Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9814749, 1 page
https://doi.org/10.1155/2023/9814749

Retraction

Retracted: Data-Driven Fatigue Damage Monitoring and Early Warning Model for Bearings

Wireless Communications and Mobile Computing

Received 10 November 2022; Accepted 10 November 2022; Published 24 January 2023

Wireless Communications and Mobile Computing has retracted the article titled "Data-Driven Fatigue Damage Monitoring and Early Warning Model for Bearings" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process and the article is being retracted with the agreement of the chief editor.

References

[1] J. Hu and S. Deng, "Data-Driven Fatigue Damage Monitoring and Early Warning Model for Bearings," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 7611670, 10 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 7611670, 10 pages
https://doi.org/10.1155/2022/7611670

Research Article

Data-Driven Fatigue Damage Monitoring and Early Warning Model for Bearings

Jie Hu[1] **and Sier Deng**[1,2]

[1]*School of Mechatronics Engineering, Northwestern Polytechnical University, Xi'an 710071, China*
[2]*School of Mechatronics Engineering, Henan University of Science & Technology, Luoyang 471003, China*

Correspondence should be addressed to Jie Hu; cacajie@mail.nwpu.edu.cn

Received 26 January 2022; Revised 22 February 2022; Accepted 24 February 2022; Published 24 March 2022

Academic Editor: Rashid A Saeed

Since the manual extraction of features is not sufficient to accurately characterize the health status of rolling bearings, machine learning algorithms are gradually being used for fault diagnosis of bearings, which can adaptively learn the required features from the input data. In this paper, *k*-nearest neighbor, support vector machines, and convolutional neural networks are successfully applied to the fault diagnosis of bearings, for the benefit of achieving the detection and early warning of bearing fatigue damage. The original samples are segmented into semioverlapping samples. When using *k*-nearest neighbor and support vector machines as early warning models, we searched their hyperparameters with random search and grid search, and the results showed that support vector machines could achieve 87.1% of bearing detection accuracy and *k*-nearest neighbor could achieve 100% of detection accuracy. When convolutional neural networks are used as the early warning model, the accuracy can reach 99.75%.

1. Introduction

Today, bearings are used in a wide range of machinery and equipment in the automotive, rolling mill, mining, metallurgy, plastics, and other industries. They are the core components for load transfer, and their reliability is particularly important. When the bearing is abnormal or malfunctions, it will seriously threaten the safe operation and normal production of mechanical equipment. Bearing, as an important supporting component, is a key component in an aeroengine. The bearings in aeroengines are subjected to high temperature, high speed, and high load. The normal operation of the rolling bearings directly affects the performance and reliability of the aeroengine. If the reliability of the rolling bearing cannot be guaranteed, it will affect the performance of the engine at a slight degree and may cause a serious aviation accident such as a plane crash. At present, the failure of rolling bearings is the main reason that affects the reliability of aeroengines. Rolling bearing failures often lead to unplanned replacement and aerial parking.

Bearing life is mainly predicted according to the structural characteristics of rolling bearings themselves and their operating conditions. The life prediction methods of rolling bearings are mainly divided into three types: life prediction methods based on mechanical model, life prediction methods based on artificial intelligence, and life prediction methods based on statistical regression. The basic principle of mechanical model-based failure prediction method is based on the theory and method of bearing fatigue crack damage expansion, using mechanical model calculation to directly derive the time required for crack expansion to failure, and then predict the life. At present, the life prediction of bearings based on mechanical model is mainly divided into two categories, which are the life prediction methods based on fracture mechanics and the life prediction methods based on damage mechanics.

Machine learning is a very popular field of research in recent decades. It mainly uses intelligent algorithms and extensive training data to obtain the evolutionary laws of the data and then predicts the test results based on the

laws obtained from the training. Currently, methods like support vector machines [1, 2], neural networks [3, 4], and neurofuzzy networks [5, 6] have been applied to lifetime prediction.

In the life prediction method of statistical regression, statistical models such as Markov models [7–10], logistic regression model [11], proportional risk model [12], and stochastic filtering method [13, 14] are mainly used to analyze the test data and establish the functional relationship between the test condition and the test life. At present, typical statistical-based bearing life prediction equations include the L-P equation, I-H equation, and Zaretsky equation.

The process of bearing fault monitoring and diagnosis is to first use sensors and data acquisition equipment to collect the information used to characterize the operating status of rolling bearings. The information collected is then processed to extract the characteristic parameters that can characterize the operation of the bearings; the characteristic parameters are evaluated with the help of specific discriminative modes, criteria and diagnostic methods to determine whether there is a fault in the equipment and the location and degree of the fault; finally, the results obtained from the state identification are used to predict the possible development trend of the operating state of the bearings.

Bearings are the basic components of the transmission system of rotating machinery, which operates at high speed for a long time; the inner ring, outer ring, and rolling elements of bearing components are prone to failure, which directly affects the safety of the whole equipment and even the whole production line. Therefore, the monitoring and diagnosis of bearings bearing monitoring and diagnosis become a critical task. Vibration signal-based analysis is the most common traditional fault diagnosis method, including wavelet analysis [15], empirical modal analysis [16], and local mean decomposition [17]. However, the feature extraction method based on vibration signal processing leads to poor monitoring and diagnosis capability and poor generalization performance in the face of alternating multiple operating conditions, severe coupling of fault information, and unknown and variable modes. Today, the application of deep learning model CNN on bearing fault diagnosis has developed rapidly, which has significantly improved the fault identification accuracy [18].

