



Practice article

Intelligent fault diagnosis of hydraulic piston pump based on deep learning and Bayesian optimization

Shengnan Tang^{a,d}, Yong Zhu^{a,b,c,*}, Shouqi Yuan^a

^a National Research Center of Pumps, Jiangsu University, Zhenjiang 212013, China

^b International Shipping Research Institute, GongQing Institute of Science and Technology, Jiujiang 332020, China

^c Key Laboratory of Fire Emergency Rescue Equipment, Ministry of Emergency Management, Shanghai 200032, China

^d Institute of Advanced Manufacturing and Modern Equipment Technology, Jiangsu University, Zhenjiang 212013, China

ARTICLE INFO

Article history:

Received 28 July 2021

Received in revised form 10 January 2022

Accepted 10 January 2022

Available online 19 January 2022

Keywords:

Hydraulic pump

Intelligent fault diagnosis

Convolutional neural network

Bayesian optimization

ABSTRACT

Hydraulic axial piston pump is broadly-used in aerospace, ocean engineering and construction machinery since it is the vital component of fluid power systems. In the light of the undiscoverability of its fault and the potential serious losses, it is valuable and challenging to complete the fault identification of a hydraulic pump accurately and effectively. Owing to the limitations of shallow machine learning methods in the intelligent fault diagnosis, more attention has been paid to deep learning methods. Hyperparameter plays an important role in a deep learning model. Although some manual tuning methods may represent good results in some cases, it is hard to reproduce due to the differences of datasets and other factors. Hence, Bayesian optimization (BO) algorithm is adopted to automatically select the hyperparameters. Firstly, the time–frequency images of vibration signals by continuous wavelet transform are taken as input data. Secondly, by setting some hyperparameters, a preliminary convolutional neural network (CNN) model is established. Thirdly, by identifying the range of each hyperparameter, BO based on Gaussian process is employed to construct an adaptive CNN model named CNN-BO. The performance of CNN-BO is verified by comparing with traditional LeNet 5 and improved LeNet 5 with manual optimization. The results indicate that CNN-BO can accomplish the intelligent fault diagnosis of a hydraulic pump accurately.

© 2022 ISA. Published by Elsevier Ltd. All rights reserved.

1. Introduction

As a pivotal power part of fluid drive system, hydraulic axial piston pump (HAPP) has been widely used in construction machinery, aerospace, forging machinery, marine and mining machinery [1–3]. Once there is failure in a hydraulic piston pump, it will produce an impact on the whole hydraulic system. More severe losses will be inflicted on the practical industrial production and personal safety [4–6]. Moreover, the distinct characteristics of its faults are sealing, hiddenness, coupling, randomness and complexity, which make it hard for feature extraction and fault identification. Therefore, it is of great essence to search for an efficient and feasible method to diagnose the faults of HAPP.

With the emergence of mechanical big data and the development of artificial intelligence in interdisciplinary subjects, intelligent methods have aroused great attention in the fields of machinery fault diagnosis [7–10]. By combining Kalman filters and artificial neural network, Cho et al. achieved the accurate and

effective fault recognition of wind turbine in different states [11]. By using empirical mode decomposition for feature extraction, an intelligent diagnosis method based on artificial neural network was utilized for rolling bearing [12]. As for data imbalance, an improved method based on support vector machine (SVM) was used for fault diagnosis of steering actuator in an automated driving vehicle [13]. Among them, grey wolf optimizer algorithm was introduced to tune the threshold. By taking the effect of data imbalance and signal heterogeneity into account, an improved SVM was developed for fault detection of braking system in high-speed train by combining weighted-feature strategy and cost-sensitive learning [14]. The above shallow models can be effective in solving some uncomplicated classification problems, while they are strongly dependent on the experience and knowledge of feature extraction.

Owing to the limitations of the generalization capability of shallow models in dealing with complex problems, deep learning (DL) based methods present the special superiority [15–17]. By using real-valued output-based diversity metric, a deep ensemble learning method was developed for gearbox and bearing fault diagnosis under the interferences of working conditions and noise [18]. As for data insufficiency, Li et al. proposed a fault diagnosis

* Corresponding author at: National Research Center of Pumps, Jiangsu University, Zhenjiang 212013, China.

E-mail address: zhuyong@ujs.edu.cn (Y. Zhu).

method of wind turbine by integrating transfer learning and convolutional autoencoder [19]. Chen et al. accomplished the fault diagnosis of gearbox by employing a convolution neural network (CNN) [20]. Chen et al. developed a modified CNN model for bearing fault diagnosis by implementing the time–frequency conversion using cyclic spectral coherence [21]. By analyzing acoustic signal, Kumar et al. applied a CNN into the failure identification towards centrifugal pump [22]. Based on vibration signal, a subset method combined deep auto-encoder with particle swarm optimization was employed for bearing fault identification [23]. A transfer auto-encoder optimized by particle swarm optimization was used to achieve the effective fault diagnosis of different bearings and gears [24]. Similarly, using particle swarm optimization for the learning of critical hyperparameter (HP), a hybrid approach was used for bearing failure identification by using CNN and SVM [25]. Fault diagnosis of converters was completed by combining a deep CNN and genetic algorithm [26]. There are still some challenges and problems in the present studies.

(1) Many studies of deep models are emphasized on the intelligent fault diagnosis of bearing, gearbox, and motor, it is still few on the HAPP. Its structure and working mechanism are more complex. The hiddenness and coupling of its failure makes it more valuable and challenging for fault diagnosis and condition monitoring.

(2) The time-consuming preprocessing of the original signals is conducted in the traditional intelligent diagnosis methods. Moreover, it requires more stringent knowledge on equipment fault mechanism and the technology of data preprocessing.

(3) The setting of hyperparameters in most deep learning models is mostly based on empirical knowledge and manual tuning. Some of them using intelligent search algorithms or evolutionary algorithms still present some limitations in the computational expense and reproducibility. Moreover, it is lacking on the application of Bayesian optimization (BO) in automatically selecting the appropriate HPs of deep models.

Therefore, the main contributions in this study are in the following:

(1) The intelligent fault diagnosis of a HAPP is explored in consideration of its special structure and mechanism. The ontology representation information of the pump is employed as the data source, which can achieve the non-destructive condition monitoring. The characteristics of the original sensor information in both time domain and frequency domain are fully utilized to save the complicated and time-consuming signal preprocessing steps.

