

Article

An Improved of Fault Diagnosis Using 1D-Convolutional Neural Network Model

Chih-Cheng Chen^{1,2}, Zhen Liu¹, Guangsong Yang^{1,*}, Chia-Chun Wu^{3,*} and Qiubo Ye¹

¹ School of Information and Engineering College, Jimei University, China; 201761000018@jmu.edu.cn

² Department of Industrial Engineering and Management, Chaoyang University of Technology, Taichung 413310, Taiwan; ccc@gm.cyut.edu.tw

³ Department of Industrial Engineering and Management, National Quemoy University, Kinmen 892, Taiwan; ccwu0918@nqu.edu.tw

* Correspondence: gsyang@jmu.edu.cn; ccwu0918@nqu.edu.tw

Abstract: The diagnosis of a rolling bearing for monitoring its status is critical for maintaining industrial equipment using rolling bearings. The traditional method of diagnosing faults of the rolling bearing has low identification accuracy, which needs artificial feature extraction to enhance the accuracy. 1D-CNN method not only can diagnose bearing faults accurately but also overcome shortcomings of the traditional methods. Different from machine learning and other deep learning models, the 1D-CNN method does not need pre-processing one-dimensional data of rolling bearing's vibration. Thus, it enhances the processing speed and improves the network structure to have a reasonable design for small sample data sets. This study proposes and tests a 1D-CNN method for diagnosing rolling bearings. By introducing the dropout operation, the method obtains high accuracy and improves the generalizing ability. The experimental results show 99.52% of the average accuracy under a single load and 98.26% under different loads.

Keywords: 1D-CNN; fault diagnosis; rolling bearing; vibration signal; single load; different load

1. Introduction

Recent industrial equipment uses many rolling bearings that playing a significant role in mechanical transmission. The failure of rolling bearing makes mechanical equipment unable to operate normally and efficiently, reduce safety, and shorten the service life. 45 - 55% of mechanical faults are caused by the faults of damaged bearings [1]. The traditional diagnosis method demands mechanical knowledge and expertise and has a time-consuming process. However, the method still cannot guarantee consistency, so does not meet the requirements of the diagnosis. Therefore, in order to ensure the reliable and normal operation of industrial equipment, the application of intelligent monitoring technology is needed [2,3]. The recognition of rolling bearings' faults is based on modeling or monitoring signals. A model-based diagnosis uses a mathematical model that simulates a real system with necessary assumptions and compares the data of the monitoring system to the mathematical model to predict the faults of rolling bearings [4]. A signal-based method extracts feature information

from time-domain [5] and frequency-domain signal [6] by using time-frequency domain analysis technology and diagnoses the faults [7]. The time-frequency analyses of signals of rolling bearing's vibration include wavelet analysis [8], SHORT-time Fourier transformation (STFT) [9], Empirical Mode Decomposition [10], and singular value decomposition [11].

The time-frequency analyses adopt a machine learning method for diagnosing rolling bearings' faults which extracts fault characteristics from collected rolling bearing vibration signals and then, divide the features into several category. Commonly used machine learning methods for the failure recognition are support vector machine (SVM) [12], k-nearest neighbor (KNN) [13], and K-means clustering. Continuous wavelet transform overcomes the shortcomings of the traditional method using Fourier transform. Thus, Wang et al. [14] designed a support vector machine (SVM) classifier based on vibration signal analysis; Georgoulas et al. [15] suggested a symbol aggregation approximation architecture to extract the required bearing fault characteristic from the azimuth signal and use k-nearest neighbor (KNN).

As classification accuracies of these methods mainly depended on the feature extraction steps, the feature extractors should be redesigned for different fault types. However, the algorithms only had simple structures and could learn non-linear relation in complex vibration signals from the rolling bearing. Thus, a deep learning method has an advantage in analyzing complex and non-stationary signals as it autonomously extracts fault features from the signals. Several researches on fault recognition of rolling bearings have used the means of deep learning recently. Yin et al. [16] extracted 38 bearing characteristics of original vibration signals using analysis methods in time domain and frequency domain and wavelet transform by using a nonlinear global algorithm. The low dimensional feature matrix of the input is processed by deep belief network (DBN) for fault feature recognition. Liu et al. [17] used the stacked sparse Auto-Encoder (SAE) to extract bearing failure characteristics from the spectrum of vibration signals with Softmax regression and classified the fault types. Liu et al. [18] adopted the RNN to sort the breakdown of rolling bearing and the noise removal automatic encoder that was based on the gating recursive unit to increase the accuracy of fault classification. Lu et al. [19] used the memory forgetting mechanism of LSTM and stacked LSTM to extract characteristics from primary signals for fault classification.

The convolutional neural network (CNN) has a superior local receptive field, weights of shared, and spatial domain sub-sampling to DBN, SAE, RNN, and LSTM. The CNN reduces the complexity and the threats of overfitting and the accuracy and efficiency of pattern recognition are improved. The CNN yields satisfactory results in pattern recognition and extracts features from signals or images. The convolutional neural network (CNN) allows two different methods to recognize the failures of the rolling bearing. The first is to use one-dimensional vibration signals (1D-CNN), while the second is to convert the original vibration signal to 2D images as the model input (2D-CNN). The 1D-CNN network is simpler than the 2D-CNN network and has lower computational complexity. Moreover, one-dimensional CNN does not need preprocessing, and the calculation speed is fast, so it is fit in real-time fault diagnosis. Ince et al. [20] applied 1D-CNN for detecting the state and early fault. A compact and adaptive 1D-CNN classifier proposed by Eren et al. [21] takes the sensor data from the time series directly as input, which was suitable for real-time fault diagnosis. Zhang et al. [22]

suggested a diagnosis of bearing's fault based on 1D-CNN, which processed the original vibration signal as the input without de-noising and achieved high accuracy even with noise and different loads. Ma et al. [23] proposed a lightweight CNN with fast training and strong transfer learning. Wang et al. [24] tried a method for fusing multimodal sensor signals (i.e. data from accelerometers and microphones) and used 1D-CNN to extract characteristics from vibration and acoustic signals which were fused. This method showed high accuracy and strong robustness.

This study investigates a fault recognition method of the rolling bearing by using the 1D-CNN (a pre-feedback 2D neural network) classifier. Convolution and pooling operations are performed alternately [25]. For a given input data, the convolution kernel extracts the features. In the supervised phase of training, the parameters of the convolution kernel are optimized by back-propagation so that the convolution kernel better extracts the appropriate features from the input data. After training the network model with the data set, the proposed model extracts appropriate classification features for the diagnosis and fault identification of rolling bearings. The availability of the suggested approach is assessed by using the data set of the rolling bearing which is supplied by CWRU. The main purpose of this study is to validate the following: (1) fast processing speed of 1D CNN as it uses the vibration data of the rolling bearing directly, (2) Dropout operation improves the classification accuracy and generalization ability, (3) Reasonable setting of convolutional layer and pooling layer. The convolution kernels of adjusted size and number extract fault features effectively, and (4) high classification accuracy for a small sample and a limited number of iterations.

The chapters of this paper are arranged as follows. Section 2 introduces the theoretical background and Section 3 presents the network model proposed in this study. Section 4 contrast the experimental results of this method with other methods using the CWRU dataset to that of other methods and proves the effectiveness. Finally, Section 5 presents the conclusion of this study.

2. Theoretical Background

Convolutional neural network (CNN) has a unique network architecture and effectively reduces the sophisticated and overfitting of a neural network. CNN is similar to the visual system of biology[26]. In the biological visual system, neurons in the visual cortex only respond to the stimulation of certain specific areas. That is, the neurons only receive local information and biological cognition of the external environment expands from local to global. Therefore, the neurons do not need to perceive the whole image in the neural network, but only perceive the local features of the image as the local information from each neuron is synthesized at the highest level of the visual cortex to get the global information of the image.

The convolution layer uses a convolution kernel and convolutes the part of the input signal (or image) and extracts the homologous characteristics. The most vital feature of convolution operation is weights of shared, that is, the same convolution kernel ergodic the input with the same parameters in the settled step size. Weights of shared cut down the parameters of the CNN, averts overfitting that is caused by too many parameters, and speeds up the processing speed. The convolution equation is as follows.

$$y^{l(i,j)} = K_i^l * X^{l(r^j)} = \sum_{j'=0}^{W-1} K_i^{l(j')} X^{l(j+j')}, \quad (1)$$

where $*$ indicates the convolution calculation, $K_i^{l(j')}$ represents the j' weight of the i th convolution kernel in the l th layer, $X^{l(r^j)}$ represents the j th partial region convolved in the l th layer, and W represents the length of the convolution operation.