After a bearing failure, some of the time domain features of the vibration signal change with the evolution of the failure, and some of the time domain features can be used to determine whether a failure has occurred. For example, the root mean square value can be used to diagnose wear-type faults, kurtosis to diagnose shock-type faults, and cliffness to diagnose early faults. For example, the root mean square value can be used to diagnose wear-related faults, kurtosis to diagnose impact-related faults, and cliffness to diagnose early faults. Although these features are simple to calculate and intuitive, the specific location of the fault cannot be determined based on these features, and the stability and sensitivity of the features cannot be guaranteed at the same time, and sensitivity cannot be guaranteed at the same time. The frequency domain signal is the signal obtained by Fourier transform of the time domain signal. Although the resonance demodulation method is widely used in the engineering field is a frequency-domain method and because the frequency-domain method does not consider the fault signal characteristics from the time domain perspective, the method can not be used to analyze the time-varying characteristics of the nonstationary signal which cannot be taken into account when using this method. The time varying characteristics of the nonstationary signal can not be taken into account when using this method [9]. In order to consider both the time domain characteristics and frequency domain characteristics of the nonstationary signal characteristics, the time-frequency characteristics of the signal can be used to analyze such signals, and a popular method for time-frequency analyses is the short-time Fourier method.

From the 1950s, the development of AI went through a "reasoning phase", in which machines were given the ability to reason logically so that they could gain intelligence, and AI can prove a number of famous mathematical properties, but they are still far from being truly intelligent because of the lack of understanding of the machines. Artificial intelligence is a new technical science studies and develops theories, methods, techniques, and applied systems for simulating, extending, and expanding human intelligence. It was first proposed by John McCarthy in 1956 and was defined as "the science and engineering of making intelligent machines." The purpose of artificial intelligence is to enable machines to think like humans and to give them intelligence. Today, the meaning of AI has been greatly expanded and is an interdisciplinary discipline. Therefore, the development of AI entered the "knowledge phase," where the knowledge of humans is concluded and imparted to make machines smart. In this time, a great variety of expert systems were introduced and a great number of results were achieved in many areas, but because of the enormous amount of human know-how, a "knowledge engineering bottleneck" emerged. It is worth noting that the goal of machine learning is to make the learned functions work well for "new samples," not just good performance on training samples. The capability to apply learned functions to new samples is termed generalization capability. In recent years, machine learning algorithms are increasingly used in the fields of voice recognition, image recognition, and NLP.

Support vector machines are a wide class of linear classifiers that perform binary classification of data in a surveillance based learning manner and solve decision bounds for the maximum marginal hyperplane of the sample under study. SVM is a sparse and robust classifier that uses a hinge loss function to compute empirical risk and adds a regularization term to the solution system to optimize structural risk. It is one of the common kernel learning methods. More and more scholars are using SVM for fatigue damage detection in bearings [19–24]. It is proved by the above scholars that the SVM algorithm can have an exciting effect on the detection of fatigue damage in bearings. The K-nearest neighbor algorithm is one of the most basic and simple machine learning algorithms. The KNN (K-nearest neighbor) method, or K-nearest neighbor method, originally proposed by Cover and Hart in 1968, is a theoretically mature method and one of the simplest machine learning

algorithms. The idea of the method is very simple and intuitive: if most of the K most similar (i.e., most neighboring) samples in the feature space belong to a certain class, then the sample also belongs to that class. The method determines the category of the sample to be classified based on the category of the nearest sample or samples. It can be used for both classification and regression. KNN performs classification by measuring the distance between different feature values. The method has a very simple and intuitive idea: if a sample falls into a category, it also belongs to that category if the majority of the K most similar samples in the characteristic domain fall into that category. The KNN algorithm is a very special kind of machine learning operation, since it does not have a learning process in the usual sense. A number of scholars have used the KNN algorithm to diagnose the faults of bearings [25–27].

Three machine learning algorithms, SVM, KNN, and convolutional neural network, are used to classify the bearing damage data. Among them, we search the hyperparameters in SVM and KNN by two search methods, random search and grid search. Judgment and warning of bearing damage through the output of the model.

Five sections are included in this article. The current state of research and the background of development of fatigue damage are described in Section 1. The flow of work in this paper is shown in Section 2. The theory and methods involved in this paper are described in Section 3. The specific experimental procedures and results are presented in Section 4. Conclusions and discussion are described in Section 5.

2. Workflow

To evaluate the fatigue damage of the bearings, we first collect the bearing dataset and then go through the established machine learning evaluator. Here, we explore the hyperparameters in both SVM and KNN machine learning methods using both random search and grid search to finally evaluate and predict the damage of the bearings. The flow is shown in Figure 1.

2.1. Support Vector Machine (SVM). Schlag et al. proposed SVM [28]. Its main core idea is an important model based on statistical theory, which can be described as a derivation on the basis of VC dimensional theory and structural risk minimization. Support vector machines have the following features:

(1) Modularity

(2) The principle of edge maximization

(3) The ability to solve problems with small samples

(4) The ability to handle higher dimensional data

(5) The ability to handle nonlinear problems by introducing penalty coefficients and kernel functions

The basic idea of support vector machine can be described by Figures 2. As shown in Figure 2, $L_1, L_2,$ and L_3 represent the optimal classification surface and its upper and lower boundaries, respectively; L_2 and L_3 are determined by the distance of points from the optimal

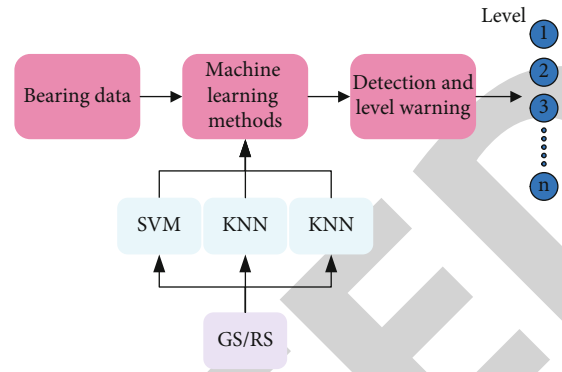

Figure 1: The workflow of this paper.

