Machine Learning Based Bearing Fault Diagnosis Using the Case Western Reserve University Data: A Review

XIAO ZHANG[ID]**1, BOYANG ZHAO**[ID]**2, (Graduate Student Member, IEEE),
AND YUN LIN**[ID]**3, (Member, IEEE)**

[1]Department of Computer Science and Technology, Shandong Technology and Business University, Yantai 264010, China
[2]School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
[3]College of Information and Communication Engineering, Harbin Engineering University, Harbin 150000, China

Corresponding author: Boyang Zhao (zby21@buaa.edu.cn)

ABSTRACT The most important parts of rotating machinery are the rolling bearings. Finding bearing faults in time can avoid affecting the operation of the entire equipment. The data-driven fault diagnosis technology of bearings has recently become a research hotspot, and the starting point of research is often the acquisition of vibration signals. There are many public data sets for rolling bearings. Among them, the most widely used public dataset is Case Western Reserve University bearing center (CWRU). This paper will start from the CWRU data set, compare and analyze some basic methods of machine learning based rolling bearing fault diagnosis, and summarize the characteristics of CWRU. First, we give a comprehensive introduction to CWRU and summarize the results achieved. After that, the basic methods and principles of machine learning based rolling bearing fault diagnosis were summarized. Finally, we conduct experiments and analyze experimental results. This paper will have certain guiding significance for the future use of CWRU for machine learning based rolling bearing fault diagnosis.

INDEX TERMS Fault diagnosis, machine learning, feature selection, classifier, CWRU.

I. INTRODUCTION

The most important parts of rotating machinery are the rolling bearings, and with the development and demand of industry, the workload of most rotating machinery is huge. Under high load, strong impact, high workload and complex environment, rolling bearings often produce faults in inner race, outer race and ball. If the fault cannot be found in time, the equipment will often shut down, resulting in huge economic losses and even safety accidents. Fault diagnosis and prediction is the core of Prognostics and health management (PHM). The main purpose of PHM in rotating machinery is to reduce use and support costs, improve the safety and integrity of rotating machinery, so as to achieve condition-based maintenance with less maintenance investment [1]. Fault diagnosis methods can be divided into three types, including model-based methods, statistical reliability-based methods, and data-driven methods [2]. The premise of fault diagnosis based on model is to know the mathematical model of the target system. This kind of diagnosis method can

penetrate into the essential performance of the target system and realize the implementation of fault prediction. But for complex dynamic systems, it is difficult to establish a mathematical model with high reliability, and the related work of fault diagnosis is greatly restricted. Information needed by fault diagnosis techniques based on statistical reliability can be found in various probability density functions (PDFs). The bathtub curve is a good example. The purpose of fault diagnosis can be achieved by processing the data collected by the sensor and combining methods such as feature engineering, deep learning or machine learning [3]. Rolling bearing fault diagnosis based on data-driven is the focus of current research, the core issue is to obtain the data. The state of the rolling bearing and the vibration signal are complementary. When the rolling bearing fails, it is often accompanied by an impulse signal [58]. Sensors are installed at different positions of the rolling bearing to collect vibration signals. The state of the bearing can be judged by directly observing these signals.

Machine learning is to extract knowledge from data, use a given sample for learning, and automatically determine and recognize the subsequent input samples, and give the

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojun Steven Li[ID].
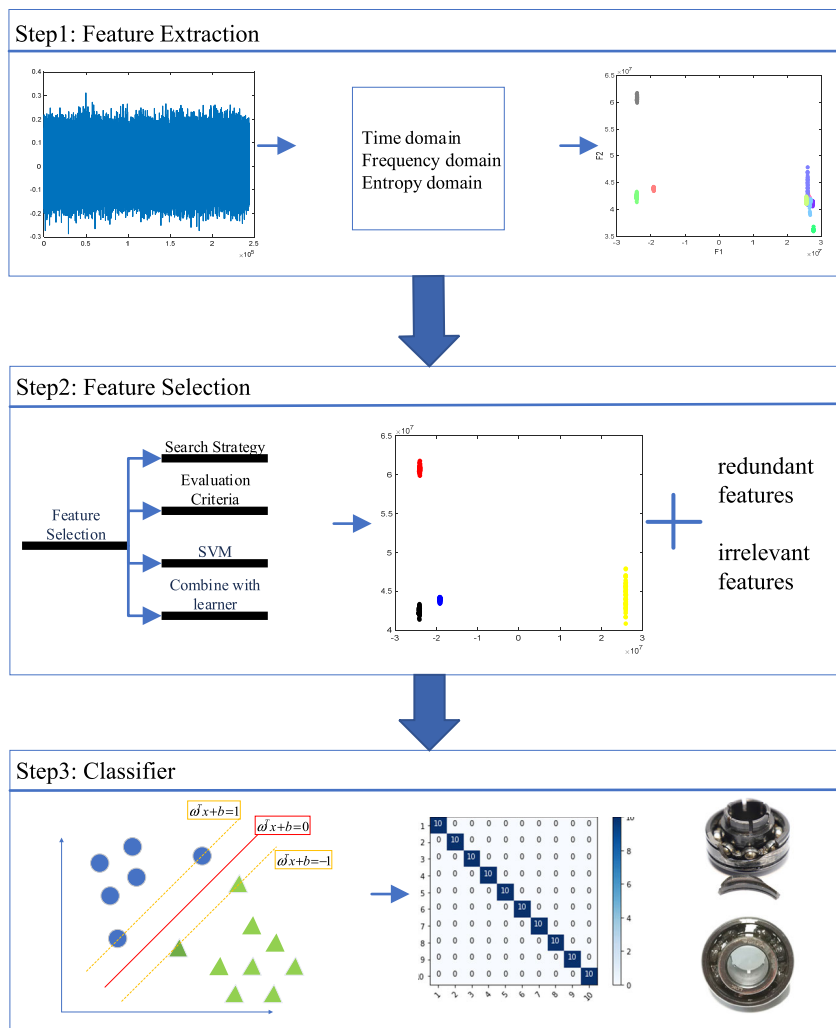
FIGURE 1. The flowchart of machine learning.

predicted results of the input samples. Machine learning can finally achieve similar outputs based on similar inputs.

It can be seen from Figure 1 that the basic process of machine learning can usually be divided into 3 parts, including feature extraction, feature selection, and classification. In the first step, multi-domain features are extracted from different domains by processing the vibration signal. In the second step, the dimensionality of the feature set is reduced according to different criteria. Extract the best feature subset with the best discriminability and the least quantity, so as to avoid the disaster of dimensionality, improve the classification accuracy, and reduce the classification time. And the last step is the classifier stage, which classifies the sample by inputting a subset of features into the classifier [4], [5].

This paper will start from the CWRU data set, compare and analyze some basic methods of machine learning based rolling bearing fault diagnosis, and summarize the characteristics of the CWRU data set. the introduction is in Section I. Section II investigates the rolling bearing dataset and discusses the characteristics of the CWRU data set. Section III

presents a review of feature extraction. Section IV summarizes feature selection. Section V summarizes the fault diagnosis classifier. Section VI is a brief review of deep learning, and discusses the existing contributions of CNN and RNN models in the field of faults. Section VII is an experiment, which compares and analyzes the algorithms mentioned in the previous parts from multiple dimensions, and discusses the results. Section VIII is the conclusion.

