Journal of Ambient Intelligence and Humanized Computing (2022) 13:4041–4056
https://doi.org/10.1007/s12652-021-03177-x

ORIGINAL RESEARCH

A bearing fault diagnosis model based on CNN with wide convolution kernels

Xudong Song[1] · Yuyang Cong[1] · Yifan Song[1] · Yilin Chen[1] · Pan Liang[1]

Received: 18 May 2020 / Accepted: 25 March 2021 / Published online: 2 April 2021

Abstract
Intelligent fault diagnosis of bearings is an essential issue in the field of health management and the prediction of rotating machinery systems. The traditional bearing intelligent diagnosis algorithms based on the combination of feature extraction and classification for signal processing require high expert experience, which are time-consuming and lack universality. Compared with traditional methods, the convolutional neural network(CNN) can extract features automatically from the original vibration time-domain signal without any preprocessing. The accuracy of intelligent fault diagnosis can be improved by utilizing the multi-layer nonlinear mapping capability of deep convolutional neural networks. In order to realize the intelligent diagnosis and improve the recognition rate, this paper adopts the strategy of widening convolution kernels to obtain a larger receptive field and proposes a network design process pattern based on this idea, in addition, obtains the convolutional neural network with wide convolution kernels (WKCNN) model through experiments. Based on the time-domain vibration signal, this paper generates more input data through expansion and adopts the wide kernels of the first two convolutional layers to quickly extract features to improve efficiency. The smaller convolution kernels are used for multi-layer nonlinear mapping to deepen the network and improve detection accuracy. The results show that WKCNN performs well in accuracy, anti-noise, and timeliness compared with other diagnostic methods.

Keywords Bearing intelligent fault diagnosis · Wide convolution kernels · Network design process pattern · WKCNN

1 Introduction

Rolling bearings are one of the most important parts of rotating machinery and equipment, but rolling bearings are easily damaged by the working environment during work, resulting in mechanical failure. According to statistics, 40% of motor failures are bearing failures (Frosini et al. 2015). As a result of the complex working environment, failure is inevitable for rolling element bearings, and the circumstances under which failure occurs are unpredictable (Chen et al. 2018). The most common way to prevent possible damage is to implement real-time monitoring of vibration when the rotating mechanism is in operation (Zhang et al. 2017b). Use the status signals collected by the sensors to apply intelligent fault diagnosis methods to identify fault types (Zhang et al. 2017b; Jayaswal et al. 2011; Yiakopoulos et al. 2011; Li et al. 2016) . The traditional intelligent fault diagnosis method can be divided into two steps: feature extraction and classification. In the field of bearing fault diagnosis, common feature extraction methods include wavelet transform (Wan and Zhang 2018), fast Fourier transform(FFT) (Safin et al. 2016), empirical pattern decomposition (Xiao et al. 2017), and so on. Common pattern classification algorithms include BP neural network (Nie et al. 2019), support vector machine(SVM) (Islam and Kim 2017; Ziani et al. 2017; Fu et al. 2020), Multi-Layer Perceptron (MLP) (Almeida et al. 2014), Deep Neural Networks(DNN) (Feng et al. 2016), Bayes classifier (Manish et al. 2018), k-nearest neighbor classifier (Kim et al. 2016), Random Forest (Xue et al. 2019) etc. Such machine learning methods are widely used to predict the type of failure. FFT-SVM (Islam and Kim 2017), FFT-MLP (Almeida et al. 2014) and FFT-DNN (Feng et al. 2016) are commonly used in fault diagnosis. FFT is a method to quickly calculate the discrete Fourier

✉ Yuyang Cong
2241447847@qq.com

Xudong Song
sxd@djtu.edu.cn

[1] Software Institute, Dalian Jiaotong University, Dalian 116028, Liaoning, China

Transform (DFT) of a sequence or its inverse transform. Fourier analysis converts a signal from its original domain (usually time or space) to a representation in the frequency domain or vice versa. SVM is a kind of generalized linear classifier which classifies data according to supervised learning. Its decision boundary is the maximum margin hyperplane to solve the learning samples. MLP is known for its ability to learn complex and nonlinear pattern features, and it is also a very commonly used classifier in fault diagnosis. The neural network is an extension of perceptron, and DNN can be understood as a neural network with many hidden layers. These three methods will be compared with our method in Sect. 5.

However, with the mechanical health monitoring entering the "big data era", the traditional intelligent diagnosis algorithm based on signal processing feature extraction and classifier needed high requirements for expert experience, which is time-consuming to design and cannot guarantee universality, and it has been unable to meet the requirements of mechanical big data.

In recent years, the convolutional neural network has achieved great success in the field of pattern recognition. The characteristic of this kind of technology is that it can automatically extract features from signals and images, replacing the cumbersome feature engineering of traditional algorithms. Deep convolutional neural networks require more training data for training than conventional algorithms due to a large number of parameters to suppress overfitting. This is also the reason why the convolutional neural network gradually stands out in the era of big data. The convolutional neural network has two main features: weights sharing and spatial pooling, which makes it very suitable for computer vision applications whose inputs are usually 2D data, but it has also been used to address natural language processing and speech recognition tasks whose inputs are 1D data (Abdel-hamid et al. 2012; Kim 2014). CNN has achieved great success in the field of image recognition. Meanwhile, CNN can also directly act on speech recognition and original time-domain vibration signals.

Zhang et al. (2017a) proposed a CNN model with two convolutional layers to diagnose the faults of bearings with a huge number of training data. In ?, firstly, the data is processed by fast Fourier transform, then the self-encoder is used for unsupervised layer by layer training, and finally, the supervised training is carried out. The algorithm can achieve a recognition accuracy of 99.68±0.22%. Feng et al. (2016) proposed an FFT-DNN fault diagnosis method, which used the preprocessed FFT spectrum image as the input of the DNN. Xu et al. (2019) proposed a novel bearing fault diagnosis method based on deep convolutional neural network (CNN) and random forest (RF) ensemble learning, which uses time-domain vibration signals are converted into two dimensional (2D) gray-scale images as the input signals.

However, the conversion of the one-dimensional signal to the two-dimensional signal will affect the spatial structure and the information related to the failure may be lost. Zhang et al. (2017b) proposed WDCNN, which can directly process one-dimensional time-domain vibration signals through the convolutional neural network, and can achieve 100% accuracy under certain conditions, and performs well under noise. Li et al. (2018)proposed a novel deep learning method for rotating machinery fault diagnosis which manages to achieve high diagnosis accuracy with small original training dataset. Jian et al. (2019) proposed a one-dimensional fusion neural network (OFNN), which combines CNN with D–S evidence theory. The method can effectively improve the cross-domain adaptive ability of the model. Although these neural network models have achieved good results in the CWRU data set, they do not mention the discussion on the selection process of the one-dimensional model structure. The contributions of this paper can be summarized below.

1. In order to further improve the accuracy and efficiency of bearing fault diagnosis, we propose the WKCNN model which is based on the characteristics of one-dimensional signals. The model performs well in the aspects of fault diagnosis accuracy, timeliness, and anti-noise interference without any pre-processing.
2. Combining the first point and experiments, a design process model of one-dimensional convolutional neural network for one-dimensional vibration signal and a model construction algorithm of WKCNN are proposed.

The remainder of this paper is as follows: Sect. 2 introduces the related work. Section 3 describes the construction process, modeling algorithm and optimization strategy of WKCNN model. In Sect. 4, we determine the structure of WKCNN model by experiments. Experiments in Sect. 5 illustrate the evaluation and analysis of WKCNN. Section 6 is the conclusion.

2 Related work

Because of the diversity of fault types that can occur in bearings, feature engineering may face limitations in designing a set of characteristic features to describe the differences between all possible fault types. Handcrafted features do not necessarily provide generalization capability and portability from one system to another or even to other failure types. They also have limited scalability due to an expert-driven manual approach. In addition, the performance of feature engineering is highly dependent on the experience and expertise of the domain experts performing the task. The quality of feature extraction greatly influences the performance of the machine learning approach to

feature extraction. As the number of monitored parameters increases, the difficulty of feature engineering for diagnostic engineers is also increasing, so people are interested in automating this process (Yan and Yu 2019), or avoiding the need for feature engineering in the first time. Deep learning has the potential to incorporate feature engineering (or at least some of it) into the end-to-end learning processes.