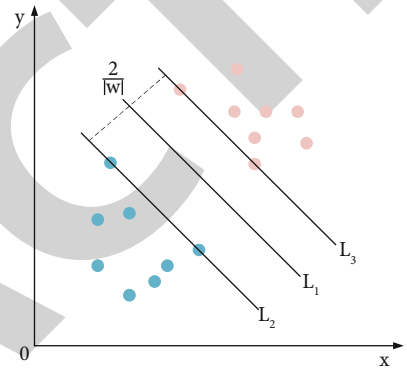

Figure 2: Schematic diagram of the basic idea of support vector machines.

classification surface in the 2-class sample data, and the distance between $L_1, L_2,$ and L_3 is $1/\|w\|$.

The relationship between the distance between the boundary and the classification surface and the number of errors can be expressed as follows:

$$\text{The number of errors} \le \left(\frac{2M}{\rho}\right)^2, \tag{1}$$

where ρ denotes the distance of all samples to the classification plane and M denotes the maximum value of the sample paradigm.

Seeking the maximum ρ becomes the optimization objective; the distance between the boundary and the classification surface with $\|w\|$ is given by

$$\rho = \frac{1}{\|w\|}|g(x)|, \tag{2}$$

where $h(x)$ is called the classification surface, in the satisfaction of the classification surface can distinguish the two types of data and under the classification of this classification surface can have the maximum distance, so that is the optimal classification surface. According to equation (2), it can be derived that ρ is maximum in the case of $\|w\|$ is minimum, and at this time, the obtained ρ is the maximum distance.

Table 1: Structural parameters of the CNN model.

Name of the layer	Parameters
Convolution layer I	Number of convolution kernels: 6; size: 5×5
Pooling layer II	Pooling method: maximum pooling; pooling domain size: 2×2; step size: 2
Convolution layer I	Number of convolution kernels: 16; size: 3×3
Pooling layer II	Pooling method: maximum pooling; pooling domain size: 2×2; step size: 2
Convolution layer III	Number of convolution kernels: 120; size: 3×3
Fully connected layer	Number of neurons: 84

When the selected 2 sample data are in the case of intermingling, the analysis shows that the error is not allowed in the search for the optimal classification surface. However, the introduction of the relaxation factor ξ can still apply in the case of errors. The classification function for separating the two types of training data is $f(x) = (w \cdot x) + b$, and the classification surface is $(w \cdot x) + b = 0$. The classification surface at this time should satisfy the following equation:

$$1 - \zeta_i \le y_i[(w \cdot x) + b] \quad (i = 1, \cdots, n). \tag{3}$$

These restrictions are listed below.

$$\begin{cases} (w, x_i) + b > +1 - \zeta_i, y_i = +1, \\ (w, x_i) + b > -1 - \zeta_i, y_i = -1, \\ 0 \le \zeta_i, \quad \forall i. \end{cases} \tag{4}$$

2.2. K-Nearest Neighbor (KNN). In the feature space, the similarity between two points is reflected by the distance. The closer the distance is, the more similar the two points are; the farther the distance is, the less similar the two points are. The KNN algorithm [29], as a supervised algorithm, also classifies samples by means of distance in essence. The specific mathematical description is as follows:

Assume that the training sample $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$. The distance between samples x_i and x_j is denoted by d:

$$d(x_i, y_i) = \cos\theta = \frac{x_i, y_i}{\|x_i\| \cdot \|y_i\|}, \tag{5}$$

where $d(x_i, x_j)$ denotes the distance of the samples and θ denotes the directional angle between the sample vectors.

The samples are filtered according to the calculation result of equation (5), and the top k samples are selected as the new neighborhood $N_k(x)$, based on the majority voting criterion, to decide whether the sample x in $N_k(x)$ belongs to y.

$$y = \arg_{Cj} \max \sum_{x_i \in N_k(x)} I, \tag{6}$$

where I denotes the indicator function and $I = 1$ when and only when $y_i = c_j$; otherwise, $I = 0$.

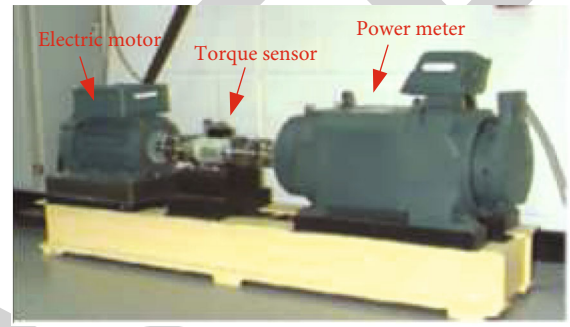

Figure 3: CWRU test bench.

2.3. Convolutional Neural Network. The characteristics of CNN [30] such as local connectivity and shared weights of convolutional layers reduce the complexity of the network model and reduce overfitting, and the model has better generalization ability. CNN includes pooling layer, activation layer, convolutional layer, softmax classification layer, and fully connected layer, and layer 1 uses a larger convolutional kernel. The structural parameters [31] are shown in Table 1.

The convolutional layer of CNN extracts different features of the input image by convolutional operations. The convolutional kernel is moved over the input feature map, and the corresponding points in the overlapping regions are multiplied and summed up, and the output feature map value is obtained after adding bias. The activation layer transforms the feature map after the convolution operation by the activation function to enhance the nonlinear expression of the model, which greatly improves the generalization ability and classification performance of the model. The main purpose of the pooling layer is to perform local feature extraction, accelerate convergence, and establish spatial and structural invariance.

2.4. Grid Search (GS) and Random Search (RS). The grid search method is an exhaustive method of searching for a given parameter value, and the parameters of the estimation function are optimized by the cross-validation method to obtain the best learning algorithm [32]. It performs a permutation of the possible values of each parameter and lists all possible combinations to generate a "grid" [33].