After the convolution operation, the activation function transforms the output value nonlinearly. The original multi-dimensional features are mapped to enhance the linear separability of the extracted features. The activation function *Tanh* and the modified linear element are used in the neural network and the expressions of the two activation functions are shown as following equations.

$$a^{l(i,j)} = \text{Tanh}(y^{l(i,j)}) = \frac{e^{y^{l(i,j)}} - e^{-y^{l(i,j)}}}{e^{y^{l(i,j)}} + e^{-y^{l(i,j)}}} \quad (2)$$

$$a^{l(i,j)} = f(y^{l(i,j)}) = \max\{0, y^{l(i,j)}\}, \quad (3)$$

where $a^{l(i,j)}$ represents the activation value of output after passing through the convolutional layer.

During the experiment of this study, the hyperbolic tangent function *Tanh* is selected as the activation function.

Reduce parameters of the CNN, accelerate the processing speed, and prevent overfitting, a pooling layer is normally appended behind the convolution layer. The maximum and mean pooling functions are used commonly. The maximum pooling function takes the local maximum as the output, while the mean pooling function takes the local mean as the output value. The two functions are shown in Eq. (4) and (5), respectively.

$$p^{l(i,j)} = \max_{(j-1)W+1 \leq t \leq jW} \{a^{l(i,t)}\} \quad (4)$$

$$p^{l(i,j)} = \frac{1}{W} \sum_{t=(j-1)W+1}^{jW} a^{l(i,t)}. \quad (5)$$

where $p^{l(i,j)}$ represents the width of the pooled area, $a^{l(i,t)}$ represents the activation function value of the t neuron in a frame i of the layer l , and W represents the width of the pooled area. The maximum pooling method is chosen in this study as it obtains position-independent features from periodic time-domain signals.

The flatten layer expands the output of the upper layer into a 1D vector, and then connects the input and output by taking this vector as the input of the full connecting layer, as shown in Figure 1. The fully connected layer also integrates different local features in the convolution layer or pooling layer. The formula for the full connection layer is shown as

$$z^{l+1(j)} = \sum_{i=1}^n W_{ij}^l a^{l(i)} + b_j^l, \quad (6)$$

where W_{ij}^l represents the weighted value of the i th neuron in the l th layer and the j th neuron in the $l+1$ layer, z_j^{l+1} represents the logit value of the j th output neuron at the $l+1$ layer, and b_j^l represents the bias value of all neurons in the layer l to the j th neuron in the layer $l+1$.

The output layer generally uses Softmax to identify and classify the extracted features. Softmax is a multi-classification form obtained by logistic regression. It is often used in multi-classification problems. The specific expression is as follows

$$q(j) = \text{softmax}(z^0(j)) = \frac{e^{z^0(j)}}{\sum_{k=1}^M e^{z^0(k)}} \quad (7)$$

where $z^0(j)$ represents the output value of the j th neuron in the output layer and M represents the sum number of categories.

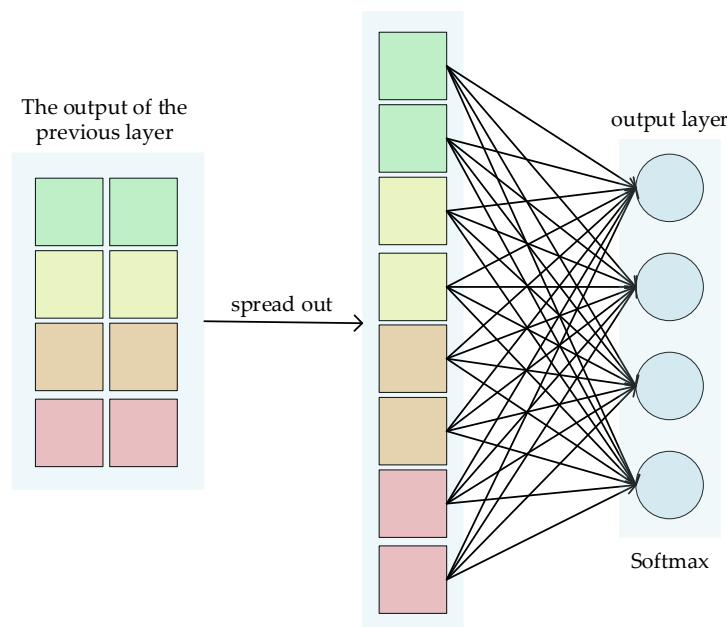


Figure 1. Flatten abridged general view.

Dropout proposed by Hinton et al. [27] reduces the overfitting and enhances the generalization ability of a neural network. The dropout algorithm sets the neurons in a certain layer of the neural network to zero at a certain probability p as shown in Figure 2. The algorithm weakens the joint adaptability of the same layer of neural nodes and improves the generalization ability. A neural network with N nodes is regarded as a set of 2^N models using the Dropout algorithm. The number of training parameters is unchanged and the optimal model is selected from the 2^N models as the best model by training.

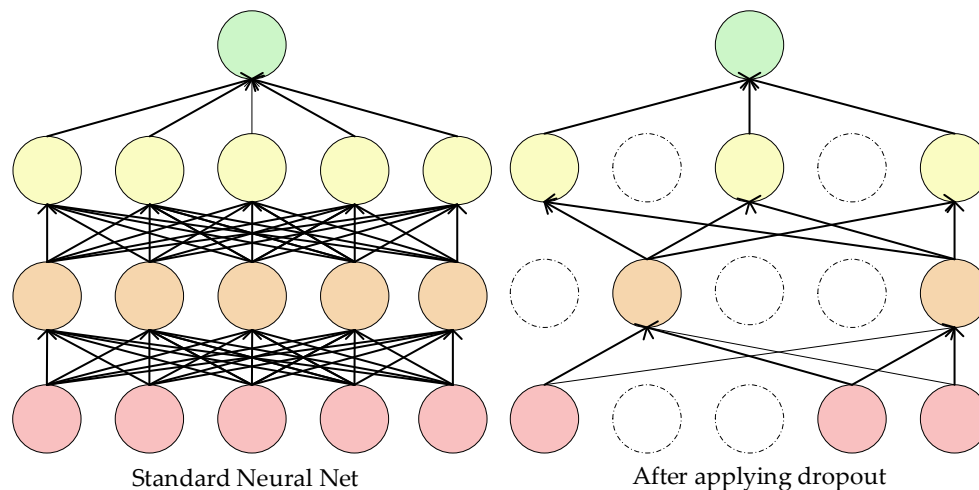


Figure 2. Dropout abridged general view.

The input data of 1D-CNN network model is in the form of one-dimensional vector, so the data form of convolution and pooling operation should also be in the form of one-dimensional, so the processing speed of the model can be accelerated.

3. Structure of 1D-CNN model

Some previous research results show that the neural network model based on 2D-CNN had high accuracy and strong generalization ability in rolling bearing fault diagnosis. However, the 2D-CNN still has the following shortcomings: (1) a long time to transform the 1D vibration signals to 2D images and (2) a slow speed to train two-dimensional convolution and pooling operations of a model. For the sake of enhance the validity and accuracy of rolling bearing failure diagnosis, a 1D-CNN model is established by using the original vibration signal.

3.1 Fault diagnosis process based on 1D-CNN

The fault identification course of the rolling bearing based on the 1D-CNN is as follows.

- (1) A sensor is installed on the corresponding position of the rolling bearing.
- (2) The one-dimensional vibration signals are collected as the original data and then, are partition into a training set and test set.
- (3) The training set is used as input to the 1D-CNN network. The model is trained to obtain the optimal model of fault vibration
- (4) The test set is input in the model and the model performance is evaluated.

The 1D-CNN was evaluated using the one-dimensional original vibration data set of the rolling bearing supplied by CWRU[28]. The data set included 1024 vibration data points and was partition into the training set and test set in the proportion of 7:3. Different loads were grouped in different fault categories as shown in Table 1.

Table 1. One-dimension original signal data set partitioning.

Fault Type	Fault		Number of Samples				
	Diameter (mm)	Fault Orientation	0HP	1HP	2HP	3HP	0123HP
Ball	0.0028	/	100	100	100	100	400
Ball	0.0056	/	100	100	100	100	400
Ball	0.0112	/	100	100	100	100	400
Inner-race	0.0028	/	100	100	100	100	400
Inner-race	0.0056	/	100	100	100	100	400
Inner-race	0.0112	/	100	100	100	100	400
Outer-race	0.0028	@3:00	100	100	100	100	400
Outer-race	0.0028	@6:00	100	100	100	100	400
Outer-race	0.0028	@12:00	100	100	100	100	400
Normal	0	/	100	100	100	100	400

Improved 1D-CNN structure

The improved 1D-CNN structure is similar to other one-dimensional CNNs and based on one-dimensional vibration signals. Adjustments are made to identify the signals so that the network structure effectively and accurately diagnoses and classifies the faults. The 1D-CNN structure incorporates 5 convolution layers, 5 pooling layers, and 2 full connection layers (Figure 3). The specific detailed network structure parameters are shown in Table 2.