As a breakthrough in the field of artificial intelligence, deep learning allows automatically processing of data, with highly nonlinear and complex feature abstracting through layers of layers, rather than using domain knowledge to hand-craft the best feature representation of data. With automatic feature learning and high-capacity modeling capabilities, deep learning provides an advanced analytical tool for intelligent manufacturing in the era of big data. It uses a cascade of nonlinear processing layers to learn the representation of data corresponding to different levels of abstraction. The hidden patterns underneath each other are then identified and predicted through end-to-end optimization. Deep learning offers great potential for promoting data-driven manufacturing applications, especially in the era of big data (Teti et al. 2010; Wu et al. 2017).

CNN is an important part of deep learning. As CNN was originally developed for image analysis, different approaches are investigated to construct two dimensional input from time series data. Time series to image encoding. Due to the success of CNN in image representation learning (Deng et al. 2009; He et al. 2016), the trend of understanding time series by translating time series into images is on the rise. By doing so, existing knowledge of image understanding and image representation learning can be directly used for Prognostics and Health Management(PHM) applications. An intuitive approach (Krummenacher et al. 2018) is to simply use the natural plot of one-dimensional time series data, signal vs time, as two-dimensional image. Alternatively, Gramian Angular Fields (GAF), Markov Transition Fields (MTF) and Recurrence Plots (RP) have been introduced in Wang and Oates (2015) and Hatami et al. (2017) as encoding approaches to translate signals to images. In addition, the time-frequency analysis can also obtain two-dimensional signal representations. The spectrum of multichannel vibration data is also studied in Janssens et al. (2016) to fit the model requirement. Park et al. (2016), the time series data is converted to a matrix for arrangement and then normalized to an image. In Wang et al. (2016), time frequency spectrum is used as the image input of the CNN model when the wavelet transform vibration signal is used. However, as we have mentioned in the introduction, converting a one-dimensional vibration signal to a two-dimensional image may result the spatial correlation in the original sequence will be destroyed and the important feature information may be lost. Therefore, our model will be built through a one-dimensional convolutional neural network.

In fact, sensor data is usually one-dimensional data by nature rather than two-dimensional images. In order to still benefit from recent progress in convolutional operations and avoid loss of important feature information, one-dimensional CNN kernels, instead of two-dimensional kernels, can be used for time series classification, anomaly detection in time series or RUL prediction. The same idea has been widely adopted for motor fault diagnosis (Ince et al. 2016), and broader PHM applications (Babu et al. 2016; Jing et al. 2017; Zhang et al. 2017a, b; Xiang et al. 2018; Jian et al. 2019). Here we mainly discuss the classification of time series. The WDCNN model proposed by Zhang et al. (2017b) is a very intelligent model using deep learning technology. WDCNN works directly on the original vibration signal without any time-consuming handmade features. Excellent diagnostic results have been obtained from the case Western Reserve University bearing data set. However, there are still some problems in this model, such as overfitting of test dataset and some details still need to be optimized. In addition, for the selection of model parameters, the theoretical description of the algorithm is also slightly inadequate. Jian et al. (2019) proposed OFNN. Experimental results show that this method can effectively improve the cross-domain adaptive capacity of the model and has a better diagnostic accuracy than other existing experimental methods. The method also uses one-dimensional convolutional neural network to establish algorithm model, and the innovation point is to realize bearing fault detection with Softmax classifier by using the class vector output synthetically determined by D–S evidence theory. However, the efficiency of this algorithm still needs to be improved in practice, and it takes a lot of time to improve the accuracy is very limited. In order to improve the accuracy and ensure the high efficiency, we proposed the WKCNN model. Moreover, the WKCNN construction algorithm is provided. In order to further verify our model, WKCNN will be compared with three traditional models FFT-SVM (Islam and Kim 2017), FFT-MPL (Almeida et al. 2014), FFT-DNN (Feng et al. 2016) and the deep learning algorithm WDCNN (Zhang et al. 2017b) without feature extraction in Sect. 5.

3 The construction method of WKCNN model

The network design process pattern of the WKCNN model is shown in Fig. 1. In order to obtain more accurate experimental results by using deep learning, we need to enhance the original experimental data to obtain more data. When determining the length of the input signal, the field size of the last pooling layer should cover the length of the input signal as much as possible so that the network can obtain more comprehensive data characteristics.

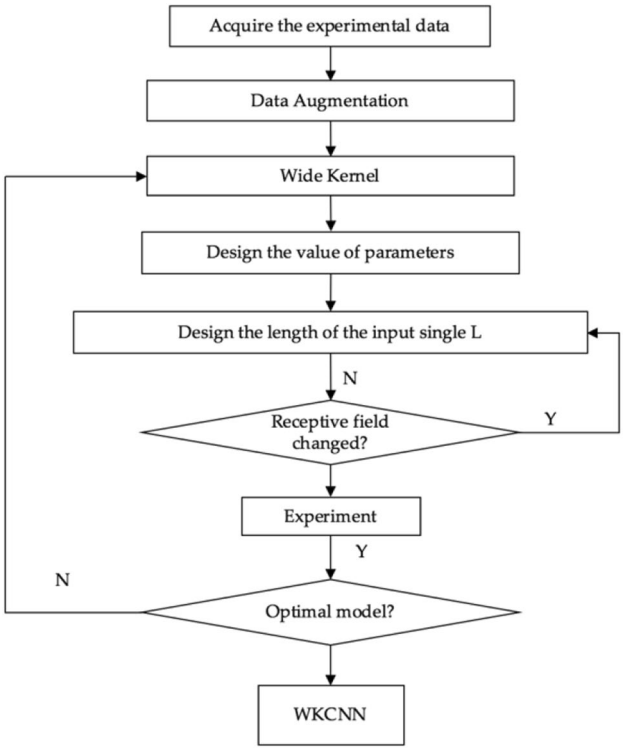


Fig. 1 Network design process pattern

The design of the network structure starts from the overall widening of the convolution kernel. The wide kernels can better suppress high-frequency noise compared with small kernels (Zhang et al. 2017b). Multilayer small wide kernels can better suppress high-frequency noise and inhibit overfitting compared with small kernels. We use K_i^l to represent the i-th convolution kernel of the l-th layer, b_i^l represents the bias of the i-th convolution kernel of the l-th layer, $x^l(j)$ to represent the local region of the j-th convolution in the l-th layer. The specific convolution operation formula is shown as follows:

$$y_i^{l+1}(j) = K_i^l * x^l(j) + b_i^l, \tag{1}$$

where the * symbol represents the dot product of the calculation kernel and the local area, $y_i^{l+1}(j)$ is the value of the j-th local area calculated from the input of the (l+1)-th layer.

In the selection of activation function, the traditional Sigmoid activation function is faced with the problem that the deep network structure is prone to gradient disappearance and the training time is too long. Therefore, the linear rectifying function (ReLU) is used in this paper. The formula is as follows:

$$z_i^{l+1}(j) = \max\left\{0, y_i^{l+1}(j)\right\}, \tag{2}$$

we use $y_i^{l+1}(j)$ to represent the output value of the convolution operation and $z_i^{l+1}(j)$ is the activation function of $y_i^{l+1}(j)$.

After obtaining the features through convolution, in order to reduce the computation, the model chooses to use the maximum pooling function to process the feature mapping results obtained from the convolution operation. When the multi-layer convolution calculation and pooling calculation are completed, the extracted features are firstly expanded smoothly and processed into a vector that can be output. After that, the full connection layer is usually used to realize random feature combination and classification. Each neuron in the full connection layer is cross-connected with the neuron representing the output feature vector of the previous layer after smooth expansion to realize the classification of local information with category distinction after convolution calculation. The calculation method of full connection layer is the same as that of ANN. The details are as follows:

$$y_j^{l+1} = \sum_{i=1}^{n} W_{ij}^l a^{l(i)} + b_j^l, \tag{3}$$

let y_j^{l+1} be the logits of the j-th output neuron in the (l+1)-th layer; W_{ij}^l represents the weight between the i-th neuron in layer l and the j-th neuron in layer l+1.b_i^j is the bias value of all neurons in layer l to the j-th neuron in layer l+1. Finally, the test results were output by Softmax function:

$$q^j = softmax\left(y_j^{l+1}\right) = \frac{e^{y_j^{l+1}}}{\Sigma_k^{10} e^{y_j^{l+1}}}, \tag{4}$$

during the experiment, the parameters such as the size of convolution kernel, step size in the model are constantly improved and adjusted. If the size of the sensing field changes, the length of the input signal needs to be redesigned to make the sensing field cover the length of the input signal as much as possible. Experiments were carried out in the aspects of accuracy, time and anti-interference, and the optimal model was approximated continuously.

