Hindawi
Shock and Vibration
Volume 2022, Article ID 5790185, 9 pages
https://doi.org/10.1155/2022/5790185

Research Article

A Stochastic Learning Algorithm for Machine Fault Diagnosis

Zhipeng Dong,[1] Yucheng Liu,[2,3] Jianshe Kang,[1] and Shaohui Zhang[2]

[1]*Army Engineering University of PLA, Shijiazhuang, China*
[2]*School of Mechanical Engineering, Dongguan University of Technology, Dongguan 523808, China*
[3]*College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China*

Correspondence should be addressed to Shaohui Zhang; zhangsh@dgut.edu.cn

Received 18 October 2021; Accepted 7 January 2022; Published 18 February 2022

Academic Editor: Changqing Shen

Industrial big data bring a large number of high-dimensional sample datasets. Although a deep learning network can well mine the internal nonlinear structure of the dataset, the construction of the deep learning model requires a lot of computing time and hardware facilities. At the same time, there are some nonlinear problems such as noise and fluctuation in industrial data, which make the deep architecture extremely complex and the recognition accuracy of the diagnosis model difficult to guarantee. To solve this problem, a new method, named stochastic learning algorithm (SL), is proposed in this paper for dimension reduction. The proposed method consists of three steps: firstly, to increase the computational efficiency of the model, the dimension of the high-dimensional data is reduced by establishing a random matrix; secondly, for enhancing the clustering influence of the sample, the input data are enhanced by feature processing; thirdly, to make the clustering effects more pronounced, the noise and interference of the data need to be processed, and the singularity value denoising method is used to denoise training data and test data. To further prove the superiority of the SL method, we conducted two sets of experiments on the wind turbine gearbox and the benchmark dataset. It can be seen from the experimental results that the SL method not only improves the classification accuracy but also reduces the computational burden.

1. Introduction

Deep neural networks have extensive applications in artificial intelligence mainly including computer vision [1–8], speech recognition [9–13], medical detection [14–19], and mechanical fault diagnosis [20–28]. Compared with human ability, the DNN model is more capable of solving this complicated problem. But it also has certain challenges. For example, to complete different tasks effectively, different DNN models need to be trained, tuning the parameters through repeating the trial and error, which optimizes the model structure [29]. Therefore, training a DNN model to effectively process specific tasks takes up days or even weeks of the entire computing cluster time [30]. In addition, the parameter optimization of the DNN model not only requires high-performance GPU, TPU, and other higher computer hardware environments but also has high requirements for datasets. Therefore, applications that require high real-time performance or data samples that lack markup are not suitable [31].

In addition to deep learning methods, shallow learning algorithms (PCA, KNN, LPP, etc.) are still largely applied in the artificial intelligence area [32–37]. Although this kind of shallow learning algorithm has the advantages of simple structure, low hardware environment configuration requirements, and relatively high computational efficiency, it also has limitations that are difficult to overcome, such as classifying data containing a large number of variables and a simplified sample set, and the problem of nonlinear nature [38]. In contrast, to overcome the shortcomings of the above algorithm, random forest (RF) was proposed and validated [39, 40]. In addition, meanwhile, RF also has many other advantages [41]. For example, it is easy to understand and simple to implement, tests fast, is highly able to handle outliers and nonlinearity, and shows good performance in parallel training and big data. As a result, it is widely used in medicine, computer vision, machine learning, and other fields, which achieved great results [42–50]. However, the number of decision trees is an important parameter of RF,

which will affect the RF's classification accuracy and computational efficiency. In this paper, a better machine learning algorithm-random learning (SL) is put forward to solve this problem. SL method uses a random mapping matrix to randomly decrease the high-dimensional data's dimensionality, improving the data after the dimensionality reduction. Moreover, the feature is denoising based on SVD to improve the classification rate. Therefore, this method has outstanding advantages: (1) the sample dimension is greatly reduced after processing; (2) the calculation efficiency of the random forest is improved; and (3) the recognition effect of random learning is ensured by the reinforcement process.

In this paper, the other sections are arranged as follows: Section 2 introduces the proposed method's theoretical research; Section 3 introduces the experimental setup and analyzes the results and the experiments; and Section 4 gives the conclusion.

2. The Proposed Method

This section describes the detailed information of the proposed stochastic learning method, and the structure of the proposed model is shown in Figure 1. The strategy of stochastic learning is introduced in Section 2.1. The basics of stochastic learning used for classification are explained in detail in Section 2.2. The implementation of the stochastic learning strategy is presented in Section 2.3. Section 2.4 introduces the proposed algorithm's specific processing process.

2.1. Strategies for the Present Stochastic Learning Method. Sample size and dimensionality would affect the machine learning method's computational efficiency directly. Inspired by the extreme learning machine from the operating mechanism, the input data is randomly reduced in dimensionality to obtain low-dimensional sample data. Unlike the extreme learning machine, the samples are not simply classified, but enhanced in two steps: firstly the sample characteristics were strengthened to obtain a good sample clustering effect; secondly, the feature denoising method is used to improve the cluster. Based on these steps, the recognition accuracy as well as the calculation efficiency of the model can be improved.

2.2. Stochastic Learning Strategy

(1) The research in this part is reference [51]. For an input dataset $\mathbf{X} = [x_1, x_2, \ldots, x_m] \in \mathfrak{R}^{n\times m}$, and if the random feature extraction layer has d nodes, then $\mathbf{H} = [h_1, h_2, \ldots, h_d] \in \mathfrak{R}^{n\times d}$ is the corresponding layer's features. The randomly extracted features are expressed by the formula

$$\mathbf{H} = (\mathbf{X} + \delta)\mathbf{W}, \tag{1}$$

where δ is the biases and W represents the input weights. $\mathbf{H} = [h_1, h_2, \ldots, h_d]$ and $\boldsymbol{W}$ are produced randomly before the start of training, and they remain fixed during the training state without any iterations.

