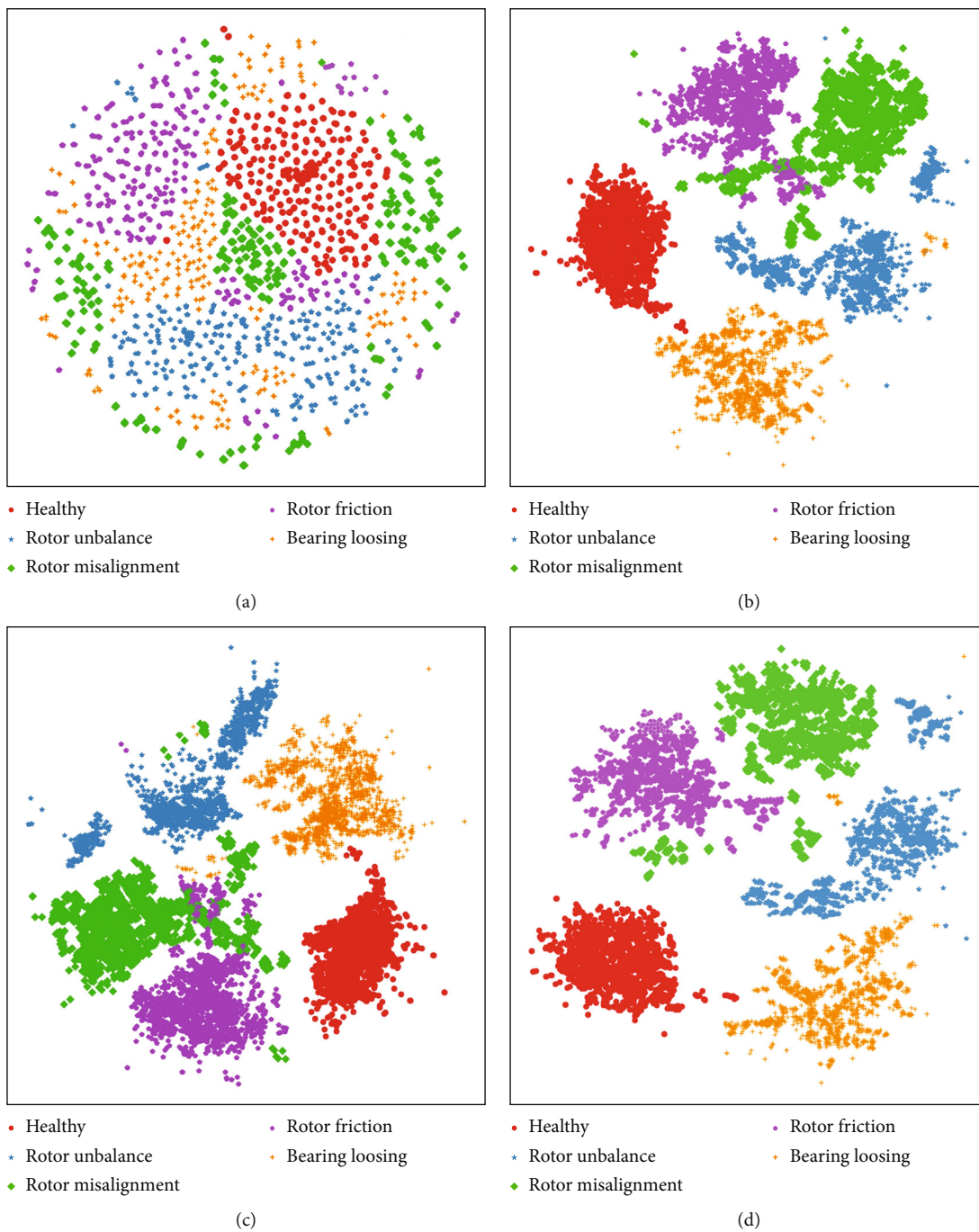


FIGURE 11: Continued.

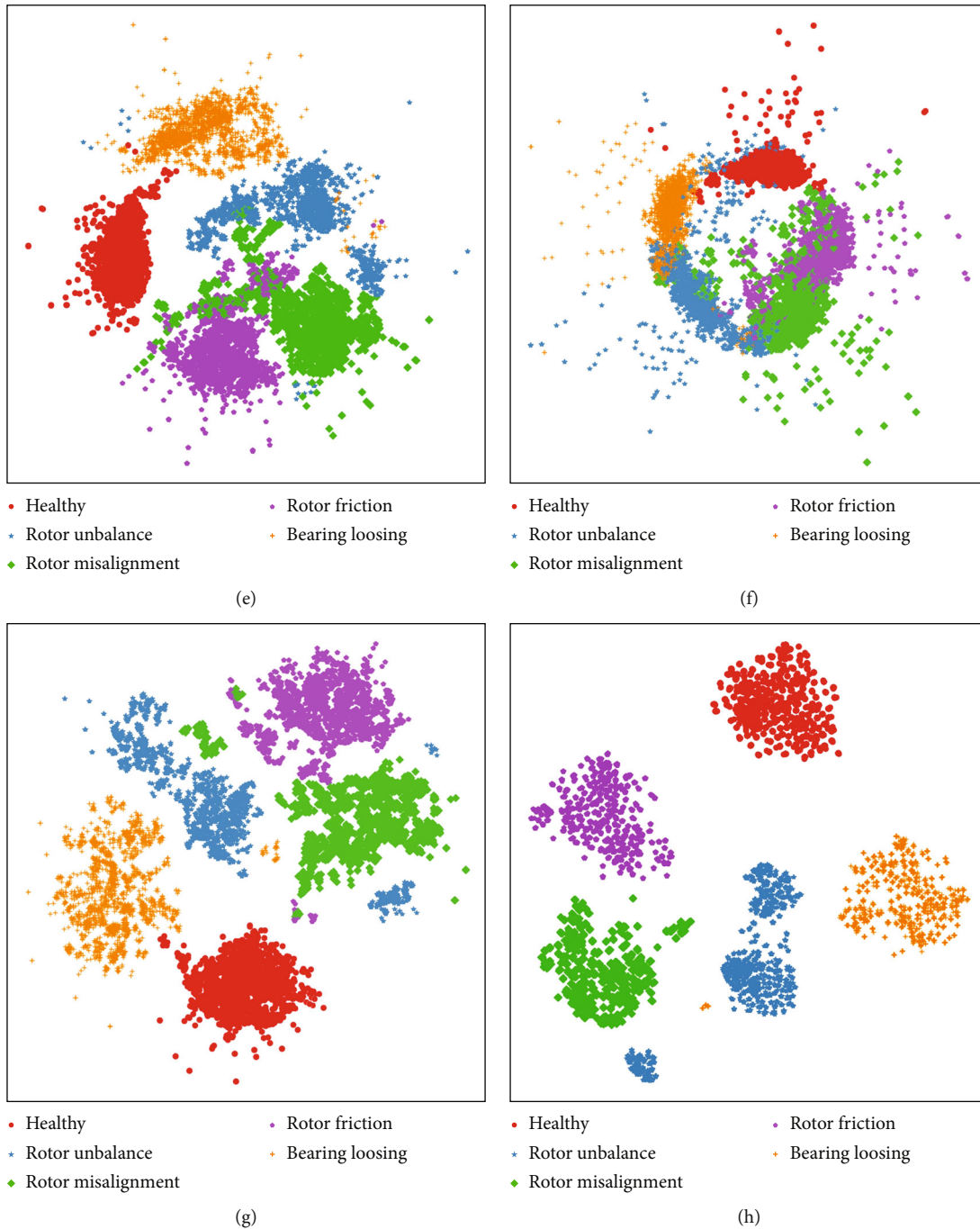


FIGURE 11: Feature visualization by t-SNE. (a) Raw sensor data. (b) Impulse+SAE extraction. (c) Kurtosis+SAE extraction. (d)Skewness+SAE extraction. (e) Shape factor+SAE extraction. (f) Clearance factor+SAE extraction. (g) Crest factor+SAE extraction. (h) Six-node compressed feature.

classification task. The contrastive divergence proposed by Hinton et al. [24] is used for the fast training of the DBN model. Firstly, the conditional probability of the hidden units contained in RBM1 is obtained by using Equation (15). The Gibbs sampling is then employed to determine the state of the hidden units. Finally, the state of the visible units can be obtained through Equation (16), which can be regarded as the reconstruction of the former hidden layer. The param-

eter $w_{i,j}$ denotes the connection weight between the i th visible node and the j th hidden node. The parameters α and β denote the biases of the visible and hidden layer, respectively.

$$p(h_i = 1|v; \theta) = 1 / \left[1 + \exp \left[-\alpha_i - \sum_{j=1}^m V_j \times w_{i,j} \right] \right], \quad (15)$$

TABLE 2: The configuration of the SAE used for feature extraction and feature fusion.