Combined with the above process, we proposed the modeling algorithm of WKCNN as shown in algorithm 1. Since we adopt deep learning technology and the strategy of broadening convolution kernels, we need to generate a large amount of experimental data through data augmentation to obtain accurate diagnosis results. Deep learning is an end-to-end algorithm based on statistics and can discover complex structures in large data sets by using back-propagation algorithms. However, its learning process is like a black box to us, and it is difficult for us to understand its internal structure. The core work of algorithm 1 is to help us determine the optimal value of each parameter and hyper-parameter based on the experiments. The breakthrough of this algorithm is a comparison between the size of the receptive field and the length of the input signal. Our theoretical basis is that in order to obtain more data feature information, the size of the receptive field should cover the length of the input

data as much as possible. The size of the receptive field is determined by the sizes of the convolution kernels. Just as the example mentioned in the following Sect. 3.2, broadening the convolution kernel is a very appropriate strategy for a one-dimensional convolutional neural network, and it can also suppress the problem of data overfitting. Therefore, we will constantly adjust the sizes of the convolution kernels and the length of the input signal in the experiment to find a model that keeps approaching perfection, which is the modeling algorithm of WKCNN.

training will consume a lot of time. In WKCNN, the total time complexity of all convolutional layers is:

$$\textbf{Time} \sim O\left(\sum_{i=1}^{d} L_i \cdot K_i \cdot C_{i-1} \cdot C_i\right) \tag{5}$$

Here i is the index of a convolution layer, and d is the depth of convolutional layers. L_i is the length of the output feature map in the i-th layer, K_i is also as know as the length of the kernels. C_{i-1} is the number of input channels of the i-th layer, and C_i is the number of kernels in the i-th layer.

Algorithm 1 WKCNN modeling algorithm

Input:
One-dimensional bearing vibration signal
Output:
The classification and model evaluation results of WKCNN
Require:
L is the length of the input signal, the shift is w, the number of sample groups is G, N is the total amount of data.
$R^{(k)}$ is the perceptive field size of the k-th layer, and the value of $R^{(0)}$ is 1, $f^{(k)}$ is the size of the current convolution kernel, $S^{(i)}$ is the step size of the i-th layer .
T is the number of data points recorded by the bearing rotation.
Method:
1. Data augmentation
$G = \frac{n-L}{w}$,
$N = G \times L$
2. The hyper-parameter and parameter of the model are determined by the experimental results
while results on test dataset can still be improved **do**
Design the size of convolution kernels and the number of layers of the network.
Design the values of parameters including the size of mini-batch, the value of epochs, etc.
Design the length of the input single L:
$R^{(k)} = R^{(k-1)} + (f^{(k)} - 1) \times \prod_{i=1}^{k-1} S^{(i)}$,
The scope of the $R^{(n)}$: $T \leq R^{(n)} \leq L$
The length of L is determined by the above formula.
program runs the WKCNN model and gets the classification results.
return results

3.1 Time complexity of WKCNN

The training time of the model is determined by the time complexity. If the time complexity is too high, the model

The time complexity of FC layers and pooling layers is not included in the above formulation. These layers usually take 5-10% of the calculation time (He and Sun 2015). As mentioned in Sect. 3.3, widening convolution kernels is

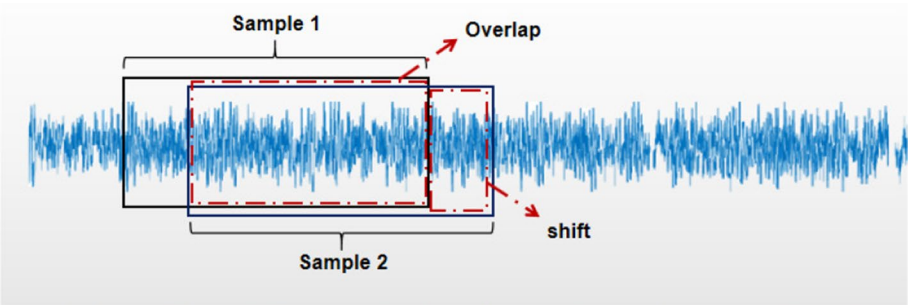


Fig. 2 Data augmentation through overlap sampling

the basic strategy for WKCNN to improve accuracy, that is, K value and receptive field increase, so L also increases accordingly. In order to reduce the time complexity, we should reduce the number of channels C on the premise of ensuring the accuracy of the model. The structure of WKCNN is shown in Table 6.

3.2 Data augmentation–data augmentation strategy

The best way to enhance the generalization of machine learning models is to use more training samples (Goodfellow et al. 2016). The purpose of data augmentation is to improve the generalization performance of deep neural network by increasing training samples.

As shown in Fig. 2, for the one-dimensional time-domain vibration signal, this paper adopts the data augmentation method of overlapping sampling, that is, when the data is sampled from the original signal, each segment of the signal overlaps with the next segment of the offset signal, which is the part of data augmentation.

Let the number of data sampling points in a file is n, the length of each training sample collected is L, and the shift is w. The number of sample groups is G and the total data quantity N can be obtained are as follows:

$$G = \frac{n - L}{w}, \tag{6}$$

$$N = G \times L, \tag{7}$$

for example, we can assume that n=10000, L=2000, and w=20. According to formula (5) and formula (6), A total of 400 sets of samples (each containing 2,000 sample points) and 800,000 sample points can be obtained, the number of points has increased by 80 times.

In the WKCNN model data augmentation experiment, there are 10 categories, in each category n is equal to 120000 and L is equal to 5200, the load is 1hp. According to the experiment and formula (5) and (6), we can get the accuracy rate, maximum sample number and maximum number of points under different shifts in Table 1. In order to avoid random sampling errors, the experiment was repeated 20 times to average. It can be seen from the Table 1 that for the training set without data augmentation, the accuracy rate was only 41.18%, due to insufficient data and inadequate fitting. Through data augmentation, the offset parameter was controlled to be between 1–5200, and the accuracy rate was between 41.18% and 99.90%. In the following experiment, we selected shift is equal to 20 with the highest accuracy. In Sect. 5.1, we proved the importance of data augmentation again through experiments which can well meet the training needs of deep neural network.

3.3 Wide Kernel—one-dimensional data processing strategy

The traditional convolutional neural network can also be used for fault diagnosis, but it is not suitable for fault diagnosis of bearing. For a two-dimensional convolutional neural network, 3×3 is the smallest size that can capture the information of pixel 8 neighborhood. Using two 3×3 convolution kernels can get the same size of the perception field as one 5×5 convolution kernel, and the number of parameters is less. In this way, not only can the depth of the network be deepened, but also a larger receptive field can be obtained with fewer parameters, and inhibiting overfitting.

However, for the one-dimensional neural network structure, the two-layer 3×1 convolution structure only gets 5×1 receptive field at the cost of 6 weights, which turns the advantage into disadvantage. Therefore, widening convolution kernel can be used as an effective strategy to improve one-dimensional convolutional neural network. In 3.1, this paper will also prove the advantages of wide convolution kernel through experiments.

3.4 Receptive field—input length selection strategy

One of the most important design basis of convolutional neural network is the receptive field, that is, the perception ranges of a neuron in its next layer. The size of the receptive field directly affects the level of the network

Table 1 Data augmentation experiment information under different shifts

Tag	1	2	3	4	5	6	7	8	9
Shift	None	1	20	50	100	200	500	1000	5200
MaxSampleNums	230	1148000	57400	22960	11480	5740	2296	1148	230
MaxPointNums	120000	5969600000	298480000	119392000	59696000	29848000	11939200	5969600	120000
TrainNums	161	7000	7000	7000	7000	3500	1400	700	161
TestNums	23	1000	1000	1000	1000	500	200	100	23
Accuracy	41.18%	99.68%	99.90%	99.81%	99.39%	94.98%	77.49%	65.25%	41.18%

layer's perception of the features of the input image, that is, whether the features acquired by a certain layer are global, or local and detailed. When the receptive field is smaller, the response is more detail-oriented. When the receptive field becomes larger, the reflected features are more holistic and global. The formula of receptive field is as follows:

$$\mathrm{R^{(k)}} = \mathrm{R^{(k-1)}} + \left(\mathrm{f^{(k)}} - 1\right) \times \prod_{\mathrm{i=1}}^{\mathrm{k-1}} \mathrm{S^{(i)}}, \tag{8}$$

$R^{(k)}$ is the perceptive field size of the k-th layer, $f^{(k)}$ is the size of the current convolution kernel, and $S^{(i)}$ is the step size of the i-th layer, and the value of $R^{(0)}$ is 1.