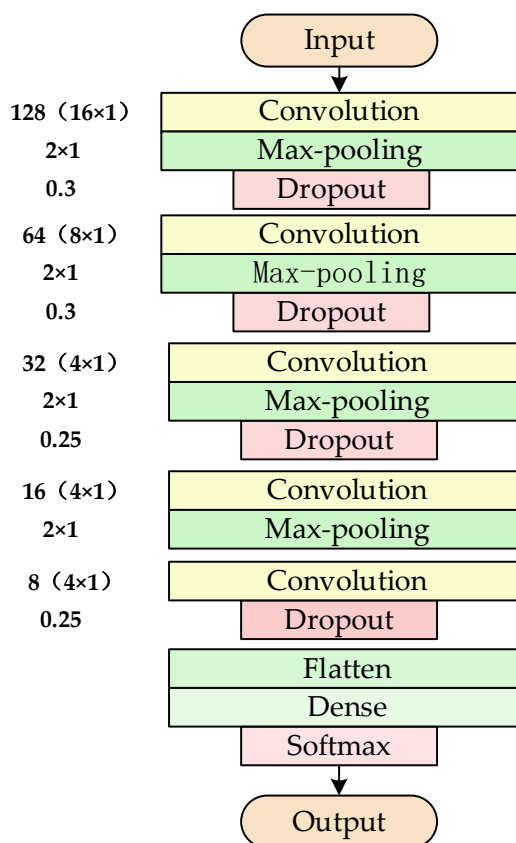


Figure 3. An improved 1D-CNN network structure for bearing fault diagnosis.

Table 2. Improved 1D-CNN network model parameter setting

Network Layer	Output Characteristic	Specific Settings
Input layer	1024×1	1024 pieces of vibration data
Conv1 layer	1009×128	128@ 16×1, stride = 1
Pool1 layer	504×128	pool size is 2×1, stride = 2
Conv2 layer	497×64	64@ 8×1, stride = 1
Pool2 layer	248×64	pool size is 2×1, stride = 2
Conv3 layer	245×32	32 @ 4×1, stride = 1
Pool3 layer	122×32	pool size is 2×1, stride = 2
Conv4 layer	119×16	16 @ 4×1, stride = 1
Pool4 layer	59×16	pool size is 2×1, stride = 2

Conv5 layer	56×8	8 @ 4×1, stride = 1
Flatten	1×448	448 neurons
Dense	1×10	10 neurons

Step 1

In each convolution layer, the appropriate number and size of the convolution kernel performs one dimensional convolution operations. The input data is the one-dimensional signal that has a length of 1024. Five convolution layers uses 128 convolution kernels of size 16 ×1 (Conv1), to Conv5), 64 of size 8×1 (Conv2), 32 of size 4×1 layer (Conv3), 16 of size 4×1 (Conv4), and 8 of size 4×1 (Conv5). *Tanh* is the hyperbolic activation function for the five convolution layers.

Step 2

The pooling layer is appended to the Conv1, Conv2, Conv3, and Conv4 and carry out a 2 × 2 max-pooling operation. The dropout operation is executed after executing the first and second pooling layers and then, the dropout ratio is set to 0.3. A dropout operation with a ratio of 0.25 is performed after the third pooling layer and that with a ratio of 0.25 after the fifth convolution layer. The dropout operation selects and deletes neurons randomly from the model to form a random subset of the neurons, solve the overfitting problem, and enhanced the generalization ability of neural network model. This does not depend on connections between neurons that have specific connections. In the flatten layer, the extracted features from the five convolution layers are extended to a one-dimensional vector. The output layer contains 10 neurons. Using Softmax as the activation function, 10 types of faults are identified after training.

4. Experimental results and analysis

4.1 Data set description

The CWRU data set is the benchmark data set and widely used in researches on diagnosing faults of the rolling bearing. As indicated in Figure 4, the experimental platform was composed of a motor, torque sensor, power tester, electronic controller. Rolling bearings of Skf6205 and Skf6203 were used in the driving end and fan end of the experimental platform. Single-point damage was machined on the bearing by electric discharging machine (EDM). The diameter of the damage was 0.0028, 0.0056, 0.0083, 0.011 and 0.0157 mm. As the outer diameter of the rolling bearing was fixed in operation, the fault data of the outer ring was real and effective. The damage points of the outer diameter of the bearing were set in the direction of 3, 6, and 12 o'clock in direction. The vibration acceleration signals of the rolling bearing are collected by acceleration sensors mounted on the fan end and the motor drive end housing. The sampling frequencies of the fan end and the driving end were 12khz and 12khz, and 48Khz. The bearing test platform uses 16-channel data recorders to collected vibration signals and a torque sensor to measured load and speed.

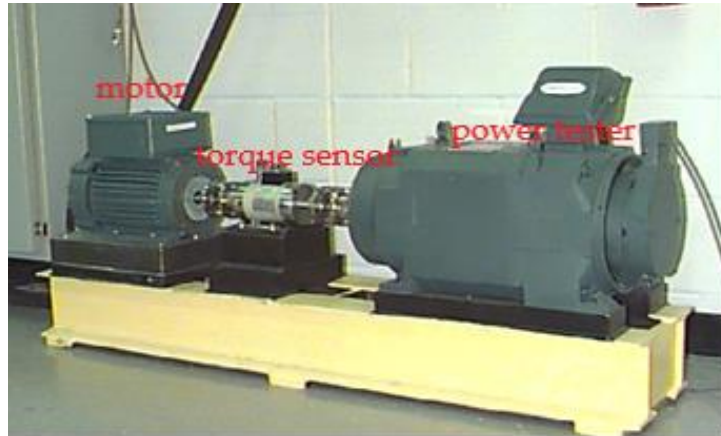


Figure 4. Rolling bearing test platform (CWRU).

In the experiment, the vibration data from the acceleration sensor at the driving end was selected at 12kHz sampling frequency. The data set included nine types of failures in the normal state of the bearing, the bearing inner ring, and the ball bearings at diameters of 0.0028, 0.0056, and 0.0083 mm. The damage points of the bearing's outer ring were in the direction of 3, 6, and 12 o'clock. The vibration signals of the 10 fault types were selected when the load was 0, 1, 2, and 3 HP with 1024 data points. Different fault types under different loads were partitioned into the training and test set in the scale of 7:3. Table 1 shows the data set of rolling bearing vibration signals. Figure 5 reveals the original vibration signals of the 10 fault types under 0 HP.

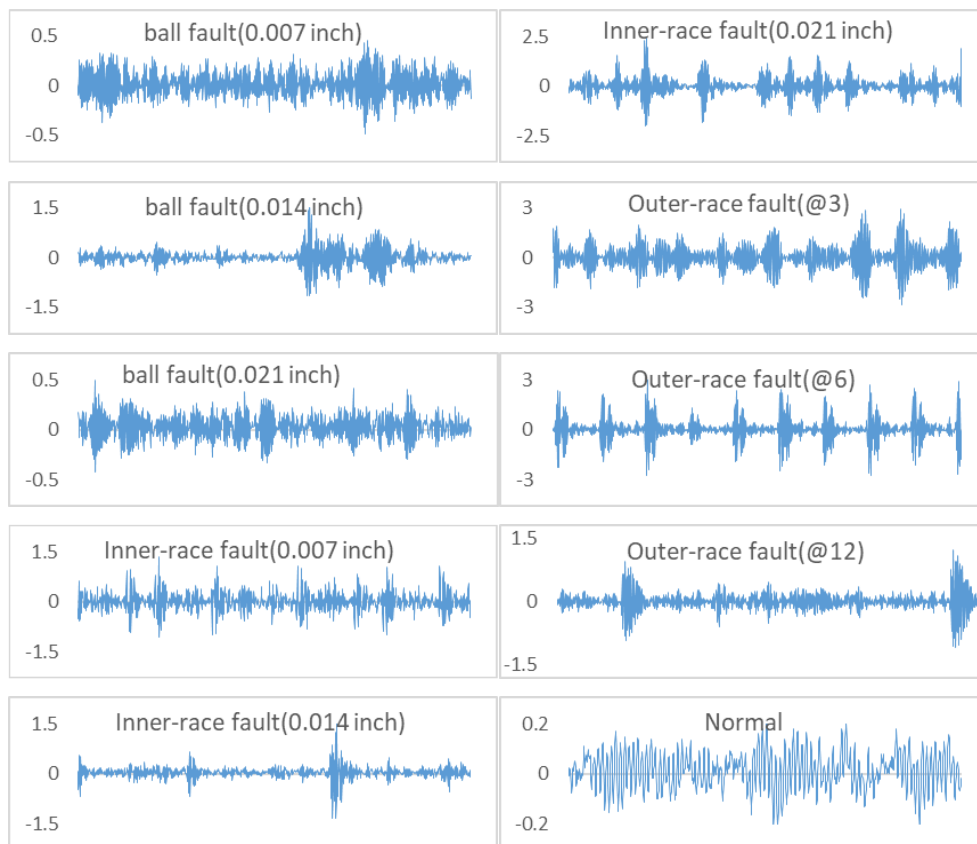


Figure 5. Sample raw vibration signals from CWRU dataset(0HP).