II. REVIEW OF DATASET
The characteristics of the mainstream rolling bearing public datasets are shown in Table 1. The datasets used to predict the remaining useful life of rolling bearings include FEMTO-ST bearing degradation data set [6], [7], IMS center bearing degradation dataset [8]–[10], and XJTU-SY bearing acceleration degradation dataset [11]. The datasets used for fault diagnosis include CWRU [12], [13], the MFPT dataset of the American Society for Mechanical Failure Prevention Technology [14], and the Paderborn bearing dataset of Paderborn University [15]. In addition, some papers also use full life

TABLE 1. The rolling bear datasets.

Literature	Dataset	Fault method	Sampling rate	Fault type
[4-5]	FEMTO-ST	Accelerated lifetime tests	25.6kHz	• 3 operating conditions • 6 learning datasets 11 test datasets
[6-8]	IMS	Accelerated lifetime tests	20kHz	• Inner race fault, ball fault, outer race fault • 3 run-to-failure experiments (including 12 bearings)
[9]	XJTU-SY	Accelerated lifetime tests	25.6kHz	• Inner race fault, ball fault, outer race fault, cage fault • 3 load conditions: 11kn, 10kn, 12kn
[10-11]	CWRU	Artificially damaged	12kHz/48kHz	• Inner race fault, ball fault, outer race fault • Fault diameter: 0.007, 0.014, 0.021 • Motor load : 0, 1, 2, 3 (HP)
[12]	MFPT	Artificially damaged/ Accelerated lifetime tests	97656Hz/48828Hz	• Inner race fault, outer race fault • Different load • 3 run-to-failure experiments
[13]	Paderborn	Artificially damaged	64kHz	• 6 Undamaged bearings • 12 Artificially damaged bearings: inner race fault, outer race fault • 2 damage level • 14 run-to-failure experiments
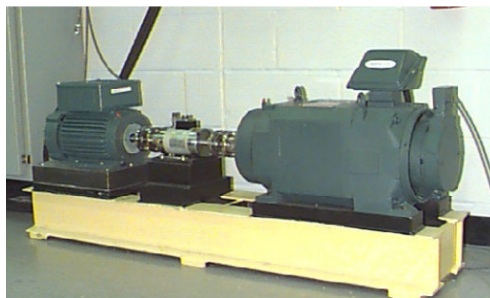
FIGURE 2. The bench of CWRU.

cycle data sets for fault diagnosis. A multi-label data set is formed by sampling data at different stages of the life cycle. Currently, the most widely used dataset is the CWRU. The faults of this dataset are man-made, with obvious features and relatively easy diagnosis. It can be used as a basic data set for algorithm testing.

The test bench of the CWRU dataset is shown in Figure 2. The test bench is composed of a 2-horsepower electric motor, a torque sensor, and a power dynamometer. Accelerometers are respectively installed on the housings of the drive end and the fan end to collect the vibration signals.

The test bench mainly records normal baseline data, drive end fault data and fan end fault data. The sampling frequency of drive end fault data is 12000 sps and 48000 sps, and the sampling frequency of the normal baseline data and fan-end fault data are both 12000 sps. Therefore, this public data set contains four categories of data. Each type of data is

mainly composed of inner race faults, ball faults and outer race faults with different load under different fault diameters. The faults of this dataset are mainly the damage of electric spark. It is a kind of artificially damage. The damage diameter includes 0.007inches, 0.014inches, 0.021inches and 0.028inches. Among them, the fault diameter of 0.007, 0.014, and 0.021 use SKF bearings, and the fault diameter of 0.028 uses NTN bearings. And it only records the inner race fault and ball fault of the drive end when the sampling frequency is 12000 sps and the fault diameter is 0.028 inches. The bearing loads are 0, 1 HP, 2 HP, and 3 HP respectively, corresponding to the different speeds. Outer race faults are divided into 6 o'clock, 3 o'clock and 12 o'clock according to the location of the fault point.

Each type of fault data is saved in a mat file, and each file contains the vibration data of the drive end and the fan end, as well as the speed. DE means drive end data, FE means fan end data, and RPM means speed.

Smith and Randall [13] proposed a CWRU-based benchmark based on three established rolling bearing fault diagnosis methods, through which new bearing fault diagnosis algorithms can be tested. Yongbo *et al.* [16] benchmarked the CWRU dataset by using various entropy and classifiers, and provided an evaluation method for subsequent new classification methods. Dhiraj *et al.* [17] discussed deep learning based bearing fault diagnosis and summarized the latest research results of deep learning based on the CWRU dataset.

In addition, for bearing fault diagnosis, the main purpose of data preprocessing is to solve the problem of data imbalance and small sample problems. The ratio of fault data to

TABLE 2. Time features.

Type	Formula	Type	Formula
Mean	$T_1 = \dfrac{1}{N_S}\sum_{i=1}^{N_S} x_i$	Std	$T_6 = \sqrt{\dfrac{1}{N_S}\sum_{i=1}^{N_S}(x_i - T_1)}$
RMS	$T_2 = \sqrt{\dfrac{1}{N_S}\sum_{i=1}^{N_S}(x_i)^2}$	Shape factor	$T_7 = \dfrac{T_2}{\dfrac{1}{N_S}\sum_{i=1}^{N_S}|x_i|}$
Kurtosis	$T_3 = (\sum_{i=1}^{N_S}(x_i - \max(x_i))^4)/N_S$	Peaking factor	$T_8 = \dfrac{\max(x_i)}{T_2}$
Peak-to-peak	$T_4 = \max(x_i) - \min(x_i)$	Pulse factor	$T_9 = \dfrac{\max(x_i)}{\dfrac{1}{N_S}\sum_{i=1}^{N_S}|x_i|}$
Var	$T_5 = \dfrac{1}{N_S}\sum_{i=1}^{N_S}(x_i - T_1)$	Margin factor	$T_{10} = \dfrac{\sum_{i=1}^{N_S}(x_i - \max(x_i))^3}{T_2^{\,3} N_S}$

TABLE 3. Frequency features.

Type	Formula
CF	$F_1 = \left(\sum_{j=0}^{N_f} f_j \times S(f_j)\right)\bigg/\sum_{j=0}^{N_f} S(f_i)$
MSF	$F_2 = \left(\sum_{j=0}^{N_f} f_j^2 \times S(f_j)\right)\bigg/\sum_{j=0}^{N_f} S(f_i)$
RMSF	$F_3 = \sqrt{\left(\sum_{j=0}^{N_f} f_j^2 \times S(f_j)\right)\bigg/\sum_{j=0}^{N_f} S(f_i)}$
VF	$F_4 = \left(\sum_{j=0}^{N_f}(f_j - P_2)^2 \times S(f_j)\right)\bigg/\sum_{j=0}^{N_f} S(f_i)$
RVF	$F_5 = \sqrt{\left(\sum_{j=0}^{N_f}(f_j - P_2)^2 \times S(f_j)\right)\bigg/\sum_{j=0}^{N_f} S(f_i)}$

health data is unbalanced, the data of different fault types are unbalanced, and the data sample size of each type is small. Jianan *et al.* [18] proposed an oversampling method named SCOTE. SCOTE transforms the multi-class data balance problem into multiple two-class data imbalance problems, and combines with multi-class LS-SVM to form a new model to solve the problem of rolling bearing fault data imbalance. Xu *et al.* [19] combined the probabilistic mixture model (PMM) and Markov Monte Carlo (MCMC) to achieve the expansion of the data set, and used the semi-supervised ladder network (SSLN) to solve the problem of fewer labeled samples.