SAE for feature extraction		SAE for feature fusion	
Number of hidden layer	1	Number of hidden layer	1
Node per layer	3-7-3	Node per layer	42-6-42
Sparsity penalty	0.02	Sparsity penalty	0.02
Number of epochs	60	Number of epochs	60
Batch size	100	Batch size	100
Weight decay	0.0001	Weight decay	0.0001

$$p(v_j = 1|h; \theta) = 1 / \left[1 + \exp \left[-\beta_j - \sum_{i=1}^{h \text{ dimension}} h_i \times w_{i,j} \right] \right]. \quad (16)$$

The gradient variation can be obtained through maximizing the log-likelihood by performing stochastic gradient descent. The updating regulations of the parameter θ can be achieved by using the following equations:

$$\Delta W_{ij} = \sigma \times \left(\langle v_j \times h_i \rangle_{p(h|v)} - \langle v_j \times h_i \rangle_{rec} \right), \quad (17)$$

$$\Delta \beta_j = \sigma \times \left(\langle v_j \rangle_{p(h|v)} - \langle v_j \rangle_{rec} \right), \quad (18)$$

$$\Delta \alpha_i = \sigma \times \left(\langle h_i \rangle_{p(h|v)} - \langle h_i \rangle_{rec} \right). \quad (19)$$

Here, σ denotes the learning rate, $\langle * \rangle_{p(h|v)}$ denotes the expectation value of the conditional distribution $p(h|v)$, and $\langle * \rangle_{rec}$ denotes the distribution expectation of the reconstructed model. The stacked RBMs are first pretrained in an unsupervised way. After the DBN pretraining process is finished, the DBN model is fine-tuned starting from the last layer by using the labeled data. The softmax layer of representing the probability of the different faulty types are added for a classification problem, and the entire model parameters are optimized by using the backpropagation algorithm.

3.2.2. Ensemble Learning Based on Bayesian Weighting. Since the individual DBN classifier only represents the machine health status from the component perspective, in order to construct the correlation network between individual components and different machine faulty conditions, the ensemble learning approach based on Bayesian weighting strategy is applied for integrating the outputs result of multiple DBN classifiers. Assuming that there are N individual DBN classifiers corresponding to the health statuses of N individual components of the equipment, the machine health status can be represented as follows.

$$P_{\text{machine}}(S = S_i) = \sum_{j=1}^N \text{Corr}(\text{component}_j) \times P_{\text{component}_j}(S = S_i | \text{DBN}_j). \quad (20)$$

Here, the $\text{Corr}(\text{component}_j)$ denotes the correlation weight between the machine and the target component

TABLE 3: The configuration of individual DBN classifier.

Number of RBM	2
Node per layer	6-10-10-10-5
Learning rate	0.1
Number of epochs	60
Batch size	100
Weight decay	0.0002

when identifying the state S_i , and it is initialized with the weight value of $1/N$, representing that all the components have the same correlation relationship with the machine at the beginning;

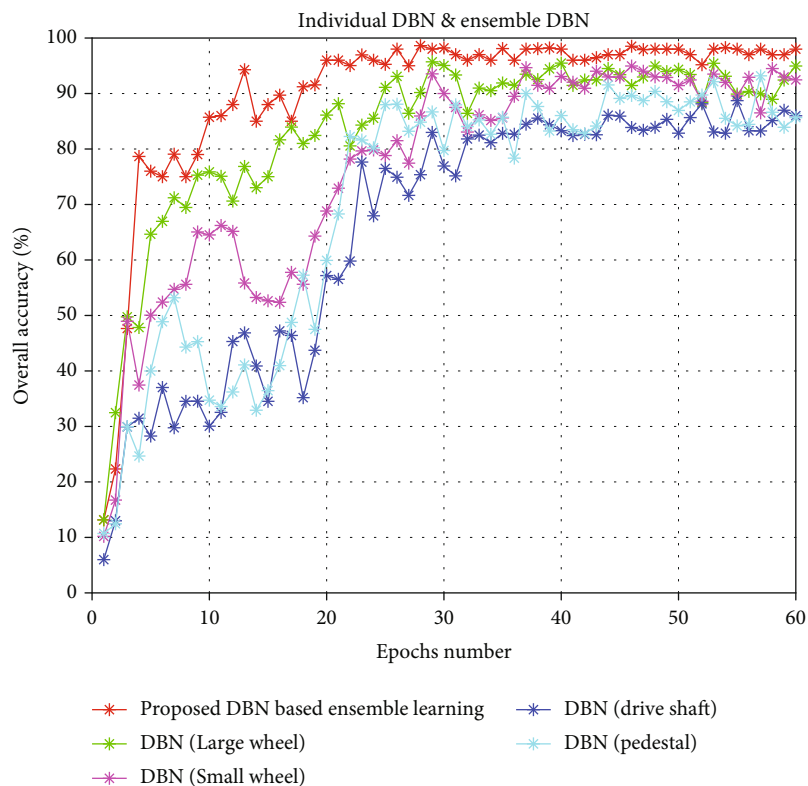
$P_{\text{component}_j}(S = S_i | \text{DBN}_j)$ denotes the output of the softmax probability of the state S_i by the j th DBN which corresponds to the j th component. The specific mathematical expression is illustrated as shown in Equation (21), where the parameter K denotes the number of the faulty types of the classification task and the Z_{S_i} denotes the raw output of the state S_i

$$P_{\text{component}_j}(S = S_i | \text{DBN}_j) = \frac{e^{z_{S_i}}}{\sum_{z=1}^K e^z}. \quad (21)$$

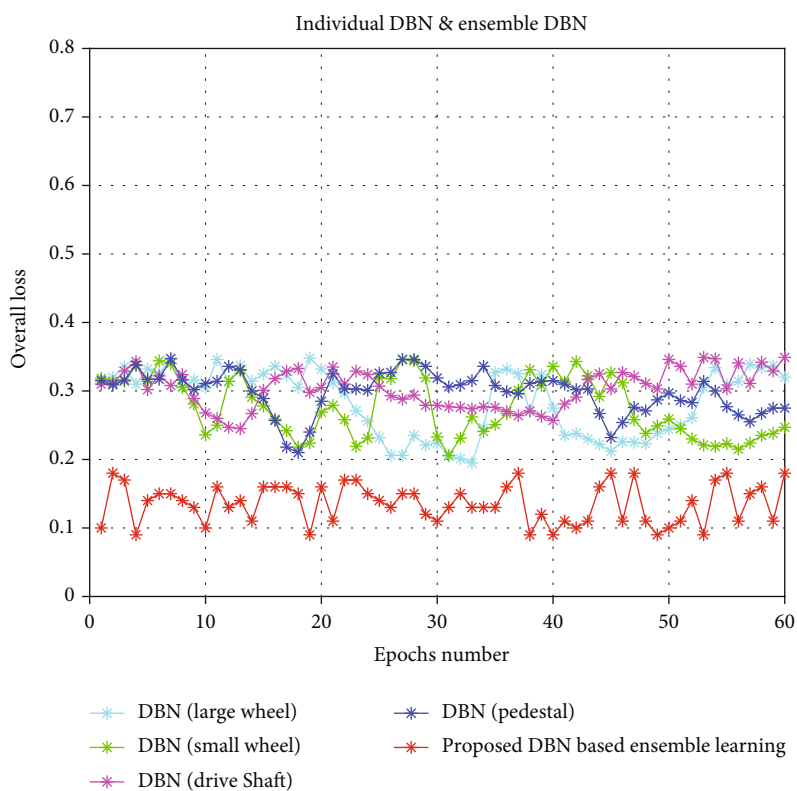
If the true state of the of the machine state is exactly S_i , the postier correlation weight can be updated as shown in Equation (22); otherwise, the weight will not be updated. It should be noted that the updated weight assignment used here is only corresponding to the classification type of the state S_i , and the different machine faulty states have different component-system weight assignment.

