Hindawi

Research Article

Bearing Fault Diagnosis of End-to-End Model Design Based on 1DCNN-GRU Network

Liu Zhiwei

Dean's Office of IT Center and Educational Affairs Office, Tianjin University of Commerce Boustead College, Tianjin 300384, China

Correspondence should be addressed to Liu Zhiwei; liuzhiwei_2008@126.com

At present, the complex and varying operating conditions of bearings make the feature extraction become difficult and lack adaptability. An end-to-end fault diagnosis is proposed. A convolutional neural network (CNN) is good at mining spatial features of samples and has the advantage of "end-to-end." Gates recurrent neural (GRU) network has good performance in processing time-dependent characteristics of signals. We design an end-to-end adaptive 1DCNN-GRU model (i.e., one-dimensional neural network and gated recurrent unit) which combines the advantages of CNN's spatial processing capability and GRU's time-sequence processing capability. CNN is applied instead of manual feature extraction to extract effective features adaptively. Moreover, GRU can learn further the features processed through the CNN and achieve the fault diagnosis. It was shown that the proposed model could adaptively extract spatial and time-dependent features from the raw vibration signal to achieve an "end-to-end" fault diagnosis. The performance of the proposed method is validated using the bearing data collected by Case Western Reserve University (CWRU), and the results showed that the proposed model had recognition accuracy higher than 99%.

1. Introduction

Bearings are vital machine components that appear in almost all rotating machinery, and the health status of bearings plays a vital role in the effective operation of the mechanical system [1]. Bearings in the mechanical equipment must undergo lousy environments, such as high speed, complicated structure, and high failure rate. Bearings are also one of the vulnerable parts of rotating machinery. Most mechanical equipment failure is caused by bearing failure; once the bearing fails, a series of failures will be triggered, which will directly affect the operation safety of the whole equipment [2]. Therefore, real-time monitoring of the state and diagnosing the bearing fault have a significant meaning.

With the development of machine learning, intelligent fault diagnosis methods have become the main approaches in mechanical diagnosis. Traditional intelligent fault diagnosis methods mainly contain feature extraction, feature selection, and fault classification [3]. The raw vibration signal sampled by the sensor contains much fault information. Extracting features related to faults from the raw

signal to diagnose bearing faults is a crucial step that affects fault classification directly. Some methods of feature extraction include frequency domain analysis and time-frequency analysis [4], Fast Fourier Transform [5], wavelet transform [6], wavelet packet transform [7], empirical mode decomposition [8], and so on. However, these conventional methods have the disadvantage of relying on handcrafted features and signal processing technologies. The robustness and extensibility of models need to be improved.

Recently, deep learning [9] and solutions to extract features from raw signals have been widely paid attention to, that combines multiple nonlinear learning layers to process raw data layer by layer and mine the association between data adaptively. Thus, this approach possesses the capacity to extract features end-to-end and avoids the complexity and uncertainty often observed in traditional feature extraction processes. That is an end-to-end algorithm structure makes the whole process without manual feature extraction. Hoang et al. transform one-dimensional vibration signals into two-dimensional images without noise reduction as the input data of CNN to diagnosis fault, which achieves very high

accuracy and has a strong character of robustness [10]. Chen et al. fused the vibration signals from the horizontal and vertical into a two-dimensional matrix and proposed a deep CNN for extracting features automatically to identify the health status of gearboxes [11]. All the above studies use CNN to extract features from vibration signals, which shows that it has excellent performance. However, vibration, pressure and other state signals gathered during machine operation are usually one-dimensional vectors [12, 13]. Therefore, some researchers seek to construct a 1DCNN model for faults diagnosis. For example, Peng et al. [14] proposed a novel deeper 1DCNN for fault diagnosis of wheel set bearings and gained good effects. Wu et al. present a method based on 1DCNN to realize fault diagnosis of rotating machinery [15]. You et al. proposed an improved ReLU-CNN model based on CNN to diagnosis mechanical faults, whose model has good performance and fast convergence rate [16]. Zhang et al. proposed an end-to-end model for bearing fault diagnosis, that extracts features with 6-layer TICNN. High accuracy is achieved in a noisy environment and even under different load by this model [17]. Guo et al. built a deep convolutional transfer learning model to diagnose bearings faults, which learns invariant features by the 1DCNN network. The proposed model was verified through six transfer experiments [18].