III. REVIEW OF FEATURE EXTRACTION

Features are essential for machine learning [57]. The current mainstream approach is to extract multi-domain features from the vibration signal of the bearings to form a multi-domain feature set. Form as many features as possible from multiple dimensions such as frequency domain, time domain, and entropy domain.

The variable in the time domain is t, and t is usually used to observe the changes in the vibration signal in the time domain. Commonly used time domain features are mean, root mean square (RMS), kurtosis, peak-to-peak, variance (Var), standard deviation (Std), shape factor, peaking factor, pulse factor, and margin factor. The specific calculation formula is shown in Table 2. [20], [21].

The variable in the frequency domain is frequency f. Through the change of the domain, observe the change of the amplitude of the vibration signal in the frequency domain with the frequency. Compared with features in time domain, the advantage of features frequency domain is intuitive. Commonly used frequency domain features are root mean square frequency (RMSF), center frequency (CF), mean square frequency (MSF), frequency variance (VF), and root frequency

variance (RVF). The specific calculation formula is shown in Table 3. [22], [23].

Entropy is used to characterize the uncertainty of the system or information. The power spectrum describes the distribution of the power of the vibration signal in the frequency domain. The singular spectrum entropy is calculated by performing singular value decomposition on the vibration signal, and the local features of the vibration signal can be obtained. The singular spectrum entropy is a fault feature in the time domain. Wavelet energy entropy is a fault feature in the time-frequency domain. Bispectral entropy is the distortion decomposition of the vibration signal in the frequency domain to describe the fault.

In addition, Dongfang *et al.* [24] solved the problem of insufficient accuracy of multi-scale entropy by improving the scale factor of multi-scale entropy. The features formed by this method can provide more precise feature vectors for classifiers. Thereby improving the accuracy of diagnosis. Keheng *et al.* [25] calculated the hierarchical entropy of the vibration signal and used the hierarchical entropy as a feature vector to input it into a classifier that combines particle swarm optimization (PSO) and SVM. It is more superior than the method that uses multi-scale entropy as the feature vector. Nayana and Geethanjali [26] extracted 12 statistical features in time domain and 6 time-dependent spectral features (TDSFs). The feature selection algorithm combining wheel-based differential evolution (WBDE) and PSO is used to process the initial feature set, and the final feature subset contains most of the features of TDSFs. Zahra *et al.* [27] proposed a fault diagnosis method for identifying the degree of failure for rolling element faults. First, EMD is used to preprocess the vibration signal, and KLD is used to further process the IMF to form a feature vector. Three classifiers DAG-SVM, KNN, and decision tree (DT) are used for comparison and verification. Rui *et al.* [28] used ensemble empirical mode decomposition (EEMD) to preprocess the vibration signal and calculate the hierarchical entropy of the sample. The improved CS-SVM is used as the classifier of the model.
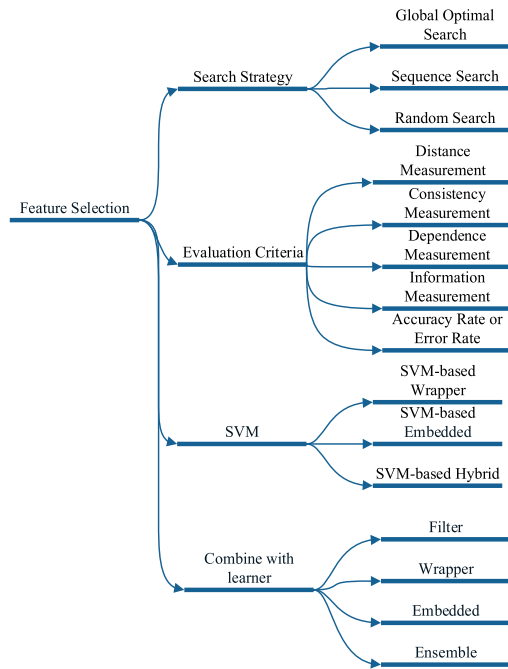
FIGURE 3. Classification of feature selection.

FIGURE 4. SVM.

IV. REVIEW OF FEATURE SELECTION

Features can be divided into related features, redundant features, and irrelevant features according to classification requirements. The purpose of feature selection is to remove redundant features and irrelevant features as much as possible, and retain related features, thereby reducing the feature vector dimension and avoiding the occurrence of dimensional disasters and overfitting. The main process of feature selection consists of four steps, namely generating feature subsets, evaluating feature subsets, stopping conditions and verification results. According to the search strategy, feature selection can be divided into three classes, namely, global optimal search, sequence search and random search. According to the evaluation criteria, the feature selection can be divided into distance measurement, consistency measurement, dependence measurement, information measurement and classification accuracy rate or classification error rate measurement. According to the combination of feature selection and learner, it can be divided into four classes, namely Filter, Wrapper, Embedded and Ensemble. Some papers also combine support vector machines with feature selection algorithms, which mainly include three categories, namely, SVM-based Wrapper, SVM-based Embedded, and SVM-based hybrid. The details are shown in Figure 3.

In the field of bearing fault diagnosis, many papers on feature selection and feature extraction have been published, and the CWRU dataset has been used for research. Xianghong *et al.* [29] proposed a feature selection model GL-mRMR-SVM, which uses the maximum correlation and minimum redundancy as the criterion for feature selection, and uses the global features in the frequency domain and time
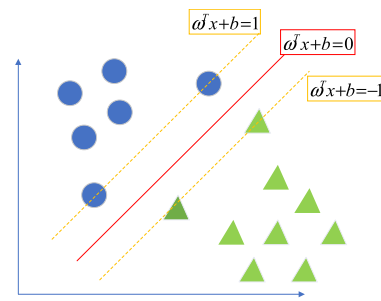
domain and the local features extracted by RNN as the original feature set, the final classifier is SVM. Xiang *et al.* [30] proposed a new envelope processing method named ICIE by improving CIE, and combining ICIE with the local mean decomposition (LMD), and proposed the ICIELMD model, which provides a new idea for the feature extraction of rolling bearings. Yue *et al.* [31] combined PCA and BP neural network to propose a new fault diagnosis model. PCA is used to reduce the dimensionality of the multi-source feature set composed of time domain, frequency domain and entropy features, and the feature subset is input into the BP neural network for fault diagnosis.

V. REVIEW OF CLASSIFIER

Classifiers are the application of some machine learning algorithms in classification. Some classic supervised classifiers include K-nearest neighbor (KNN), naive Bayes classifier, support vector machine (SVM), grey relational degree (GRD), and decision tree. Some classic unsupervised classifiers include clustering methods including k-means clustering, DBSCAN, and agglomerative clustering. The parameters of the classifier will be improved according to different problems to improve the generalization ability of the classifier.