4.2 Parameters of improved 1D-CNN model

A neural network is a multi-layer composite function in mathematics. A neural network is a linear function without an activation function. When the samples are not linearly separable, the activation function should introduce non-linear factors. Common activation functions are *Tanh* and *Relu*. *Tanh* effectively expands the effect of features in the continuous cycle when the difference of eigenvalues is obvious. *Relu* makes the output of neurons with a negative input value zero, which reduces the interdependence among parameters and speeds up the calculation. To explore the influence of *Tanh* and *Relu* on the 1D-CNN network, their activation functions are used in the experiment under the load of 3 HP. Figure 6 and Figure 7 show the curve of loss function and accuracy in the training process.

The curves of the training loss by *Tanh* and *Relu* in Figure 6 show a constant decrease in 30 epochs. The training loss of *Tanh* decreases faster than *Relu* and converges to zero. The accuracies of *Tanh* and *Relu* keep increasing. *Tanh* reaches the accuracy of 1 faster than *Relu* and has higher accuracy in the whole epochs. The result represents that *Tanh* is more suitable activation function than *Relu* for the improved 1D-CNN. Figure 5 shows 10 fault types under 0 HP have different vibration curves and *Relu* is not suitable for training with a small number of samples.

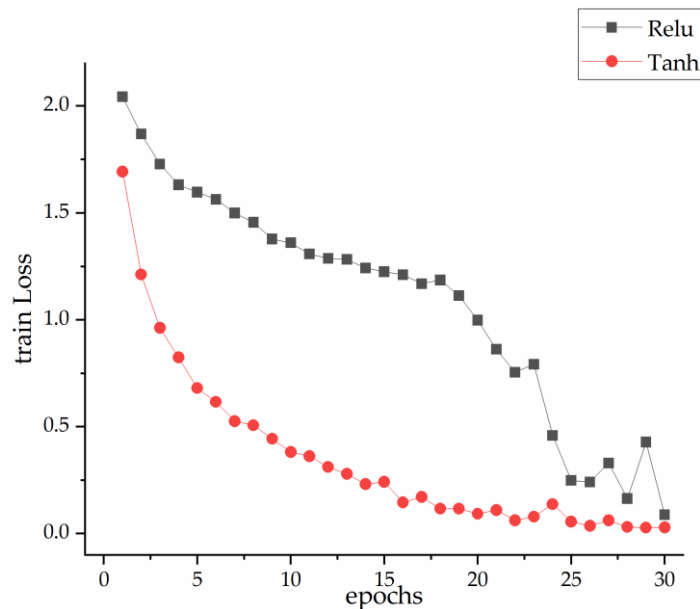


Figure 6. Loss function during the training set with *Tanh* and *Relu*.

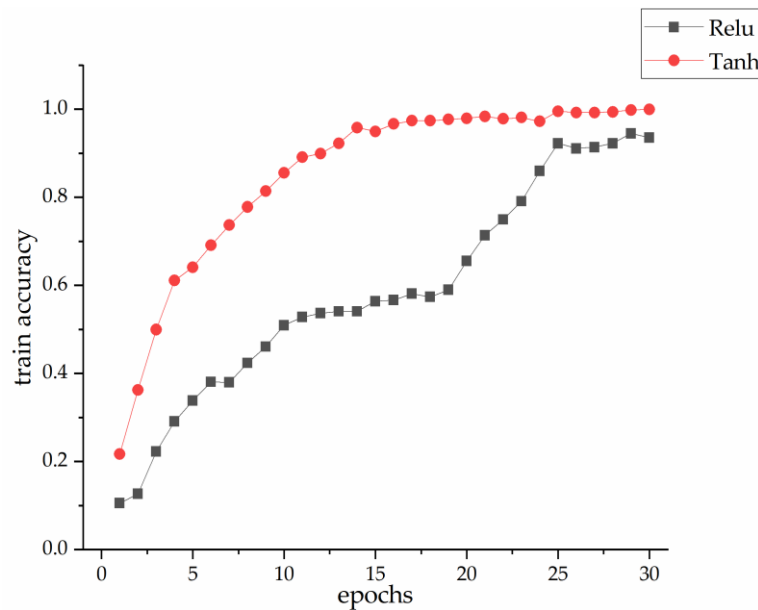


Figure 7. Accuracy during the training set with Tanh and ReLU.

When training the network model, a batch size not merely impact the training speed but also the accuracy. A large batch-size can expedite progress the training process, but it requires a large memory space of a computer. Thus, the small batch-size is appropriate in training despite slow operation speed and noise. The occurrence of noise prevents the training process from falling into the local optimum. Choosing a suitable batch-size is important for the model. In this study, the batch-sizes of 10, 20, 30, 40, 50, 60, 70, and 80 as under 0123 HP were chosen. The experimental result suggests that the improved 1D-CNN is available in any batch-size and has an accuracy higher than 99% (Figure 8). Figure 8 shows, The best accuracy, 99.75% is obtained for the batch-size of 50. The final result is the average accuracy in five training sessions.

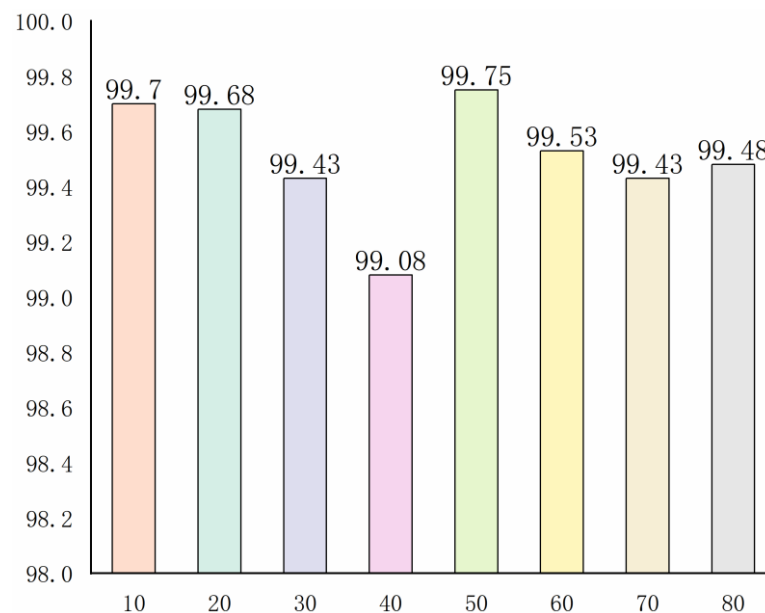


Figure 8. Influence of different batch_size on model accuracy.

Overfitting occurs commonly in neural networks. There are several methods to solve the outfitting among which dropout is most simple and effective. The improved 1D-CNN in this

experiment is trained on small data sets. To avoid the overfitting problem, dropout is added to the network model as shown in Figures 9 and 10, which show the loss and accuracy of the test set with or without dropout under 1 HP.

The loss and accuracy of the test set proved that the improved 1D-CNN performance improved with the added of dropout layer. To prove the impact of dropout on the model performance, five experiments were conducted under loads of 0, 1, 2, 3, 0123 HP (Figure 11). The accuracy with dropout increases from 98.13% to 99.53% under 0 HP, from 97.27% to 98.33% under 1 HP, and from 99.16% to 99.75% under 0123 HP. Therefore, the dropout improves the performance of the improved 1D-CNN under different loads.