(2) Different working conditions are set up, and different wear degrees of the same fault type are included in the analysis. The performance of the proposed method is explored from different perspectives. By utilizing the feature extraction ability of CNN for high-dimensional information and high-precision recognition ability under supervised learning, the self-learning and classification of time–frequency features for the HAPP are realized.

(3) The adaptive global optimization ability of BO algorithm is integrated to enhance diagnosis accuracy. It can present great reproducibility, good convergence and strong robustness. The constructed method can be extended to other pumps and other rotating machinery.

The reminder of this work is organized in the following. Section 2 outlines basic theory of CNN and BO. Section 3 introduces the procedure of the proposed method. The diagnosis method combining BO and CNN is called CNN-BO. Section 4 details the steps of vibration signal acquisition. In Section 5, the comparative analysis of main results is provided. Finally, Section 6 provides some conclusions.

2. Theoretical basis

2.1. Convolutional neural network

CNN is a special representative of feed-forward neural networks and presents the potent capability in mining and extracting useful features in a supervised learning way [27,28]. The structure of a typical CNN comprises input layer, output layer, convolutional layer (Conv), pooling layer, rectified linear unit layer (ReLU) and fully-connected layer (FC).

The convolution layer plays a critical role in feature extraction for CNN. The input is convolved by each convolutional kernel by conducting dot product. The size of convolutional kernel can be understood as the local receptive field. The Conv operation can be formulated by [29,30],

$$X_v^l = F\left(\sum_{u \in M_v} X_u^{l-1} * K_v^l + B_v^l\right) \quad (1)$$

where, X denotes the input of CNN, M_v is used to select the input feature maps, l represents the l th layer of a convolutional neural network, K_v denotes the convolutional kernel, and X_v^l is attained from $(l-1)$ th and denotes the v th feature maps. The convolution operation is represented by $(*)$. B_v^l is the bias corresponding to the j th output of the l th Conv in the convolution process. F is an activation function and can promote the capability of CNN in solving the nonlinear problems. As one of widely used unsaturated activation functions in CNN, ReLU presents the superiority in computing and resolving gradient vanishing.

Pooling layer can select more beneficial information from the features provided by Conv. It can reduce the parameters at the same time of maintaining the invariance of image feature. The corresponding operation can be calculated by [31,32],

$$a_{v-s}^l = F(W_v^l \text{down}(M_v^{l-1}) + B_v^l) \quad (2)$$

among these, F denotes an activation function, the weight is denoted by W_v^l , B_v^l presents the bias, M_v^{l-1} is used to denote the feature map in the $(l-1)$ th layer. Pooling operation is expressed by $\text{down}(\cdot)$, which is often the calculation of the maximum or the average value.

FC is employed to transform the obtained feature maps into a 1D vector. Softmax classifier is used to obtain the predicted results.

2.2. Bayesian optimization algorithm

HP tuning plays a significant role in the construction of deep learning models. It can be completed by manual tuning and automatic methods. Although good results can be obtained by manual optimization, it mainly relies on the experience and possesses undesired reproducibility. Random search and grid search can tune HPs automatically, while some noneffective spaces may be inevitable since it cannot obtain the previous searched information. As representatives of evolutionary algorithms, genetic algorithm and particle swarm optimization can accomplish the global optimization, but it is not preferable on account of the computation and convergence.

In the light of the deficiencies of the above methods, BO has attracted great attention in parameter adjusting owing to its distinctive advantages. The difference of BO is to consider the previous parameter information by using Gaussian progress (GP) to continually update the prior. And BO shows small number of iterations and fast convergence speed. Moreover, BO still remains robust in solving the non-convex problems, and can effectively avoid local optimality. Owing to the good convergence and robustness, it is beneficial to employ BO to achieve the optimization of HPs accurately and efficiently [33,34]. This paper plays an

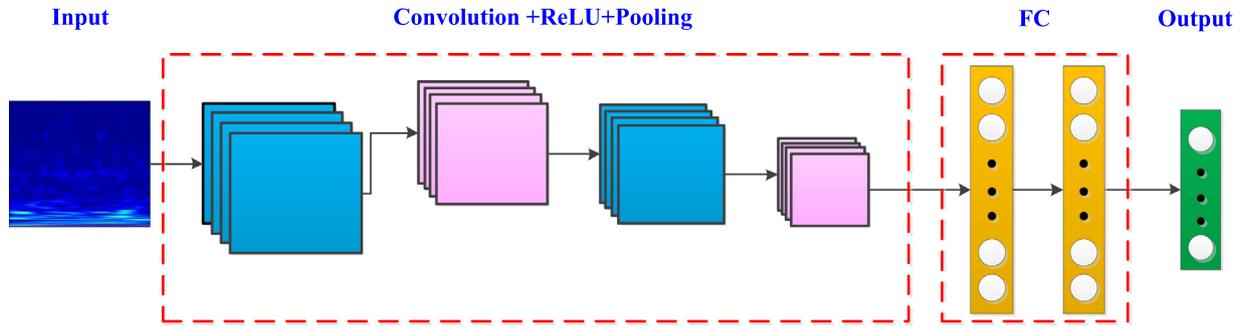


Fig. 1. The structure of convolutional neural network.

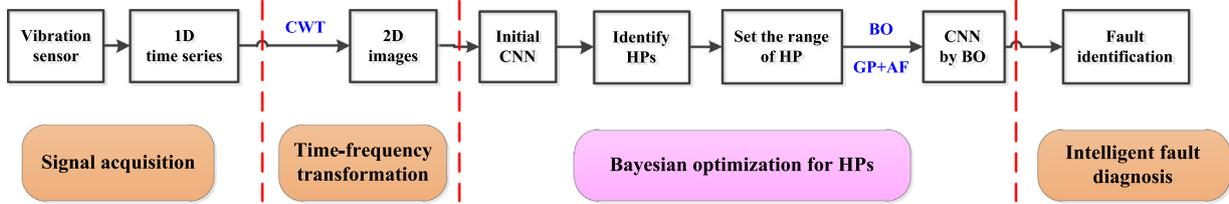


Fig. 2. The procedure of fault diagnosis using CNN-BO.