Random search [34] is analogous to grid search, but instead of searching all values in its search space, RS selects samples between the top and bottom bounds of the search

Figure 4: Data overlap sampling.

Table 2: Sample-specific composition information.

Type of sample	Length of sample	Number of sample	Type label
Normal	1000	100	0
Inner ring damage (7 mils)	1000	100	1
Outer ring damage (7 mils)	1000	100	2
Rolling body damage (7 mils)	1000	100	3
Inner ring damage (14 mils)	1000	100	4
Outer ring damage (14 mils)	1000	100	5
Rolling body damage (14 mils)	1000	100	6
Inner ring damage (21 mils)	1000	100	7
Outer ring damage (21 mils)	1000	100	8
Rolling body damage (21 mils)	1000	100	9

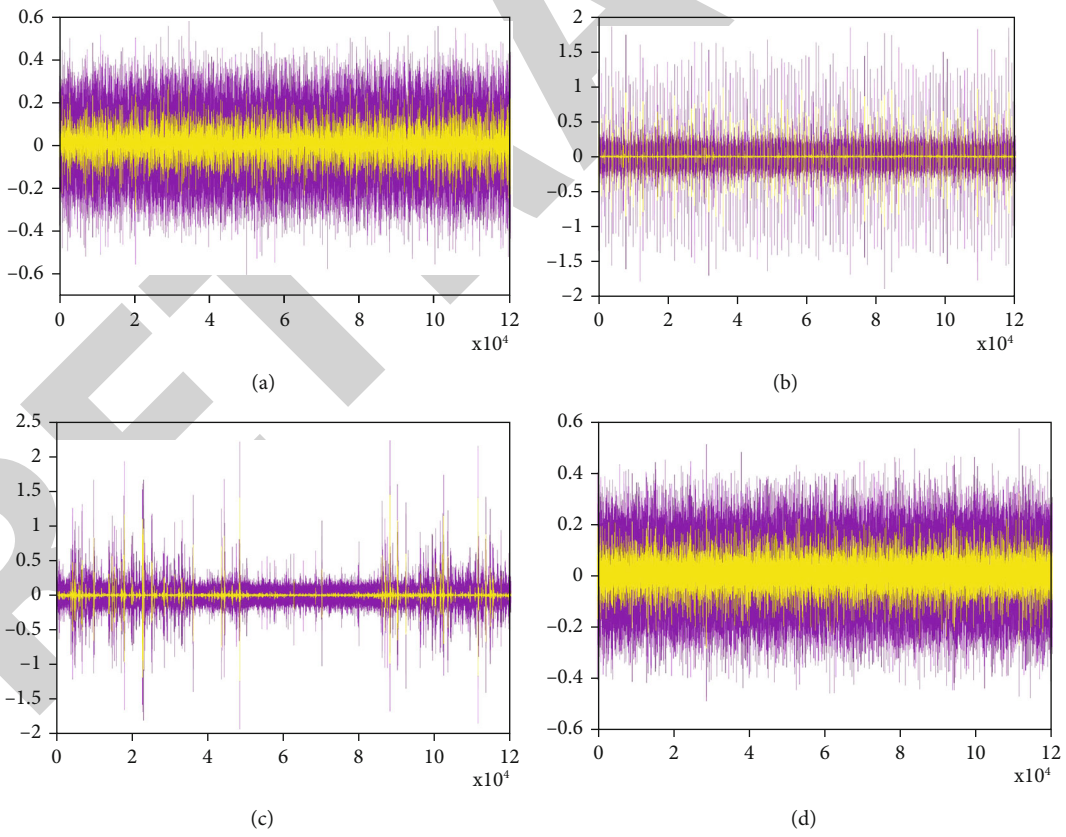

Figure 5: Wavelet denoising effect comparison before and after. (a)–(d) denote the signals of four different depth defects.

space. The primary advantage of RS is that it can be easily parallelized and resource allocated because each evaluation is performed independently. Random search is a method of using random numbers to find the optimal solution of a function approximation, which is different from the violent search method of grid search. RS is an optimization method to perform an unacceptable time-consuming procedure with huge data size.

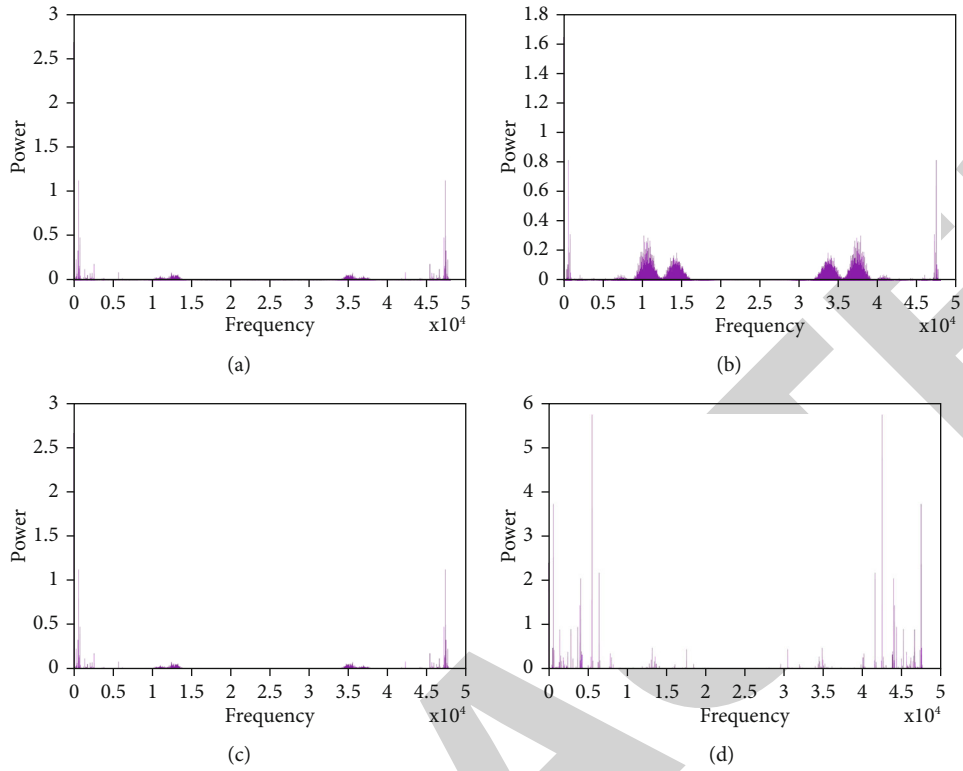

Figure 6: Spectrogram of the denoised signal.