$$\text{Corr}(\text{component}_j)_{\text{postier}} = \frac{P_{\text{component}_j}(S = S_i | \text{DBN}_j)}{\sum_{j=1}^N P_{\text{component}_j}(S = S_i | \text{DBN}_j)}. \quad (22)$$

3.3. The Combination of the SAE and the DBN-Based Ensemble Learning. The overall framework of the proposed approach is illustrated in Figure 8. First, the sensors monitoring the same component are categorized, and six classical time-domain features are extracted and normalized from these sensor data. Second, the SAE-based feature fusion is used for obtaining the compressed representation of the six normalized features which is also regarded as the



(a)



(b)

FIGURE 12: The performance of single individual classifier and the proposed DBN-based ensemble learning. (a) Testing accuracy curves. (b) Testing loss curves.

TABLE 4: The average accuracy of each condition of individual classifier and the proposed ensemble learning.

Classifier	Healthy	Rotor unbalance	Rotor misalignment	Rotor friction	Bearing loosing	Total accuracy	Variance
Proposed ensemble learning	97.33%	95.45%	96.42%	96.57%	95.40%	96.78%	0.66
DBN (large wheel)	93.01%	83.29%	84.17%	87.64%	92.87%	90.6%	15.98
DBN (small wheel)	91.23%	86.11%	88.92%	84.09%	92.02%	89.8%	11.29
DBN (drive shaft)	84.24%	82.19%	81.94%	80.35%	87.38%	83.73%	7.32
DBN (pedestal)	87.22%	84.50%	82.99%	85.68%	86.44%	86.29%	2.77

TABLE 5: The comparison of different prediction model.

Input	Classifier	Total accuracy	Variance
Six-node compressed feature	Proposed DBN-based ensemble learning (Bayesian weighting)	96.78%	0.66
	DBN-based ensemble learning (winner takes all)	95.32%	0.87
Six-node compressed feature (concatenated)	Single CNN	91.33%	2.06
	Single DBN	90.87%	2.35
	Single BPNN	91.64%	2.51
	Single SVM	91.03%	1.93

TABLE 6: The comparison of different model input.

Classifier	Input	Total accuracy
Proposed DBN-based ensemble learning (Bayesian weighting)	Six-node compressed feature	96.78%
	Impulse+SAE extraction	84.31%
	Kurtosis+SAE extraction	86.33%
	Skewness+SAE extraction	82.46%
	Shape factor+SAE extraction	83.92%
	Clearance factor+SAE extraction	85.37%
	Crest factor+SAE extraction	86.87%
	Raw sensor data without SAE extraction	55.24%

comprehensive representation of the target component. Third, the extracted compressed features representing the statuses of different components are sent to multiple individual DBN classifiers for the machine faulty identification from the component perspective. Finally, the multiple outputs of DBNs are aggregated through the Bayesian weighting strategy, and the correlation degree between the individual component and the whole machine system is constructed. The correlation weight is constantly updated based on the posterior true label during the training process. During the testing process, the ensemble weighting strategy applied on different DBN classifiers is defined based on the mean value of the correlation weight obtained during the training process. The general procedure of the proposed methodology can be illustrated in Algorithm 1.

4. Methodology Evaluation

4.1. Case Study I: Equipment Faulty Identification for Wind Turbine Gearbox

4.1.1. *Data Description and Experimental Set up.* The performance of the proposed approach is evaluated on the

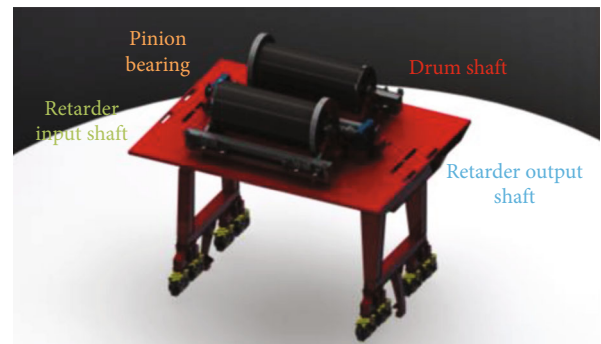


FIGURE 13: The component-system structure of the crane trolley.

TABLE 7: The sensor deployment of crane trolley.

Subsystem	Monitoring sensor number
Pinion bearing	PB-01; PB-02; and PB-03
Drum shaft	DS-01; DS-02
Retarder input shaft	RIS-01; RIS-02
Retarder output shaft	ROS-01; ROS-02; and ROS-03

TABLE 8: Description of datasets of Case Study II.

Datasets	Speed range (rpm)	Sample number in each condition	Condition status
Training/testing	75-100/75-100	750/750	Healthy
		150/150	Wheel biting
		150/150	Wire rope over winding
		150/150	Brake failure

multisensor datasets of the wind turbine gearbox provided by the enterprise of Shanghai electric group. As shown in Figure 9, the wind turbine gearbox mainly consists of four main components namely large wheel, small wheel, drive shaft, and pedestal. Each component is monitored by three vibration sensors from the vertical, horizontal, and lateral directions. The sampling frequency is set as 36 kHz, and the sampling time is set as 1 s under the load speed ranging from 800 rpm to 1600 rpm. The time window is set as 10, where each sample contains 10 sampling points. Therefore, there are 3600 samples in total, in which half are for training and half are for testing. There are five kinds of gearbox conditions namely healthy, rotor unbalance, rotor misalignment, rotor friction, and bearing loosening. The specific details of the datasets of Case Study I are illustrated as shown in Table 1.

4.1.2. Feature Extraction and Normalization. In this paper, six kinds of features including the impulse factor, kurtosis, skewness, shape factor, clearance factor, and crest factor are extracted and normalized from each sample collected by sensors monitoring the same subsystem. Take the component of “driving shaft” as an example; the raw signal and the six kinds of normalized features are extracted from three vibration sensors from vertical, horizontal, and lateral directions, respectively, as shown in Figures 10(a)–10(g). It can be found that in Figure 10 that the amplitude of the raw signals and all the six extracted and normalized features of impulse, kurtosis, skewness, shape factor, clearance factor, and crest factor varies obviously due to the different monitoring locations of sensors which may increase not only the difficulty of feature selection but also the uncertainty in faulty identification. Therefore, it is necessary to find an effective method to merge the information collected from the three vibration sensors to obtain the comprehensive representative status of the target component, and in this, paper the proposed two-stage sparse autoencoder is introduced owing to its powerful feature compression and feature reconstruction ability.