A. SUPPORT VECTOR MACHINE (SVM)

SVM is a linear classifier that solves binary Classification problems. It realizes data classification by finding the maximum interval hyperplane. SVM is to find a hyperplane like $\omega^T x + b = 0$. As shown in Figure 4.

The optimization purpose of SVM is to make the distance between the support vector and the hyperplane as large as possible under the premise of correct classification, that is, to find the maximum interval plane, the optimization problem of SVM is transformed into formula 1.

$$\max \frac{1}{\|\omega\|} \quad s.t. \ y_i(\omega^T x_i + b) \geq 1 \quad (1)$$

Many papers use the CWRU dataset to study the application of SVM in the fault diagnosis of rolling bearing. Jianan *et al.* [18] proposed an oversampling method named SCOTE and used SVM as a verification classifier. SCOTE transforms the multi-class data balance problem into multiple two-class data imbalance problems, and combines with multi-class LS-SVM to form a new model to solve the problem
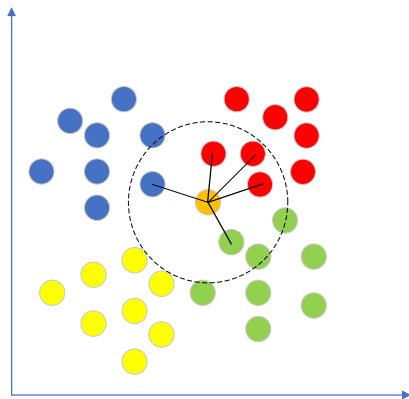
FIGURE 5. KNN.

of rolling bearing fault data imbalance. Yang *et al.* [32] used wavelet packet transform (WPT) to preprocess the data, obtain the energy distribution of the signal, and extract the feature to form a feature vector, and use the improved particle swarm algorithm (IPSO) proposed in the paper to optimize the parameters of the SVM. Wentao *et al.* [33] proposed a new distance-based feature selection method, by introducing a group identification matrix to obtain the coefficient of each feature. The verification classifier is SVM.

B. K-NEAREST NEIGHBOR (KNN)

KNN is a supervised classification algorithm for multi-class classification, and it is necessary to grasp the labels of existing samples in advance. For an unknown sample, we need to calculate the distance between the unknown sample and all existing samples, and select k samples with the closest distance, and then judge the class of the unknown sample based on the number of various types of samples in the k samples. The basic principle is shown in Figure 5.

The first key point of the basic principle of the KNN is to quantify the features of the training set. Since the distance between the existing sample and the unknown sample is calculated, it is necessary to make sure that the features contained in each sample are quantified as numbers. The second key point of the KNN algorithm is data normalization. The value range of the sample feature data has a direct impact on the calculation of the distance, so it is necessary to normalize each feature data to a specific value range. The third key point of the KNN algorithm is to determine the distance function. The existing distance functions include Euclidean distance, cosine distance, Hamming distance and Manhattan distance. Among them, the most widely used is Euclidean distance, which is shown in formula 2.

$$d(r, R) = \sqrt{\sum_{i=1}^{n} (r_i - R_i)^2} \qquad (2)$$

Many papers use the CWRU dataset to study the application of KNN in bearing fault diagnosis. Xin *et al.* [34] used singular values as the input of the model, and combined graph

theory with SVD, and proposed a new method of graph modeling. Validation with KNN classifier proves the effectiveness of this method in early fault diagnosis. Qingfeng *et al.* [35] proposed a weighted KNN fault diagnosis method named WKNN, and used the ReliefF feature selection algorithm to process the feature subset formed by the multi-domain feature set as the input of WKNN, which solved the poor generalization ability of the classifier under variable working conditions.

Ensemble learning is to combine multiple basic classifiers to complete a classification task, thereby improving the overall generalization ability. Among them, Boosting is a serial ensemble learning method. Through training weak classifiers in turn, a strong classifier is finally formed by weighting each weak classifier. Bagging reduces the variance of classification by averaging the results of multiple classifiers [36], [37].

VI. REVIEW OF DEEP LEARNING

Deep learning is a research direction of machine learning. At the same time, with the expansion of deep learning, more and more papers use deep learning to solve some problems in the bearing fault diagnosis. Commonly used networks are Auto-encoders [38]–[42], Convolutional neural networks (CNN), generative adversarial networks [33]–[45], recurrent neural networks (RNN), deep transfer learning [46]–[48].

Convolutional neural networks (CNN) are roughly the same as other neural networks. They use superimposed layers to construct the network, except that the CNN has more convolutional layers and pooling layers. As shown in Figure 6, it is a simple CNN model. The input of the convolutional layer is the input feature map, and the output of the convolutional layer is the output feature map. The convolution layer mainly performs convolution operations, which is a bit similar to filter calculation in image processing. The convolution layer can ensure that the shape and size of the input data remain unchanged. The pooling layer is to compress the data in the high and long directions. With the promotion of CNN, more and more papers have improved the accuracy of rolling bearing fault diagnosis by improving CNN. Fuzhou *et al.* [49] proposed a CNN based sigmoid multitask fault diagnosis model. Unlike other bearing fault diagnosis using CNN network, this model can identify the type and the extent of the fault. In the verification of the CWRU data set, the average accuracy rate can reach 96%. Yangyang *et al.* [50] proposed a fault diagnosis model by combining singular value decomposition and one-dimensional convolutional neural network (SVD-1DCNN), which has a good effect on fault diagnosis and classification. Afrasiabi *et al.* [51] proposed a new fault diagnosis model for rolling bearings by accelerating and compressing CNN, and compared it with SVM, ANN, and LVQ to verify the superiority of the proposed model. Mengyu *et al.* [52] combined sequential tracking and 1DCNN, and proposed a new fault diagnosis model named OT-1DCNN. It solves the problem that the bearing speed changes with time under working conditions, and improves the accuracy of fault diagnosis in
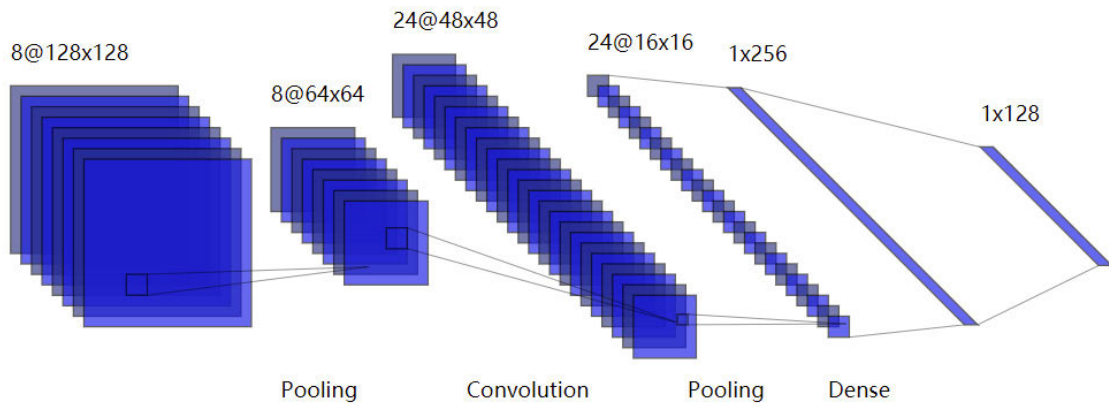
FIGURE 6. CNN.

different speed. Yang *et al.* [53] proposed a new hybrid model based on deep learning. First, continuous wavelet changes are used to convert vibration signals into time-frequency images, and then CNN is used to extract image features which are input to the gcForest classifier. Yinghua *et al.* [54] proposed a ID-DCNN model, and used this network to automatically extract features. Finally, the softmax and T-SNE algorithms are used to verify the feasibility of the proposed method.