Table 3: The best SVM model using GS.

Accuracy	C	Kernel function
0.871429	14.96281926	Linear
0.857143	4.928685152	Linear
0.857143	9.653530095	RBF
0.857143	23.53466656	Linear
0.857143	7.521708519	Linear

Table 4: The best SVM model using RS.

Accuracy	C	Kernel function
0.871429	13.48907516	Linear
0.871429	10.56731338	Linear
0.871429	3.133206637	RBF
0.871429	11.00524368	Linear
0.857143	11.56794508	Linear

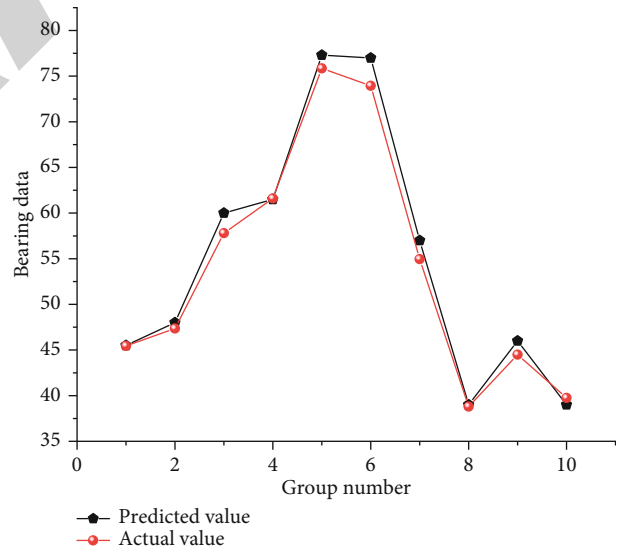

Figure 7: A comparison of predicted and true data.

3. Experimental Verification

3.1. Data Preparation. In order to verify feasibility of the model, a sample of the open bearing database from Case Western Reserve University was selected for experimental validation, and the experimental rig is shown in Figure 3. The experimental rig consists of a 2 hp motor (left side), a torque sensor (middle), a power meter (right side), and electronic control equipment. The experimental rig consists of ing the EDM technique with fault diameters of 7, 14, and 21 mils. For the experiments, acceleration sensors were mounted on the drive side and fan side of the motor housing to collect vibration signals from the faulty bearing.

The number of data samples is playing an essential part in the generalization ability of the model, but the failure data of rolling bearings are comparatively small and the number of samples needs to be increased. The overlap sampling method can maintain the periodicity and time-varying nature of the vibration signal, so the data samples are increased using the overlap sampling method, and the principle is shown in Figure 4.

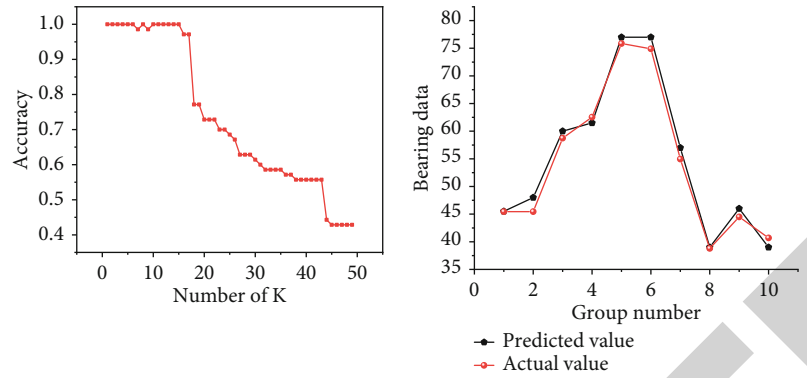

Figure 8: The classification results and comparison of predicted and true data. (a) denotes the classification results; (b) denotes the comparison of predicted and true data.

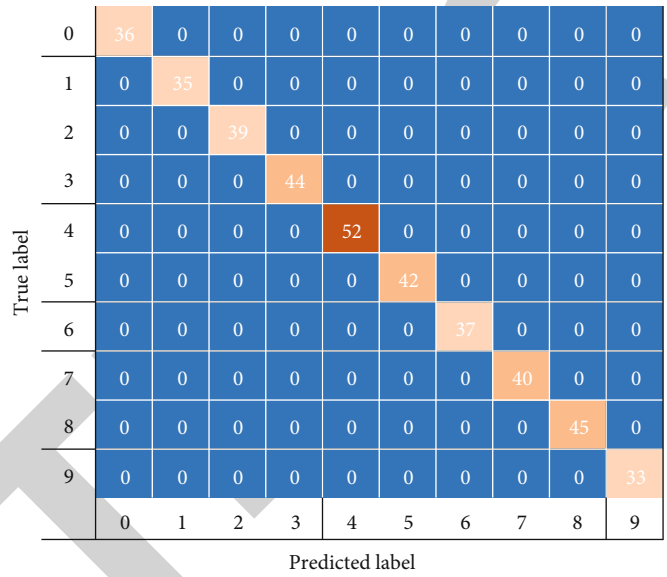

Figure 9: Test set recognition results.