After determining the convolution kernel size and the step size of each layer, we can calculate the size of the receptive field of each layer. Because the vibration signal is periodic, and the phase value of each input signal may not be the same. Therefore, the field size of the last pooling layer should be at least greater than the number of sampling points recorded in one cycle. Let the field size of the last pooling layer at the input layer is $R^{(n)}$,T is the number of data points recorded by the bearing rotation, and L is the length of the input signal.Therefore, $\mathrm{T} \le \mathrm{R^{(n)}} \le \mathrm{L}$ should be used as the design criterion.

3.5 Batch normalization—inhibit overfitting strategy

The normalization layer of batch processing is to reduce the displacement of internal covariance, inhibit overfitting, and enable rapid learning (increase the learning rate). The BN layer is usually added after the convolution layer or the fully connected layer and before the activation function. In this paper, a batch normalization layer is added between the convolution layer and the activation layer. Specifically, the mean value of the data distribution is 0 and the variance is 1. In mathematical terms, as shown below.

$$\mu_{\mathrm{B}} = \frac{1}{\mathrm{m}} \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{x_i}, \tag{9}$$

$$\sigma_{\mathrm{B}}^2 = \frac{1}{\mathrm{m}} \sum_{\mathrm{i=1}}^{\mathrm{m}} \left(\mathrm{x_i} - \mu_{\mathrm{B}}\right)^2, \tag{10}$$

$$\hat{\mathbf{x}}_{\mathrm{i}} = \frac{\mathbf{x}_{\mathrm{i}} - \mu_{\mathrm{B}}}{\sqrt{\sigma_{\mathrm{B}}^2 + \varepsilon}}, \tag{11}$$

$$\mathrm{y_i} = \gamma \hat{\mathrm{x}}_{\mathrm{i}} + \beta, \tag{12}$$

here for the set of m input data B= $\{x_1, x_2, \ldots, x_m\}$ to find the mean μ_B and the variance σ_B^2 . Where ε is a small value, (for example, 10e−7, etc.), γ and β are parameters.

3.6 Adam—avoid local optimality strategy

For shallow neural networks, SGD (Stochastic Gradient Descent) can converge to the global optimum. However, for WKCNN, due to the deep layers and too many parameters, it is easy to fall into local optimization if the parameters are not selected well. Therefore, this paper adopts Adam (adaptive moments) algorithm (Kingma and Ba 2014). Adam is an adaptive learning rate optimization algorithm, which dynamically adjusts the learning rate of each parameter by using the first-order moment estimation and second-order moment estimation of the gradient. The main advantage of Adam is that after bias correction, the estimation of the first moment (momentum term) and second moment (non-central) initialized from the origin is modified, so that the learning rate of each iteration is within a certain range. Adam is usually robust to the selection of parameters, so it is very helpful to the parameter adjustment of neural network.

4 Determine the structure of WKCNN model by experiments

This section introduces the experimental data and experimental environment, in addition determines the size of WKCNN's wide convolution kernel and the number of model layers through two experiments.

4.1 Data source

The experimental data used in this paper is from the rolling bearing database center of Case Western Reserve University(CWRU) in the United States. The CWRU bearing data sampling system is shown in Fig. 3(Fig. 3 cited

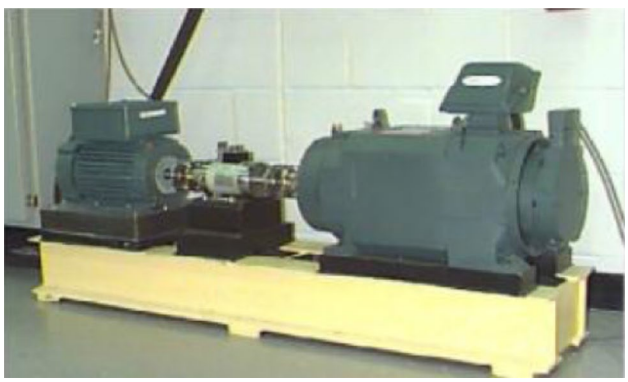


Fig. 3 CWRU data sampling system

from https://csegroups.case.edu/bearingdatacenter/pages/apparatus-procedures).

The data used in the experimental tests is the fan bearing failure data of model SKF6205 in the CWRU database. The sampling frequency is 12 kHz, and the load characteristics are 1HP, 2HP, 3HP (HP, Horsepower). The bearings diagnosed have three types of defect positions, namely ball damage, outer race damage and inner race damage. The diameters of the damage are 0.007 inch, 0.014 inch and 0.021 inch respectively, and there are a total of 9 damage states.

From the analysis of a single data file in CWRU, it can be seen that the data volume of each file is about 120000 sampling points. That is, according to the sampling frequency of the sampling system of 12 kHz and the speed of 1800 r/min, the data collection duration can be calculated to be about 10 seconds, and it can also be calculated that about 400 sampling points are collected for each rotation of the bearing. Therefore, in order to avoid the influence of uncertainty caused by accidental factors, this paper used 5200 sampling points of data, that is, the data generated by 13 rounds of bearing rotation, to make a single sample for training. Since the data collected by the bearing is periodic, in order to make full use of the data and avoid overfitting in the training process the data enhancement method mentioned in 2.1 is used to expand the data.

4.2 Model determination experiments

Using Tensorflow and Keras framework to build WKCNN fault diagnosis model in Python3.7 environments. As shown in Table 2, during training, the size of mini-batch is 256 in a range from 32 to 1024, the learning rate of Adam algorithm is 0.001 ranging from 0.0001 to 1, and epochs is 20, and the test restult under different epochs are shown in Fig. 8. The length of input data in each group is 5200, the training data set is 7000, and the test data set is 1000. The fault diagnosis classification involved in this experiment is divided into 10 types, 9 types of fault bearings and one fault-free bearing. The specific label classification is shown in Table 3.

Table 2 The optimal value and ranges of experimental parameters

Parameters	Scope	Accuracy	Optimal value
Mini-batch	32 ~ 1024	96.20% ~ 99.90%	256
Adam	0.0001 ~ 1	63.80% ~ 99.90%	0.01
epochs	1 ~ 100	28.60% ~ 99.90%	20
Training data set	70 ~ 14000	40.00% ~ 99.90%	7000
Test data set	10 ~ 2000	40.00% ~ 99.90%	1000

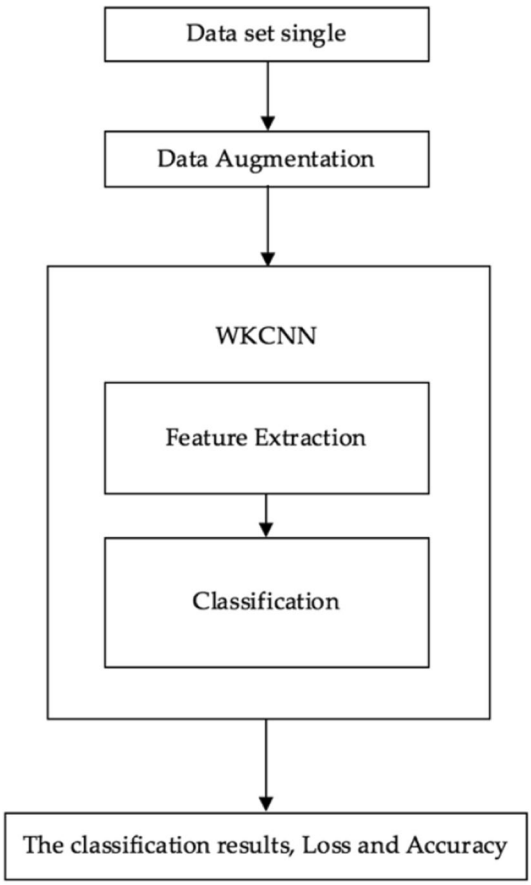


Fig. 4 Fault diagnosis flow

The whole process includes data augmentation, model training (feature extraction and classification) and model diagnosis. The flow of WKCNN model determination is shown in Fig. 4, the model of WKCNN is shown in Fig. 6 and the structure of the WKCNN is shown in Table 6.

4.2.1 The test results under different size of kernels to prove the importance of wide convolution kernel

Only 1HP load is used in this experiment. The network model contains five convolution layers, five pooling layers.