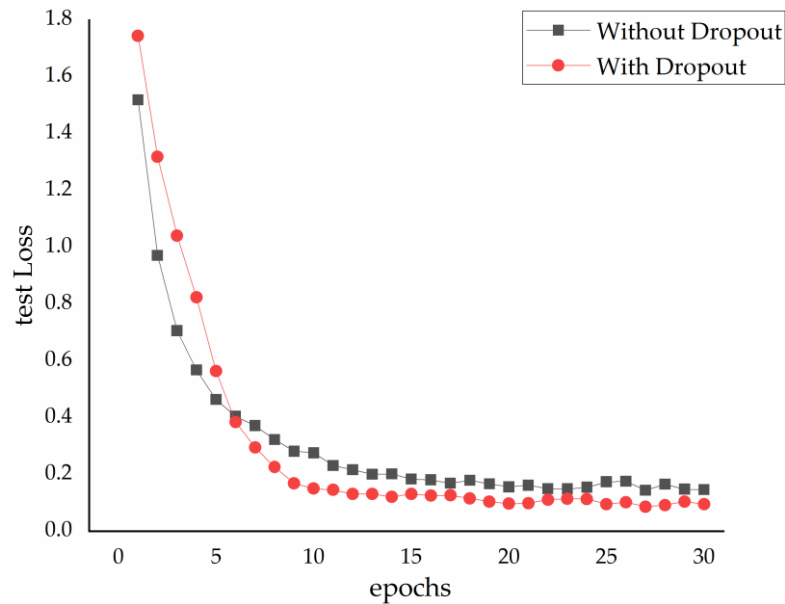


Figure 9. Testing loss of the test set with Dropout and without Dropout.

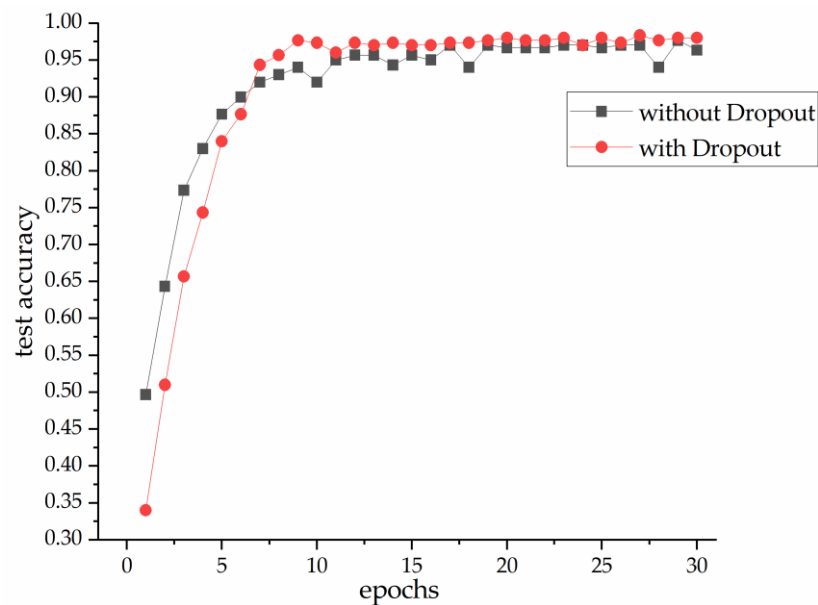


Figure 10. Accuracy of the test set with Dropout and without Dropout.

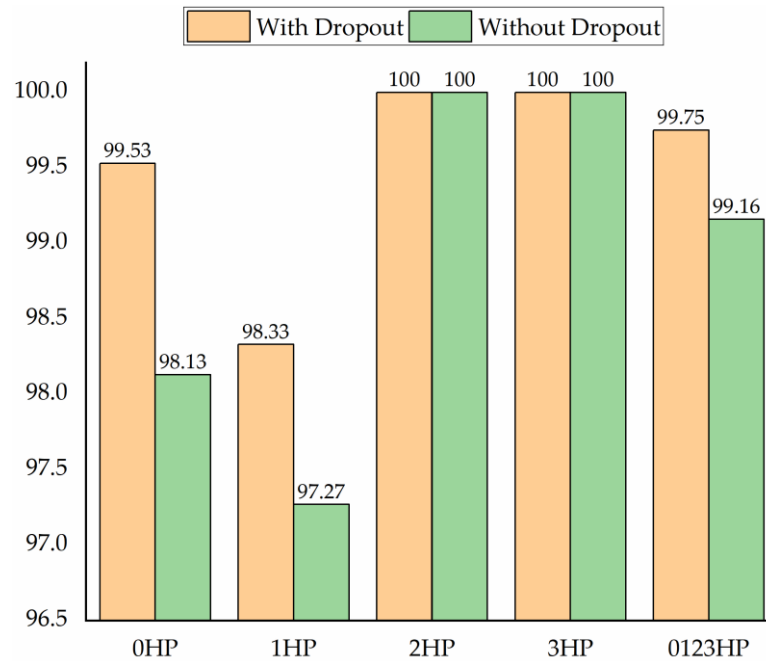


Figure 11. Under different loads with Dropout and without Dropout.

4.3 Contrast with other fault diagnosis methods

The improved 1D-CNN model is compared with the experiments of five different models based on machine learning and deep learning, so as to prove the effectiveness of the improved 1D-CNN model in the fault diagnosis of rolling bearings. The five models are LSTM, MLP, SVM, Random forest, and KNN. The datasets in Table 1 were used by the six models. Experiments were carried out under loads of 0, 1, 2, 3, 0123 HP. Parameters of each model are as follows:

(1) Improved 1D-CNN model

The learning ratio was set to 0.001. The optimizer was Adam that combines the advantages of degrade and rmsprop algorithm and has high computing efficiency and low memory requirement. The loss function is categorical_crossentropy, the batch-size was set to 50, and the iteration time was 30.

(2) LSTM

The first layer of the LSTM had 32 neurons with *Tanh* as the activation function. The second layer had 32 neurons in the full connection layer with *Relu* as the activation function. The third layer had 10 neurons and was classified by Softmax. The learning ratio was set to 0.001 and the optimizer is Adam. The loss function was categorical_crossentropy. The batch-size was 50 and the iteration time was 30.

(3) MLP model

The first, second, third, and fourth layers were the whole connective layer with 300, 400, 200, and 100 neurons, respectively. The activation function was *Relu*. Batch normalization and dropout were adopted with a probability of 0.4 in each full connection layer. The fifth layer was the output layer with 10 neurons and was classified by Softmax. The learning ratio was set to 0.002 and the optimizer is Adam. The loss function was categorical_crossentropy. The batch-size was 20 and the iteration time was 40.

(4) SVM

Gaussian kernel function (RBF) and grid search method were used to find the optimal regularization parameters c (the higher c , the easier to overfit.) and gamma (to control the width of Gaussian kernel). The 5-fold cross-validation method was used, too.

(5) Random forest model

100 decision trees were used to construct the random forest model. A Gini coefficient was used as the division standard of a decision tree. The number of samples for dividing nodes was 2, and the partial sampling data of training samples were used to fit the decision tree.

(6) KNN model

This model searched for the optimal n -grid search method_neighbors and used 10-fold cross-validation method.

The experimental results of various rolling bearing fault diagnosis methods are shown in Table 3, 4, and Figure 12. Table 3 shows the accuracies of the six models under different loads and the average accuracy of the improved 1D-CNN network in different loads is 99.52%. The improved 1D-CNN's average accuracy is 67.12, 36.42, and 23.54% higher than KNN, random forest, and SVM, respectively. The performance of KNN, random forest, and SVM based on machine learning is not as good as the proposed model. When the KNN algorithm is used, the data set needs to be preprocessed. In this study, the original vibration signals were directly inputted into SVM without any processing, but SVM did not perform well on the data set even with more feature points. Random forest is sensitive to noise, resulting in overfitting of the training process.

The average accuracy of the improved 1D-CNN is 14.93% and 24.69% higher than that of LSTM and MLP. The average accuracies of LSTM (84.59%) and MLP (74.83%) are higher than random forest (63.1%), KNN (32.4%), and SVM (75.98%). These suggest that the deep learning method performs better than machine learning. Some nonlinear relations in complex vibration signals of rolling bearings cannot be learned by machine learning methods, but deep learning has great advantages in the analysis of complex nonlinear non-stationary signals. In the experiments, all methods were trained on the small sample data set, there was a difference between the highest and the lowest accuracy. However, the improved 1D-CNN model has only a difference of 1.67%.

Table 3. Accuracy of six different models.