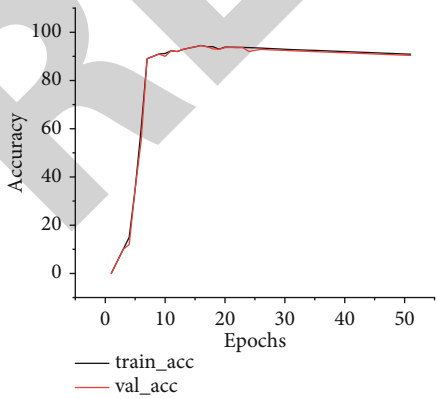

Figure 10: Accuracy curve of model training and verification.

Table 5: Recognition accuracy of different models (%).

Model	Accuracy
DNN	95.46
DBN	95.78
CNN-4	95.43
CNN	99.75

Ten bearing states with sampling frequency of 12 kHz, load of 2 hp, and fault diameters of 7 mils, 14 mils, and 21 mils were selected for the experiments, and the specific sample composition information is shown in Table 2. We can see from it that the experimental data consisted of 1000 samples with a sample length of 1000. 1999 subsamples were formed by a semioverlapping sample cut before training and divided into 1280 training subsamples, 320 validation subsamples, and 399 test subsamples. First, we

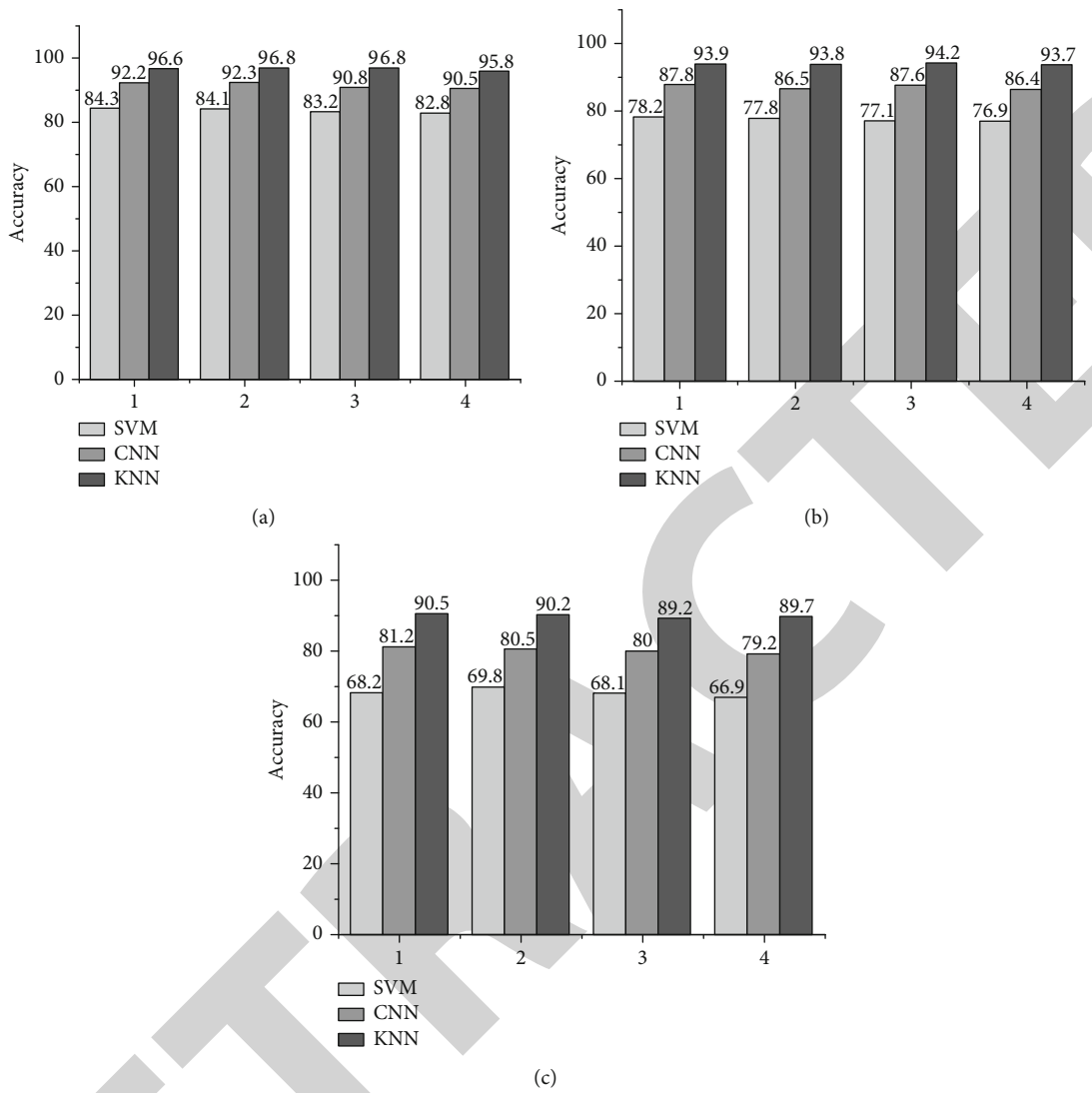

Figure 11: Comparison of SVM, CNN, and KNN accuracies when SNR is -3,-6, and -9; (a) denotes a signal-to-noise ratio of -6; (b) denotes a signal-to-noise ratio of -6; (c) denotes a signal-to-noise ratio of -9.

perform wavelet denoising of the data. Figure 5 shows the comparison of the effect after wavelet denoising. The denoised spectrum is displayed in Figure 6.

4. Experimental Results and Analysis

4.1. Validation Experiments with SVM for Hyperparameter Search. In the support vector machine, the penalty coefficient C and the kernel function type $f(x)$ are commonly used. We preset the penalty coefficients within (0~50) and use linear kernel, radial basis function (RBF), polynomial kernel, and sigmoid kernel as kernel functions. We let the iteration number be set to 20 and use GS and RS methods to search for kernel functions and penalty coefficients.