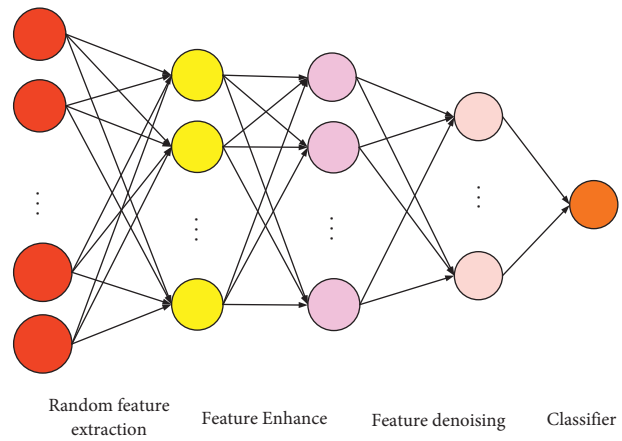

Figure 1: The proposed SL method.

(2) The research in this part is reference [51]. Since the input data seems to be nonlinear, in order to improve classification, an aggregate method of the activation function is introduced. For the random features $\mathbf{H} = [h_1, h_2, \ldots, h_d]$, through the activation function, the dataset can be transformed into

$$\begin{aligned} Q(\mathrm{H}) &= \frac{1}{1 + e^{-\mathbf{H}}}, \\ g &= Q(\mathbf{H})\mathrm{W}_{im}, \end{aligned} \tag{2}$$

where $\mathbf{W}_{im}$ represents a weight vector connecting the improved classification operation and the random feature output, g represents the better classification layer's output, and $\boldsymbol{Q}$ is an activation function. Assuming that the training samples contain c categories, q_{i} represents the *i-th* category's sample amount ($i = 1, 2, \ldots, \mathrm{c}$), and p_{i} represents the *i-th* category's centrality, then

$$\begin{aligned} p_i &= \frac{1}{q_i}\sum_{j=1}^{N} d_j \begin{cases} d_j = h(j)x(j) \in i-th \text{ category}, \\ d_j = 0 \text{ otherwise}, \end{cases} \\ t(k) &= \sum_{i=1}^{c} z_i \begin{cases} z_i = p_i t(k) \in i-th \text{ category}, \\ z_i = 0 \text{ otherwise}, \end{cases} \end{aligned} \tag{3}$$

where the center matrix is represented by t, $t \in \mathrm{R}^{n\times d}$. In order to get the connection weight vector $\mathbf{W_{im}}$, the improved classification's output and the random features can be converted as

$$\begin{aligned} \mathbf{t} &= QW_{im}, \\ W_{im} &= \left(Q^T Q\right)^{-1} Q^T \mathbf{t}. \end{aligned} \tag{4}$$

To guarantee the mapping's stability which is between the input and the target while preventing the overfitting of the function, regularization processing

is performed. The regularization formula is as follows:

$$W_{im} = \left[Q(H)^T Q(H) + \lambda e\right]^{-1} Q(H)^T \mathbf{t}, \quad (5)$$

where e represents the unit diagonal matrix and λ represents the penalty coefficient.

Therefore, the output dataset $\boldsymbol{Y}$ is expressed as

$$Y = Q(H) W_{im}. \quad (6)$$

The above expression includes the model-building process. Regarding the test datasets, it is assumed that the input dataset is $Z = [z_1, z_2, \ldots, z_m]$. The random feature is expressed by the following formula:

$$P = (Z + \delta) W. \quad (7)$$

With regard to random features $P = [p_1, p_2, \ldots, p_d]$, through the activation function Q, the dataset can be transformed into

$$Q(\mathbf{P}) = \frac{1}{1 + e^{-\mathbf{P}}}. \quad (8)$$

For the test dataset, the output of the improved classification layer is expressed as follows:

$$\mathbf{T} = Q(\mathbf{P}) \mathbf{W_{im}}. \quad (9)$$

(3) Although the above process can reduce the impact of noise and data fluctuation on clustering results, it is impossible to avoid overfitting completely, and noise still exists in the transformation process. Therefore, training samples and test samples are denoised at the same time. After dimension reduction, the dimension of the sample is greatly reduced and the computational complexity is greatly reduced. Traditional data mining methods only rely on training samples to build models, and test samples cannot be processed.

First, based on the phase space reconstruction method, each feature vector of $\boldsymbol{Y}$ can be converted into an $n \times m$ matrix $\boldsymbol{F}$ where m represents the phase space's embedding dimension, which can be decided by the mutual information and false nearest neighbor approaches, respectively [19, 20]. Then, the SVD-based is applied for exact factorization of the matrix $\boldsymbol{F}$.

$$F = U \Lambda V^T, \quad (10)$$

where $\boldsymbol{V^T}$ represents the complex unitary matrix or an $m \times m$ real, U represents the complex unitary matrix or an $n \times n$ real, and $\boldsymbol{\Lambda}$ represents a diagonal matrix with dimension $n \times m$ in which the diagonal entries of the matrix $\boldsymbol{F}$ are singular value.

The energy concentration of features and noise is reflected by the distribution of singular values. Among them, the useful feature corresponds to a larger value, and the first k useful features are retained, while the noise corresponds to a smaller value, and the noise is set to zero, a new diagonal matrix $\boldsymbol{\Lambda'}$ can be obtained, and then no-free features can be obtained through the inverse singular value transformation.

$$F' = U \Lambda' V^T. \quad (11)$$

Finally, the no-noise features $\boldsymbol{F'}$ is transformed into the new feature by inverse space reconstruction and the new features space can be rewritten as $\boldsymbol{Y'}$.