Table 3 Description of 12 kHz rolling element bearing datasets

Fault location	none	Inner race			Outer race			Ball		
Tag	0	1	2	3	4	5	6	7	8	9
Loss of diameter(inch)	None	0.007	0.014	0.021	0.007	0.014	0.021	0.007	0.014	0.021
Train(group)	700	700	700	700	700	700	700	700	700	700
Test(group)	100	100	100	100	100	100	100	100	100	100

Table 4 Results of the first two convolution layers with different sizes

First layer	Second layer	Accuracy (%)
5	5	79.29
16	5	83.30
32	5	86.89
64	5	88.50
128	5	89.29
5	16	86.89
16	16	89.80
32	16	95.20
64	16	99.90
128	16	97.69
5	32	88.89
16	32	91.89
32	32	95.10
64	32	98.50
128	32	97.50

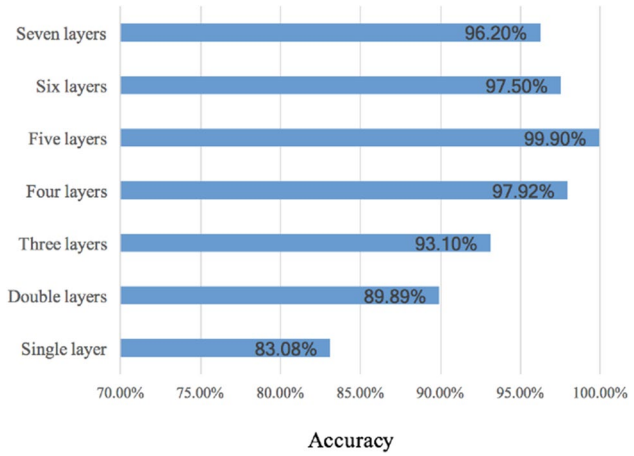


Fig. 5 Accuracy under different layer structure

The width of convolution kernels of the last three layers are 5, and the width of pooling layers are all 2. The necessity of using wide convolution kernels is proved by changing the size of kernels in the first two convolution layers. It can be seen from Table 4 that the accuracy rate is higher when the width of the first two layers of convolution kernel becomes large. For example, when the width of the first two-layer convolution kernel is both 5, the accuracy rate is only 79.29%. When the first-layer convolution kernel size is 128 and the second-layer size is 32, the accuracy rate can reach 97.50%. However, it is not that the wider the convolution kernel is, the higher the accuracy will be. When the width of the convolution kernel is too large, the time domain resolution will be reduced, resulting in the loss of some details. When the convolution kernel is too small, it is difficult to capture the medium-low frequency features, which may be interfered by high-frequency noise in industry. It can be seen from the experiment that when the size of the first two layers of convolution kernels is 64 and 16 respectively, the diagnostic accuracy is the highest, which can reach 99.90%.

4.2.2 The test results under different numbers of layers to determine the optimal number of layers

This experiment illustrates the optimality of the five-layers structure (five convolutional layers, five pooling layers and a fully connected layer) by changing the number of WKCNN layers. In this experiment, single-layer structure to seven-layer structure are used for comparison. As shown in Fig. 5, five-layer structure can reach 99.90% accuracy, which is higher than other structures. In Table 5, the time for WKCNN to diagnose a signal is 0.442ms, which can well meet the real-time demand. Through this experiment, it can be concluded that appropriately deepening the network layer of WKCNN can obtain stronger feature extraction ability than the shallow layer, but excessive layers will lead to overfitting.

4.3 Structure of the WKCNN Model

In this paper, the model WKCNN (Wide Kernel Convolutional Neural Networks) is proposed by the experiments of 3.2.1 and 3.2.2, as shown in Fig. 6. The network consists of five convolutional layers, five pooling layers, one fully connected layer and one Softmax layer. The time-domain vibration signal (after data augmentation) passes through the first convolution layer and then enters the Batch Normalization layer (BN layer) and the ReLU activation layer, becoming a set of feature maps, and then performs the down-sampling operation through the maximum pooling. As the number of layer increases, the width of output signal decreases. The signal classification operation is completed by the fully connected layer. Flatten expands all the data features obtained in the last pooling layer, and the Dense layer performs nonlinear transformation to extract the correlation among these features. Finally, use the Softmax function to output 10 different health states that meet the experimental requirements.

Table 5 The processing time of single signal

Layers	Single layer	Double layer	Three layer	Four layer	Five layer	Six layer	Seven layer
Time(ms)	0.367	0.403	0.420	0.432	0.442	0.454	0.476

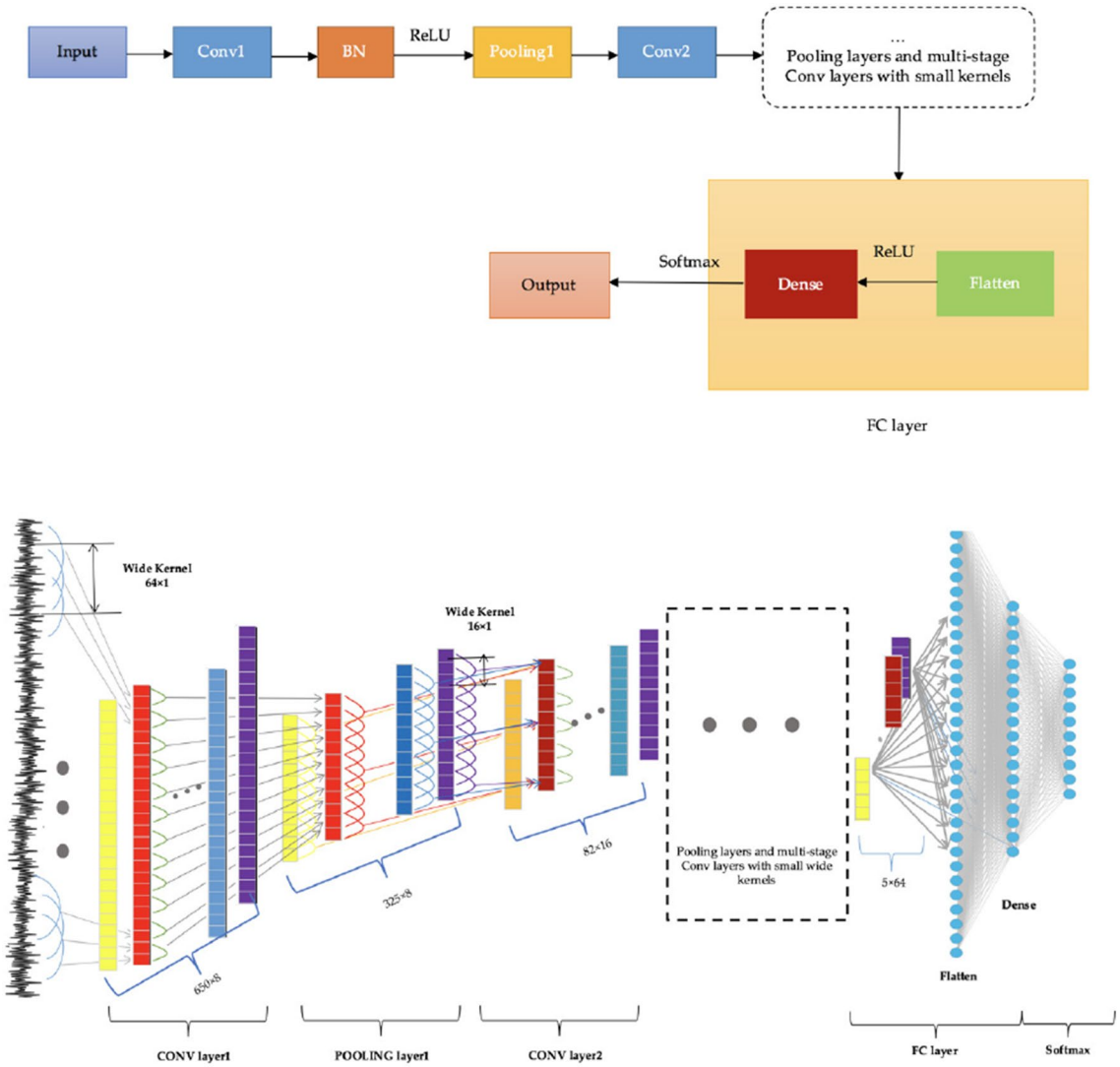


Fig. 6 The model of WKCNN

As shown in Table 6, the size of the first convolution kernel is 64*8 (width, depth, the height is none), the second layer is 16*16, the third and fourth layers are 5*32, and the fifth layer is 5*64. The area size (width) of the pooling layer is all selected as 2. The number of hidden neurons in the fully connected layer is 64. In the process of back propagation, Adam optimization algorithm is selected to update the weight, so as to minimize the value of the loss function.

The first two layers of WKCNN are large convolution kernels, the purpose of which is to extract short-time features, and its function is similar to that of short-time Fourier transform. The difference is that the window of the short-time Fourier transform is the sine function, while the first two layers of the large convolution kernel of WKCNN are trained by the optimization algorithm. The advantage is that it can automatically learn the features that are diagnostic oriented, and automatically remove the features that are not helpful for diagnosis. There is not only high accuracy but also greatly improves the learning speed of the model. In order to enhance the expressive power of WKCNN, except for the first two layers, the convolution kernel of the other convolution layers adopts a small convolution kernel with width of 5. Since there are few parameters in the small convolution kernel, it is beneficial to deepen the network and inhibit overfitting.