Method	Highest accuracy (%)	Lowest accuracy (%)	Mean (%)
1D-CNN	100	98.33	99.52
LSTM	92.44	76.22	84.59
MLP	84.39	66.78	74.83
SVM	81.28	70.32	75.98
RandomForst	67.88	57.11	63.10
KNN	34	29	32.40

Table 4 and Figure 12 show the average accuracies of different models under different loads. The improved 1D-CNN had the accuracies of 99.53%, 98.33%, 100%, 100%, and 99.75% under different loads with an average accuracy of 99.52% that was higher than those of the other five models. Moreover, the standard deviation of the improved 1D-CNN was only 0.62%, which is lower than 5.3%, 6.99%, 4.05%, 3.93%, and 1.85% of the other five methods. These results prove the effectiveness of the improved 1D-CNN in fault diagnosis under different loads.

Table 4. Accuracy and Std of different models under different loads.

Method		0 HP	1 HP	2 HP	3 HP	0123 HP
1D-CNN	Accuracy (%)	99.53	98.33	100	100	99.75
	Std (%)			0.62		
LSTM	Accuracy (%)	92.44	76.22	83.33	83.78	87.17
	Std (%)			5.30		
MLP	Accuracy (%)	74.44	66.78	80.89	67.66	84.39
	Std (%)			6.99		
SVM	Accuracy (%)	75.75	70.32	81.28	72.98	79.57
	Std (%)			4.05		
RandomForst	Accuracy (%)	57.11	61.35	67.88	62.24	66.92
	Std (%)			3.93		
KNN	Accuracy (%)	29	33	34	34	32
	Std (%)			1.85		

4.4 Performances under different loads

In the practical application, the fault samples of equipment under various loads are difficult to collect. Thus, the faults of rolling bearings are collected under a certain load. However, training the model for diagnosing the faults requires the fault data under different loads. In the research of this paper, the generalization ability of the improved 1D-CNN under different loads was investigated and the result was compared to that of Shufflenet V2, MobileNet, ICN [29], DFCNN [30], and PFC-CNN [31]. The specific experimental contrast results are shown in Table 5 and Figure 13. ("1HP → 2HP" means using 1HP as a training set and 2HP as a test set.)

Table 5 shows that the highest and the lowest accuracy of the improved 1D-CNN is 100% and 94%. The lowest accuracy was lower than that of Shufflenet V2 (96.3%) and ICN (94.17%). The average accuracy was 98.26%, which was higher than that of Shufflenet V2 (97.36%), MobileNet (94.38%), ICN (97.07%), DFCNN (90.05%), and PFC-CNN (93.31%). The

results validate the improved 1D-CNN is effective in diagnosing the bearing faults under different loads.

Table 5. Accuracy contrast between different loads on different models.

	1D-CNN	Shufflenet V2 [29]	MobileNet [29]	ICN [29]	DFCNN [30]	PFC-CNN [31]
Highest accuracy (%)	100	99.4	98.4	99.8	/	97
Lowest accuracy (%)	94	96.3	90	94.17	/	90
Average (%)	98.26	97.36	94.38	97.07	90.05	93.31

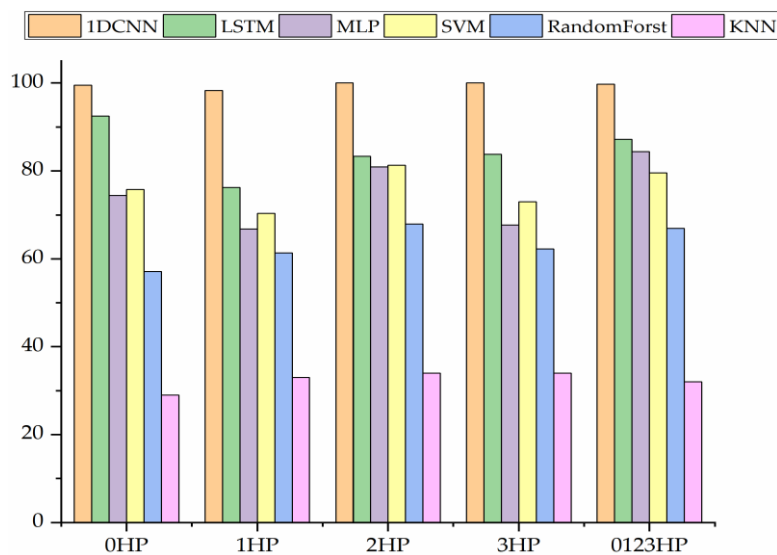


Figure 12. Accuracy of different models under different loads.

The accuracies of the diagnosis under various loads with different methods are shown in Figure 13. Except for the case of 2HP \rightarrow 1HP, the improved 1D-CNN model had higher accuracies than Shufflenet V2 and ICN. In other cases, the proposed model showed effectiveness in diagnosing the faults under cross-loads.

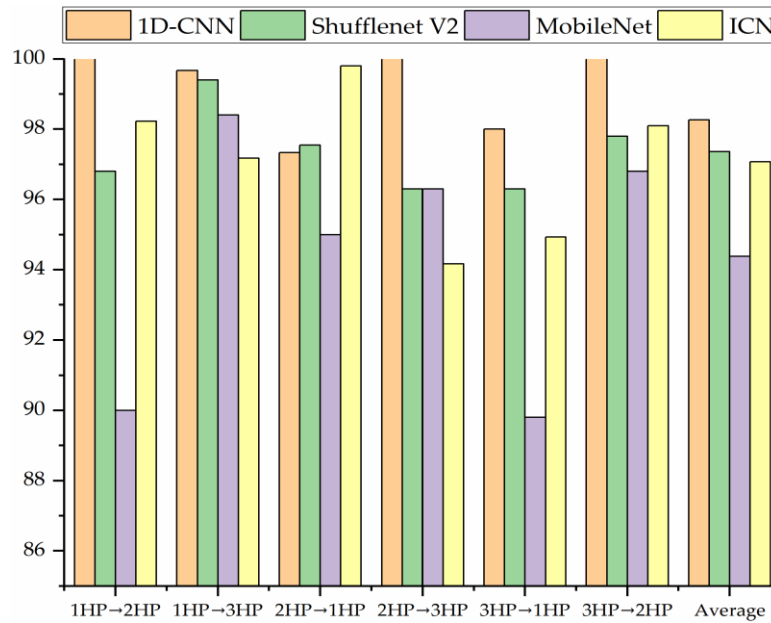


Figure 13. Accuracy of different models across loads.

4.5 Visual analysis of validity of 1D-CNN model

To more intuitively assess the accuracy of the improved 1D-CNN model bearing fault diagnosis, the results under different loads are summarized in the confusion matrix. Table 6 shows the bearing status represented by each number in Figure 14. The confusion matrix presents the predicted result of the samples on the horizontal axis and the actual label of the samples on the vertical axis. 5% of the ball fault of 0.0056 mm were incorrectly predicted as the ball fault (0.0028 mm) under 0 HP and 12% of the ball fault (0.0084 mm) were incorrectly predicted as ball fault (0.0028 mm) under 1HP. Under the load of 0 and 1 HP, there were errors in the diagnosis and prediction of the ball fault as the characteristic information of the ball fault under lower loads is masked by noise. In other cases, the improved 1D-CNN has an appropriate prediction.

Table 6. The numerical ID of the rolling bearing status.

Bearing state	Ball 0.0028	Ball 0.0056	Ball 0.0112	Inner 0.0028	Inner 0.0056	Inner 0.0112	Outer @3 0.0028	Outer @6 0.0028	Outer @12 0.0028	Normal 0
ID	0	1	2	3	4	5	6	7	8	9