Compared with CNN, 1DCNN uses the raw data as input directly without processing, thus avoiding features loss or distortion. In addition, the vibration signal is time-series and contains abundant time-dependent properties. RNN (recurrent neural network) has good performance in processing time-dependent characteristics of signals. For the networks based on RNN, LSTM (long short-term memory network) and GRU(gated recurrent unit) have applied to fault diagnosis. Yu et al. proposed a novel algorithm based on stacked LSTM for bearing fault diagnosis, features extracted automatically by LSTM. Experimental results show that the accuracy is up to 99% [19]. Hinchi et al. constructed lifetime a prediction model of rolling bearing, which can reflect the degradation trend of the rolling bearing well [20]. Rui et al. designed an enhanced GRU network and applied it on the generated sequence of local features to learn the representation. Experiments on tool wear prediction, gearbox fault diagnosis, and incipient bearing fault detection verify the effectiveness of this model [21].

This paper presents an end-to-end adaptive framework named 1DCNN-GRU for bearing fault diagnosis. Combine advantages of CNN and GRU to extract features from raw data and achieve end-to-end fault diagnosis. Extract features by CNN to replace manual screening and characterization and then input the features into GRU to extract temporal characteristics further. The features extracted jointly are used for fault diagnosis. A series of experiments with the rolling bearing dataset from CWRU demonstrate that the proposed method has high accuracy, practicability, and feasibility.

2. Related Theories

2.1. Convolutional Neural Network (CNN).
CNN is a typical feed forward neural network inspired by biological neural processes, which is powerful for processing spatial data. The network is structured by a convolutional layer and a pooling layer. The topological features embedded in the input data are extracted, through convoluting and pooling the input data layer by layer. A typical convolutional neural network structure is shown in Figure 1, which can mainly divide into five steps: input layer, convolutional layer, pooling layer, full-connected layer, and output layer.

Generally, the convolutional layer contains a set of filters. We can combine with each filter using the input volume to extract the local l function from the local input area. The convolutional layer convolutes the input volume by the kernel, to generate the characters of input data. We can express the output of convolution as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l\right), \tag{1}$$

where f is an activation function, M_j is an input, x_i^{l-1} is the input to layer l, x_j^l is the output of layer l, k is the kernel, and b is the bias.

The pooling layer is followed behind the convolutional layer, also named the subsampling layer. The feature extraction is obtained by the convolution layer as the input of the pooling layer. It is not modified by back propagation. The max-pooling mathematics model is described as follows:

$$P_i^{l+1}(j) = \max\left\{Q_i^l(t)\right\}$$
$$(j-1)W + 1 < t_x \le jW. \tag{2}$$

Here, $(j-1)H + 1 < t_y \le jH$ $P_i^{l+1}(j)$ is the corresponding to the neuron in layer $l+ 1$, $Q_i^l(t)$ is the corresponding in the i_{th} frame of a layer to the t_{th} activations, W denotes width, and H denotes height.

2.2. 1D Convolutional Neural Network.
The kernel of a typical convolutional neural network is usually two-dimensional. Convolution operates a feature graph through a sliding window in the width and height directions; multiply and sum the corresponding positions. Operate a feature signal through a sliding window in one (width or height) direction while performing one-dimensional convolution. The input of one-dimensional CNN is one-dimensional data, like some text and time series data samples usually, the kernel is one dimensional, and the output of convolution and pooling is also one-dimensional data. The structure of one-dimensional CNN is shown in Figure 1.