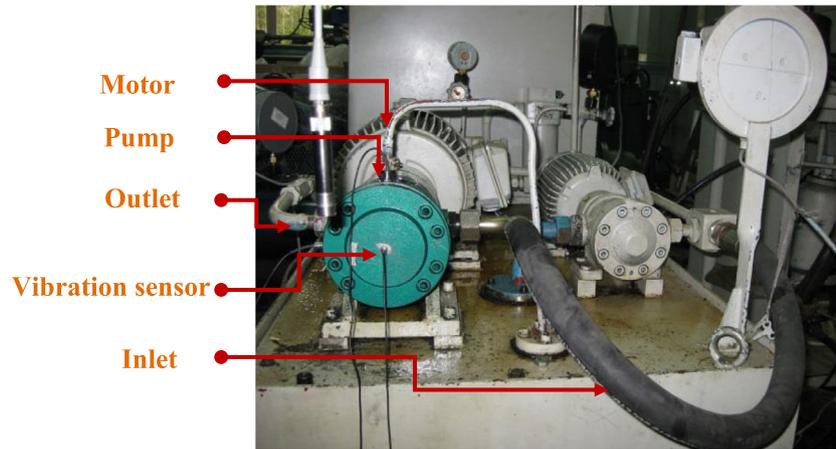


Fig. 3. The test bench of a hydraulic piston pump.

emphasis on BO based on GP, which is used for fitting the objective function. GP and acquisition function are considered as two significant functions in BO [35].

GP can be used for parameter tuning, non-linear regression and classification [36,37]. It is critical to choose the suitable mean function and covariance function. If the prior distribution conforms to joint Gaussian, it can be represented by [38,39]:

$$f(n_{1:t}) \sim \text{Normal}(a(n_{1:t}), C^\theta(n_{1:t}, n_{1:t})) \quad (3)$$

where, f denotes a smooth function, $n_{1:t}$ represents a finite input collection, the mean vector is represented by $a(n_{1:t})_i = a(n_i)$; and the covariance matrix is denoted by $C^\theta(n_{1:t}, n_{1:t})_{ij} = c^\theta(n_i, n_j)$, which is parameterized by θ .

When Matérn 5/2 is selected as covariance function, and a zero-mean function is employed, the corresponding covariance can be expressed by:

$$c_{M5/2}^\theta(n_i, n_j) = \exp(-\sqrt{5}r)(1 + \sqrt{5}r + \frac{5}{3}r^2) \quad (4)$$

among them, $r = (n_i - n_j)^T \text{diag}(\theta^2)^{-1}(n_i - n_j)$, diag denotes a diagonal matrix. The covariance functions is presented with a d length-scale HP θ_i . It is essential to mine the underlying functional

relationships via flexible control of the smoothness of the Matérn 5/2.

Acquisition function is a kind of heuristics for the evaluation of an effective point according to the current model, including the improvement based methods, the entropy based strategy and some combined methods. The strategy of probability of improvement can only reflect the improved probability, but cannot represent the amount of improvement. The method of upper confidence bound is sensitive to parameter β . As for entropy search strategy, it requests a large amount of computation and the approximate technology. Compared with the above functions, EI is a very prevailing acquisition function on the basis of improvement strategy. Its main superiorities are the few parameters and simple calculation. It integrates the probability of improvement and presents the improved amount as well. Moreover, it can balance between depth and width. Moreover, the factor of noise is considered in noisy EI, which is much closer to reality. It is beneficial to use EI for balancing the exploration and exploitation.

EI can be expressed by [39]:

$$\alpha_{EI}(X; \theta, D) = E[\max(0, f(X) - f(X^*))] \quad (5)$$

where, D denotes a dataset, $f(X^*)$ represents the current best value, and maximizing $f(X)$ is the goal.

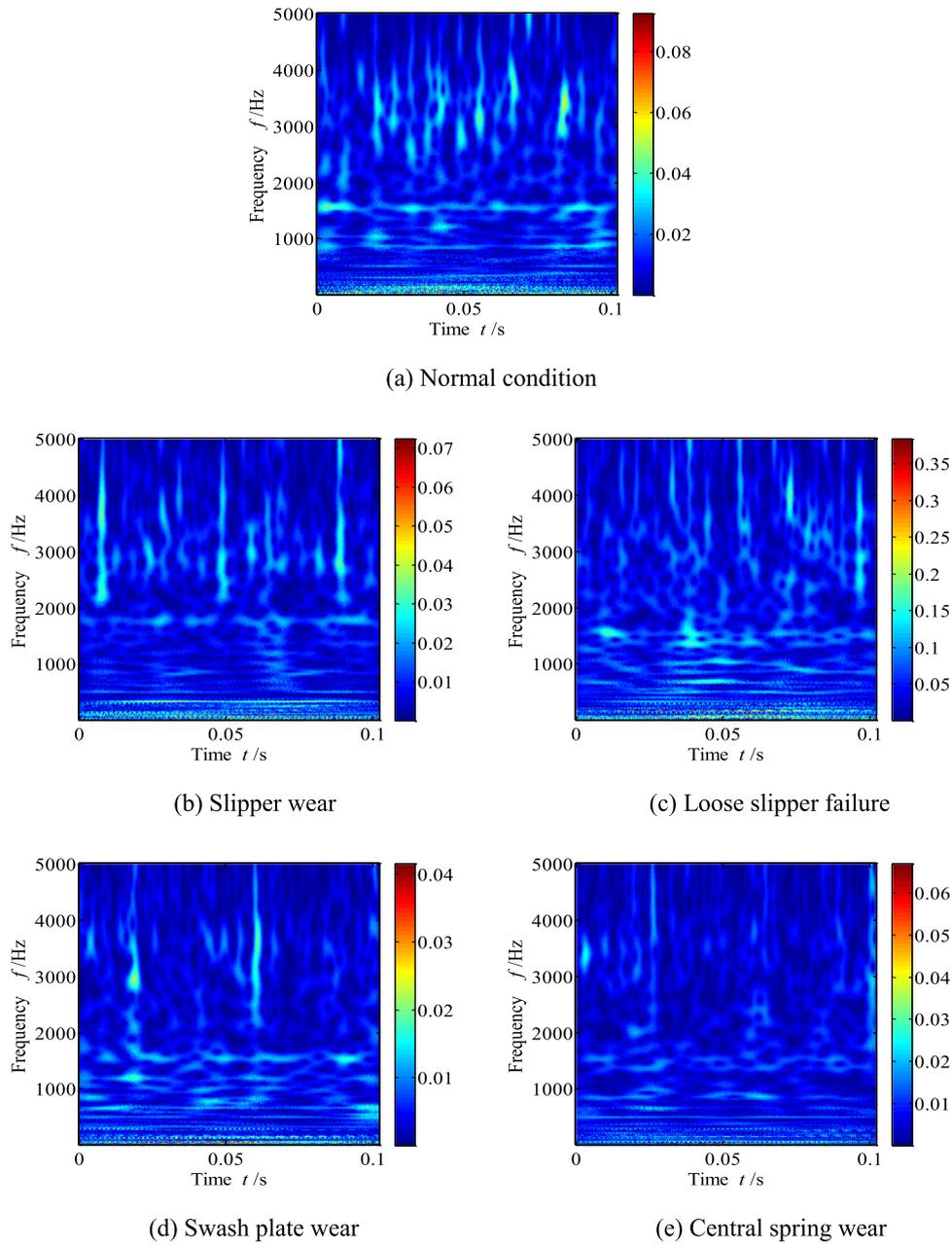


Fig. 4. CWT representations of vibration signal.