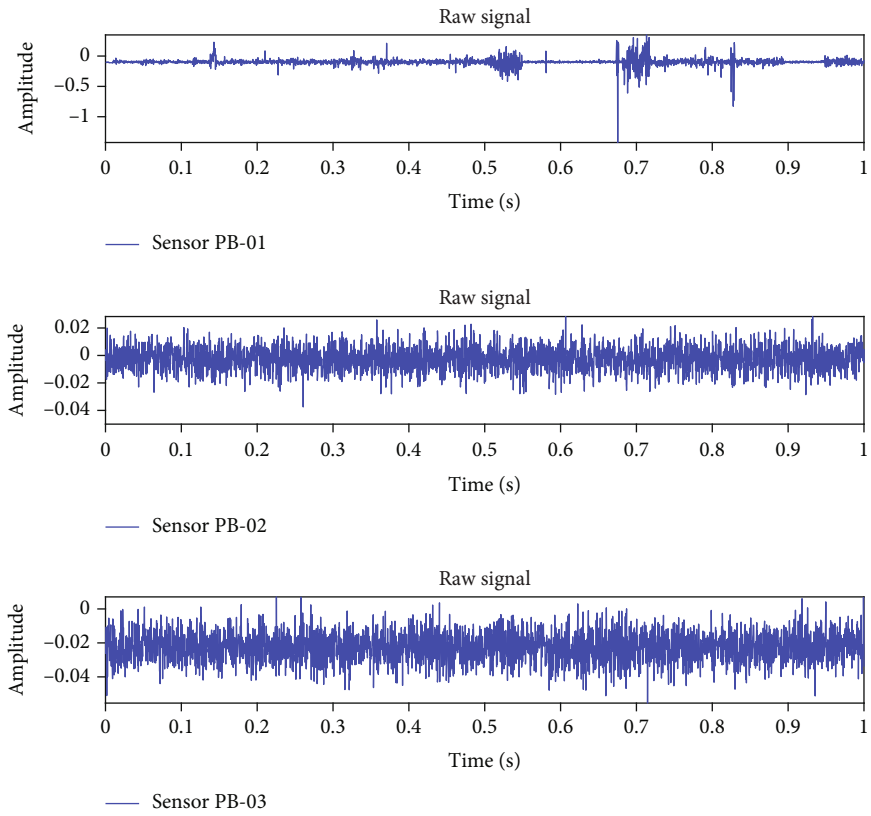
4.1.3. The Feature Fusion. In order to demonstrate the robustness of the fused feature and to visualize the feature representation ability, the raw sensor data without SAE extraction, the six statistical features being further extracted by SAE and the six-node compressed composite feature fused by the proposed two-stage SAE model are represented on a 2-D feature map as shown in Figures 11(a)–11(g) by using the t-SNE (t-distributed stochastic neighbor embedding) technology. As shown in Figure 11(a), the t-SNE fails to separate the five turbine gearbox conditions with raw sen-

sor data of the multisensor. Most of the samples are mixed with each other, and it can only be distinguishable between the healthy and unhealthy state which may greatly influence the accuracy of the faulty identification model. In Figures 11(b)–11(g), most of the samples are correctly classified by the six extracted and normalized features fused by one stage SAE with less samples being mistakenly classified, where only a few marginal samples of rotor unbalance, rotor friction, and rotor misalignment are mistakenly mixed with each other, whereas in Figure 11(h), all the five conditions are perfectly separated clearly compared with the raw sensor data illustrated in Figure 11(a) and the six extracted and normalized features fused by one stage SAE illustrated from Figures 11(b)–11(g). Therefore, the proposed six-node compressed feature can greatly improve the data clustering ability because the compressed feature contains the compressed information of the six features rather than a single feature, which can be more effective when dealing with the nonstationary characteristics of faulty signals.

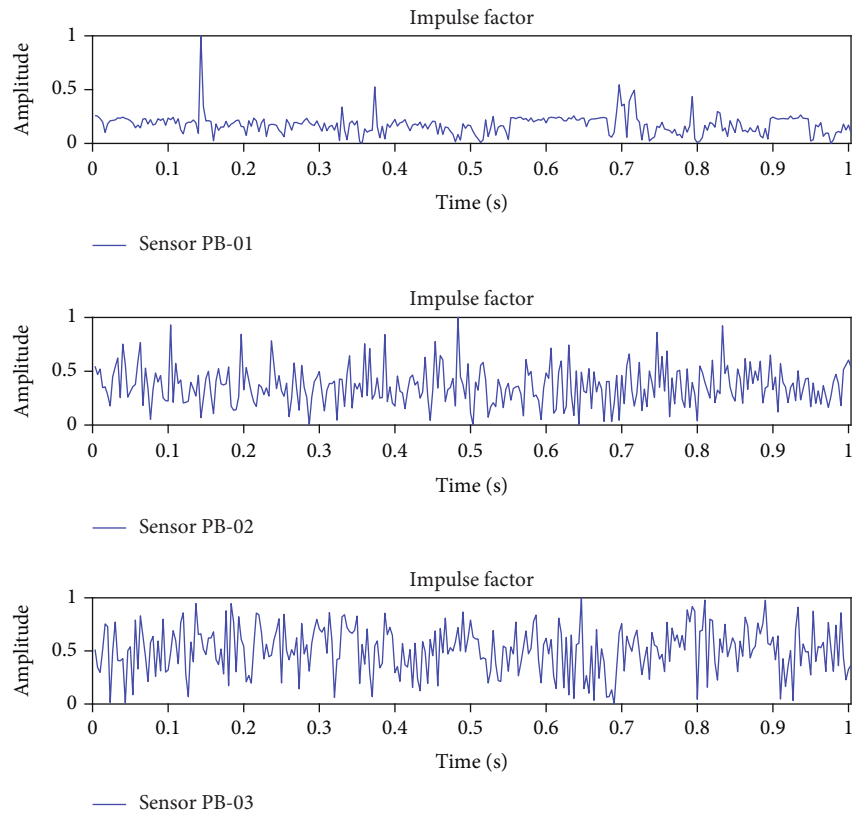
The configuration of the SAEs used for feature extraction and feature fusion is illustrated as shown in Table 2.

4.1.4. DBN-Based Ensemble Learning. The six-node compressed feature indicating the comprehensive representation of the four subsystems are extracted and used as the input of five individual DBN classifiers. The configuration of a single individual DBN classifier is illustrated as shown in Table 3. The DBN used as the individual classifier has five layers in total, where the node number of the input layer node is set as 6, which has the same dimension of the six-node compressed feature. The node number of the output layer is set as 5 according to the type number of the turbine gearbox health conditions. The overall testing accuracy curves and overall loss curves of each individual DBN and the proposed ensemble learning networks are illustrated as shown in Figures 12(a) and 12(b). It can be found that the proposed ensemble learning approach outperforms other five individual DBN classifiers in terms of the overall testing accuracy and the overall testing loss.

The experiment is repeated 10 times, and the average accuracy in predicting each kind of conditional status and the total accuracy is illustrated as shown in Table 4. It can be found that the proposed DBN-based ensemble learning approach achieves the highest average prediction accuracy in terms of predicting each kind of health condition and the total accuracy. The “variance” is used as the metric to evaluate stability of the faulty prediction model. From the variance evaluation of each classifier, it can be found that the performance of the proposed ensemble learning approach remains stable in predicting each kind of

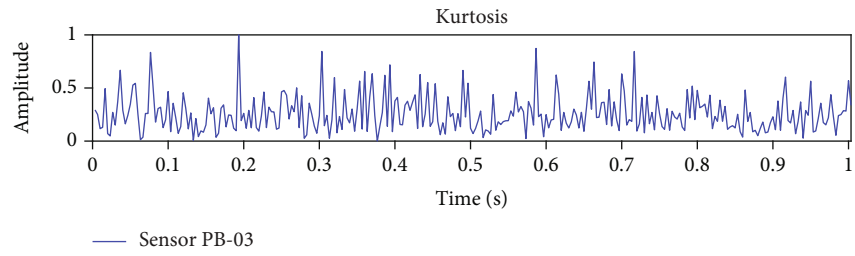
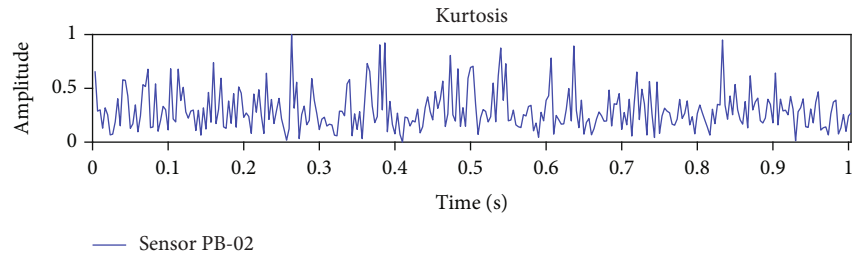
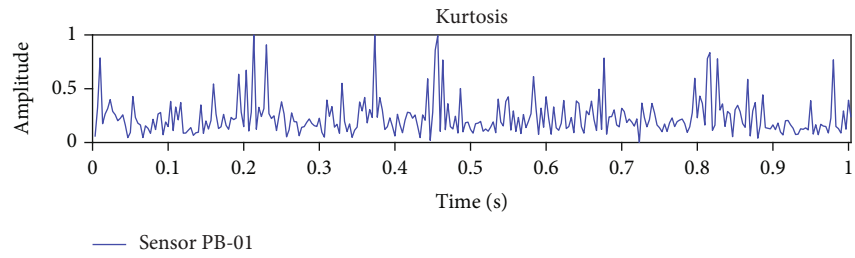