The purpose of recurrent neural networks (RNN) is to link past data with current data to achieve "memory persistence". Unlike convolutional neural networks, RNN has a ring structure, which allows data to be maintained. In terms of parameter settings, RNN is also different from other neural network. The parameters of each layer of a neural network are different, and the parameters used by each layer of the RNN are the same at each moment. A special architecture of RNN is the long and short-term memory network model (LSTM). Many papers also apply RNN to the bearing fault diagnosis. Alex and Martin [55] combined CNN and RNN, and proposed a model named RNN-WDCNN, which solved the problem of domain adaptation and high-frequency noise under actual working conditions. Amin *et al.* [56] combined end-to-end CNN and LSTM, and proposed a new type of recurrent neural network named CRNN. Using the data of CWRU as the input of the model, the fault type of the bearing can be identified with the highest accuracy in a short time.

VII. EXPERIMENTS DATA AND DISCUSSION
This experiment will conduct experiments from 3 parts of the CWRU data set, including 12k Drive End Data, 48k Drive End Data and Fan End Data. The machine learning classifier and deep learning network will be used to analyze the CWRU data set, and the results will be compared and analyzed.

A. DATA DESCRIPTION
This experiment will compose three data sets from the three aspects of 12k Drive End Data, 48k Drive End Data and Fan End Data to carry out the experiment. Each dataset is composed of the health data under 0 load and the data of inner race fault, ball fault, and outer race fault (6 o'clock) with fault diameters of 0.007, 0.014, and 0.021 respectively. Each type of sample set contains 10 types of data.

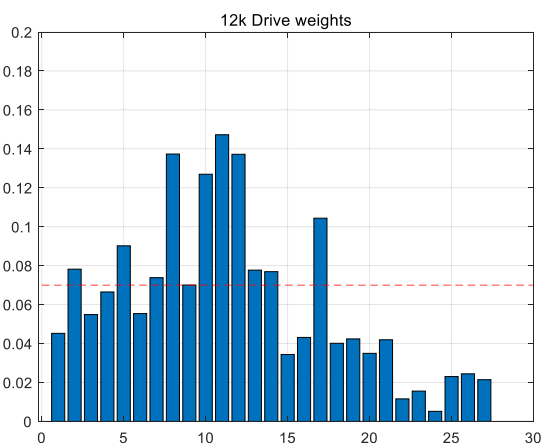### B. EXPERIMENT 1: THE INFLUENCE OF FEATURE SELECTION
This experiment chooses ReliefF feature selection algorithm which is based on Euclidean distance. It is a kind of feature weighting algorithm.

The dataset is processed to form a sample set containing 500 samples. Each type of sample consists of 50 samples, including 40 training samples and 10 test samples, and each sample contains 2048 sampling points. A multi-domain feature set containing 27 features is extracted from the four dimensions of the time domain, frequency domain, and entropy energy of the rolling bearing vibration signal, as shown in Table 4. The marginal spectrum energy feature is the marginal spectrum energy entropy of the top six IMFs obtained by decomposing the vibration signal through EMD.
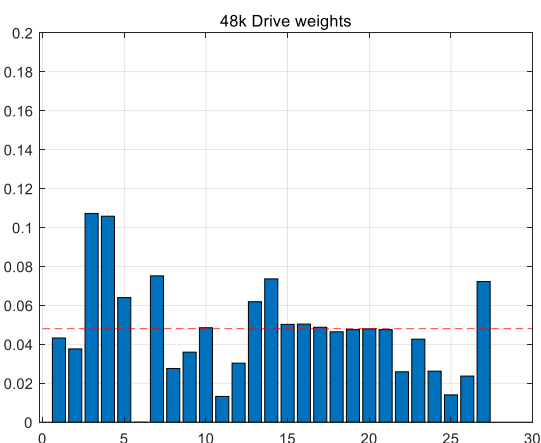
The three feature sets of CWRU are respectively used as the input of the ReliefF algorithm, and the weight scores of each feature of the three feature sets are shown in Figure 7, Figure 8, and Figure 9. The criterion is the feature weight score. It can be found that the sensitivity of each CWRU data set to the feature is not the same. The weight score of the frequency domain feature of the 12k Drive data set is relatively high, and the weight score of the energy feature is relatively low. The top 11 features include T2, T5, T7, T8, T9, T10, F1, F2, F3, F4, H2. The weight score of the time domain features of the 48k Drive data set is relatively high. The top 11 features include T3, T4, T5, T7, T10, F3, F4, F5, H1, H2, E6. The energy features of the 12k Fan data set have a relatively high weight score, and the frequency domain feature has a relatively low weight score. The top 11 features

TABLE 4. Feature set.

Time	Time	Frequency	Entropy	Marginal spectrum energy
Mean T1	Std T6	CF F1	Power Spectrum H1	IMF1 E1
RMS T2	Shape factor T7	MSF F2	Power Spectrum H2	IMF2 E2
Kurtosis T3	Peaking factor T8	RMSF F3	Singularity Spectrum H3	IMF3 E3
Peak-to-peak T4	Pulse factor T9	VF F4	Singularity Spectrum H4	IMF4 E4
Var T5	Margin factor T10	RVF F5	Wavelet Energy H5	IMF5 E5
			Bispectral entropy H6	IMF6 E6

FIGURE 9. The weight score of 12k fan.

FIGURE 10. The accuracy of feature subsets and feature sets.
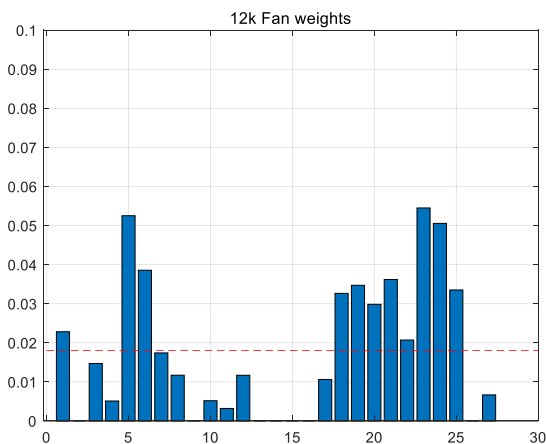

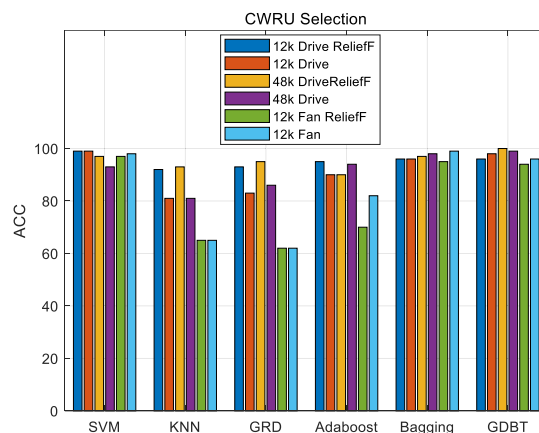**FIGURE 7.** The weight score of 12k drive.