Suppose that each result of the function is destructed by σ -sub-Gaussian noise, the expression combining with GP can be:

$$\alpha_{EI}(X; \theta, D) = \sigma_t(X; \theta, D)[x\Phi(x) + \phi(x)] \quad (6)$$

where, $x = \frac{\mu_t(X; \theta, D) - f(X^*)}{\sigma(X; \theta, D)}$, Φ denotes the cumulative distribution function, and ϕ denotes the probability density function of the standard normal distribution.

Further, considering the potential effect of stochastic noise, the corresponding EI can be represented by:

$$\alpha_{EI}(X; \theta, D) = \sigma_t(X; \theta, D)[s\Phi(s) + \phi(s)] \quad (7)$$

among them, $s = \frac{\mu_t(X; \theta, D) - \mu(\theta^*)}{\sigma(X; \theta, D)}$, and the best expected value calculated from the mean results are expressed by $\mu(\theta^*)$.

3. Proposed diagnosis method

3.1. Construction of CNN

As a typical CNN model, LeNet 5 has been proved to be successful in handwritten digit recognition and image classification [40]. A deep learning framework is built for HAPP fault diagnosis on the basis of conventional LeNet 5. Fig. 1 presents the infrastructure of CNN composed of two Conv and two FC. The smaller convolutional kernel is preliminarily designed. ReLU activation function is employed to provide the non-linear learning ability for CNN. Maxpooling is used for dimension reduction by down-sampling. Softmax regression function is set in the output layer to achieve the prediction and classification.

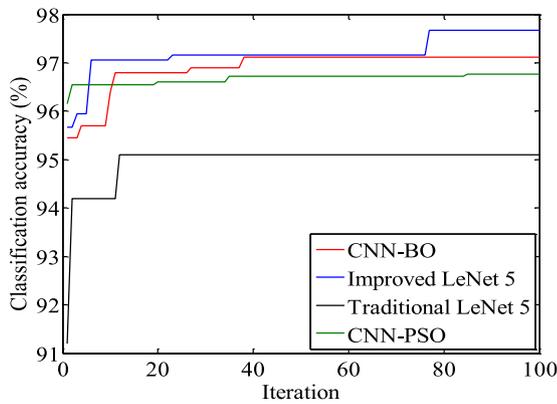


Fig. 5. Iteration process of different models.

Table 1

Key parameters of the main components.

Component	Parameter name	Parameter setting
Pump	Type	MCY14-1B
	Theoretical displacement	10 ml/r
	Nominal pressure	31.5 MPa
	Rated speed	1470 r/min
	Rotary frequency	24.5 Hz
Drive motor	Type	Y132M-4

3.2. Implementation of the proposed method

Bayesian optimization algorithm is used for the selection of HPs, and the proposed diagnosis method is presented in Fig. 2. The detailed steps are in the following:

- (1) Vibration sensor is used for signal acquisition.
- (2) The 2D CWT images of 1D time series are input into CNN.
- (3) An improved CNN is constructed by setting initial HPs based on traditional LeNet 5.
- (4) HPs optimization is implemented by using Bayesian algorithm. The objective function is the real function relationship between the performance of the CNN model and HPs. The HP combination obtained from the acquisition function in each iteration is utilized for CNN training and testing. The group that makes the test accuracy as high as possible will be selected. Firstly, main HPs to be optimized are identified, such as learning rate (LR), epoch, kernel size and number, batch size and so on. Secondly, an appropriate range is set for each HP. Thirdly, Bayesian optimization loop is performed according to the evaluate function. GP is used to model the function relationship between the performance of CNN and HPs. To decrease the effect of additional results, noisy EI is selected as acquisition function to guide the follow-up sampling. Finally, the optimized CNN model is verified by using the optimal HP combinations.
- (5) The fault classification of HAPP is intelligently accomplished by integrating a deep CNN and Bayesian optimization.

4. Experimental data acquisition

The test bench of the HAPP with seven plungers is used for collecting typical fault data, as displayed in Fig. 3. The key parameters of the major components in the experiment are shown in Table 1, mainly including a hydraulic pump and a drive motor. The signal collection is completed in Yanshan University. The sampling frequency of 10 kHz is used in the experiment.

The Piezoelectric accelerometer of YD72-D is employed for vibration signal acquisition. It is located in the end cover of the pump shell by direct pasting. Five different health states are taken

into account, loose slipper failure (sx), swash plate wear (xp), without any fault (zc), central spring wear (th) and slipper wear (hx) respectively. Among them, zc means it is in normal operation. The other four failures occur on the slipper, swash plate and center spring. Loose slipper failure is triggered by the deformation and the large gap of the plunger ball head and ball socket sink. Once the gap of plunger ball head and slipper socket is more than the limit value, high pressure oil in plunger cavity could leak. Metal contact friction happens between slipper and swash plate, which will result in slipper ablation and swash plate wear. The center spring maintains a contact seal between the cylinder block and the valve plate, and the preload of the spring also squeezes the return plate to push the plunger and slipper back. Spring wear or break will make the plunger return unsuccessful or unable to return, which leads to internal leakage, the pump oil shortage, and even action hysteresis of the hydraulic system actuator. Signals under five different working pressures are acquired.

5. Results and discussion

5.1. Input data

The measured vibration signals were changed into 2D distributions by using CWT. The wavelet basis function is ComplexMorlet. The center frequency of 3 and the bandwidth of 3 are set in the experiment. Choose the length of the scale sequence as 256.