Table 6 The structure of WKCNN

No.	Layer type	Kernel size/ stride	Kernel number	Output size (width× depth)	Padding
1	Conv1	64/8	8	650 × 8	Yes
2	Pooling1	2/2	8	325 × 8	No
3	Conv2	16/4	16	82 × 16	Yes
4	Pooling2	2/2	16	41 × 16	No
5	Conv3	5/1	32	41 × 32	Yes
6	Pooling3	2/2	16	20 × 32	No
7	Conv4	5/1	32	20 × 32	Yes
8	Pooling4	2/2	64	5 × 64	No
9	Conv5	5/1	64	5 × 64	Yes
10	Pooling5	2/2	64	5 × 64	No
11	Flatten			320	
12	Dense			64	
13	Softmax			10	

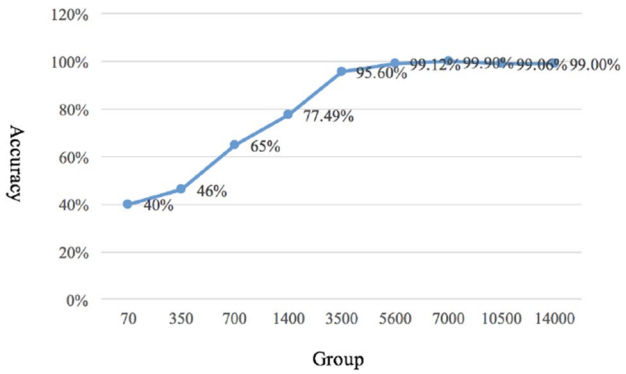


Fig. 7 The recognition rate of WKCNN under different training samples

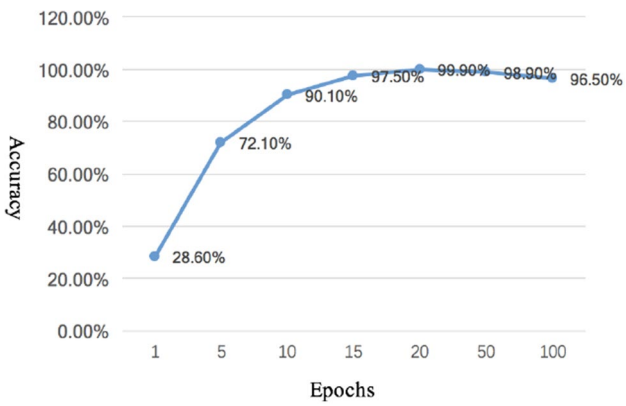


Fig. 8 The recognition rate of WKCNN under different epochs

5 Validation of the WKCNN model

5.1 The test results under different number of training datasets

WKCNN adopts the strategy of deep learning and broadening convolution kernels. To train a large number of parameters, sufficient training sample data is the prerequisite in WKCNN. In order to investigate how much training data is sufficient and how well does WKCNN perform in different data volumes, different sizes of training data are fed to train the network. In this experiment, the WKCNN model was trained on the training samples of 70, 350, 700, 1400, 3500, 5600, 7000, 10500, 14000 groups(The samples are randomly selected). In the training process, the load is 1hp, the size of mini-batch is 256, the learning rate of Adam algorithm is 0.001, and the epochs is 20.

The experimental results are shown in Fig. 7. When the training sample is 14000, the recognition accuracy rate is up to 99%, while when the training sample is 70 times, the accuracy rate is only 40%. The experimental results illustrate the influence of the number of training samples on the diagnostic accuracy. When the number of training samples exceeds 3500, the accuracy can reach more than 95.6%, and it is not difficult to find that with the increase of sample size, the accuracy rate increases significantly at the beginning, but after reaching a certain threshold, the improvement rate of accuracy begins to slow down, and then slowly declines after reaching the peak. With 7000 sample size, the accuracy is 99.90% which is the peek of the curve, while the number of samples exceeds 7000, the accuracy drops slightly, remaining around 99%. Through this experiment, it is not difficult to draw a conclusion that 7000 sets of training data is an ideal value. Not only has the highest accuracy rate, but also meets the requirements of WKCNN in terms of quantity and scale. When the training sample data exceeds 7000, the accuracy of the training sample will increase, and the overfitting problem leads to the decrease of the accuracy of the test set sample. In other words, underfitting occurs when the training dataset is too small, and overfitting may occur when the training dataset is too large. Thus, in the following experiments, the WKCNN model is trained with 7000 samples.

5.2 The test results under different epochs

One epoch means that the whole data set is passed forward and backward only once in the neural network. For a large training dataset in a neural network, only one transmission is not enough to obtain accurate experimental results. The complete dataset needs to be transmitted multiple times

in the same neural network, with the number of epochs increases, more number of times the weight are changed in the neural network and the test dataset goes from underfitting to optimal to overfitting. Therefore, we need to determine the optimal epochs under the given parameters which the number of the training dataset is 7000, the test data set is 1000, the load is 1hp, the size of mini-batch is 256, the learning rate of Adam algorithm is 0.001.

In this experiment, the values of epochs in the WKCNN model were adjusted to 1, 5, 10, 15, 20, 50, 100. In Fig. 8, it shows that the accuracy increases by 68.90% when the epochs rise from 1 to 15. With 20 epochs, the accuracy peaks at 99.90% . When the epochs go up to 50, the recognition rate is 98.90%, while the epochs is 100, the accuracy is only 96.50%.

Obviously, the experimental results well confirm our previous conclusions. When epochs is between 1 and 20 times, the model is still underfitting. After more than 20 times, the error of the training dataset decreases and that of the test dataset increases which is overfitting. Therefore, we can determine that 20 is the optimal number of epochs. If the number of epochs is too small, it will lead to underfitting, but if the number of epochs is too high, it will lead to overfitting. Compared with Li et al. (2018) (1000 epochs), WKCNN can achieve a high recognition rate with less epochs. It can be inferred that WKCNN can learn features in one-dimensional data faster.

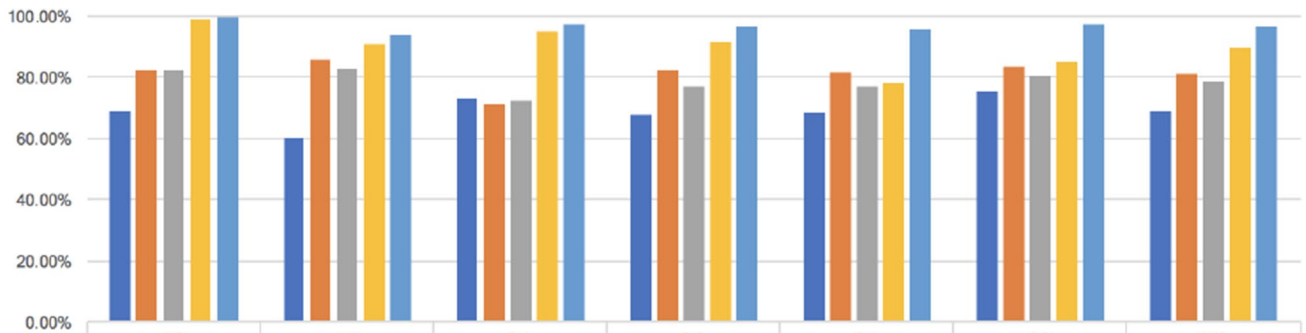
5.3 The test results under different loads and variable loads conditions

In this experiment, we studied the accuracy of WKCNN under different loads and variable loads conditions. In the training process, the number of the training dataset is 7000, the test dataset is 1000, the size of mini-batch is 256, the learning rate of Adam algorithm is 0.001, and the epochs is 20.

The fault recognition rate of WKCNN was tested under loads of 0HP, 1HP, 2HP and 3HP respectively. As shown in Table 6, it can be seen that the recognition rate of convolutional neural network on each data set reached over 99.90%, and the accuracy of test results of this model on 2HP and 3HP could reach 100%.

Next, training set A, B and C represents the training data under the load of 1-3HP respectively. The variable load capacity of WKCNN is compared with three traditional classification methods and a new deep learning method which do not need feature extraction. The classification methods of SVM (Islam and Kim 2017; Ziani et al. 2017; Fu et al. 2020), MLP (Almeida et al. 2014) and DNN (Feng et al. 2016) that need pretreatment are compared. The data is transformed fast Fourier transform(FFT). The last choice of comparison method is WDCNN (Zhang et al. 2017b), which also does not need preprocessing. There are two comparison points for the variable load problem in this experiment. The first is the advantages of deep learning methods compared with traditional methods, and the second is the advantages of WKCNN compared with another efficient deep learning algorithm.