FIGURE 8. The weight score of 48K drive.

include T1, T5, T6, H3, H4, H5, H6, E1, E2, E3, E4. It can be found from the figures that there are some features with a weight score of 0, indicating that these features are irrelevant features. The feature selection is used to remove irrelevant features and reduce the dimensionality of the feature set, thereby avoiding dimensional disasters.

C. EXPERIMENT 2: PERFORMANCE OF FAULT DIAGNOSIS WITH DIFFERENT MACHINE LEARNING CLASSIFIERS

The classifiers selected in this experiment are SVM, KNN, Grey relational degree (GRD), Adaboost, Bagging and GDBT classifiers. What's more, GRD can compare the dynamic changes of rolling bearings. The degree of closeness between the test sample and various types of faults can be judged by the geometric similarity between the different state data rows of the rolling bearing. The basic classifiers of Boosting are set to SVM and DT respectively to form Adaboost and GDBT classifiers. And set the basic classifier of Bagging to SVM. The top 11 features in the weight score of each feature set are respectively formed into a fault feature subset, and the feature set and the processed feature subset are respectively used as the input of the classifiers. The classification accuracy of each classifier is shown in the figure 10.

As shown in Figure 10, the fault diagnosis classifier based on machine learning performs well on the three data sets of

TABLE 5. The running time of CNN and LSTM.

	CNN	LSTM
12k Drive	32	3549
48k Drive	35	4834
12k Fan	35	4501

CWRU as a whole. Among them, the SVM classifier has the best fault diagnosis effect, which may be because the fault diagnosis problem of rolling bearings is a small sample problem, and SVM is a classifier suitable for solving the small sample classification problem. The overall fault classification effect of Adaboost and Bagging, which use SVM as the basic classifier, is also very good. It also shows that SVM has a very good effect in solving the problem of small sample classification.

In Figure 10, the blue, yellow, and green histograms indicate that the ReliefF process is included in the fault diagnosis. It can be found that the ReliefF has the effect of improving the classification accuracy of the classifier. It shows that the ReliefF removes the irrelevant features of the feature set.

D. EXPERIMENT 3: PERFORMANCE OF FAULT DIAGNOSIS WITH DEEP LEARNING

In addition, this article will compare the performance of simple deep learning models and machine learning classifiers on the CWRU dataset. This paper will use two basic models of LSTM and CNN.

The window size of the convolutional layer is 64×1 and the step size is 16×1. The convolutional layer is followed by a pooling layer with a window size of 2×1 and a scale factor of 2. There are several convolutional layers and pooling layers, the size of the convolution kernel is 3×1, and the step size is 1×1. The window size of the pooling layer is 2×1, and the scaling factor is 2. The number of convolutional layers and pooling layers is determined by the subsequent optimization process. There are several fully connected layers, the number of neurons is 32, and the specific number of layers is also determined by subsequent optimization. Finally, there is a fully connected output layer, and the activation function is the softmax.

The output dimension of LSTM model is set to 32, the function used for the loop time step is sigmoid. And the activation function is the Tanh.

The accuracy of the three data sets under the CNN and LSTM are shown in Figure 11, and the running time is shown in Table 5.

The overall fault diagnosis accuracy of the CNN and LSTM models perform well. The accuracy of the CNN model in the two datasets are close to 100%, which are higher than the accuracy of the LSTM model. Compared with machine learning classifiers, the accuracy is not much different. The advantage of the deep learning model in fault diagnosis is that it does not require manual feature extraction and feature
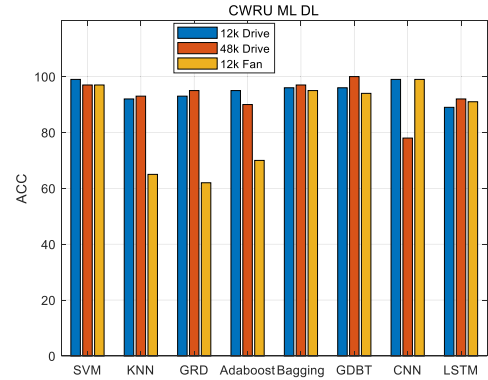

FIGURE 11. The accuracy of CNN and LSTM.

selection. But the disadvantage of the deep learning model is that the running time of the model is too long, especially the LSTM model.

VIII. CONCLUSION

This paper firstly investigated the current mainstream rolling bearing public datasets. Subsequently, the contributions made by the CWRU dataset in the field of machine learning are summarized. Finally, experiments were carried out on three datasets of CWRU and the results were discussed.. Get the following conclusions:

- The mainstream rolling bearing fault diagnosis data set mainly has two fault methods, one is the artificially damaged and the other is the accelerated lifetime test. Among them, the CWRU dataset is the most widely used in fault diagnosis. The advantage of the CWRU dataset is that the data is simple and can be used for initial verification of the fault diagnosis model.
- Machine learning based bearing fault diagnosis can be divided into three steps. The first step is to extract features, the second step is feature selection, and the third step is classifier recognition. The feature extraction part mainly extracts the sensitive features of the vibration signal in different domains to form a feature set with strong discrimination. According to different discrimination criteria or the combination with the learner, the feature selection algorithm can be divided into different categories. The most widely used machine learning classifiers are SVM and KNN. Many papers have proposed different improved classifiers based on SVM or KNN.
- CWRU contains four types of data: 12k Drive End Fault Data, 48k Drive End Fault Data, Fan End Fault Data, and benchmark data. Different data sets have different sensitivity to the features of different domains.
- Based on the CWRU dataset, the SVM has the best classification effect. In addition, the ensemble learning based on the SVM classifier also has a good classification effect. SVM is a classifier suitable for small sample problems.

- In the bearing fault diagnosis, a simple deep learning model does not require manual feature extraction and feature selection, but it requires longer running time and higher requirements for hardware equipment.

It is worth noting that the sample data of the CWRU data set is evenly distributed, and the fault types are all single faults. In actual working conditions, the data is often unevenly distributed and most of the fault types are compound faults. Therefore, the CWRU data set is only suitable for preliminary verification of the fault diagnosis algorithm. For further verification of the algorithm, a full life cycle data set is required.

REFERENCES

[1] S.-K. Zeng, M. G. Pecht, and J. Wu, "Status and perspectives of prognostics and health management technologies," *Acta Aeronaut. Astronaut. Sinica*, vol. 25, no. 5, p. 626, 2005.

[2] Y. Peng, D. Liu, and X. Peng, "A review: Prognostics and health management," *J. Electron. Meas. Instrum.*, vol. 24, no. 1, pp. 1–9, Mar. 2010.

[3] P. Yu and L. Datong, "Data-driven prognostics and health management: A review of recent advances," *Chin. J. Sci. Instrum.*, vol. 35, no. 3, pp. 481–495, 2014.

[4] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.

[5] S. Manikandan and K. Duraivelu, "Fault diagnosis of various rotating equipment using machine learning approaches—A review," *Proc. Inst. Mech. Eng. E, J. Process Mech. Eng.*, vol. 235, no. 2, pp. 629–642, Apr. 2021.