Likewise, the clustering effect of test samples can be greatly improved by denoising test samples directly here.

2.3. Proof. Assuming the input dataset can be represented as follows:

$$\mathbf{X} = \overline{\mathbf{X}} + \Delta \mathbf{X}, \quad (12)$$

where $\overline{\mathbf{X}}$ represents the difference ideal dataset while $\Delta \mathbf{X}$ represents the perturbation dataset. Then,

$$t = \overline{\mathbf{X}}. \quad (13)$$

For the random features $\mathbf{H} = [h_1, h_2, \ldots, h_d]$, the dataset is represented as

$$\mathbf{H} = (\overline{\mathbf{X}} + \Delta \mathbf{X} + \delta) \mathbf{W} = \overline{\mathbf{H}} + \Delta \mathbf{H}, \quad (14)$$

where $\overline{\mathbf{H}}$ represents the difference ideal dataset while ΔH represents the perturbation dataset.

For the random features $\mathbf{H} = [h_1, h_2, \ldots, h_d]$, through the activation function, the dataset can be converted to

$$Q(\mathbf{H}) = Q(\overline{\mathbf{H}} + \Delta \mathbf{H}). \quad (15)$$

As described in (5), the relationship is expressed by the formula as follows:

$$t = Q(\mathbf{H}) \mathbf{W_{im}} = \overline{\mathbf{X}}. \quad (16)$$

Thus,

$$\mathbf{X} = \overline{\mathbf{X}} + \Delta \mathbf{X} \xrightarrow{W} \mathbf{H} \xrightarrow{Q} Q(\mathbf{H}) \xrightarrow{\mathbf{W_{im}}} \overline{\mathbf{X}}. \quad (17)$$

That is

$$\Delta \mathbf{X} \xrightarrow{W} \mathbf{H} \xrightarrow{Q} Q(\mathbf{H}) \xrightarrow{\mathbf{W_{im}}} 0. \quad (18)$$

End.

2.4. Detailed Description of the Proposed Algorithm. First, reduce the dimensionality of the input dataset, then optimize the transformation by improving the classification, and then improve the internal class. Relying on the artificial intelligence network driven by big data can effectively increase the fault diagnosis's recognition accuracy and reduce the calculation time, improve the calculation efficiency, and overcome the shortcomings of depending on the physical field's knowledge to extract features manually.

On the basis of the above ideas, the schematic diagram of the proposed SL method in actual fault diagnosis is shown in Figure 2. Here are the concrete steps:

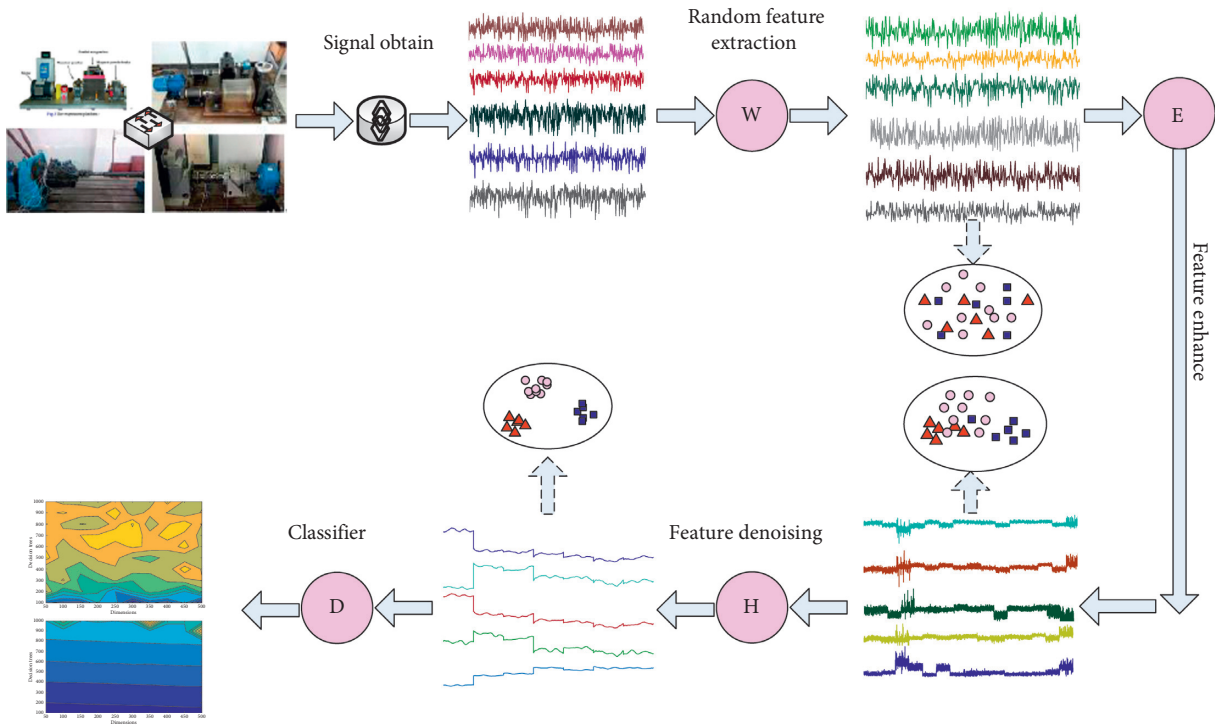

Figure 2: Overview of the proposed SL method.