(a)

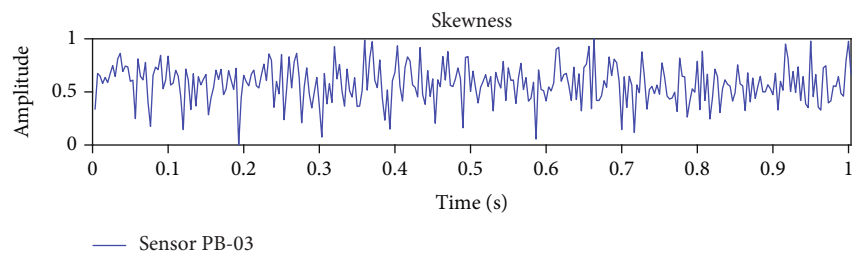
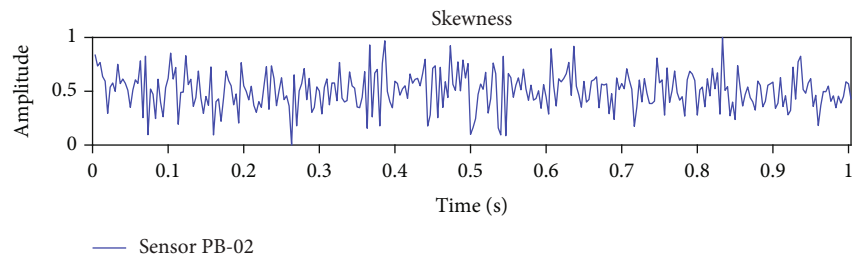
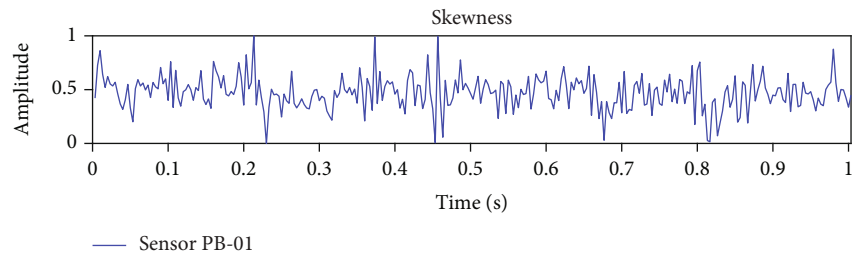


(b)

FIGURE 14: Continued.

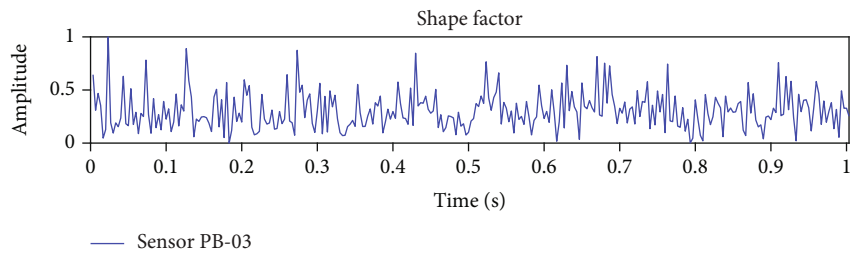
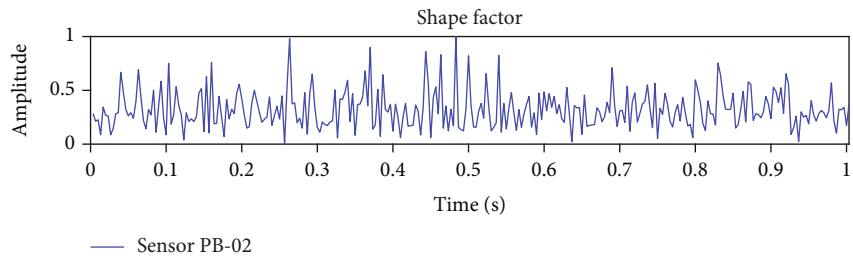
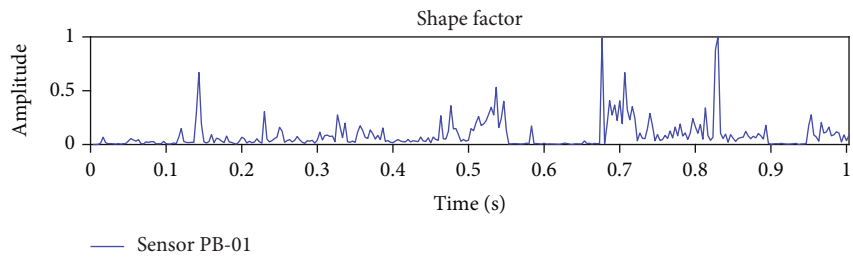


(c)

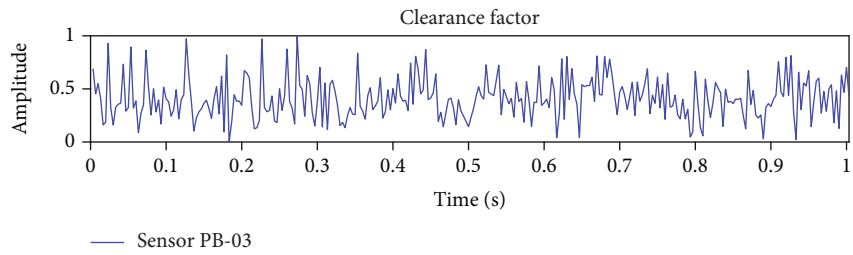
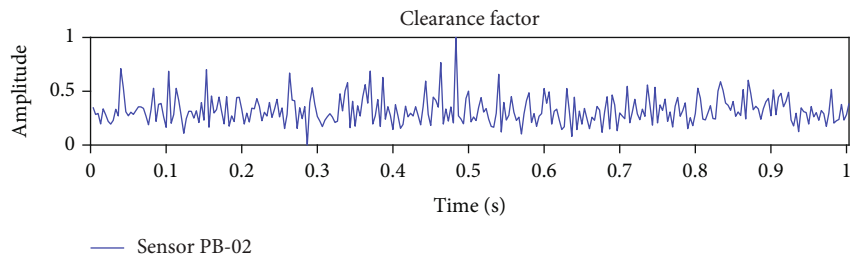
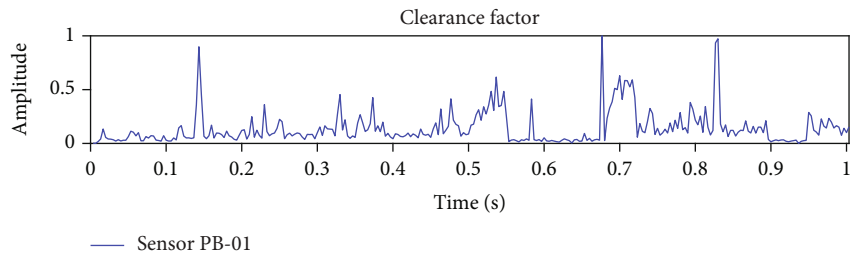


(d)

FIGURE 14: Continued.



(e)



(f)

FIGURE 14: Continued.