The CWT images are shown in Fig. 4. There is no remarkable difference among the time–frequency representations in varying conditions. It is almost impossible to differentiate each type only based on the features from such similar wavelet images. Therefore, it is crucial to probe an effective and feasible method to extract the valid information and accomplish the fault classification.

The whole datasets include 6000 time–frequency images, 1200 images for each type of five conditions. The image is resized via the transform strategy before input into the network. For train samples, random horizontal flip is employed to gain more useful information. Train dataset and test dataset are established by randomly splitting, 70% and 30% respectively.

5.2. Identification of hyperparameters

The HPs of the above CNN are further optimized by BO via employing the identified ranges. The preliminary settings of HPs involve the input of 64×64 , ReLU activation function, and Max pooling of 2×2 . The HPs to be optimized are mainly LR, batch size, max epoch, kernel size of Conv1 and Conv2, kernel number of Conv1 and Conv2, and kernel number of FC1 and FC2. The range of each HP is based on Improved LeNet 5 with manual optimization.

With the same ranges, the optimized HP groups by using BO and PSO are presented in Table 2. The HPs of the proposed CNN model are detailed in Table 3, including the parameters of the input and output layer. The HPs of traditional LeNet 5 and Improved LeNet 5 are listed in Table 4.

The whole optimization experiment includes 100 trials. For the three different CNN models, BO is conducted to adaptively learn the HPs. To validate the performance of the BO algorithm, PSO is used for comparison. The iteration process of different models is presented in Fig. 5. CNN-BO converges in less than 40 iterations and the classification accuracy achieves over 97%. The accuracies of traditional LeNet 5 and improved LeNet 5 are 95.05% and 96.67%. The accuracy of the proposed method is greatly higher than that of traditional LeNet 5 and presents faster convergence in comparison with improved LeNet 5. The CNN model by PSO reaches the optimized accuracy of less than 97% in

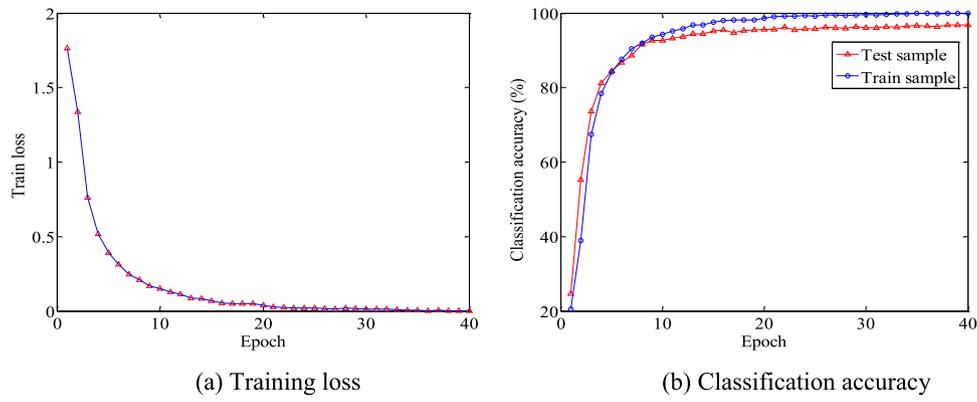


Fig. 6. Curves of training loss and accuracy with CNN-BO.

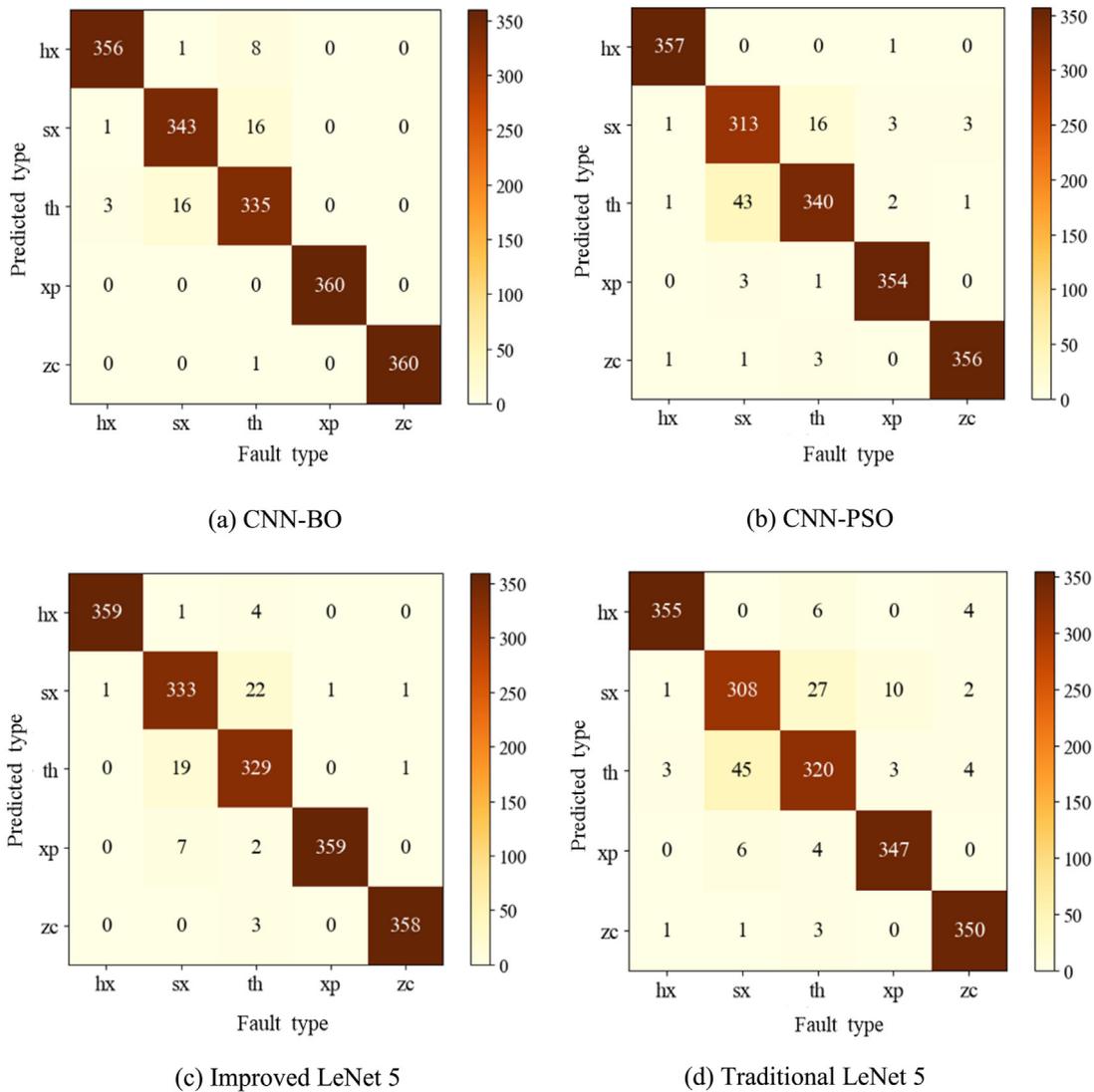


Fig. 7. Confusion matrix of different models.

more than 80 iterations. The whole process with PSO takes about 120 h, which is much longer than the 5 h consumed by the similar process with BO in 100 iterations. It can be implied that good convergence and classification performance are accomplished by employing BO algorithm.