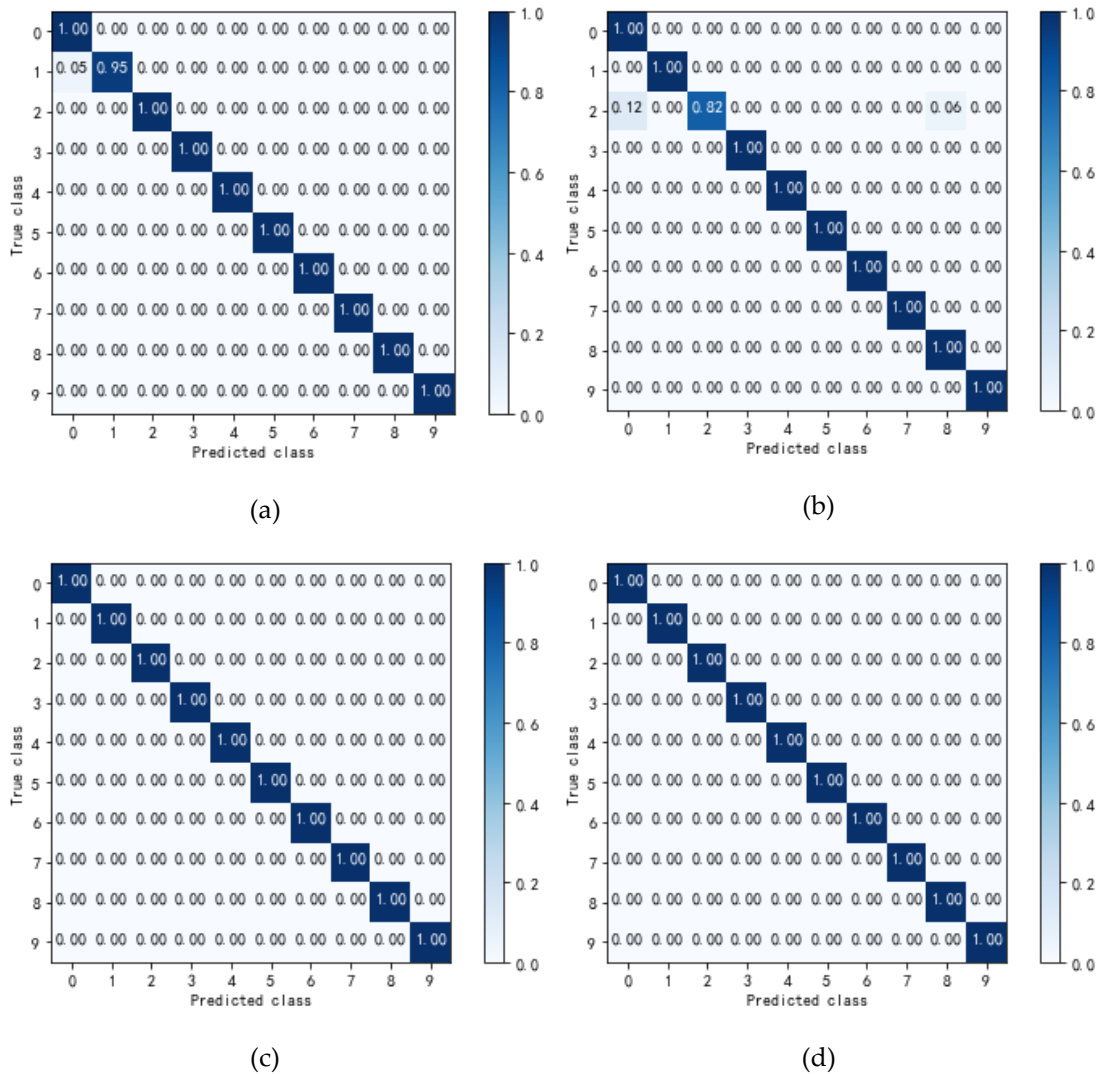


Figure 14. confusion matrix under different loads (a) 0 HP, (b) 1 HP, (c) 2 HP, and (d) 3 HP

t-SNE is a dimension reduction algorithm based on manifold learning, which is different from the traditional PCA and MMD methods. t-SNE uses normalized Gauss collated high-dimensional spatial data features for similarity modeling. At the same time, t-distribution is used to model the similarity of low-dimensional spatial data. KL distance narrows the distance distribution of high- and low-dimensional space and allows visualizing high-dimensional data into two-dimensional or three-dimensional graphics. To prove that the improved 1D-CNN distinguishes different fault types, the t-SNE visualization algorithm is used (Figure 15). By reducing the data dimension in the prediction of results under different loads [32], the improved 1D-CNN visualizes the characteristics of the CWRU dataset. Table 6 shows the status of the bearings represented by the numbers (Figure 15). Figure 15 also shows that the bearing failure characteristics representing the alike fault type are gathered together. Various types of bearing faults are separated, which shows that the improved 1D-CNN effectively distinguishes fault characteristics under different loads.

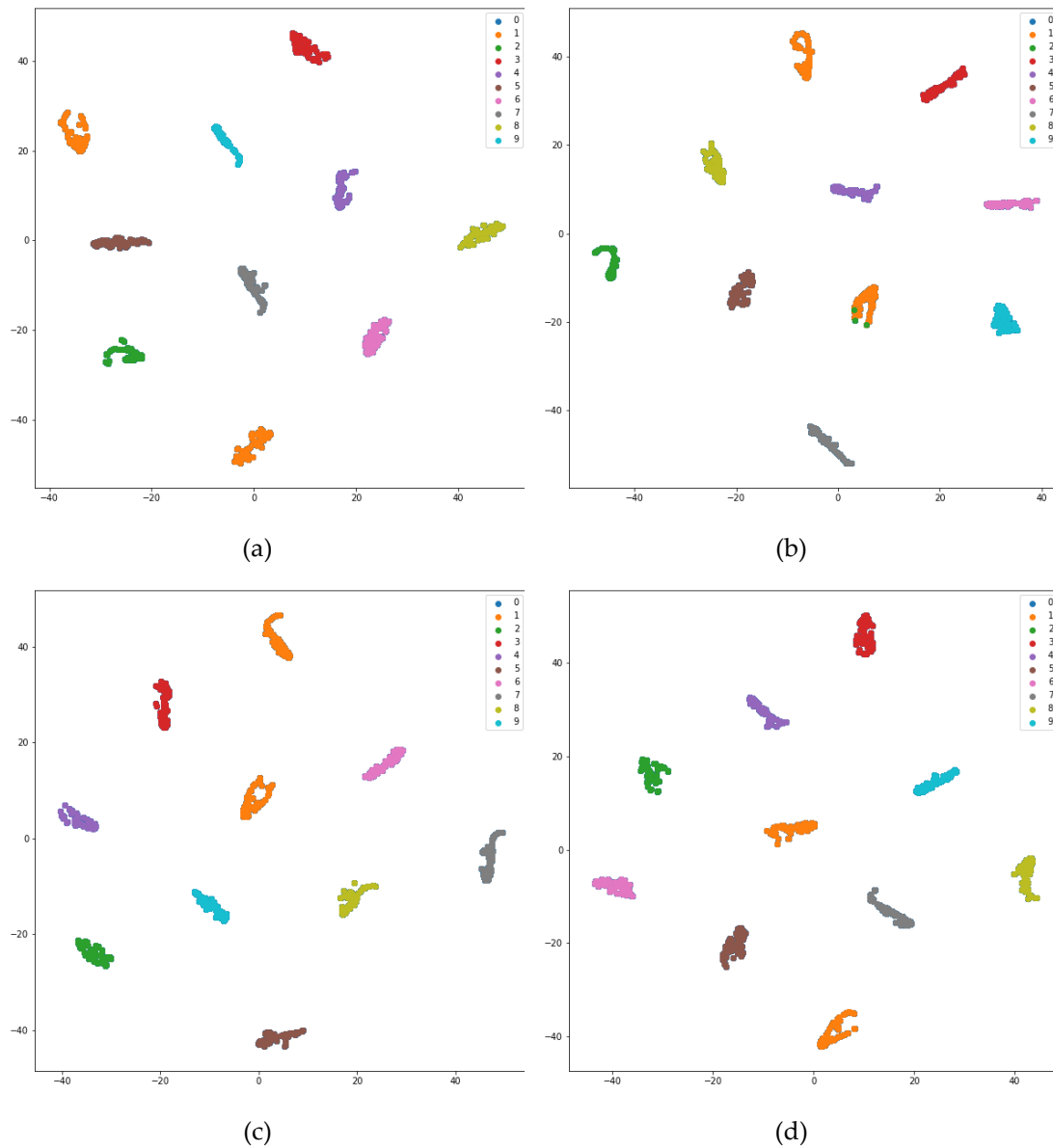


Figure 15. t-SNE under different loads (a) 0 HP, (b) 1 HP, (c) 2 HP, and (d) 3 HP

5. Conclusion

A new method for diagnosing faults of the rolling bearing is proposed by using an improved 1D-CNN. The vibration data set of the rolling bearing supplied by Case Western Reserve University (CWRU) is used to verify the model. The following conclusions can be drawn from a series of experiments in this paper.

- (1) The average accuracy of the improved 1D-CNN is 99.52% under a single load and 98.26% under different loads. The improved 1D-CNN shows improved accuracy, enhanced processing speed without preprocessing.
- (2) The improved 1D-CNN model has advantages in analyzing complex and non-stationary signals and is prevail over the traditional machine learning method in terms of accuracy, speed, and use of computer memory.

- (3) The improved 1D-CNN model uses a local receptive field, weight sharing, and spatial domain subsampling, which decreases the complicity of the network and the risk of overfitting, and speeds up the training speed.
- (4) Adding dropout operation improves the accuracy of the improved 1D-CNN model and enhances generalization ability.