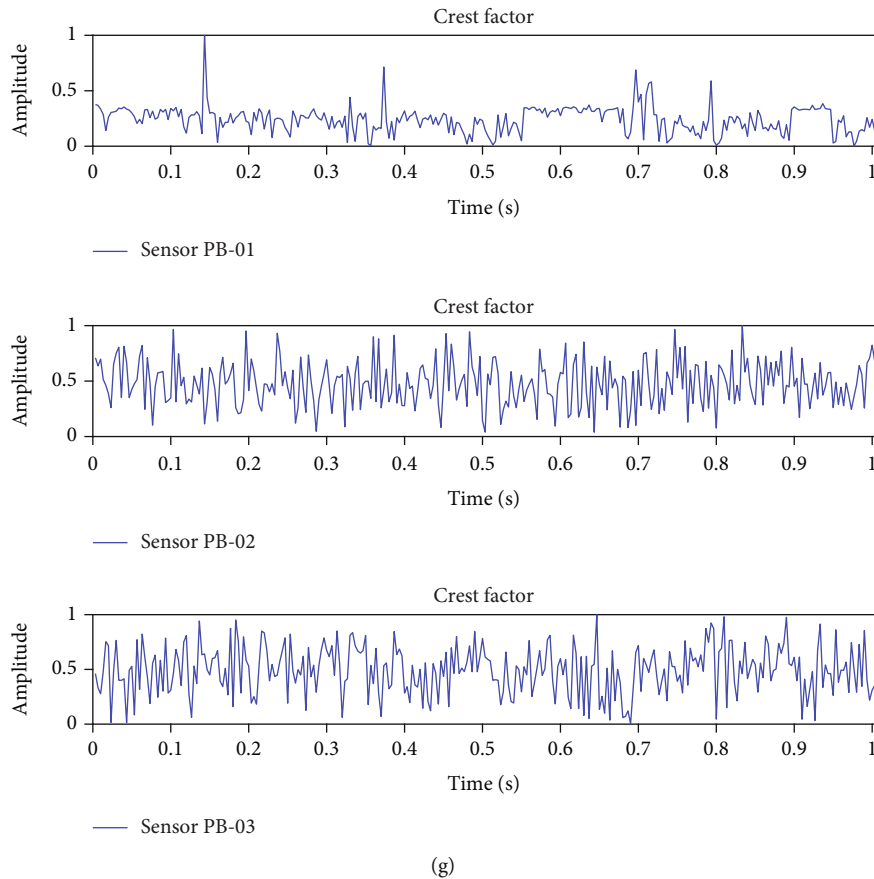


FIGURE 14: Six extracted normalized features of pinion bearing of (a) raw signal, (b) impulse factor, (c) kurtosis, (d) skewness, (e) shape factor, (f) clearance factor, and (g) crest factor.

conditional status while the performance of individual DBN classifier varies heavily in terms of the prediction of certain faulty type. The reason should be that different components have different sensitive degree to different machine faulty conditions, indicating the necessity of constructing a component-system affiliation network.

4.1.5. Comparison with Other Methods. In order to evaluate the effectiveness of the proposed two-stage SAE-based feature fusion and the DBN-based ensemble learning, the comparison experiments are carried out 10 times from the following two aspects, respectively.

(1) Proving the Effectiveness of the Proposed DBN-Based Ensemble Learning Classifier. In order to evaluate the proposed DBN-based ensemble learning classifier, we fix the model input as the six-node composite feature. The classifiers including the proposed DBN-based ensemble learning classifier based on Bayesian weighting, DBN-based ensemble learning classifier on winner takes all, single CNN, single DBN, single BPNN, and single SVM classifiers are used for comparison. It should be noted that when using single classifier, the six-node compressed features of different components are concatenated as a categorical input. The simulation result is illustrated as shown in Table 5. It can be found that the proposed DBN-based ensemble learning approach

based on Bayesian weighting strategy achieves the highest total accuracy of 96.78% while ensemble learning based on “winner takes all” strategy achieves the total accuracy of 95.32%. The total accuracy obtained by the rest four single classifiers are beyond 92%. Therefore, it can be concluded that the ensemble learning classifier can be very effective when dealing with the faulty prediction task of heavy machine with complex component-system structure.

From the stability evaluation of the different classifiers, it can be found that the proposed ensemble learning approach with the Bayesian weighting strategy achieves the lowest variance value of 0.66, indicating the model accuracy stability in predicting each kind of machine conditions. Moreover, it can be found that the ensemble learning approach with the “winner takes all” being used for comparison outperforms the other four single classifiers in terms of the prediction stability with the variance value of 0.87. The reason should be that although the ensemble learning model based on “winner takes all” is totally based on the principal of majority voting which cannot reflect the component-system affiliation relationship of the heavy equipment, it can reflect the component-system structure of the heavy machine which cannot be achieved by a single classifier.

(2) Proving the Effectiveness of the Proposed Two-Stage SAE-Based Feature Fusion. In order to evaluate the proposed two-

TABLE 9: The configuration of the SAE used for feature fusion.

Subsystem	Node per layer	Sparsity penalty	Batch sizes	Number of hidden layer	Weight decay
Pinion bearing	3-7-3	0.02	100	1	0.0001
Drum shaft	2-5-2	0.02	100	1	0.0001
Retarder input shaft	2-5-2	0.02	100	1	0.0001
Retarder output shaft	3-7-3	0.02	100	1	0.0001

TABLE 10: The configuration of the SAE used for feature compression.

Subsystem	Node per layer	Sparsity penalty	Batch sizes	Number of hidden layer	Weight decay
Pinion bearing	42-6-42	0.02	100	1	0.0001
Drum shaft	30-6-30	0.02	100	1	0.0001
Retarder input shaft	30-6-30	0.02	100	1	0.0001
Retarder output shaft	42-6-42	0.02	100	1	0.0001

stage SAE-based feature fusion, we fix the classifier as the proposed DBN-based ensemble learning classifier based on the Bayesian weighting strategy. The raw sensor data, the six statistical features with SAE extraction, and the six-node compressed features are used as the input for comparison. The simulation result is illustrated as shown in Table 6. It can be found that the proposed ensemble learning approach can achieve the highest total accuracy of 96.78% with the model input of the six-node compressed feature. The six statistical features with SAE extraction can achieve the accuracy ranging from 82% to 87% which can be inferior to the six-node compressed feature. The raw sensor data achieve the lowest total accuracy of 55.24%. The conclusion is consistent with the visualization effect of t-SNE illustrated in Figure 11.

4.2. Case Study II: Faulty Prognostic for Industrial Port Crane Trolley

4.2.1. Data Description and Experimental Set up. In Case Study II, the performance of the proposed approach is evaluated on the multisensor datasets of the crane trolley provided by the enterprise of the CSIC (China Shipping Industry Company). As shown in Figure 13, the crane trolley mainly consists of four subsystems namely the pinion bearing, drum shaft, retarder input shaft, and retarder output shaft. There are 10 vibration sensors with the sensor series number from PB-01 to ROS-03 installed inside the crane trolley, and the sensor deployment is illustrated as shown in Table 7. The sampling frequency in Case Study II is set to 24 KHz, and the sampling time is set as 1 s. The time window is set to 10, and there are 2400 samples in total containing 10 sampling points in each.

The specific details of the datasets of Case Study II are illustrated as shown in Table 8. There are four kinds of crane trolley conditions namely wheel biting, wire rope over winding, brake failure, and healthy.

4.2.2. Feature Extraction and Normalization. Same as the Case Study I, six kinds of features are extracted and normalized from sensors monitoring the same subsystem.