5.3. Analysis of diagnosis results

To avoid random influences, ten repeated experiments are carried out. From Fig. 6, the CNN model converges in epoch 40 and the training loss reaches a smaller value approaching zero. The test and train accuracy only presents slight distinction.

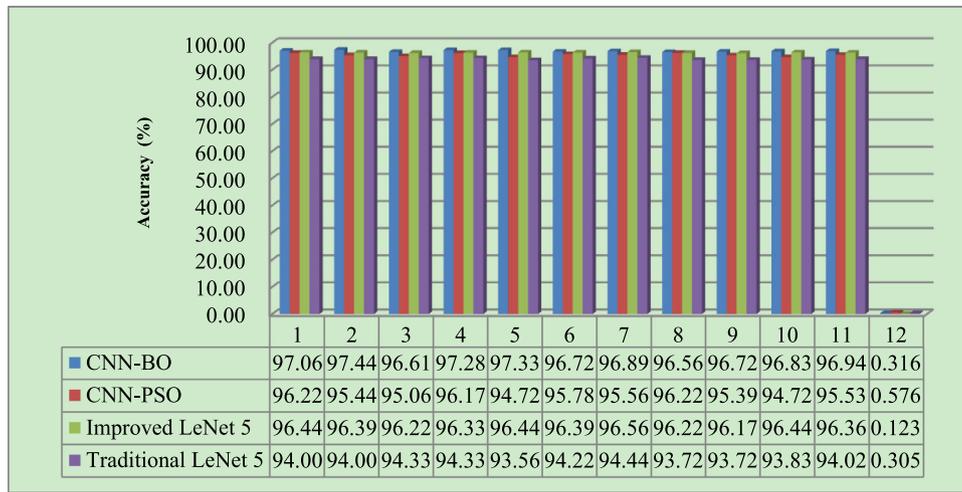


Fig. 8. Accuracy of different CNN models in 10 trials. Group 11 denotes the average value, and group 12 presents the standard error.

Table 2

Best hyperparameter group by using BO and PSO.

No.	Name of hyperparameters	Range	BO result	PSO result
1	LR	[0.001, 0.01]	0.01	0.0064
2	Batch size	[24, 56]	35	52
3	Max epoch	[30, 60]	40	33
4	Kernel size of Conv1	[3, 7]	3	3
5	Kernel size of Conv2	[3, 7]	3	5
6	Kernel number of Conv1	[6, 20]	20	9
7	Kernel number of Conv2	[10, 30]	16	18
8	Output of FC1	[80, 160]	123	80
9	Output of FC2	[60, 100]	77	40

Table 3

Hyperparameters of the proposed CNN.

No.	Layer	Parameter
1	Input	64 × 64 × 3
2	LR	0.01
3	Batch size	35
4	Max epoch	40
5	Conv1	20 × 3 × 3
6	Conv2	16 × 3 × 3
7	Max pooling 1	20 × 2 × 2
8	Max pooling 2	16 × 2 × 2
9	FC1	123 × 1
10	FC2	77 × 1
11	Output	5 × 1

Table 4

Hyperparameters of traditional and improved models.

No.	Name	Traditional LeNet 5	Improved LeNet 5
1	Input	32 × 32 × 3	64 × 64 × 3
2	LR	0.009	0.009
3	Batch size	32	42
4	Max epoch	40	40
5	Conv1	6 × 5 × 5	9 × 3 × 3
6	Conv2	16 × 5 × 5	18 × 5 × 5
7	Max pooling 1	6 × 2 × 2	9 × 2 × 2
8	Max pooling 2	16 × 2 × 2	18 × 2 × 2
9	FC1	120 × 1	120 × 1
10	FC2	84 × 1	84 × 1
11	Output	5 × 1	5 × 1

The statistical results of confusion matrix are shown in Fig. 7. It can be found that an improvement in the fault types of sx and th is achieved by CNN-BO compared with three different models.

Table 5

Classification precision of each type with different models.

Types	CNN-BO	CNN-PSO	Improved LeNet 5	Traditional LeNet 5
zc	100.00	98.61	99.17	98.56
xp	100.00	98.89	97.55	97.20
hx	97.53	99.72	98.63	97.26
sx	95.28	93.15	93.02	88.51
th	94.63	87.86	94.27	85.33

Classification precision is the ratio of labels predicted to be correct to the labels that are considered to be correct by the model. It can be used for the analysis of classification result for each condition. As displayed in Table 5, the precision of both zc and xp reaches up to 100%. In comparison to the other three models, CNN-BO represents the higher precision on sx and th. Compared with traditional LeNet 5, the precision of CNN-BO for sx increases by 6.77%, and 9.30% for th. The precisions of CNN-BO for five states are higher than those of CNN-PSO, and there is an increase of 6.77% for th.

The average accuracy of ten experiments in test dataset is shown in Fig. 8. For CNN-BO, no obvious difference is observed among the ten results. Compared with traditional LeNet 5, CNN-BO presents the remarkable enhancement. A slight improvement of accuracy can be obtained compared to Improved LeNet 5. Compared with CNN-PSO, CNN-BO shows much stronger stability and higher accuracy. It can be indicated that the performance of the model is promoted via automatic optimization of HPs by using BO.

To explore the results learned by CNN-BO, *t*-distributed stochastic neighbor embedding (*t*-SNE) is exploited for the reduction and visualization. The horizontal axis and vertical axis represent the two dimensions of *t*-SNE embedding space, namely Component 1 and Component 2. Each point denotes a testing data. It can be seen from Fig. 9 that scattered spots are observed from the feature distribution of input. Although some features in types of zc and xp begin to get together, most features in different conditions present almost uniform distribution from the information extracted by Conv 1 and Conv 2. Through the learning of FC, feature clusters are formed in the same types, and obvious classes are distinct in this stage. CNN-BO successfully achieves the fault classification of a HAPP using the time–frequency images of vibration signals.