Step 1: install the sensor at the position where the vibration is most direct, conduct experiments under different working conditions, and collect raw data

Step 2: process the sensor's full-channel data to generate sample dataset and corresponding labels required for training and testing

Step 3: the training dataset is used to train the SL model

Substep 1: reduce the dimensionality of the input dataset to a low dimension by using the method of random feature extraction

Substep 2: Aggregate nonlinear mapping matrices using random features extracted from samples and sample informative labels to shorten the distance between samples of the same type

Substep 3: the training samples after improve classification was denoised by SVD

Substep 4: input the information labels and training samples in order to build the classifiers model

Step 4: the testing dataset is used to test the trained SL model

Step 5: finally, the fault diagnosis accuracy is outputted

3. Stochastic Learning Algorithm for Condition Recognition

Fault diagnosis experiments are carried out based on the wind turbine gearbox and bearing benchmark dataset, respectively, to verify the proposed SL method's effectiveness. The experimental results of the two experiments are analyzed to verify whether the proposed method can increase the calculation efficiency and diagnosis accuracy.

In the experiments, determine the effect of model parameters (such as the number, step size, and dimension of decision trees) of the proposed stochastic learning method on computation time and classification accuracy is determined. The decision trees range from 100 to 500, the step size by 50, and the dimensions are set according to the input dataset.

3.1. Fault Diagnosis for Wind Turbine Gearbox

3.1.1. Failure Experiment Setup. We conduct experiments on the transmission platform that is named DDS which is designed by SpectraQuest Inc. (company website "http://www.pinxuntech.com/"), and the experimental transmission is shown in Figure 3. The experimental device mostly consists of a drive motor, a two-stage planetary gearbox, two grade parallel shaft gearboxes, and a magnetic powder brake. In the experiment, four typical gear failures were studied, namely, surface wear, tooth cracks, chipped tooth, and missing tooth. At the same time, in order to ensure the unity of the experiment, the transmission platform adopts ordinary gears for the experiment. The four most typical gear failures are discussed, including surface wear, cracked teeth, chipped teeth, and missing teeth.

The main component of the drivetrain diagnostics simulator is a gearbox, and it is also the place where the drivetrain diagnosis simulator is prone to failure. In the experiment, the secondary sun gear of the planetary gearbox was diagnosed by using the acceleration sensor to collect the

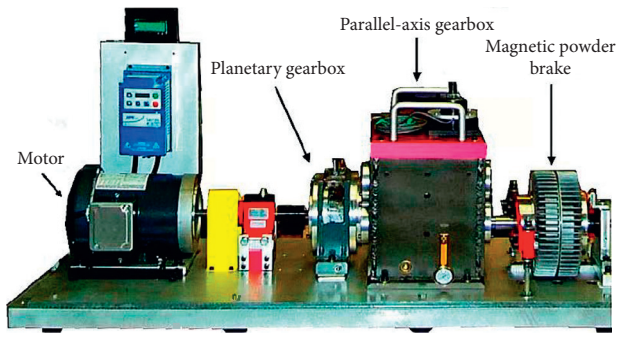

Figure 3: The experiment platform.

vibration signals of different fault types of faults in the transmission process. Table 1 shows the fault state settings and working condition settings of the gearbox.

During the experiment, a total of 5 patterns were set up, 1 normal pattern and 4 different failure patterns. The vibration signal in each pattern was collected by the acceleration sensor. We performed a wavelet transform on the collected vibration signals to extract impulse signals. According to the literature [52], features are extracted from the original signal and impulse signal, respectively, and a total of 50 features are counted. We can obtain 2000×50 feature samples in these experiments. During model training, the dataset is divided, 50% of the samples are applied for model training, and 50% of the samples are applied for testing.

The experimental outcomes of the SL algorithm are indicated in Figure 4. We conclude from Figure 4(a) that the data dimensions and the decision trees quantity will influence the SL algorithm's classification accuracy, but the sensitivity of the recognition accuracy to the dimensions is below the number of decision trees. Figure 4(b) illustrates that decision trees quantity mainly influences the calculation time of the SL algorithm, and the data dimension has a small effect on the calculation time. Therefore, in order to ensure that the algorithm model has high classification accuracy and, at the same time, high computational efficiency, the parameters of the SL algorithm should be set to multiple decision trees and fewer dimensions. From the yellow dot area in Figure 4(b), we can see that the SL algorithm can classify up to 92%, and its corresponding calculation time is relatively low, which is 15 seconds.

3.1.2. Comparison of Different Methods of Diagnosis. The data collected in the experiment is used for different algorithm models for fault diagnosis, the recognition accuracy of different methods is compared, and the diagnostic performance of the proposed SL method is further discussed. The classification accuracy and calculation time of different methods are shown in Table 2. In Table 2, the average recognition accuracy rates of SVM, ESN, SAE + ESN, and SAE + Softmax are 22.10%, 76.98%, 44.73%, and 62.00%, respectively. It is unacceptable in actual engineering applications. The corresponding std is 0, 0.679, 1.988, and 2.770, respectively. In addition, as a shallow learning algorithm, the SVM method must have a lower computational efficiency than the deep learning algorithm SAE. The calculation time of SAE + ESN and SAE + Softmax exceeds 12 s, while the calculation time of SVM and ESN is 0.96 s and 0.68 s, respectively. The classification accuracy and calculation time that are about the SL method are shown in Figure 5. The blue in the picture shows the recognition accuracy, and the red shows the calculation time. With a different number of decision trees, the identification accuracy obtained by the SL method is more than 90%, ranging from 90.08% to 90.28%, far exceeding the recognition accuracy of other algorithms. The greater the number of decision trees, the longer the calculation time. The calculation time of the SL method ranges from 3.25 s to 14.55 s. From Table 2 and Figure 5, compared with other methods, the SL method has the highest computational efficiency. The proposed SL method is preceded by other methods in recognition accuracy and calculation time in Table 2 and Figure 5.