[6] P. Nectoux, R. Gouriveau, and K. Medjaher, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," Tech. Rep., 2012.

[7] H. Li and Y. Wang, "Rolling bearing reliability estimation based on logistic regression model," in *Proc. Int. Conf. Quality, Rel., Risk, Maintenance, Saf. Eng. (QR2MSE)*, Jul. 2013, pp. 1730–1733.

[8] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "A data-driven failure prognostics method based on mixture of Gaussians hidden Markov models," *IEEE Trans. Rel.*, vol. 61, no. 2, pp. 491–503, Jun. 2012.

[9] J. Yu, "Health condition monitoring of machines based on hidden Markov model and contribution analysis," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2200–2211, Aug. 2012.

[10] J. Yu, "Local and nonlocal preserving projection for bearing defect classification and performance assessment," *IEEE Trans. Ind. Electron.*, vol. 59, no. 5, pp. 2363–2376, May 2012.

[11] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Trans. Rel.*, vol. 69, no. 1, pp. 401–412, Mar. 2020.

[12] A. Boudiaf, A. Moussaoui, A. Dahane, and I. Atoui, "A comparative study of various methods of bearing faults diagnosis using the case western reserve university data," *J. Failure Anal. Prevention*, vol. 16, no. 2, pp. 271–284, Feb. 2016.

[13] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64, pp. 100–131, Dec. 2015.

[14] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock Vib.*, vol. 2017, pp. 1–17, Oct. 2017.

[15] C. Lessmeier, J. K. Kimotho, and D. Zimmer, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," in *Proc. Eur. Conf. Prognostics Health Manage. Soc.*, 2016.

[16] Y. Li, X. Wang, S. Si, and S. Huang, "Entropy based fault classification using the case western reserve university data: A benchmark study," *IEEE Trans. Rel.*, vol. 69, no. 2, pp. 754–767, Jun. 2020.

[17] D. Neupane and J. Seok, "Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review," *IEEE Access*, vol. 8, pp. 93155–93178, 2020.

[18] J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, and D. Huang, "New imbalanced bearing fault diagnosis method based on sample-characteristic oversampling technique (SCOTE) and multi-class LS-SVM," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 107043.

[19] X. Ding, X. Lu, D. Wang, Q. Lv, and H. Zhai, "Rolling bearing fault diagnosis based on probabilistic mixture model and semi-supervised ladder network," *Adv. Mech. Eng.*, vol. 12, no. 12, Dec. 2020, Art. no. 168781402097774.

[20] W. Chenxi, N. Liwei, J. Rong, W. Xing, and L. Junan, "Intelligent identification of bearing faults using time domain features," in *Proc. 4th Int. Conf. Digit. Manuf. Autom. (ICDMA)*, Jun. 2013, pp. 713–716.

[21] M. Wang, N.-Q. Hu, L. Hu, and M. Gao, "Feature optimization for bearing fault diagnosis," in *Proc. Int. Conf. Quality, Rel., Risk, Maintenance, Saf. Eng. (QR2MSE)*, Jul. 2013, pp. 1738–1741.

[22] H. Helmi and A. Forouzantabar, "Rolling bearing fault detection of electric motor using time domain and frequency domain features extraction and ANFIS," *IET Electr. Power Appl.*, vol. 13, no. 5, pp. 662–669, May 2019.

[23] T. Xu, J. Ji, X. Kong, F. Zou, and W. Wang, "Bearing fault diagnosis in the mixed domain based on crossover-mutation chaotic particle swarm," *Complexity*, vol. 2021, pp. 1–13, Apr. 2021.

[24] D. Zhao, S. Liu, S. Cheng, X. Sun, L. Wang, Y. Wei, and H. Zhang, "Dense multi-scale entropy and it's application in mechanical fault diagnosis," *Meas. Sci. Technol.*, vol. 31, no. 12, Dec. 2020, Art. no. 125008.

[25] K. Zhu, X. Song, and D. Xue, "A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm," *Measurement*, vol. 47, pp. 669–675, Jan. 2014.

[26] B. R. Nayana and P. Geethanjali, "Improved identification of various conditions of induction motor bearing faults," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 1908–1919, May 2020.

[27] Z. Mezni, C. Delpha, D. Diallo, and A. Braham, "Bearing fault severity classification based on EMD-KLD: A comparative study for incipient ball faults," in *Proc. Prognostics Health Manage. Conf. (PHM-Besancon)*, May 2020, pp. 257–262.

[28] R. Wang, Z. Zhang, Z. Xia, J. Miao, and Y. Guo, "A new approach for rolling bearing fault diagnosis based on EEMD hierarchical entropy and improved CS-SVM," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Qingdao)*, Oct. 2019, pp. 1–6.

[29] X. Tang, Q. He, X. Gu, C. Li, H. Zhang, and J. Lu, "A novel bearing fault diagnosis method based on GL-mRMR-SVM," *Processes*, vol. 8, no. 7, p. 784, Jul. 2020.

[30] X. Li, J. Ma, X. Wang, J. Wu, and Z. Li, "An improved local mean decomposition method based on improved composite interpolation envelope and its application in bearing fault feature extraction," *ISA Trans.*, vol. 97, pp. 365–383, Feb. 2020.

[31] S. Yue, X. Aidong, W. Kai, H. Xiaojia, G. Haifeng, and Z. Wei, "A novel bearing fault diagnosis method based on principal component analysis and BP neural network," in *Proc. 14th IEEE Int. Conf. Electron. Meas. Instrum. (ICEMI)*, Nov. 2019, pp. 1125–1131.

[32] Y. Shao, X. Yuan, C. Zhang, and C. Liu, "A novel deep stacking least squares support vector machine for rolling bearing fault diagnosis," in *Proc. 32nd Chin. Control Decis. Conf. (CCDC)*, 2020, pp. 2758–2763.

[33] W. Mao, L. Wang, and N. Feng, "A new fault diagnosis method of bearings based on structural feature selection," *Electronics*, vol. 8, no. 12, p. 1406, Nov. 2019.

[34] X. Wen, G. Lu, J. Liu, and P. Yan, "Graph modeling of singular values for early fault detection and diagnosis of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 145, Nov. 2020, Art. no. 106956.

[35] Q. Wang, S. Wang, B. Wei, W. Chen, and Y. Zhang, "Weighted K-NN classification method of bearings fault diagnosis with multi-dimensional sensitive features," *IEEE Access*, vol. 9, pp. 45428–45440, 2021.

[36] F. Chen, M. Cheng, B. Tang, B. Chen, and W. Xiao, "Pattern recognition of a sensitive feature set based on the orthogonal neighborhood preserving embedding and AdaBoost_SVM algorithm for rolling bearing early fault diagnosis," *Meas. Sci. Technol.*, vol. 31, no. 10, Oct. 2020, Art. no. 105007.

[37] X. Yan, Z. Sun, J. Zhao, Z. Shi, and C.-A. Zhang, "Fault diagnosis of active magnetic bearing–rotor system via vibration images," *Sensors*, vol. 19, no. 2, p. 244, Jan. 2019.

[38] Y. Lin, Y. Li, X. Yin, and Z. Dou, "Multisensor fault diagnosis modeling based on the evidence theory," *IEEE Trans. Rel.*, vol. 67, no. 2, pp. 513–521, Jun. 2018.