According to Fig. 9, we can find that the accuracy of the three traditional intelligent diagnosis methods under different load conditions is lower than the adaptive feature extraction method based on convolutional neural network, which is mainly due to the poor applicability of the manually designed extracted features and the non-linear expression ability of SVM, which limits the recognition rate under different load conditions. Although MLP and DNN have relatively strong fitting ability, their generalization



	A-B	A-C	B-A	B-C	C-A	C-B	AVG
FFT-SVM	68.60%	60.00%	73.20%	67.60%	68.40%	75.32%	68.85%
FFT-MLP	82.10%	85.60%	71.50%	82.40%	81.80%	83.20%	81.10%
FFT-DNN	82.20%	82.60%	72.30%	77.00%	76.90%	80.30%	78.55%
WDCNN	99.20%	91.00%	95.10%	91.50%	78.10%	85.10%	90.00%
WKCNN	99.60%	94.00%	97.10%	96.50%	95.80%	97.30%	96.72%

Fig. 9 Comparison of different methods under variable load condition

ability is low, so the diagnostic accuracy under different loads needs to be further improved. WKCNN relies entirely on the one-dimensional convolutional neural network to automatically extract and classify features in an end-to-end manner without excessive manual intervention. Compared to traditional methods, the process of feature extraction is eliminated, and the hidden features retained in the sample can be better discovered. Compared to WDCNN, each convolution kernel of WKCNN is wider. So we can get a larger receptive field which the features obtained by WKCNN are more global, thus the overfitting can be suppressed more effectively. From the change of working conditions, the diagnostic accuracy between 1 HP and 3 HP was significantly lower than that between 2 HP. This shows that the greater the load change, the greater the signal difference in the same health state.

5.4 The test results under white noise

In the actual operation of bearing, there is external interference noises usually. The noise of the diagnosed signal is generally additive Gaussian white noise (Gondal et al. 2014), and the signal-to-noise ratio(SNR) is the standard of evaluating the strength of the noise. Let P_S and P_N represent the energy of signal and noise respectively, the definition of SNR is as follows:

$$\mathrm{SNR(dB)} = 10\log_{10}\left(\frac{P_S}{P_N}\right), \tag{13}$$

according to formula (12), the larger the noise, the smaller the SNR. When the signal and the noise energy are the same, the SNR is 0. Therefore, in this experiment, the training set was added with Gaussian additive white noise with SNR value of 0-10db to detect the noise resistance of WKCNN. As shown in Table 7, WKCNN was compared with FFT-SVM (Islam and Kim 2017), FFT-MLP (Almeida et al. 2014) and FFT-DNN (Feng et al. 2016) and WDCNN(Zhang et al. 2017b). In this experiment, the training data set is 7000, the test data set is 1000, the size of mini-batch is 256 and epochs is 20.

We compare the accuracy of FFT-MLP and FFT-DNN in the traditional methods in the case of the highest noise of 0dB, which is only 41.50% and 58.52% respectively, and the accuracy of FFT-SVM is 89.50%. However, the two methods of deep learning achieve more than 98% recognition accuracy without any denoising pre-processing which proves again that compared with traditional methods deep learning methods learn more hidden features through an end-to-end approach and more powerful learning ability. Compared with WDCNN, the performances of WKCNN from 0–10 dB are slightly higher than WDCNN. The reason is that its overall wider convolution kernels can effectively avoid overfitting.

Table 7 The recognition rate of WKCNN under different loads

Load	0HP	1HP	2HP	3HP
Accuracy	99.90%	99.90%	100%	100%

5.5 Performance evaluation of WKCNN

When considering whether a model is appropriate, it is not sufficient to rely solely on accuracy rate (Hamori et al. 2018). We usually use Precision, Recall, F1-Measure and receiver operating characteristic(ROC) curve to evaluate a model. We set the training data set to 7000, the test data set to 1000. The size of the small batch is 256 and the epochs is 20.

As we mentioned in 3.1, the bearings diagnosed have three types of defect positions, namely ball damage, outer race damage and inner race damage. The diameters of the damage are 0.007 inch, 0.014 inch and 0.021 inch respectively, and there are a total of 9 damage states with a normal state that correspond exactly to the 10 categories of states in Table 8. As shown in Table 7, we can find that the values of Precision, Recall and F1-measure of each type. All values except B021 and IR007 reached 100%. In the bottom half of Table 7, the two averages of Macro Avg and Weight Avg are given. Macro Avg means to average each category's Precision, Recall and F1-Measure sum. Weight Avg is an improvement on Macro Avg, considering the proportion of samples for each category in the total sample. The experimental result shows that the average values of Precision,

Table 8 Comparison of noise resistance of different methods

Model methods	SNR					
	0 dB (%)	2 dB (%)	4 dB (%)	6 dB (%)	8 dB (%)	10 dB (%)
FFT-MLP	41.50	50.34	78.65	92.35	97.24	99.38
FFT-DNN	58.52	70.24	85.34	95.78	98.20	99.50
FFT-SVM	89.50	96.38	97.23	98.52	99.00	99.52
WDCNN	98.77	99.49	99.67	99.80	99.81	99.88
WKCNN	98.90	99.52	99.77	99.82	99.86	99.89

Table 9 Bearing performance diagnosis with precison, recall and F1-measure

Bearing classification	Precision (%)	Recall (%)	F1-measure (%)	Support
B007	100	100	100	100
B014	100	100	100	100
B021	96	93	94	100
IR007	93	96	95	100
IR014	100	100	100	100
IR021	100	100	100	100
OR007	100	100	100	100
OR014	100	100	100	100
OR021	100	100	100	100
Normal	100	100	100	100
Avg				
Macro avg	99	99	99	1000
Weight avg	99	99	99	1000

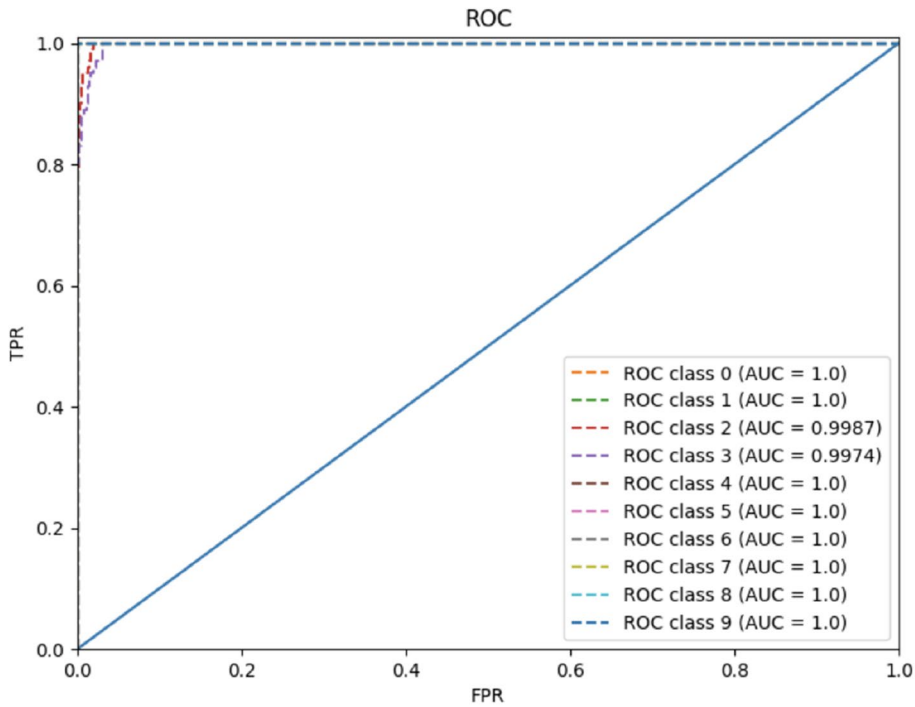


Fig. 10 ROC curve for WKCNN

Recall rate and F1-Measure are all up to 99% under two average methods.

Figure 10 displays ROC curve with area under the curve(AUC) for the WKCNN, the vertical axis corresponds to the true positive ratio, whereas horizontal axis corresponds to the false positive ratio. The blue line is the random 50%/50% classification. A good model is one that shows a high true positive rate value and low false positive value. When one curve is completely enveloped by another, it can be asserted that the latter performs better than the former. It is found that, in the ten classifications, the AUC of class 2 is 0.9987 and that of class 3 is 0.9974. The rest of the classifications reached 1, thus proving the high performance of WKCNN in diagnosing bearing faults (Table 9).