3.2. Fault Diagnosis for the Benchmark Dataset

3.2.1. Failure Experiment Setup. To further approve not only the effectiveness but also the superiority of the SL method, the benchmark data is used to evaluate the proposed SL model. The benchmark dataset is a rolling bearing fault dataset offered with the Case Western Reserve University Bearing Data Center (CWRU) [53]. For the benchmark dataset, we consider the bearings under normal conditions and the bearings with 3 different faults, including outer ring faults, memory faults, and ball faults. Each fault includes 3 defect levels (0.18 mm, 0.36 mm, and 0.53 mm wide grooves). During the course of this experiment, the sensor sampling frequency was set to 48 kHz, and the data were collected under a load of "1," "2," and "3" HP and the acquisition time of each group of vibration signals last for 10 s. A total of 10 sets of bearing data in different states were obtained in the experiment, including 1 set of healthy bearings and 9 sets of faulty bearings, called NR and fault 1 to fault 9.

The dataset is extracted from the original signal, 800 small sample sets are collected for each load, each small sample set has a total of 300 data, each group of bearing data has 2400 small sample sets, and the benchmark dataset uses a total of 24,000 small samples set. Therefore, the size of the original data in the high-dimensional space is 24000×300, and the data plan used to train the model is the same as in Section 3.1.

Figure 6 has displayed the experimental results of the SL algorithm. As shown in Figure 6, the SL algorithm has a recognition accuracy of over 98% under the conditions of different numbers of decision trees and diverse data dimensions, and the algorithm is not sensitive to changes in dimensions and decision trees. When the amount of decision trees is 400 and the dimension is 100, the classification accuracy of the SL method almost reaches 100% at the highest. Generally speaking, the identification accuracy of the algorithm is influenced by input dimensions and decision trees quantity, but the former is more influential, and the recognition accuracy is positively correlated with the amount of decision trees.

Table 1: Planetary gearbox conditions.

Failure pattern	Failure type	Load (Nm)
A	Health	0/1.4/2.8/25.2
B	Surface wear	0/1.4/2.8/25.2
C	Cracked tooth	0/1.4/2.8/25.2
D	Chipped tooth	0/1.4/2.8/25.2
E	Missing tooth	0/1.4/2.8/25.2

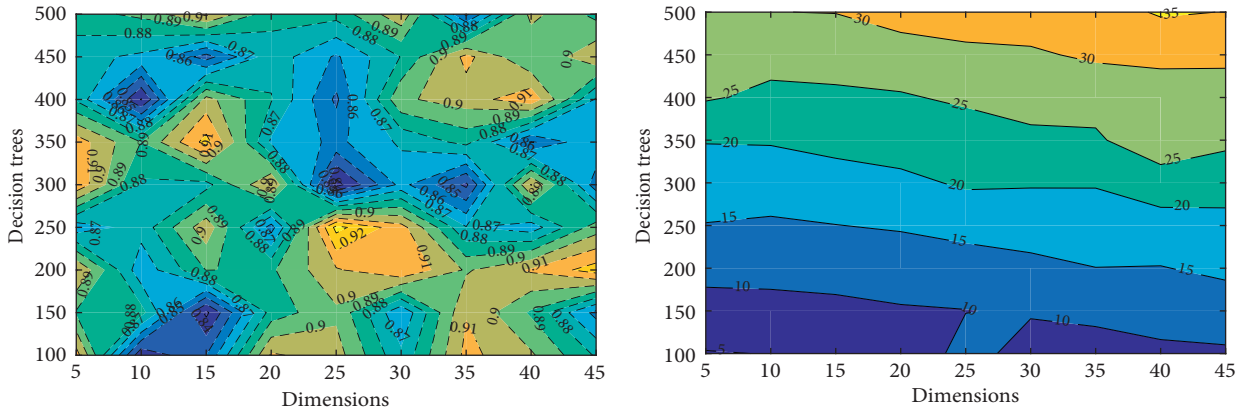

Figure 4: The experiment results for wind turbine gearbox based on SL classifier. (a) Recognition accuracy rate; (b) computing time.

Table 2: Diagnosis results of different approaches.

Approach	Accuracy		Time (s)
	avg (%)	std	
SVM	22.10	0	0.96
ESN	76.98	0.679	0.68
SAE + ESN	44.73	1.988	12.64
SAE + Softmax	62.00	2.770	12.15

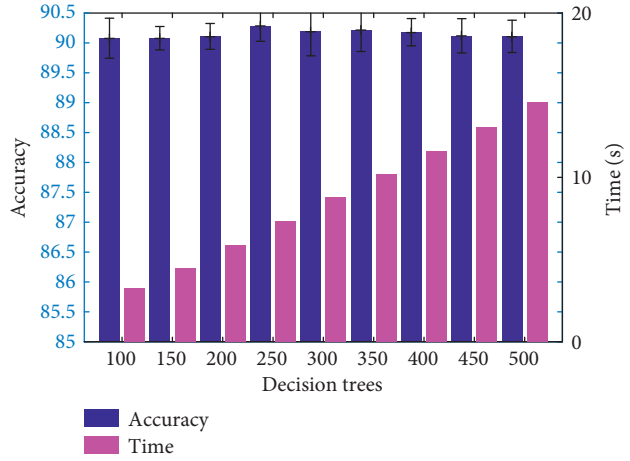

Figure 5: The experiment results for the wind turbine gearbox based on the SL classifier.