[39] W. Mao, W. Feng, Y. Liu, D. Zhang, and X. Liang, "A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 150, Mar. 2021, Art. no. 107233.

[40] W. Mao, W. Feng, X. Liang, and X. Zhang, "A novel deep output kernel learning method for bearing fault structural diagnosis," *Mech. Syst. Signal Process.*, vol. 117, no. 2, pp. 293–318, 2019.

[41] M. Cui, Y. Wang, X. Lin, and M. Zhong, "Fault diagnosis of rolling bearings based on an improved stack autoencoder and support vector machine," *IEEE Sensors J.*, vol. 21, no. 4, pp. 4927–4937, Feb. 2021.

[42] S. Zhang, F. Ye, B. Wang, and T. G. Habetler, "Semi-supervised bearing fault diagnosis and classification using variational autoencoder-based deep generative models," *IEEE Sensors J.*, vol. 21, no. 5, pp. 6476–6486, Mar. 2021.

[43] S. Dixit and N. K. Verma, "Intelligent condition-based monitoring of rotary machines with few samples," *IEEE Sensors J.*, vol. 20, no. 23, pp. 14337–14346, Dec. 2020.

[44] M. Xu and Y. Wang, "An imbalanced fault diagnosis method for rolling bearing based on semi-supervised conditional generative adversarial network with spectral normalization," *IEEE Access*, vol. 9, pp. 27736–27747, 2021.

[45] T. Zheng, L. Song, J. Wang, W. Teng, X. Xu, and C. Ma, "Data synthesis using dual discriminator conditional generative adversarial networks for imbalanced fault diagnosis of rolling bearings," *Measurement*, vol. 158, Jul. 2020, Art. no. 107741.

[46] H. Yin, Z. Li, J. Zuo, H. Liu, K. Yang, and F. Li, "Wasserstein generative adversarial network and convolutional neural network (WG-CNN) for bearing fault diagnosis," *Math. Problems Eng.*, vol. 2020, pp. 1–16, May 2020.

[47] M. Wang, Y. Lin, Q. Tian, and G. Si, "Transfer learning promotes 6G wireless communications: Recent advances and future challenges," *IEEE Trans. Rel.*, vol. 70, no. 2, pp. 790–807, Jun. 2021.

[48] Y. Lin, Y. Tu, and Z. Dou, "An improved neural network pruning technology for automatic modulation classification in edge devices," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5703–5706, May 2020.

[49] F. Feng, C. Wu, J. Zhu, S. Wu, Q. Tian, and P. Jiang, "Research on multitask fault diagnosis and weight visualization of rotating machinery based on convolutional neural network," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 42, no. 11, pp. 1–14, Nov. 2020.

[50] Y. Wang, S. Huang, J. Dai, and J. Tang, "A novel bearing fault diagnosis methodology based on SVD and one-dimensional convolutional neural network," *Shock Vib.*, vol. 2020, pp. 1–17, Jan. 2020.

[51] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Real-time bearing fault diagnosis of induction motors with accelerated deep learning approach," in *Proc. 10th Int. Power Electron., Drive Syst. Technol. Conf. (PEDSTC)*, Feb. 2019, pp. 155–159.

[52] M. Ji, G. Peng, J. He, S. Liu, Z. Chen, and S. Li, "A two-stage, intelligent bearing-fault-diagnosis method using order-tracking and a one-dimensional convolutional neural network with variable speeds," *Sensors*, vol. 21, no. 3, p. 675, Jan. 2021.

[53] Y. Xu, Z. Li, S. Wang, W. Li, T. Sarkodie-Gyan, and S. Feng, "A hybrid deep-learning model for fault diagnosis of rolling bearings," *Measurement*, vol. 169, Feb. 2021, Art. no. 108502.

[54] Y. Yinghua, L. Doliang, and L. Xiaozhi, "Fault diagnosis based on one-dimensional deep convolution neural network," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Aug. 2020, pp. 5630–5635.

[55] A. Shenfield and M. Howarth, "A novel deep learning model for the detection and identification of rolling element-bearing faults," *Sensors*, vol. 20, no. 18, p. 5112, Sep. 2020.

[56] A. Khorram, M. Khalooei, and M. Rezghi, "End-to-end CNN + LSTM deep learning approach for bearing fault diagnosis," *Appl. Intell.*, vol. 51, pp. 736–751, Feb. 2021.

[57] Y. Ying, J. Li, Z. Chen, and J. Guo, "Study on rolling bearing on-line reliability analysis based on vibration information processing," *Comput. Electr. Eng.*, vol. 69, pp. 842–851, Jul. 2018.

[58] J. Li, Y. Ying, Y. Ren, S. Xu, D. Bi, X. Chen, and Y. Xu, "Research on rolling bearing fault diagnosis based on multi-dimensional feature extraction and evidence fusion theory," *Roy. Soc. Open Sci.*, vol. 6, no. 2, Feb. 2019, Art. no. 181488.

XIAO ZHANG was born in Yantai, Shandong, China, in 1981. He received the Ph.D. degree in computer science from the University of Twente, The Netherlands, in 2014. He is currently an Associate Professor with Shandong Technology and Business University. His research interests include artificial intelligence, big data analysis, fault-tolerant computing, and blockchain technology.

BOYANG ZHAO (Graduate Student Member, IEEE) was born in Harbin, Heilongjiang, China, in 1999. He received the M.S. degree in electronic and information engineering from Harbin Engineering University, in 2021. He is currently pursuing the B.S. degree in electronic and information with Beihang University, Beijing, China. His research interests include fault diagnosis, machine learning, and deep learning.

YUN LIN (Member, IEEE) received the B.S. degree from Dalian Maritime University, Dalian, China, in 2003, the M.S. degree from the Harbin Institute of Technology, Harbin, China, in 2005, and the Ph.D. degree from Harbin Engineering University, Harbin, in 2010. He was a Research Scholar with Wright State University, USA, from 2014 to 2015. He is currently a Full Professor with the College of Information and Communication Engineering, Harbin Engineering University. He had published more than 200 international peer-reviewed journal articles/conference papers, such as the IEEE Transactions on Industrial Informatics, IEEE Transactions on Communications, IEEE Internet of Things Journal, IEEE Transactions on Vehicular Technology, IEEE Transactions on Cognitive Communications and Networking, *TR*, INFOCOM, GLOBECOM, ICC, VTC, and ICNC. His current research interests include fault detection and recognition, machine learning and data analytics over wireless networks, signal processing and analysis, cognitive radio and software defined radio, artificial intelligence, and pattern recognition. In addition, he served as the General Vice Chair for VTC-2021 Fall, the General Chair for ADHIP 2020, the TPC Chair for MOBIMEDIA 2020, ICEICT 2019, and ADHIP 2017, and a TPC Member for GLOBECOM, ICC, ICNC, and VTC. He had successfully organized several international workshops and symposia with top-ranked IEEE conferences, including INFOCOM, GLOBECOM, DSP, and ICNC, among others. He is serving as an Editor for the IEEE Transactions on Reliability, *Digital Communications and Networks*, *KSII Transactions on Internet and Information Systems*, and *International Journal of Performability Engineering*.

● ● ●