In total, we searched for 40 combinations. The search is first performed by the GS method for hyperparameters in SVM. The top ten best combinations are shown in Table 3; we can see that the highest accuracy can reach 0.87 when using the SVM classifier, and the corresponding kernel function is linear with penalty factor $C = 14.96281926$. From the table, we can also see that the linear kernel performs the best.

The RS method is also used to find the hyperparameters of the support vector machine at the same time. The parameters are set the same as before. Table 4 shows the top ten optimal models. The maximum accuracy of SVM classification under RS search condition reaches 87.1%, and the corresponding function was a linear kernel function with penalty factor $C = 13.48907516$. From Tables 3 and 4, we can see that the classification accuracies obtained using different search methods are similar due to the decision of the performance of the SVM itself. Figure 7 represents the comparison of the predicted values with the true values after using SVM.

4.2. Validation Experiments with KNN for Hyperparameter Search. We studied the effect of KNN algorithm in defective deep classification by is. K is used as the hyperparameter to be fined. Again, we used the two search algorithms

mentioned above and preset the range of K within (1 to 50). Because their classification accuracy is fixed, only the time to search for the same combination will differ. Figure 8(a) shows the results when different characteristics are used as input. Figure 8(b) represents the comparison of the predicted values with the true values after using KNN.

4.3. Validation Experiments with CNN. The confusion matrix provides a clearer picture of the identification of the rolling bearing condition in the test set, as shown in Figure 9. The confusion matrix shows that 1 of the 45 samples of rolling element damage with a failure diameter of 7 mils in the test set was misidentified as a failure diameter of 7 mils in the outer ring. The remaining nine rolling bearing states are correctly classified. This shows that the proposed model has a high recognition rate of bearing faults.

The accuracy curves of the training and validation sets after training are shown in Figure 10. The accuracy of the training set converges to 1 as the epochs increase during the training of the model; the accuracy of the validation set fluctuates during the training process, but the overall convergence trend is good. This indicates that the model can better learn the feature relationship between vibration signals during the training process.

In order to verify the recognition performance of CNN models with other fault diagnosis models, deep neural networks (DNN), deep confidence networks (DBN), and convolutional neural networks with four layers (CNN-4) are selected for comparison tests. Among them, DNN and DBN are composed of input layer, two hidden layers, and output layer with the number of neurons of 1000, 100, 100, and 10, respectively. CNN-4 is composed of four convolutional layers, four maximum pooling layers, and fully connected layers. The first convolutional kernel has a size of 64×1 and a move step of 16, the remaining convolutional kernels have a size of 3×1 and a move step of 1, and the maximum pooling layer has a size of 2×1 and a move step of 2. A fully connected layer consists of 32 neurons. Softmax classifier was used for all comparison models. Table 5 shows the comparison results.

From the table, it can be seen that the DNN and DBN models have poor integrated recognition of the bearing state. Compared with the DNN and DBN models, the CNN-4 model has a greater advantage in diagnostic accuracy and stability than the DNN and DBN models. The reason is that it retains important parameters through local connectivity and weight sharing, which reduces a large number of weights involved in training computation, and achieves good learning effect. However, compared to DBNs, CNNs stack multiple residual modules together and perform well. The CNN model achieves a combined recognition rate of up to 99.75% for bearing states by stacking multiple residual modules on top of each other and performing residual and convolution operations on them. Therefore, the features learned by the CNN model have good classification characteristics and can be well used in the fault diagnosis of bearings.

By addition of Gaussian white noise with various signal-to-noise ratios to the original signal, the resulting mixed signal is used as the input to the model as a way to verify the prediction accuracy of the model for noise-containing data. Figures 11 shows the accuracy comparison of the SVM, CNN, and KNN models on the data sets A, B, C, and D under four working conditions when the signal-to-noise ratios are -3, -6, and -9 dB, respectively.

As can be seen from the figure, CNN and KNN models can achieve more than 75% recognition accuracy in all cases, which still has a significant advantage over SVM. When the signal-to-noise ratio is high, both CNN and KNN models can achieve more than 90% accuracy, and as the signal-to-noise ratio decreases (the power of the noise increases), recognition accuracy decreases to different degrees.

5. Results

For the shortcomings of traditional fault diagnosis methods such as low diagnosis efficiency and long time-consuming, SVM, KNN, and CNN are successfully applied to the fault diagnosis of rolling bearings, so as to achieve the detection and early warning of bearing fatigue damage. Firstly, the original samples are segmented by semioverlapping samples, secondly, the segmented subsamples are divided into training subsamples, validation subsamples and test subsamples, and finally, the fault classification of rolling bearings is achieved by SVM, KNN, and CNN. When the early warning models of SVM and KNN were used, we searched their hyperparameters with random search and grid search, and the results showed that SVM could achieve 87.1% of bearing detection accuracy; KNN could achieve 100% of detection accuracy. When CNN is used as the early warning model, the accuracy can reach 99.75%.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of 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.

References

[1] M. Tukan, C. Baykal, D. Feldman, and D. Rus, "On coresets for support vector machines," *Theoretical Computer Science*, vol. 890, pp. 171–191, 2021.

[2] P. R. M. Júnior, T. E. Boult, J. Wainer, and A. Rocha, "Open-set support vector machines," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 125, pp. 1–14, 2021.

[3] W. Yan, H. Cao, and Z. Cui, "Tibetan text classification based on RNN," *Journal of Physics: Conference Series*, vol. 1848, no. 1, p. 12139, 2021.

[4] Y. Hu, Y. Wang, and H. Wang, "A decoding method based on RNN for OvTDM," *China Communications*, vol. 17, no. 4, pp. 1–10, 2020.