3.2.2. Comparison of Different Methods of Diagnosis. As described above, the corresponding comparison results are exhibited in Table 3 and Figure 7. The average classification accuracy rate of SVM, ESN, SAE + ESN, and SAE + Softmax is $26.28\% \pm 0.673$, $39.07\% \pm 1.27$, $37.14\% \pm 0.902$, and $10.02\% \pm 0.0289$, respectively. Furthermore, the SVM and ESN method must have a lower computational efficiency than the deep learning algorithm SAE. The computing times of SAE + ESN and SAE + Softmax are about 290 s, while that of SVM and ESN is 171.55 s and 7.95 s, respectively. As

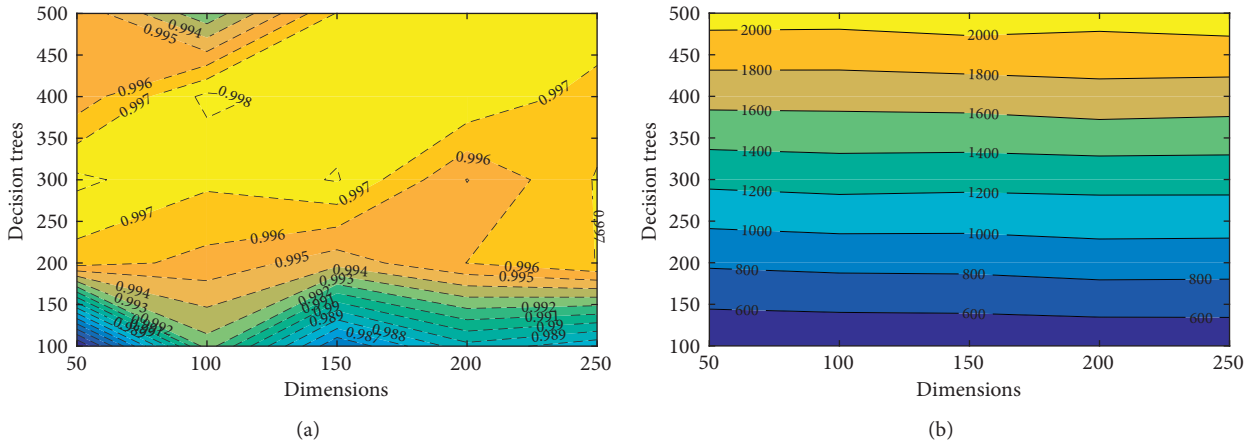

Figure 6: The experiment results for benchmark dataset based on SL classifier. (a) Recognition accuracy rate; (b) computing time.

Table 3: Diagnosis results of different approaches.

Approach	Accuracy		Time (s)
	avg (%)	std	
SVM	26.28	0.673	171.55
ESN	39.07	1.27	7.95
SAE + ESN	37.14	0.902	297.53
SAE + Softmax	10.02	0.0289	292.90

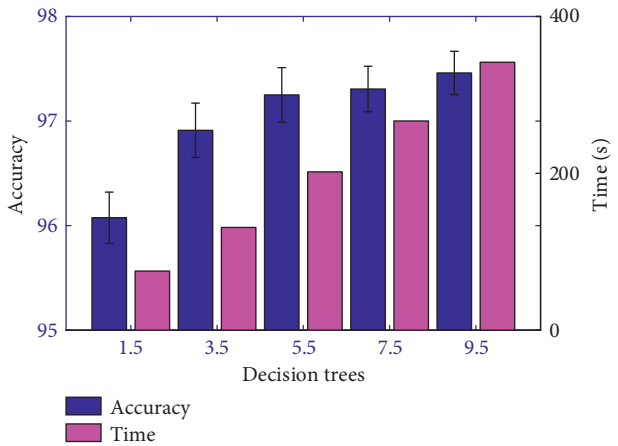

Figure 7: The experiment results for benchmark dataset based on RF classifier.

shown in Figure 7, the SL method classification accuracy and calculation time under different amounts of decision trees. The blue in the figure shows the recognition accuracy, and the red shows the calculation time. Figure 7 shows that the average recognition accuracy rate is 96.08% to 97.46% and increasing the number of decision trees can improve the classification accuracy. Meanwhile, the calculation time is also increasing. The calculation time of the SL method ranges from 75.36 s to 341.7 s.

4. Conclusions

This paper proposes a random learning dimensionality reduction algorithm method and uses it for machine fault state recognition. The classification accuracy and operation efficiency of the dimension reduction algorithm is affected by the size of the dimensions. In the stochastic learning method (SL), random feature extraction is performed on the input high-dimensional data through a random mapping matrix. After random feature extraction, low-dimensional feature data will be obtained for model training. Therefore, after feature extraction, the input SL model sample dimension is largely decreased, and the calculation efficiency is greatly improved. The use of information-enhanced data for training guarantees the classification effect of the SL algorithm.

Data Availability

The data used to support the findings of this study are obtained from previously reported studies. These prior

studies (and datasets) are cited at relevant places within the text as references.

Conflicts of Interest

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

Acknowledgments

This work was supported in part by the special projects in Key Fields of Ordinary Colleges and Universities in Guangdong Province (2020ZDZX3029) and Dongguan Science and Technology Commissioner Project (20201800500212and 20201800500282).

References

[1] Q.-K. Tran and S.-k. Song, "Computer vision in precipitation nowcasting: applying image quality assessment metrics for training deep neural networks," *Atmosphere*, vol. 10, no. 5, p. 244, 2019.

[2] R. Patel and S. Patel, "A comprehensive study of applying convolutional neural network for computer vision," *International Journal of Advanced Science and Technology*, vol. 29, no. 6s, pp. 2161–2174, 2020.

[3] A. Z. da Costa, H. E. H. Figueroa, and J. A. Fracarolli, "Computer vision based detection of external defects on tomatoes using deep learning," *Biosystems Engineering*, vol. 190, pp. 131–144, 2020.

[4] C.-T. Wu, P. van Beek, P. Schmidt, J. P. Moreira, and T. R. Gardos, "Evaluation of semi-frozen semi-fixed neural network for efficient computer vision inference," *Electronic Imaging*, vol. 2021, no. 17, p. 213, 2021.