2.3. Gate Recurrent Unit Network.
The recurrent neural network is a special network, which is proposed based on the view that "human cognition is based on experiences and memories." Compared with CNN, there is an association between each time step calculation in RNN, which not only considers the input of the previous moment but also gives the network "memory" of the previous content. So RNN is good at capturing the long-term dependence of input sequences. GRU is an improved version of RNN by the gates, which can overcome the vanishing gradient problem of
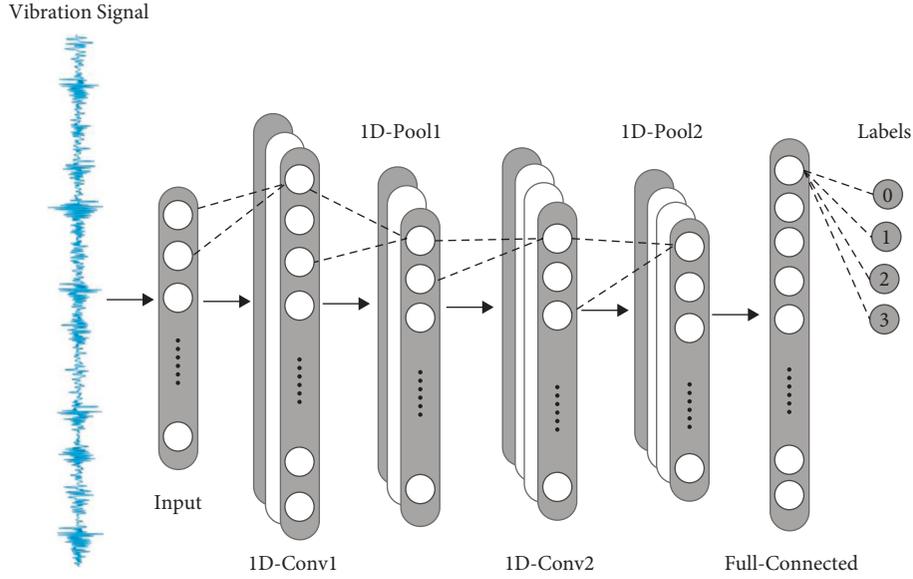
Vibration Signal

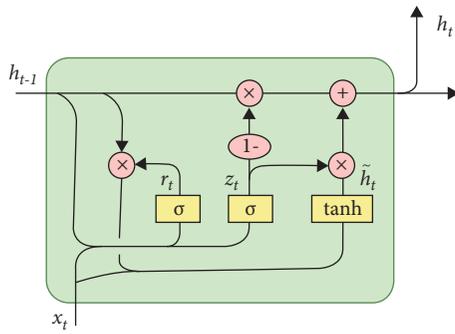FIGURE 1: The structure of one-dimensional convolution.

FIGURE 2: The internal structure diagram of GRU.

RNN. A typical GRU consists of a hidden state, reset gate, and update gate, and the basic structure is shown in Figure 2.

The basic GRU process is as follows: update gate z_t. The update gate acts on the output of the hidden layer at the previous time h^{t-1} and the input of the current time x^t, and the logical value is the state of the gate. The calculation process is as follows:

$$z_t = \sigma\left(W_z x_t + U_z h_{t-1} + b_z\right), \qquad (3)$$

where W_z and U_z are all the weight matrices which can be obtained through learning. σ is an activation function and b_z is the bias weight.

Reset gate is r_t. Similarly, reset gate r_t acts on the output of the hidden layer at the previous time h^{t-1} and the input of the current time x^t, and the logical value is the state of the gate. The calculation process is as follows:

$$r_t = \sigma\left(W_r x_t + U_r h_{t-1} + b_r\right), \qquad (4)$$

where W_r and U_r are the weight matrices, b_r and σ are similar with (3) and instant information of the current time $\tilde{h}_t$. After getting the state of the gate, the reset gate judges the importance of current input and previous output, then

decides the proportion of past activation to realize information reset. Updating the $\tilde{h}_t$ with the activation function $\tan h$.

$$\tilde{h}_t = \tanh\left(W_h x_t + r_t \odot \left(U_h h_{t-1}\right) + b_h\right). \qquad (5)$$

Here, W_h and U_h are the weight matrices, $\odot$ denotes an elementwise multiplication, and b_h is the bias.

The output of the current hidden layer is controlled by the update gate, which can perform two operations: forgetting and selective memory; forgetting the output of the previous moment; and selecting memory of instant information of the current time. Finally, the current activation is computed as follows:

$$h_t = z_t \odot h_{t-1} + \left(1 - z_t\right) \odot \tilde{h}_t. \qquad (6)$$

3. The Proposed 1DCNN-GRU Network

The main framework of the proposed 1DCNN-GRU model for bearing fault diagnosis is shown in Figure 3. The model which we design mainly includes four parts: data processing with a nonoverlap sliding window; data input based on the raw processed data; feature extraction based on 1DCNN and GRU; faults classifier based on GRU, and sigmoid is the activation function of the probability of classifying output. In order to improve the adaptability and accuracy of bearing fault diagnosis, the proposed model is designed by 1DCNN and GRU. The main effect of 1DCNN is to perform preliminary feature extraction on the raw signal, training to fitting, and preliminary screening score is high, so effective features can be quickly screened from the raw signal.