[5] S. Sinha and A. Paul, "Neuro-fuzzy based intrusion detection system for wireless sensor network," *Wireless Personal Communications*, vol. 114, no. 1, pp. 835–851, 2020.

[6] P. V. de Campos Souza, "Fuzzy neural networks and neuro-fuzzy networks: a review the main techniques and applications used in the literature," *Applied Soft Computing*, vol. 92, p. 106275, 2020.

[7] O. Danisman and U. Uzunoglu Kocer, "Hidden Markov models with binary dependence," *Physica A: Statistical Mechanics and its Applications*, vol. 567, p. 125668, 2021.

[8] I. L. MacDonald, "A coarse-grained Markov chain is a hidden Markov model," *Physica A: Statistical Mechanics and its Applications*, vol. 541, p. 123661, 2020.

[9] J. K. Grewal, M. Krzywinski, and N. Altman, "Markov models – hidden Markov models," *Nature Methods*, vol. 16, no. 9, pp. 795-796, 2019.

[10] C. Puerto-Santana, P. Larranaga, and C. Bielza, "Autoregressive asymmetric linear Gaussian hidden Markov models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, pp. 1–1, 2021.

[11] E. C. Zabor, C. A. Reddy, R. D. Tendulkar, and S. Patil, "Logistic regression in clinical studies," *International Journal of Radiation Oncology ∗ Biology ∗ Physics*, vol. 112, no. 2, pp. 271–277, 2022.

[12] P. P. Rejani and P. G. Sankaran, "Modeling and analysis of proportional hazards competing risks cure rate model," *Journal of the Indian Society for Probability and Statistics*, vol. 21, no. 1, pp. 175–185, 2020.

[13] M. E. Orchard and G. J. Vachtsevanos, "A particle filtering-based framework for real-time fault diagnosis and failure prognosis in a turbine engine," in *Conference on Control & Automation*, 2008.

[14] J. Wang and R. X. Gao, *Multiple Model Particle Filtering for Bearing Life Prognosis*, Prognostics & Health Management, 2013.

[15] J. Chen, Z. Li, J. Pan et al., "Wavelet transform based on inner product in fault diagnosis of rotating machinery: a review," *Mechanical Systems and Signal Processing*, vol. 70-71, pp. 1–35, 2016.

[16] M. Van, H. J. Kang, and K. S. Shin, "Rolling element bearing fault diagnosis based on non-local means de-noising and empirical mode decomposition," *Science Measurement & Technology Iet*, vol. 8, no. 6, pp. 571–578, 2014.

[17] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *Journal of the Royal Society Interface*, vol. 2, no. 5, pp. 443–454, 2005.

[18] Z. Wei, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 3, p. 425, 2017.

[19] J. D. S. Souza, M. V. L. D. Santos, R. S. Bayma, and A. L. A. Mesquita, "Analysis of window size and statistical features for SVM-based fault diagnosis in bearings," *IEEE Latin America Transactions*, vol. 19, no. 2, pp. 243–249, 2021.

[20] U. Parmar and D. H. Pandya, "Experimental investigation of cylindrical bearing fault diagnosis with SVM," *Materials Today: Proceedings*, vol. 44, pp. 1286–1290, 2021.

[21] X. Jia, B. Xiao, Z. Zhao, L. Ma, and N. Wang, "Bearing fault diagnosis method based on CNN-LightGBM," in *IOP Conference Series: Materials Science and Engineering*, vol. 1043no. 2-IOP Publishing.

[22] C. He, T. Wu, R. Gu, and H. Qu, "Bearing fault diagnosis based on wavelet packet energy spectrum and SVM," *Journal of Physics: Conference Series*, vol. 1684, no. 1, 2020.

[23] B. A. Qiu, P. Sun, and L. Li, "An intelligent bearing fault diagnosis method based on SF-SVM," in *IOP Conference Series: Materials Science and Engineering*, vol. 1210, 2021no. 1.

[24] X. Li and L. Fu, "Rolling bearing fault diagnosis based on MEEMD sample entropy and SSA-SVM," *Journal of Physics: Conference Series*, vol. 2125, no. 1, p. 12003, 2021.

[25] S. W. Fei, "The hybrid method of VMD-PSR-SVD and improved binary PSO-KNN for fault diagnosis of bearing," *SHOCK AND VIBRATION*, vol. 2019, Article ID 4954920, 7 pages, 2019.

[26] M. Peña, M. Cerrada, X. Alvarez et al., "Feature engineering based on ANOVA, cluster validity assessment and KNN for fault diagnosis in bearings," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 6, pp. 3451–3462, 2018.

[27] R. V. Sanchez, P. Lucero, R. E. Vasquez, M. Cerrada, J. C. Macancela, and D. Cabrera, "Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and KNN," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 6, pp. 3463–3473, 2018.

[28] S. Schlag, M. Schmitt, and C. Schulz, "Faster support vector machines," *Journal of Experimental Algorithmics (JEA)*, vol. 26, pp. 1–21, 2021.

[29] S. Zhang, "Challenges in KNN classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 1457, p. 1, 2021.

[30] D. Moolchandani, A. Kumar, and S. R. Sarangi, "Accelerating CNN Inference on ASICs: a survey," *Journal of Systems Architecture*, vol. 113, p. 101887, 2021.

[31] Y. Zhang, K. Xing, R. Bai, D. Sun, and Z. Meng, "An enhanced convolutional neural network for bearing fault diagnosis based on time–frequency image," *Measurement*, vol. 157, p. 107667, 2020.

[32] G. Zhang, R. Pan, J. Zhou, L. Wang, and W. Gao, "Spectrophotometric color matching for pre-colored fiber blends based on a hybrid of least squares and grid search method," *Textile Research Journal*, vol. 1256, 2021.

[33] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.

[34] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of machine learning research*, vol. 13, 2012.