[5] F. K. Gustafsson, M. Danelljan, and T. B. Schon, "Evaluating scalable bayesian deep learning methods for robust computer vision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 318-319, Seattle, Washington D. C., USA, June 2020.

[6] R. Yang, S. K. Singh, M. Tavakkoli et al., "CNN-LSTM deep learning architecture for computer vision-based modal frequency detection," *Mechanical Systems and Signal Processing*, vol. 144, Article ID 106885, 2020.

[7] N. Hussain, M. A. Khan, M. Sharif et al., "A Deep Neural Network and Classical Features Based Scheme for Objects Recognition: An Application for Machine inspection," *Multimedia Tools and Applications*, pp. 1–23, 2020.

[8] G. Mukherjee, B. Tudu, and A. Chatterjee, "A convolutional neural network-driven computer vision system toward identification of species and maturity stage of medicinal leaves: case studies with Neem, Tulsi and Kalmegh leaves," *Soft Computing*, vol. 25, pp. 1–20, 2021.

[9] G. Li, S. Liang, S. Nie, W. Liu, and Z. Yang, "Deep neural network-based generalized sidelobe canceller for dual-channel far-field speech recognition," *Neural Networks*, vol. 141, pp. 225–237, 2021.

[10] C. H. H. Yang, J. Qi, S. Y. C. Chen et al., "Decentralizing feature extraction with quantum convolutional neural network for automatic speech recognition," in *Proceedings of the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6523–6527, IEEE, Toronto, ON, Canada, June 2021.

[11] X. Cui, W. Zhang, U. Finkler, G. Saon, M. Picheny, and D. Kung, "Distributed training of deep neural network acoustic models for automatic speech recognition: a comparison of current training strategies," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 39–49, 2020.

[12] Z. Wu, D. Zhao, Q. Liang, Y. Jiahui, G. Anmol, and P. Ruoming, "Dynamic sparsity neural networks for automatic speech recognition," in *Proceedings of the ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 6014–6018, IEEE, Toronto, ON, Canada, June 2021.

[13] J. Guglani and A. N. Mishra, "DNN based continuous speech recognition system of Punjabi language on Kaldi toolkit," *International Journal of Speech Technology*, vol. 24, no. 1, pp. 41–45, 2021.

[14] J. Li, Y. Pei, A. Yasin, A. Yasin, S. Ali, and T. Mahmood, "Computer vision-based microcalcification detection in digital mammograms using fully connected depthwise separable convolutional neural network," *Sensors*, vol. 21, no. 14, p. 4854, 2021.

[15] K. M. Black, H. Law, A. Aldoukhi, J. Deng, and K. R. Ghani, "Deep Learning Computer Vision Algorithm for Detecting Kidney Stone composition," *BJU Int*, vol. 125, 2020.

[16] A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," *Computer Methods and Programs in Biomedicine*, vol. 196, Article ID 105581, 2020.

[17] S. Ghosh, A. Bandyopadhyay, S. Sahay, R. Ghosh, I. Kundu, and K. C. Santosh, "Colorectal histology tumor detection using ensemble deep neural network," *Engineering Applications of Artificial Intelligence*, vol. 100, Article ID 104202, 2021.

[18] S. Albahli, "A deep neural network to distinguish covid-19 from other chest diseases using x-ray images," *Current medical imaging*, vol. 17, no. 1, pp. 109–119, 2021.

[19] S. Sabut, O. Pandey, B. S. P. Mishra, and M. Monalisa, "Detection of ventricular arrhythmia using hybrid time–frequency-based features and deep neural network," *Physical and Engineering Sciences in Medicine*, vol. 44, no. 1, pp. 135–145, 2021.

[20] H. Han, L. Xu, X. Cui, and Y. Fan, "Novel chiller fault diagnosis using deep neural network (DNN) with simulated annealing (SA)," *International Journal of Refrigeration*, vol. 121, pp. 269–278, 2021.

[21] C. D. Nguyen, A. E. Prosvirin, C. H. Kim, and J. Kim, "Construction of a sensitive and speed invariant gearbox fault diagnosis model using an incorporated utilizing adaptive noise control and a stacked sparse autoencoder-based deep neural network," *Sensors*, vol. 21, no. 1, p. 18, 2021.

[22] F. Zhou, T. Sun, X. Hu, and T. Wang, "A sparse denoising deep neural network for improving fault diagnosis performance," *Signal, Image and Video Processing*, vol. 15, pp. 1–10, 2021.

[23] D. T. Hoang, X. T. Tran, M. Van, and H. J. Kang, "A deep neural network-based feature fusion for bearing fault diagnosis," *Sensors*, vol. 21, no. 1, p. 244, 2021.

[24] K. Zhou, C. Yang, J. Liu, and Q. Xu, "Dynamic Graph-Based Feature Learning with Few Edges Considering Noisy Samples for Rotating Machinery Fault diagnosis," *IEEE Transactions on Industrial Electronics*, 2021.

[25] C. Yang, K. Zhou, and J. Liu, "SuperGraph: Spatial-Temporal Graph-Based Feature Extraction for Rotating Machinery diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 69, 2021.

[26] X. Li, J. Cheng, H. Shao, and B. Cai, "A fusion CWSMM-based framework for rotating machinery fault diagnosis under strong interference and imbalanced case," *IEEE Transactions on Industrial Informatics*, 2021.

[27] Z. He, H. Shao, Z. Ding, and H. Jiang, "Modified Deep Auto-Encoder Driven by Multi-Source Parameters for Fault Transfer Prognosis of aero-engine," *IEEE Transactions on Industrial Electronics*, vol. 69, 2021.