The data training process of the model is shown in Figure 3. First, the raw signal data is obtained from the drive end accelerometer, and then we analyzed the vibration frequency, fluctuation amplitude, and vibration amplitude period of the signal to acquire the ideal truncated sampling
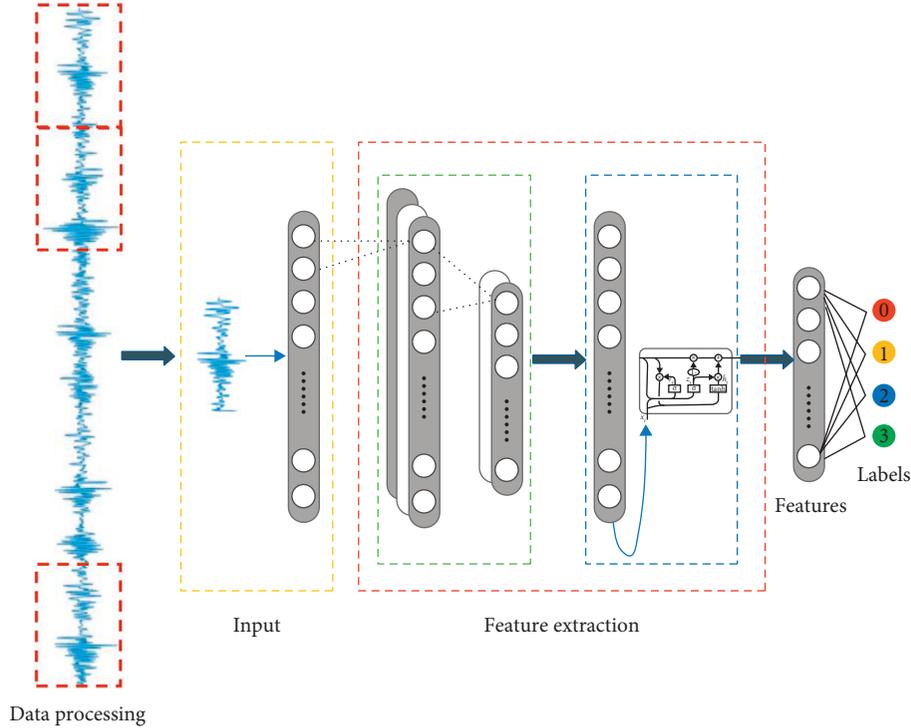
FIGURE 3: The structure of the proposed 1DCNN-GRU model.

window size. Through the mechanism of the sliding window to intercept raw acceleration signal sequence to generate preliminary characteristics as the input of the neural network. Next, the CNN network extract features from the input data automatically without any handcrafted features, which has the advantages of high efficiency. However, the CNN network has a poor ability to obtain the temporal correlation features. Then, the GRU network was subsequently designed to overcome the disadvantages of the CNN network, and the high-weight features learned by CNN as the input of the GRU network for learning and training. The features with correlation among vibration signal were further learned and then perform better correlation explanation. Finally, output the probability of the category through the sigmoid function to diagnosis the health status.

In order to solve the optimization problem of the proposed model, PMSprop, Adam, and Adadelta are used. The classification loss function used mean-squared-logarithmic-error (MSLE). The proposed 1DCNN-GRU model predicts $\widehat{y}$, and the loss function is defined as follows:

$$L(y, \widehat{y}) = \frac{1}{N} \sum_{i=0}^{N} \left(\log(y_i + 1) - \log(\widehat{y}_i + 1) \right)^2. \tag{7}$$

3.1. Validation of the Proposed 1DCNN-GRU Network

3.1.1. Data Description. We can obtain the raw fault data of rolling bearing through Case Western Reserve University (CWRU). There are four health types as follows: respectively, normal (NO), rolling ball fault (BF), inner fault (IF), and outer fault (OF). We select the raw vibration signals from the

TABLE 1: Description of the experimental samples.

Fault type	Load (horsepower)	Sample length	Number of samples	Label
NO	1/2/3	1024	500	0
IF	1/2/3	1024	500	1
OF	1/2/3	1024	500	2
BF	1/2/3	1024	500	3

load of 1, 2, and 3 horsepower randomly. Then, nonoverlapping sampling was used to process the original signals, and samples of each type are obtained, Table 1 shows some of the experimental samples and the each health vibration signal is shown in Figure 4. The dataset contains 500 samples of each type, of which 60% are randomly selected for training, 20% for verification, and 20% for testing.