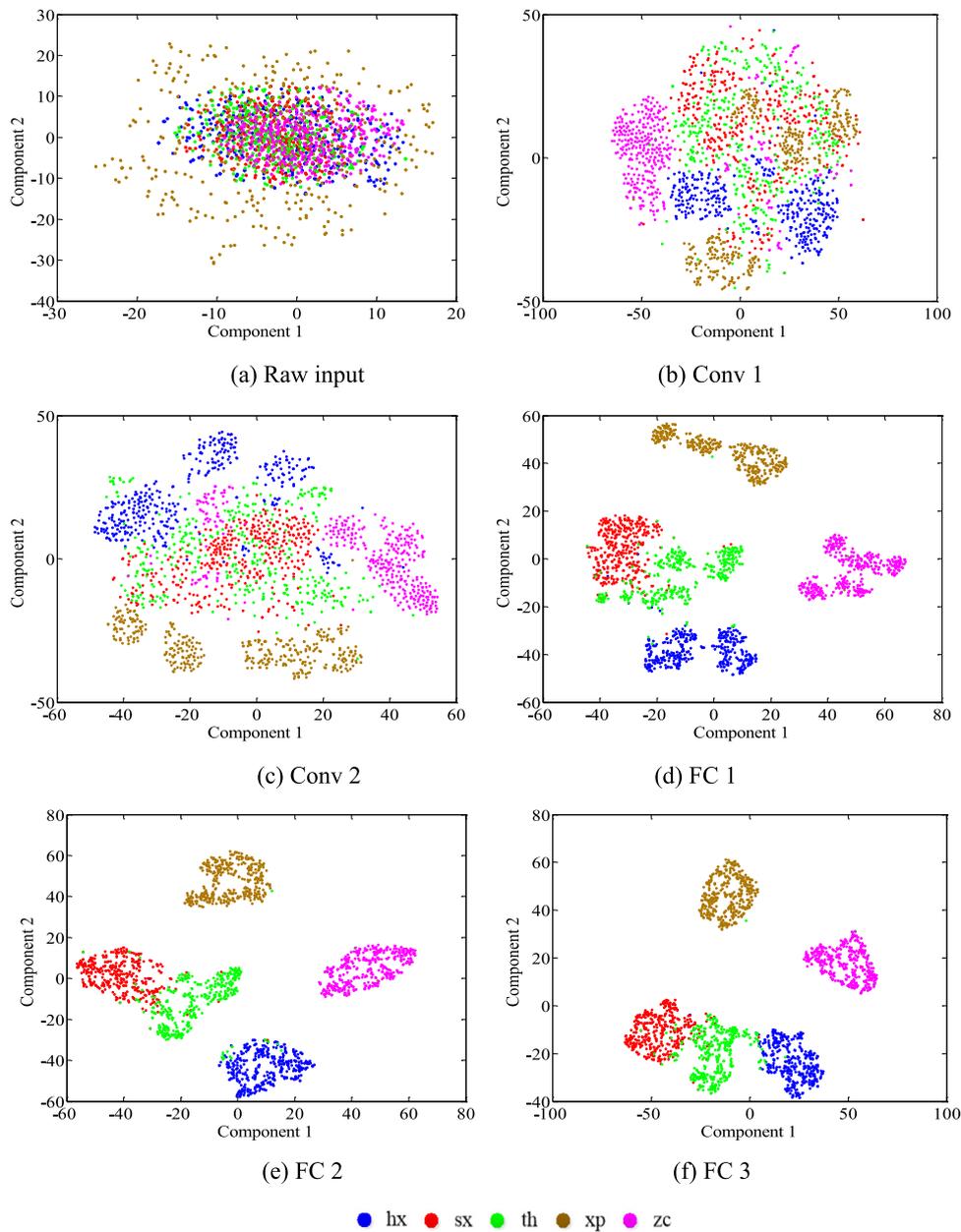


Fig. 9. Feature representations visualized by *t*-SNE.

6. Conclusions

By integrating the deep learning theory and Bayesian optimization algorithm, an improved method is constructed to intelligently realize the fault identification of hydraulic pump.

Bayesian optimization algorithm achieves the adaptive learning of model hyperparameters, and presents faster convergence and better robustness compared with PSO. The constructed CNN-BO attains the average classification accuracy of 96.94%, which increases by 2.92% compared with that of traditional LeNet 5. Moreover, it presents the remarkable improvements for loose slipper failure and central spring wear. The classification precisions of the swash plate wear and normal condition reach up to 100%. The precisions of CNN-BO for five states are higher than those of CNN-PSO, especially an increase of 6.77% is realized for central spring wear. The classification performance is further demonstrated by the visualization of *t*-SNE. Therefore, the useful

information implied in the time–frequency features of original signal can be effectively learned by CNN-BO and the different conditions of the pump are accurately distinguished.

In future work, the enhancement will be focused on the Gaussian function in the Bayesian optimization to better model the objective function. The present study is emphasized on the exploitation of single vibration signal from a hydraulic piston pump, furthermore, the research will be explored based on multiple source fused features.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The article was funded by National Key R&D Program of China (2020YFC1512402), National Natural Science Foundation of China (52175052), Open Foundation of the Key Laboratory of Fire Emergency Rescue Equipment (2020XFZB07), and the Youth Talent Development Program of Jiangsu University. We express our gratitude to Wanlu Jiang and Siyuan Liu for the support in the experiment, who are both in Yanshan University.