[28] C. Wang and Z. Xu, "An Intelligent Fault Diagnosis Model Based on Deep Neural Network for Few-Shot Fault diagnosis," *Neurocomputing*, vol. 456, 2021.

[29] H. Tong, R. C. Qiu, D. Zhang, H. Yang, Q. Ding, and X. Shi, "Detection and classification of transmission line transient faults based on graph convolutional neural network," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 3, pp. 456–471, 2021.

[30] M. A. Asghar, M. J. Khan, M. Rizwan, R. M. Mehmood, and S.-H. Kim, "An innovative multi-model neural network approach for feature selection in emotion recognition using deep feature clustering," *Sensors*, vol. 20, no. 13, p. 3765, 2020.

[31] X. Zhao, M. Jia, and Z. Liu, "Semisupervised graph convolution deep belief network for fault diagnosis of electormechanical system with limited labeled data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5450–5460, 2020.

[32] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "PCA-based feature reduction for hyperspectral remote sensing image classification," *IETE Technical Review*, vol. 38, no. 4, pp. 377–396, 2021.

[33] T. R. Gadekallu, D. S. Rajput, M. P. K. Reddy et al., "A novel PCA-whale optimization-based deep neural network model for classification of tomato plant diseases using GPU," *Journal of Real-Time Image Processing*, vol. 18, no. 4, pp. 1383–1396, 2021.

[34] M. P. Uddin, M. A. Mamun, M. I. Afjal, and M. A. Hossain, "Information-theoretic feature selection with segmentation-based folded principal component analysis (PCA) for hyperspectral image classification," *International Journal of Remote Sensing*, vol. 42, no. 1, pp. 286–321, 2021.

[35] R. Ran, Y. Ren, S. Zhang, and B. Fang, "A novel discriminant locality preserving projections method," *Journal of Mathematical Imaging and Vision*, vol. 63, no. 5, pp. 541–554, 2021.

[36] Y.-L. He, Y. Zhao, X. Hu, X.-N. Yan, Q.-X. Zhu, and Y. Xu, "Fault diagnosis using novel AdaBoost based discriminant locality preserving projection with resamples," *Engineering Applications of Artificial Intelligence*, vol. 91, Article ID 103631, 2020.

[37] W. M. Shaban, A. H. Rabie, A. I. Saleh, and M. A. Abo-Elsoud, "A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier," *Knowledge-Based Systems*, vol. 205, Article ID 106270, 2020.

[38] Y. Zhou, K. Xu, F. He, and D. He, "Nonlinear fault detection for batch processes via improved chordal kernel tensor locality preserving projections," *Control Engineering Practice*, vol. 101, Article ID 104514, 2020.

[39] B. P. O. Lovatti, M. H. C. Nascimento, Á. C. Neto, E. V. R. Castro, and P. R. Filgueiras, "Use of Random forest in the identification of important variables," *Microchemical Journal*, vol. 145, pp. 1129–1134, 2019.

[40] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Systems with Applications*, vol. 134, pp. 93–101, 2019.

[41] E. Izquierdo-Verdiguier and R. Zurita-Milla, "An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, Article ID 102051, 2020.

[42] C. Iwendi, A. K. Bashir, A. Peshkar et al., "COVID-19 patient health prediction using boosted random forest algorithm," *Frontiers in Public Health*, vol. 8, p. 357, 2020.

[43] J. Li, Y. Tian, Y. Zhu et al., "A multicenter random forest model for effective prognosis prediction in collaborative clinical research network," *Artificial Intelligence in Medicine*, vol. 103, Article ID 101814, 2020.

[44] G. L. Watson, D. Xiong, L. Zhang et al., "Fusing a Bayesian case velocity model with random forest for predicting COVID-19 in the US," 2020.

[45] D. V. Urista, D. B. Carrué, I. Otero et al., "Prediction of antimalarial drug-decorated nanoparticle delivery systems with random forest models," *Biology*, vol. 9, no. 8, p. 198, 2020.

[46] Y. Chen, W. Zheng, W. Li, and Y. Huang, "Large group activity security risk assessment and risk early warning based on random forest algorithm," *Pattern Recognition Letters*, vol. 144, pp. 1–5, 2021.

[47] D. Sun, S. Shi, H. Wen, J. Xu, X. Zhou, and J. Wu, "A hybrid optimization method of factor screening predicated on GeoDetector and Random Forest for Landslide Susceptibility Mapping," *Geomorphology*, vol. 379, Article ID 107623, 2021.

[48] L. Dong, H. Du, F. Mao et al., "Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning technique—subtropical area for example," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 113–128, 2019.

[49] O. Guehairia, A. Ouamane, F. Dornaika, and A. Taleb-Ahmed, "Feature fusion via Deep Random Forest for facial age estimation," *Neural Networks*, vol. 130, pp. 238–252, 2020.

[50] J. Upadhyay, A. Rawat, D. Deb, V. Muresan, and M.-L. Unguresan, "An RSSI-based localization, path planning and computer vision-based decision making robotic system," *Electronics*, vol. 9, no. 8, p. 1326, 2020.

[51] J. Luo, Y. Liu, S. Zhang, and J. Liang, "Extreme random forest method for machine fault classification," *Measurement Science and Technology*, vol. 32, no. 11, Article ID 114006, 2021.

[52] X. Wang, Y. Qin, Y. Wang, S. Xiang, and H. Chen, "Reltanh: an activation function with vanishing gradient resistance for sae-based dnns and its application to rotating machinery fault diagnosis," *Neurocomputing*, vol. 363, 2019.

[53] K. A. Loparo, "Bearing Data center, Case Western Reserve University," 2014, http://csegroups.case.edu/bearingdatacenter/pages/download-data-file.