3.1.2. Model Training and Testing

(1) Optimizer and Learning Rate. Selecting different optimizers and learning rates plays a vital role in improving the training speed and classification accuracy for different models and classification tasks. Therefore, for the model in this paper, the alternative optimizers include RMSprop, Adam, and Adadelta, meanwhile considering the influence of different learning rates on the actual rate of convergence, different learning rates are applied for model training. To test the performance of the proposed model, each group of optimizers and learning rates used in training is independent. The accuracy with different groups is shown in Table 2. The loss change of different optimizers and learning rates is shown in Figure 5.
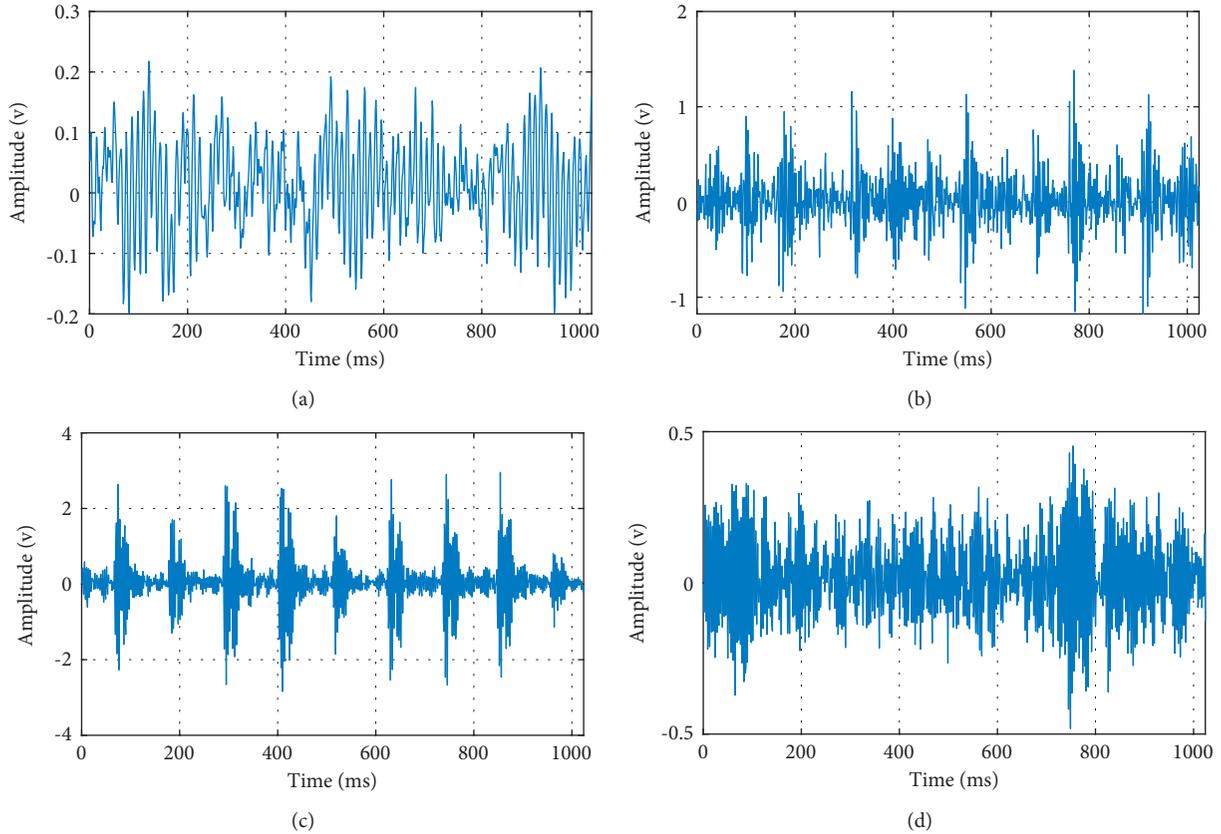
(a)

(b)

(c)

(d)

FIGURE 4: Vibration signals of bearings with different types. (a) NO. (b) IF. (c) OF. (d) BF.

TABLE 2: Accuracy of different optimizers and learning rates.