In the actual industrial environment, the vibration data of the bearing will be disturbed by great noise, and the vibration data of the rolling bearing will be more and more complicated. The results of this study lead future researches to develop an accurate method of diagnosing the faults of the rolling bearing in complex bearing vibration data with noise with the integration of deep learning method.

Author Contributions: Data curation, Chih-Cheng Chen; Formal analysis, Chih-Cheng Chen; Funding acquisition, Chih-Cheng Chen; Investigation, Chih-Cheng Chen; Methodology, Chih-Cheng Chen; Project administration, Chih-Cheng Chen, Guangsong Yang, Chia-Chun Wu and Qiubo Ye; Resources, Chih-Cheng Chen, Guangsong Yang and Chia-Chun Wu; Software, Zhen Liu and Chia-Chun Wu; Supervision, Zhen Liu, Guangsong Yang and Chia-Chun Wu; Validation, Zhen Liu; Visualization, Zhen Liu and Guangsong Yang; Writing – original draft, Zhen Liu.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgments: This study was supported by the National Science Council of Taiwan under grant MOST 107-2221-E-507-002-MY3.

References

1. Nandi, S.; Toliyat, H.A.; Li, X. Condition Monitoring and Fault Diagnosis of Electrical Motors—A Review. *IEEE Transactions on Energy Conversion* 2005, 20, 719–729, doi:10.1109/TEC.2005.847955.
2. Lei, Y.; Lin, J.; Zuo, M.J.; He, Z. Condition monitoring and fault diagnosis of planetary gearboxes: A review. *Measurement* 2014, 48, 292–305, doi:10.1016/j.measurement.2013.11.012.
3. Fan, L.; Wang, S.; Wang, X.; Han, F.; Lyu, H. Nonlinear dynamic modeling of a helicopter planetary gear train for carrier plate crack fault diagnosis. *Chinese Journal of Aeronautics* 2016, 29, 675–687, doi:10.1016/j.cja.2016.04.008.
4. Gao, Z.; Cecati, C.; Ding, S.X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches. *IEEE Transactions on Industrial Electronics* 2015, 62, 3757–3767, doi:10.1109/TIE.2015.2417501.
5. Filippetti, F.; Bellini, A.; Capolino, G. Condition monitoring and diagnosis of rotor faults in induction machines: State of art and future perspectives. In *Proceedings of the 2013 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD)*; 2013; pp. 196–209.
6. Kral, C.; Habetler, T.G.; Harley, R.G. Detection of mechanical imbalances of induction machines without spectral analysis of time-domain signals. *IEEE Transactions on Industry Applications* 2004, 40, 1101–1106, doi:10.1109/TIA.2004.830762.
7. Li, D.Z.; Wang, W.; Ismail, F. An Enhanced Bispectrum Technique With Auxiliary Frequency Injection for Induction Motor Health Condition Monitoring. *IEEE Transactions on Instrumentation and Measurement* 2015, 64, 2679–2687, doi:10.1109/TIM.2015.2419031.

8. Zhang, X.; Liu, Z.; Wang, J.; Wang, J. Time–frequency analysis for bearing fault diagnosis using multiple Q-factor Gabor wavelets. *ISA Transactions* 2019, 87, 225–234, doi:10.1016/j.isatra.2018.11.033.
9. Gao, H.; Liang, L.; Chen, X.; Xu, G. Feature extraction and recognition for rolling element bearing fault utilizing short-time Fourier transform and non-negative matrix factorization. *Chin. J. Mech. Eng.* 2015, 28, 96–105, doi:10.3901/CJME.2014.1103.166.
10. Li, Y.; Xu, M.; Huang, W.; Zuo, M.J.; Liu, L. An improved EMD method for fault diagnosis of rolling bearing. In *Proceedings of the 2016 Prognostics and System Health Management Conference (PHM-Chengdu)*; 2016; pp. 1–5.
11. Li, H.; Liu, T.; Wu, X.; Chen, Q. Research on bearing fault feature extraction based on singular value decomposition and optimized frequency band entropy. *Mechanical Systems and Signal Processing* 2019, 118, 477–502, doi:10.1016/j.ymssp.2018.08.056.
12. Chen, R.; Huang, D.; Zhao, L. Fault Diagnosis of Rolling Bearing Based on EEMD Information Entropy and Improved SVM. In *Proceedings of the 2019 Chinese Control Conference (CCC)*; 2019; pp. 4961–4966.
13. Appana, D.K.; Islam, Md.R.; Kim, J.-M. Reliable Fault Diagnosis of Bearings Using Distance and Density Similarity on an Enhanced k-NN. In *Proceedings of the Artificial Life and Computational Intelligence*; Wagner, M., Li, X., Hendtlass, T., Eds.; Springer International Publishing: Cham, 2017; pp. 193–203.
14. Wang, X.; Zheng, Y.; Zhao, Z.; Wang, J. Bearing Fault Diagnosis Based on Statistical Locally Linear Embedding. *Sensors* 2015, 15, 16225–16247, doi:10.3390/s150716225.
15. Georgoulas, G.; Karvelis, P.; Loutas, T.; Stylios, C.D. Rolling element bearings diagnostics using the Symbolic Aggregate approximation. *Mechanical Systems and Signal Processing* 2015, 60–61, 229–242, doi:10.1016/j.ymssp.2015.01.033.
16. Yin, A.; Lu, J.; Dai, Z.; Li, J.; Ouyang, Q. Isomap and Deep Belief Network-Based Machine Health Combined Assessment Model. *SV-JME* 2016, 62, 740–750, doi:10.5545/sv-jme.2016.3694.
17. Rolling Bearing Fault Diagnosis Based on STFT-Deep Learning and Sound Signals Available online: <https://www.hindawi.com/journals/sv/2016/6127479/> (accessed on Nov 12, 2020).
18. Liu, H.; Zhou, J.; Zheng, Y.; Jiang, W.; Zhang, Y. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Transactions* 2018, 77, 167–178, doi:10.1016/j.isatra.2018.04.005.
19. A Novel Hierarchical Algorithm for Bearing Fault Diagnosis Based on Stacked LSTM Available online: <https://www.hindawi.com/journals/sv/2019/2756284/> (accessed on Nov 12, 2020).
20. Ince, T.; Kiranyaz, S.; Eren, L.; Askar, M.; Gabbouj, M. Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks. *IEEE Transactions on Industrial Electronics* 2016, 63, 7067–7075, doi:10.1109/TIE.2016.2582729.
21. A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier | *Journal of Signal Processing Systems* Available online: <https://dl.acm.org/doi/abs/10.1007/s11265-018-1378-3> (accessed on Nov 12, 2020).
22. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing* 2018, 100, 439–453, doi:10.1016/j.ymssp.2017.06.022.
23. Ma, S.; Cai, W.; Liu, W.; Shang, Z.; Liu, G. A Lighted Deep Convolutional Neural Network Based Fault Diagnosis of Rotating Machinery. *Sensors* 2019, 19, 2381, doi:10.3390/s19102381.

24. Wang, X.; Mao, D.; Li, X. Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network. *Measurement* 2020, 108518, doi:10.1016/j.measurement.2020.108518.
25. Hubel, D.H.; Wiesel, T.N. Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 1959, 148, 574–591.
26. Lawrence, S.; Giles, C.L.; Ah Chung Tsoi; Back, A.D. Face recognition: a convolutional neural-network approach. *IEEE Transactions on Neural Networks* 1997, 8, 98–113, doi:10.1109/72.554195.
27. Dropout: a simple way to prevent neural networks from overfitting: *The Journal of Machine Learning Research*: Vol 15, No 1 Available online: <https://dl.acm.org/doi/abs/10.5555/2627435.2670313> (accessed on Nov 12, 2020).
28. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mechanical Systems and Signal Processing* 2015, 64–65, 100–131, doi:10.1016/j.ymssp.2015.04.021.
29. Liu, H.; Yao, D.; Yang, J.; Li, X. Lightweight Convolutional Neural Network and Its Application in Rolling Bearing Fault Diagnosis under Variable Working Conditions. *Sensors* 2019, 19, 4827, doi:10.3390/s19224827.
30. Zhang, J.; Sun, Y.; Guo, L.; Gao, H.; Hong, X.; Song, H. A new bearing fault diagnosis method based on modified convolutional neural networks. *Chinese Journal of Aeronautics* 2020, 33, 439–447, doi:10.1016/j.cja.2019.07.011.
31. Zhu, X.; Luo, X.; Zhao, J.; Hou, D.; Han, Z.; Wang, Y. Research on deep feature learning and condition recognition method for bearing vibration. *Applied Acoustics* 2020, 168, 107435, doi:10.1016/j.apacoust.2020.107435.
32. Maaten, L. van der; Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 2008, 9, 2579–2605.