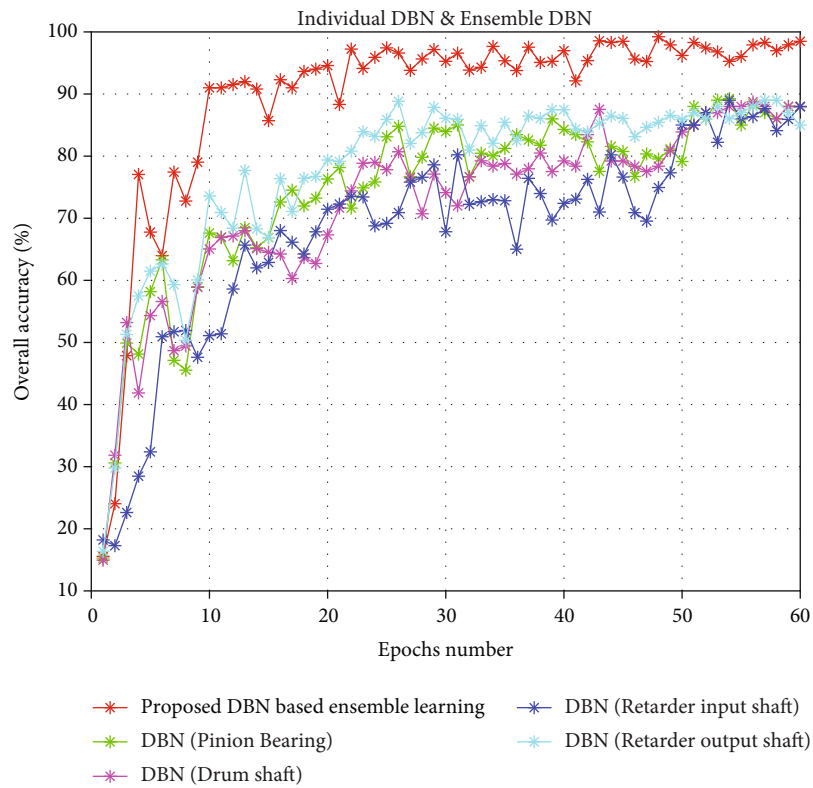
TABLE 11: The configuration of individual DBN classifier used in Case Study II.

Number of RBM	2
Node per layer	6-10-10-10-4
Learning rate	0.1
Number of epochs	60
Batch size	100
Weight decay	0.0002

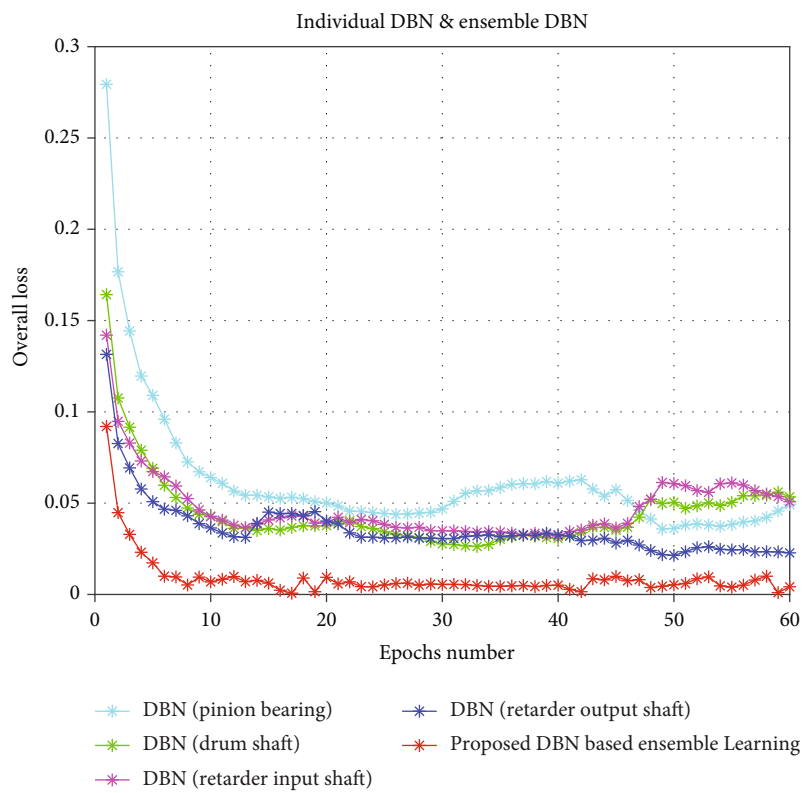
Take the pinion bearing as an example, six kinds of normalized features are extracted from the “sensor PB-01,” “sensor PB-02,” and “sensor PB-03,” respectively, as shown in Figures 14(a)–14(g).

4.2.3. The Feature Fusion and Feature Construction. Same as the Case Study I, the six extracted features from the sensor signal are fused by SAE to obtain the compressed feature representing the comprehensive health status of the target component. The configuration of the SAE of the four subsystems used for feature extraction and feature fusion is illustrated as shown in Tables 9 and 10.

4.2.4. DBN-Based Ensemble Learning. Same as the Case Study I, the six-node compressed feature indicating the comprehensive representation of the four subsystems are extracted and used as the input of four individual DBN classifiers. The configuration of a single individual DBN classifier used in Case Study II is illustrated as shown in Table 11. Since there are four kinds of crane trolley condition, the unit number of the output layer is set as 4. The overall accuracy curves and overall loss curves of each individual DBN and the proposed ensemble learning networks are illustrated as shown in Figures 15(a) and 15(b). It can be also found that in Case Study II, the proposed ensemble learning approach outperforms other four individual DBN classifiers in terms of the overall testing accuracy and the overall testing loss.



(a)



(b)

FIGURE 15: The performance of single individual classifier and the proposed DBN-based ensemble learning. (a) Testing accuracy curves. (b) Testing loss curves.

TABLE 12: The average accuracy of each condition of individual classifier and the proposed ensemble learning in Case Study II.

Classifier	Healthy	Wheel biting	Wire rope over winding	Brake failure	Total accuracy	Variance
Proposed ensemble learning	96.88%	96.54%	96.87%	98.35%	97.02%	0.65
DBN (pinion bearing)	85.53%	81.77%	88.59%	74.15%	84.02%	38.84
DBN (drum shaft)	82.87%	88.96%	85.58%	79.17%	83.51%	17.20
DBN(retarder input shaft)	82.11%	73.01%	77.84%	85.23%	80.83%	28.17
DBN (retarder output shaft)	84.85%	91.40%	84.79%	89.20%	86.21%	10.81

TABLE 13: The comparison of different prediction models.

Input	Classifier	Total accuracy	Variance
Six-node compressed feature	Proposed DBN-based ensemble learning (Bayesian weighting)	97.02%	0.65
	DBN-based ensemble learning (winner takes all)	93.06%	1.35
Six-node compressed feature (concatenated)	Single CNN	88.76%	8.74
	Single DBN	86.31%	7.81
	Single BPNN	86.97%	5.93
	Single SVM	87.79%	8.85

TABLE 14: The comparison of different model input.

Classifier	Input	Total accuracy
Proposed DBN-based ensemble learning (Bayesian weighting)	Six-node compressed feature	97.02%
	Impulse+SAE extraction	83.51%
	Kurtosis+SAE extraction	82.20%
	Skewness+SAE extraction	85.67%
	Shape factor+SAE extraction	83.91%
	Clearance factor+SAE extraction	81.49%
	Crest factor+SAE extraction	85.43%
	Raw sensor data without SAE extraction	62.33%

The experiment is repeated 10 times, and the average accuracy in predicting each kind of conditional status and the total accuracy is illustrated as shown in Table 12. It can be found that the proposed ensemble learning approach outperforms the other four individual DBN classifiers in terms of the average prediction accuracy in each condition. Moreover, the proposed ensemble learning approach outperforms the other four DBN classifiers in terms of the stability in predicting the four different kinds of crane trolley conditions based on the variance evaluation in Case Study II.

4.2.5. Comparison with Other Methods. Same as the Case Study I, the comparison experiments are carried out 10 times and evaluated from the input aspect and prediction aspect, respectively.

(1) Proving the Effectiveness of the Proposed DBN-Based Ensemble Learning Classifier. In comparison experiment of prediction aspect of Case Study II, we fix the model input as the six-node composite feature as shown in Table 13. It can be also found that the proposed DBN-based ensemble learning approach based on Bayesian weighting outperforms all the other four single classifiers and the ensemble learning

approach based on (winner takes all) with the highest total accuracy of 97.02%.

Moreover, it can be found that both ensemble learning models outperforms the other four single models in terms of the stability evaluated with lower variance which is consistent with the conclusion of Case Study I.

(2) Proving the Effectiveness of the Proposed Two-Stage SAE-Based Feature Fusion. In comparison experiment of the input aspect of Case Study II, we fix the classifier as the proposed DBN-based ensemble learning classifier based on the Bayesian weighting strategy as shown in Table 14. It can be also found in Case Study II that the proposed ensemble learning approach can achieve the highest total accuracy of 97.02% with the model input of the six-node compressed feature. The six fused features of impulse, kurtosis, skewness, shape factor, clearance factor, and the crest factor can achieve the inferior average accuracy ranging from 81% to 86%. The raw sensor data achieve the lowest total accuracy of 62.33%.