Learning rates	0.001		0.01		0.1	
Optimizers	Train	Test	Train	Test	Train	Test
			Accuracy (%)			
RMSprop	0.998	0.992	0.987	0.965	0.241	0.237
Adam	0.993	0.992	0.995	0.985	0.432	0.487
Adadelta	0.312	0.237	0.723	0.715	0.992	0.986

Table 2 shows that the experimental results of RMSprop and Adam optimizer are similar, and the best classification results appear when the learning rate is 0.001. With the increase of learning rate, the classification results decreased seriously. It is the direct opposite of the Adadelta optimizer, and the best classification effect is achieved when the learning rate is equal to 1. Considering the accuracy and loss, the RMSprop optimizer is finally selected, and the learning rate is set to 0.001.

(2) Batch Size. Before the optimizer performs a weight update, we can obtain the batch size which is the number of training instances through observing, which can affect the model's generalization performance. Compared with the learning rate, the batch size is less sensitive to the model, but the batch size is also a critical parameter to further improve the performance of the model. Increasing the batch size in an appropriate range can reduce the training time and
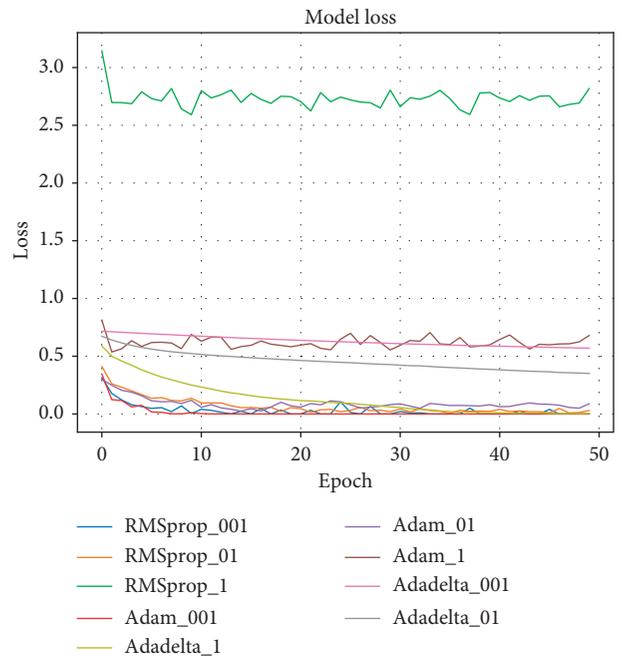

FIGURE 5: Loss change of different optimizers and learning rates.

contribute to the stability of model convergence. However, performance appears to have a downward trend when the batch size is too large. Table 3 shows that the model which we

TABLE 3: Performance of model w.r.t different batch size.

Batch size	Converge time (s)	Training accuracy	Testing accuracy
50	41.9s	1	0.9850
100	29.3	1	0.9775
200	24.9	1	0.9850
300	25.9	0.9987	0.9725

TABLE 4: Performance of model w.r.t different dropout value.

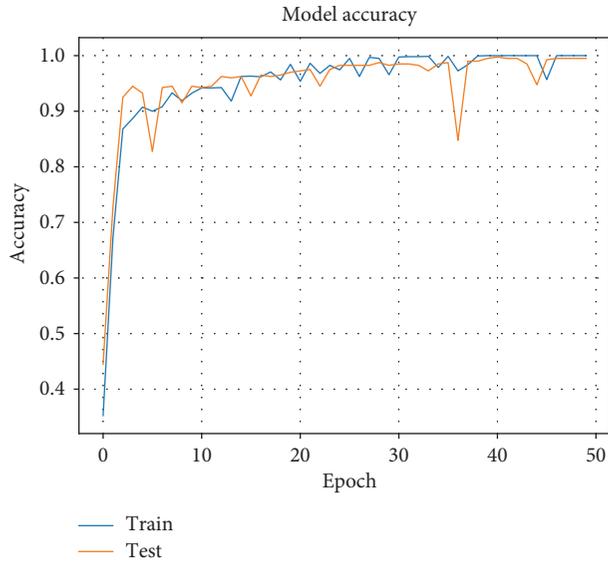| Dropout	Converge time (s)	Training accuracy	Testing accuracy
0.1	25.5	0.9981	0.9775
0.2	24.9	0.9981	0.9700
0.3	25.4	0.9956	0.9750
0.4	25	0.9987	0.9825
0.5	25.4	0.9987	0.9850

FIGURE 6: Accuracy change of training and testing data.

design can achieve the best classification and the shortest training time when batch size equals 200.