6 Conclusions

In order to improve the accuracy and efficiency of bearing fault diagnosis, we use the most popular deep learning technology and combine the characteristics of one-dimensional vibration signal to propose the WKCNN model. At the same time, we summarize the generation algorithm of WKCNN model and the design process of one-dimensional convolutional neural network.

The core idea of WKCNN model is to widen the convolution kernels. On this basis, deep learning and other optimization algorithms are combined, which will be the most suitable optimization algorithm for one-dimensional convolution network.

In order to verify the superiority of WKCNN algorithm, firstly, we compared the accuracy changes of WKCNN in different training samples and epochs. We draw a conclusion that, for deep learning training, a relatively large amount of training data is needed. However, if the amount of training data is too large, overfitting will occur; if the amount of training data is too small, underfitting will occur. Epochs is also the same, the appropriate value is very important. In these two experiments, 99.90% accuracy of fault diagnosis can be achieved under appropriate parameters.

Secondly, we compare WKCNN with three traditional and efficient classification algorithms SVM, MLP, DNN(feature extraction technology uses fast Fourier transform) and WDCNN algorithm, which also uses deep learning technology and does not require feature extraction. Next, we compare the accuracy of five algorithms under variable loads and noise conditions respectively, and draw two conclusions. The first one is that, compared with traditional methods, WKCNN automatically extracts and classifies features end-to-end through one-dimensional convolutional neural network, without too much manual intervention. Compared with the traditional methods, this method eliminates the process of feature extraction and better reveals the hidden features retained in the sample. The second point is compared to other convolutional neural network, each convolution kernel of WKCNN is wider. In this way, a larger receptive field can be obtained, and the features obtained by WKCNN are more global, thus inhibiting overfitting more effectively.

Finally, in terms of the efficiency of the model, compared with the traditional method, the feature extraction of data is not required, and the end-to-end method is directly used to classify the data. The processing time of a single signal is 0.442 ms, which provides real-time guarantee for the arrival of the era of big data.

In the future research, this method will be combined with other methods and extended to the complex classification and regression problems such as health status assessment and residual life prediction of rotating machinery. In addition, parallel computing technology will be studied to improve the efficiency of the method.

Acknowledgements This work has been supported by Liaoning Provincial Natural Science Foundation of China (No. 2019-ZD-0105).

References

Abdel-hamid O, rahman Mohamed A, Jiang H, Penn G (2012) Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition

Almeida LFD, Bizarria JW, Bizarria FC, Mathias MH (2014) Condition-based monitoring system for rolling element bearing using a generic multi-layer perceptron. J Vib Control. https://doi.org/10.1177/1077546314524260

Babu GS, Zhao P, Li XL (2016) Deep convolutional neural network based regression approach for estimation of remaining useful life

Chen Y, Peng G, Xie C, Zhang W, Li C, Liu S (2018) Acdin: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis. Neurocomputing. page S092523121830300X

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. Proc of IEEE Comput Vis Pattern Recognit 248–255

Feng J, Lei Y, Jing L, Xin Z, Na L (2016) Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech Syst Signal Process 72–73:303–315

Frosini L, Harlisca C, Szabo L (2015) Induction machine bearing fault detection by means of statistical processing of the stray flux measurement. IEEE Trans Ind Electron 62(3):1846–1854

Fu W, Shao K, Tan J, Wang K (2020) Fault diagnosis for rolling bearings based on composite multiscale fine-sorted dispersion entropy and svm with hybrid mutation sca-hho algorithm optimization. IEEE Access 99:1

Gondal I, Amar M, Wilson C (2014) Vibration spectrum imaging: a novel bearing fault classification approach. IEEE Trans Ind Electron 62:9

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. The MIT Press, Cambridge

Hamori S, Kawai M, Kume T, Murakami Y, Watanabe C (2018) Ensemble learning or deep learning? application to default risk analysis

Hatami N, Gavet Y, Debayle J (2017) Classification of time-series images using deep convolutional neural networks

He K, Sun J (2015) Convolutional neural networks at constrained time cost. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

He K, Zhang X, Ren S, Sun J(2016) Identity mappings in deep residual networks

Ince T, Kiranyaz S, Eren L, Askar M, Gabbouj M (2016) Real-time motor fault detection by 1-d convolutional neural networks. IEEE Trans Industr Electron 63(11):7067–7075

Islam MMM, Kim JM (2017) Time–frequency envelope analysis-based sub-band selection and probabilistic support vector machines for multi-fault diagnosis of low-speed bearings. J Ambient Intell Hum Comput

Janssens O, Slavkovikj V, Vervisch B, Stockman K, Loccufier M, Verstockt S, Rik VDW, Van Hoecke S (2016) Convolutional neural network based fault detection for rotating machinery. J Sound Vib 2:331–345

Jayaswal P, Verma S, Wadhwani A (2011) Development of ebp-artificial neural network expert system for rolling element bearing fault diagnosis. J Vib Control 17(8):1131–1148

Jian X, Li W, Guo X, Wang R (2019) Fault diagnosis of motor bearings based on a one-dimensional fusion neural network. Sensors 19:1

Jing L, Zhao M, Li P, Xu X (2017) A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox

Kim Y (2014) Convolutional neural networks for sentence classification. Eprint Arxiv

Kim J, Khan S, Ali U, Sharif Islam Md R (2016) Distance and density similarity based enhanced k-nn classifier for improving fault diagnosis performance of bearings. Shock Vib

Kingma D, Ba J (2014) Adam: a method for stochastic optimization. Comput Sci

Krummenacher G, Ong CS, Koller S, Kobayashi S, Buhmann JM (2018) Wheel defect detection with machine learning. IEEE Trans Intell Transport Syst 2:1176–1187

Li Y, Xu M, Wei Y, Huang W (2016) A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. Measurement

Li X, Zhang W, Ding Q, Sun J-Q (2018) Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. J Intell Manuf

Manish K, Saini A, Aggarwal (2018) Detection and diagnosis of induction motor bearing faults using multiwavelet transform and naive bayes classifier. Int Trans Electr Energy Syst

Nie M, Zhao Q, Bi S, Xu Y, Shen T (2019) Apple external quality analysis based on bp neural network. International Conference on Industrial Artificial (Intelligence)

Park JK, Kwon BK, Park JH, Kang DJ (2016) Machine learning-based imaging system for surface defect inspection. Int J Precis Eng Manuf Green Technol 3(3):303–310

Safin NR, Prakht VA, Dmitrievskii VA, Dmitrievskii AA (2016) Stator current fault diagnosis of induction motor bearings based on the fast fourier transform. Russian Electr Eng 87(12):661–665

Teti R, Jemielniak K, O'Donnell G, Dornfeld D (2010) Advanced monitoring of machining operations. CIRP Ann Manuf Technol 59(2):717–739

Wan S, Zhang X (2018) Teager energy entropy ratio of wavelet packet transform and its application in bearing fault diagnosis. Entropy 20(5):388

Wang Z, Oates T (2015) Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. Workshops at the Twenty-ninth Aaai Conference on Artificial (Intelligence)

Wang J, Zhuang J, Duan L, Cheng W (2016) A multi-scale convolution neural network for featureless fault diagnosis. In: 2016 International Symposium on Flexible Automation (ISFA)

Wu D, Jennings C, Terpenny J, Gao RX, Kumara S (2017) A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests. J Manuf Sci Eng 139(7):071018

Xiang L, Zhang W, Qian D (2018) Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks. IEEE Trans Ind Electron 9:1

Xiao Y, Fei D, Ding E, Wu S, Fan C (2017) Rolling bearing fault diagnosis using modified lfda and emd with sensitive feature selection. IEEE Access 99:1

Xu G, Liu M, Jiang Z, Söffker D, Shen W (2019) Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. Sensors 19:5

Xue X, Li C, Cao S, Sun J, Liu L (2019) Fault diagnosis of rolling element bearings with a two-step scheme based on permutation entropy and random forests. Entropy 21:1

Yan W, Yu L (2019) On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach

Yiakopoulos CT, Gryllias KC, Antoniadis IA (2011) Rolling element bearing fault detection in industrial environments based on a k-means clustering approach. Expert Syst Appl 38(3):2888–2911

Zhang W, Peng G, Li C (2017a) Rolling element bearings fault intelligent diagnosis based on convolutional neural networks using raw sensing signal

Zhang W, Peng G, Li C, Chen Y, Zhang Z (2017b) A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors 17:425

Ziani R, Felkaoui A, Zegadi R (2017) Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized fisher's criterion. J Intell Manuf 28(2):405–417