The comparison result of Case Study II is consistent with the result of Case Study I, indicating the effectiveness of the proposed faulty identification approach also being applicable in the practical use of industrial crane trolley of Case Study II.

5. Conclusion and Future Work

5.1. Main Contribution of the Proposed Paper. In this paper, a novel hybrid faulty prognostic approach based on sparse autoencoder- and deep belief network-based ensemble learning is proposed. The main contribution is summarized as follows:

- (1) Introducing the SAE-based feature extraction method so that the features extracted from different sensors are merged into one stream, releasing the influence caused by different sensor deployment
- (2) Proposing the SAE-based feature fusion method so that the six kinds of extracted features are compressed to construct a composite feature which is regarded as the health indicator of a target component. The constructed composite feature has better robust ability when dealing with the nonstationary faulty vibration signal
- (3) Proposing the DBN-based ensemble learning so that the machine internal component-system relationship can be well represented. The outputs of these DBN classifiers are aggregated by using the Bayesian weighting strategy which represent the affiliated degree between the health statuses of the target component and faulty probability of the certain type of machine faulty status

The proposed hybrid faulty prognostic approach is evaluated on two case studies of wind turbine gearbox and the industrial port crane trolley. Both case studies demonstrate that the proposed hybrid faulty identification approach outperforms other traditional faulty identification methods in terms of the prediction accuracy and prediction stability when dealing with the industrial heavy machine, indicating the necessity of applying the two-stage SAE- and the DBN-based ensemble learning.

5.2. Future Work. Although the proposed hybrid faulty identification method has been well evaluated in the proposed two case studies, there are still some issues needed to be considered. Firstly, the computation complexity of the proposed Bayesian weighting strategy used in the ensemble learning process should be taken into account. In the future, some more suitable weighting strategies should be designed which can not only release the computation burden but also reflect the component-system relationship. Moreover, the proposed hybrid faulty identification approach should be expected to be evaluated in other heavy machines such as vehicle system and aircraft engine system which also have a highly complex internal component-system structure.

Data Availability

The raw/processed sensor data of the two case studies are offered by the enterprises of “Shanghai Electric Group” and the “China Shipping Industrial Company,” respectively. It cannot be shared due to the corporate confidentiality.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The research has been financially supported by the Science and Technology Innovation 2030—the Significant Project of a New Generation of Artificial Intelligence (Grant No. 2018AAA0101800)—and the Natural Science Foundation of Shanghai (Grant No. 19ZR1461500).

References

- [1] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, “Residual life predictions from vibration-based degradation signals: a neural network approach,” *IEEE Transactions on Industrial Electronics*, vol. 51, no. 3, pp. 694–700, 2004.
- [2] A. Malhi, R. Yan, and R. X. Gao, “Prognosis of defect propagation based on recurrent neural networks,” *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 3, pp. 703–711, 2011.
- [3] B. Kilundu, X. Chimentin, and P. Dehombreux, “Singular spectrum analysis for bearing defect detection,” *Journal of Vibration and Acoustics*, vol. 133, no. 5, article 051007, 2011.
- [4] P. K. Varshney, “Multisensor data fusion,” *Electronics & Communication Engineering Journal*, vol. 9, no. 6, pp. 245–253, 1997.
- [5] T. Wang and H. Sun, “Design method of data acquisition in intelligent sensor based on web data mining clustering technology,” in *International Conference on Education*, Shenyang, China, 2015.
- [6] C. Axenie and J. Conradt, “Cortically inspired sensor fusion network for mobile robot egomotion estimation,” *Robotics and Autonomous Systems*, vol. 71, pp. 69–82, 2015.
- [7] K. Peng, R. Jiao, J. Dong, and Y. Pi, “A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter,” *Neurocomputing*, vol. 361, no. Oct.7, pp. 19–28, 2019.
- [8] J. Wu, Y. Su, Y. Cheng, X. Shao, C. Deng, and C. Liu, “Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system,” *Applied Soft Computing*, vol. 68, pp. 13–23, 2018.
- [9] J. Gokulachandran and K. Mohandas, “Comparative study of two soft computing techniques for the prediction of remaining useful life of cutting tools,” *Journal of Intelligent Manufacturing*, vol. 26, no. 2, pp. 1–14, 2013.
- [10] H. Yan, K. Liu, X. Zhang, and J. Shi, “Multiple sensor data fusion for degradation modeling and prognostics under multiple operational conditions,” *IEEE Transactions on Reliability*, vol. 65, no. 3, pp. 1416–1426, 2016.
- [11] T. P. Banerjee and S. Das, “Multi-sensor data fusion using support vector machine for motor fault detection,” *Information Sciences*, vol. 217, no. Complete, pp. 96–107, 2012.
- [12] L. Xu, C. Tan, X. Li, Y. Cheng, and X. Li, “Fuel-type identification using joint probability density arbiter and soft-computing techniques,” *IEEE Transactions on Instrumentation & Measurement*, vol. 61, no. 2, pp. 286–296, 2012.
- [13] J. Wang, J. Xie, R. Zhao, L. Zhang, and L. Duan, “Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing,” *Robotics and Computer-Integrated Manufacturing*, vol. 45, pp. 47–58, 2017.

- [14] B. Y. Sun, X. M. Zhang, J. Li, and X. M. Mao, "Feature fusion using locally linear embedding for classification," *IEEE Transactions on Neural Networks*, vol. 21, no. 1, pp. 163–168, 2010.
- [15] J. Ni, C. Zhang, and S. X. Yang, "An adaptive approach based on KPCA and SVM for real-time fault diagnosis of HVCBs," *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1960–1971, 2011.
- [16] N. Saravanan, V. N. S. K. Siddabattuni, and K. I. Ramachandran, "Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM)," *Applied Soft Computing*, vol. 10, no. 1, pp. 344–360, 2010.
- [17] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 6, pp. 1517–1525, 2004.
- [18] M. Aminian and F. Aminian, "Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, no. 2, pp. 151–156, 2000.
- [19] C. Li, R. V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mechanical Systems & Signal Processing*, vol. 76–77, pp. 283–293, 2016.
- [20] T. Junbo, L. Weining, A. Juneng, and W. Xueqian, "Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder," in *The 27th Chinese Control and Decision Conference (2015 CCDC)*, Qingdao, China, 2015.
- [21] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 1926–1935, 2017.
- [22] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2727–2736, 2018.
- [23] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2017.
- [24] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [25] V. T. Tran, F. Althobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4113–4122, 2014.
- [26] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Reliability Engineering & System Safety*, vol. 115, no. jul., pp. 124–135, 2013.
- [27] Z. Chen, X. Zeng, W. Li, and G. Liao, "Machine fault classification using deep belief network," in *2016 IEEE international instrumentation and measurement technology conference proceedings*, Taipei, Taiwan, 2016.
- [28] G. Brown, J. Wyatt, R. Harris, and Y. Xin, "Diversity creation methods: a survey and categorisation," *Information Fusion*, vol. 6, no. 1, pp. 5–20, 2005.
- [29] F. Fernandez-Navarro, P. A. Gutierrez, C. Hervás-Martínez, and X. Yao, "Negative correlation ensemble learning for ordinal regression," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 24, no. 11, pp. 1836–1849, 2013.
- [30] J. Yang, X. Zeng, S. Zhong, and S. Wu, "Effective neural network ensemble approach for improving generalization performance," *Transactions on Neural Networks & Learning Systems*, vol. 24, no. 6, pp. 878–887, 2013.
- [31] R. Coop, A. Mishtal, and I. Arel, "Ensemble learning in fixed expansion layer networks for mitigating catastrophic forgetting," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 24, no. 10, pp. 1623–1634, 2013.
- [32] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multi-objective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2306–2318, 2017.
- [33] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: a review and future trends," *Information Fusion*, vol. 52, pp. 1–12, 2019.
- [34] A. Ng, "Sparse auto-encoder," *CS294A Lecture notes*, vol. 72, pp. 1–19, 2011.