(3) Dropout. Dropout is one of the forms for regularization. We can verify the influence in the different dropout values; we further experiment on the bearing data. Table 4 shows that the different dropout values have a weakly influence on the converge time and accuracy of the proposed model. Finally, the dropout value is set to 0.5, because the model has the highest accuracy when dropout equals to 0.5.

3.1.3. Results Analysis. Keras framework is selected to train the data after building the fault diagnosis model using 1DCNN-GRU. The iterations, batch size, learning rate of the PMSprop algorithm, and dropout value are set at 50, 200, and 0.01 to training. The accuracy and loss of training set and testing set change with training epochs are shown in Figures 6 and 7.
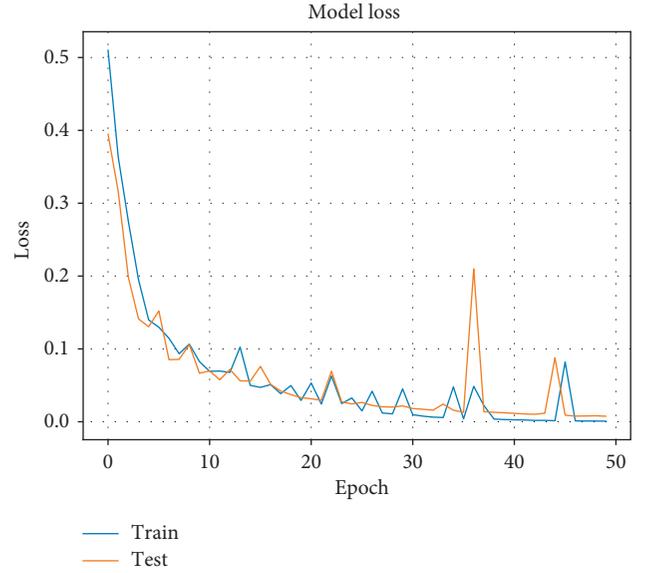

FIGURE 7: Loss change of training and testing data.

TABLE 5: Accuracy of classification models.

Methods	Average training accuracy (%)	Average testing accuracy (%)
SVM	79.61	69.29
Bayesian	68.93	70.10
CNN	97.86	95.76
LSTM	98.35	96.67
1DCNN-GRU	99.98	98.96

According to Figures 6 and 7, the proposed model has high accuracy stability. After the first eight training epochs, the loss value of the testing data decreased rapidly, and the model fitted quickly. After the 8th training, the loss value decreased slowly, the slope of the losing curve was close to 0, and the model can complete convergence. The accuracy of the training set and testing set increased rapidly in the first four training epochs, which reached 1.0 nearly after the 30th training iteration. With increasing the number of epochs, the accuracy curve is very smooth, and the curves of the training set and testing set tend to coincide.

In order to compare the diagnosis results of the proposed model with the current mainstream intelligent fault diagnosis algorithms, this paper experimented based on the representative traditional machine learning methods SVM and Bayesian and the deep learning methods CNN and LSTM. The time and frequency domain feature parameters include mean, variance, standard deviation, maximum, minimum, RMS, absolute mean, root mean square, waveform factor, kurtosis factor, pulse factor, margin factor, skewness, and kurtosis as the input of SVM and Bayesian for fault diagnosis. The input of CNN and LSTM is the same as the proposed method to perform the end-to-end fault diagnosis. The accuracy of different methods is presented in Table 5. The accuracy of 1DCNN-GRU is higher than other methods.

4. Conclusions

In this paper, an end-to-end bearing diagnosis classification model based on 1DCNN and GRU are proposed. We test and verify the following four types of faults experiments:

(1) The proposed model avoided the dependence of traditional feature extraction methods on professional knowledge and realized the end-to-end bearing fault diagnosis, which reduced the complexity of the diagnosis process.

(2) GRU was applied in tandem with 1DCNN, it can not only extract spatial features of vibration signal but also further learn the characteristics of time-sequenced, so that have better data expression ability. Meanwhile, it avoids the dependence of traditional feature extraction methods on professional knowledge and reduces the complexity of the diagnosis process.

(3) The experiments on the CWRU dataset show that the average diagnosis accuracy of the proposed model is higher than SVM, Bayesian, CNN, and LSTM, and it is suitable for the accurate diagnosis of rolling bearing faults.