References

- [1] Suo X, Jiang Y, Wang W. Hydraulic axial plunger pump: gaseous and vaporous cavitation characteristics and optimization method. *Eng Appl Comput Fluid* 2021;15:712–26.
- [2] Bergada JM, Kumar S, Watton J. Axial piston pumps, new trends and development. Nova Science Publishers; 2012.
- [3] Ma Z, Wang S, Liao H, Zhang C. Engineering-driven performance degradation analysis of hydraulic piston pump based on the inverse Gaussian process. *Qual Reliab Eng Int* 2019;35:2278–96.
- [4] Guo S, Chen J, Lu Y, Wang Y, Dong H. Hydraulic piston pump in civil aircraft: current status, future directions and critical technologies. *Chin J Aeronaut* 2020;33:16–30.
- [5] Zhu Y, Li G, Wang R, Tang S, Su H, Cao K. Intelligent fault diagnosis of hydraulic piston pump combining improved lenet-5 and PSO hyperparameter optimization. *Appl Acoust* 2021;183:108336.
- [6] Ye S, Zhang J, Xu B, Zhu S, Xiang J, Tang H. Theoretical investigation of the contributions of the excitation forces to the vibration of an axial piston pump. *Mech Syst Signal Process* 2019;129:201–17.
- [7] Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi AK. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech Syst Signal Process* 2020;138:106587.
- [8] Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX. Deep learning and its applications to machine health monitoring. *Mech Syst Signal Process* 2019;115:213–37.
- [9] Wang B, Zhang X, Xing S, Sun C, Chen X. Sparse representation theory for support vector machine kernel function selection and its application in high-speed bearing fault diagnosis. *ISA Trans* 2021;118:207–18.
- [10] Zhu Y, Li G, Wang R, Tang S, Su H, Cao K. Intelligent fault diagnosis of hydraulic piston pump based on wavelet analysis and improved AlexNet. *Sensors* 2021;21:549.
- [11] Cho S, Choi M, Gao Z, Moan T. Fault detection and diagnosis of a blade pitch system in a floating wind turbine based on Kalman filters and artificial neural networks. *Renew Energ* 2021;169:1–13.
- [12] J.B.A.N. Fnaiech BCM, Fnaiech F. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl Acoust* 2015;89:16–27.
- [13] Shi Q, Zhang H. Fault diagnosis of an autonomous vehicle with an improved SVM algorithm subject to unbalanced datasets. *IEEE Trans Ind Electron* 2021;68:6248–56.
- [14] Liu J, Hu Y, Yang S. A SVM-based framework for fault detection in high-speed trains. *Measurement* 2021;172:108779.
- [15] Tang S, Yuan S, Zhu Y. Deep learning-based intelligent fault diagnosis methods toward rotating machinery. *IEEE Access* 2020;8:9335–46.
- [16] Xu Z, Li C, Yang Y. Fault diagnosis of rolling bearings using an improved multi-scale convolutional neural network with feature attention mechanism. *ISA Trans* 2021;110:379–93.
- [17] Tang S, Zhu Y, Yuan S. An improved convolutional neural network with an adaptable learning rate towards multi-signal fault diagnosis of hydraulic piston pump. *Adv Eng Inf* 2021;50:101406.
- [18] Pang S, Yang X, Zhang X, Sun Y. Fault diagnosis of rotating machinery components with deep ELM ensemble induced by real-valued output-based diversity metric. *Mech Syst Signal Process* 2021;159:107821.
- [19] Li Y, Jiang W, Zhang G, Shu L. Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data. *Renew Energy* 2021;171:103, e115.
- [20] Chen R, Huang X, Yang L, Xu X, Zhang X, Zhang Y. Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform. *Comput Ind* 2019;106:48, e59.
- [21] Chen Z, Mauricio A, Li W, Gryllias k. A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks. *Mech Syst Signal Process* 2020;140:106683.
- [22] Kumar A, Gandhi CP, Zhou Y, Kumar R, Xiang J. Improved deep convolution neural network (CNN) for the identification of defects in the centrifugal pump using acoustic images. *Appl Acoust* 2020;167:107399.
- [23] Zhang Y, Li XY, Gao L, Li P-G. A new subset based deep feature learning method for intelligent fault diagnosis of bearing. *Expert Syst Appl* 2018;110:125–42.
- [24] Shao H, Ding Z, Cheng J, Jiang H. Intelligent fault diagnosis among different rotating machines using novel stacked transfer auto-encoder optimized by PSO. *ISA Trans* 2020;105:308–19.
- [25] Xue F, Zhang W, Xue F, Li D, Xie S, Fleischer J. A novel intelligent fault diagnosis method of rolling bearing based on two-stream feature fusion convolutional neural network. *Measurement* 2021;176:109226.
- [26] Ke L, Zhang Y, Yang B, Luo Z, Liu Z. Fault diagnosis with synchrosqueezing transform and optimized deep convolutional neural network: An application in modular multilevel converters. *Neurocomputing* 2021;430:24–33.
- [27] Peng S, Huang H, Chen W, Zhang L, Fang W. More trainable inception-ResNet for face recognition. *Neurocomputing* 2020;411:9–19.
- [28] Tang S, Yuan S, Zhu Y. Data preprocessing techniques in convolutional neural network based on fault diagnosis towards rotating machinery. *IEEE Access* 2020;8:149487–149496.
- [29] Chen Z, Li C, Sanchez R-V. Gearbox fault identification and classification with convolutional neural networks. *Shock Vib* 2015;2015:1–10.
- [30] Goodfellow I, Bengio Y. A courville, deep learning. Cambridge, MA, USA: The MIT Press; 2016.
- [31] Cheng Y, Lin M, Wu J, Zhu H, Shao X. Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network. *Knowl Based Syst* 2021;216:106796.
- [32] Tang S, Yuan S, Zhu Y. Convolutional neural network in intelligent fault diagnosis toward rotatory machinery. *IEEE Access* 2020;8:86510–9.
- [33] Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. *Proc Adv Neural Inf Process Syst* 2012;2951–9.
- [34] Abbasimehr H, Paki R. Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization. *Chaos Solitons Fractals* 2021;142:110511.
- [35] Frazier PI. A tutorial on Bayesian optimization. 2018, [Online]. arXiv: 1807.02811. Available: <https://arxiv.org/abs/1807.02811>.
- [36] Peng S, Feng Q. Reinforcement learning with Gaussian processes for condition-based maintenance. *Comput Ind Eng* 2021;(158):107321.
- [37] Magnant C, Giremus A, Grivel E, Ratton L, Joseph B. Bayesian non-parametric methods for dynamic state-noise covariance matrix estimation: application to target tracking. *Signal Process* 2016;127:135–50.
- [38] Shahriari B, Swersky K, Wang Z, Adams RP, Freitas ND. Taking the human out of the loop: A review of Bayesian optimization. *Proc IEEE* 2016;104:148–75.
- [39] Wang Z, de Freitas N. Theoretical analysis of Bayesian optimization with unknown Gaussian process hyper-parameters. 2014, [Online]. Available: arXiv:1406.7758.
- [40] Cun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86:2278–324.