In future work, in order to avoid the influence of the overfitting phenomenon on the accuracy of fault diagnosis, we will study the optimization of model parameters.

Data Availability

The data used to support the findings of this study are included within the article. Because it is a numerical simulation example, readers can get the same results as this article by using the LMI toolbox of Matlab and the theorem given in this article.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] Q. T. Tran, S. D. Nguyen, and T. I. Seo, "Algorithm for estimating online bearing fault upon the ability to extract meaningful information from big data of intelligent structures," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3804–3813, 2019.

[2] A. Rai and S. H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rollingelement bearings," *Tribology International*, vol. 96, pp. 289–306, 2016.

[3] H. D. Shao, H. K. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mechanical Systems and Signal Processing*, vol. 102, pp. 278–297, 2018.

[4] J. H. Cai, W. W. Hu, and X. Ch Wang, "Rotor fault diagnosis based on Higher-order statistics," *Journal of Vibration, Measurement & Diagnosis*, vol. 33, no. 2, pp. 298–301, 2013.

[5] H. C. Lin and Y. C. Ye, "Reviews of bearing vibration measurement using fast Fourier transform and enhanced fast Fourier transform algorithms," *Advances in Mechanical Engineering*, vol. 11, no. 1, Article ID 168781401881675, 2019.

[6] N. BeSsous, S. E. Zouzou, W. Bentrah, S. Sbaa, and M. Sahraoui, "Diagnosis of bearing defects in induction motors using discrete wavelet transform," *International Journal of System Assurance Engineering and Management*, vol. 9, no. 2, pp. 335–343, 2018.

[7] S. Djaballah, K. Meftah, K. Khelil, M. Tedjini, and L. Sedira, "Detection and diagnosis of fault bearing using wavelet packet transform and neural network," *Frattura ed Integrità Strutturale*, vol. 13, no. 49, pp. 291–301, 2019.

[8] X. Zhang and J. Zhou, "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines," *Mechanical Systems and Signal Processing*, vol. 41, no. 1-2, pp. 127–140, 2013.

[9] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[10] D. T. Hoang and H. J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognitive Systems Research*, vol. 53, pp. 42–50, 2019.

[11] H. Chen, N. Hu, Z. Cheng, L. Zhang, and Y. Zhang, "A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes," *Measurement*, vol. 146, pp. 268–278, 2019.

[12] Y. X. Chou, B. L. Xu, Y. Gu, R. L. Zhang, L. G. Wang, and Y. Jin, "A fast mathematical morphology filter on one dimensional sampled signal," in *Proceedings of the 2017 Int'l Conf. on Control, Automation and Information Sciences (ICCAIS)*, pp. 233–238, Chiang Mai, Thailand, November 2017.

[13] D. C. Yu, Z. G. Yu, X. C. Chen et al., "Compliance control with vector decomposer using one-dimensional pressure sensor for robotichand," in *Proceedings of the Chinese Automation Congress (CAC)*, pp. 445–450, Jinan, China, October 2017.

[14] D. D. Peng, Z. L. Liu, H. Wang, Y. Qin, and L. Jia, "A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains," *IEEE Access*, vol. 7, pp. 10278–10293, 2019.

[15] X. Wu, Z. K. Peng, J. Ren, C. Cheng, W. Zhang, and D. Wang, "Rub-impact fault diagnosis of rotating machinery based on 1-D convolutional neural networks," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8349–8363, 2020.

[16] W. You, C. Shen, D. Wang, L. Chen, X. Jiang, and Z. Zhu, "An intelligent deep feature learning method with improved activation functions for machine fault diagnosis," *IEEE Access*, vol. 8, pp. 1975–1985, 2020.

[17] W. Zhang, C. Li, Peng, Y. Chen, and Z Zhang, "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*, vol. 100, pp. 439–453, 2018.

[18] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 9, pp. 7316–7325, 2019.

[19] L. Yu, J. Qu, F. Gao, and Y. Tian, "A novel hierarchical algorithm for bearing fault diagnosis based on stacked LSTM," *Shock and Vibration*, vol. 2019, Article ID 2756284, 2019.

[20] A. Z. Hinchi and M. Tkiouat, "Rolling element bearing-remaining useful life estimation based on a convolutional long-short-term memory network," *Procedia Computer Science*, vol. 127, pp. 123–132, 2018.

[21] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1539–1548, 2018.