Research Article

Research on Intelligent Fault Diagnosis of Rolling Bearing Based on Improved ShufflenetV2-LSTM

Shaohui Ning¹, Yansong Wang¹, Wenan Cai², Zhenlin Zhang³, Yukun Wu¹, Yonglei Ren¹, and Kangning Du¹

¹School of Mechanical Engineering, Taiyuan University of Science and Technology, Taiyuan, Shanxi 030024, China
²School of Mechanical Engineering, Jinzhong University, Shanxi 030619, China
³North Automatic Control Technology Institute, Shanxi 030006, China

Correspondence should be addressed to Shaohui Ning; ningshaohui@tyust.edu.cn

Received 8 December 2021; Accepted 7 January 2022; Published 18 March 2022

Abstract

The timing characteristics of the fault vibration signal of the rolling bearing are ignored by the ShufflenetV2 network. For the bearing fault problem of multiple working conditions, the fault diagnosis signal is extracted by the feature, which cannot be performed efficiently and accurately. The ShufflenetV2 network has a deep number of layers and a large amount of parameters, which causes the network to be prone to overfitting. In response to the above problems, an improved ShufflenetV2-LSTM intelligent fault diagnosis system is proposed, which is a model in which the long short-term memory network (LSTM) layer and the Dropout layer are serially embedded in the ShufflenetV2 network. This method preserves the ability of the ShufflenetV2 network to extract features, and the advantage of the ability of LSTM to strengthen the data sequence is also inherited, so this improves the accuracy of fault diagnosis. The generalization ability of the model can be enhanced by Dropout, which can effectively suppress the degree of overfitting. In addition, this paper also optimized the activation function and optimizer in the model to make up for the additional time cost brought by the Dropout layer, so that the robustness of the system is improved and fault diagnosis can be analyzed efficiently. Experimental analysis shows that the diagnosis accuracy of the proposed algorithm is as high as 97% and early failures of rolling bearings can be effectively identified in real time.

1. Introduction

The rolling bearing in the mechanical transmission system is a key component, and its failure causes the equipment to be eccentrically loaded or the machine is destroyed. However, in actual production activities, due to internal factors (such as parts wear) or environmental factors (such as rapid temperature change), the occurrence of system failure is inevitable. Therefore, the early fault diagnosis of the rolling bearing is of great significance [1, 2]. A large amount of knowledge storage in signal processing is needed in the field of traditional fault diagnosis [3]. With the development of intelligent fault diagnosis methods, the ability of shallow machine learning structures to learn complex feature relationships is very limited [4]. All the above deep learning algorithms have been introduced in the field of bearing fault diagnosis in recent years, making it more intelligent [5].

Deep learning is a branch of machine learning where multiple layers of data processing units are assembled into a deep architecture to extract multiple levels of data abstraction, and each layer automatically learns higher level of data representations from the output of the previous layer [6]. With the significant advantage of automatically learning data representations, it is considered as an advanced technique in big data analysis [7]. Therefore, deep learning-based intelligent diagnostic methods are more flexible and capable to deal with difficulties in real-world applications than traditional machine learning-based methods [8].

Most of the researches have realized the detection of bearing composite fault and fault degree through the analysis of vibration signals on experimental data or real-world
data. For instance, Zhang C install accelerometers on the running parts of high-speed trains to collect vibration signals and use adaptive deep filtering technology to realize composite fault detection of train bearings [9].

In 2012, the classic Alexnet network was the first to be applied to image classification and achieved excellent results [10]. Then, a large number of researchers applied the CNN model to the field of mechanical fault diagnosis and achieved remarkable success [11]. For example, Lu et al. proposed to convert one-dimensional vibration signals into two-dimensional gray maps and then use CNN for classification [12]. Li et al. proposed an intelligent fault diagnosis method for rolling element bearings based on the deep distance metric learning [13]. The VGG-16 network is used to process grayscale image signals [14]. The ShufflenetV2 network was used to process the two-dimensional time-frequency diagram of the fault signal, and it was found that the diagnosis result was better than other networks through a large number of experiments [15]. However, in complex working conditions and many types of faults, the diagnostic effect of a single CNN model is not ideal [16].

With the continuous improvement and development of intelligent fault diagnosis research, it is no longer enough to diagnose simple faults. In complex multiconditions or multiposition rolling bearing faults, the concept of time sequence needs to be introduced [17]. In other words, the ability to diagnose faults can be improved by realizing the connection of time information before and after. LSTM has the ability to remember information and transmit long-term time series information [18]. It is very effective to deal with a large amount of fault vibration data. The combination of whale algorithm and Adam optimizer optimizes LSTM, and the IWOA-LSTM algorithm is used for fault diagnosis [19]. After the signal is denoised by improved VMD, the variance is calculated. It inputs into LSTM network for rolling bearing life prediction [20]. However, the methods proposed in the above documents require feature preprocessing on the collected vibration signals, which does not meet the current requirements for real-time performance [21].

In response to the above problems, the lightweight ShufflenetV2 network is selected as the main body, in this article, and LSTM is embedded in it, constructing an improved ShufflenetV2-LSTM intelligent fault diagnosis system. In view of the fact that LSTM strengthens the time sequence of the data in order to learn hidden features, combining the ability to extract features and the advantages of lightweight networks, ShufflenetV2 deep structure makes up for the shortcomings of the single CNN model. This paper also uses Dropout layer to suppress the model overfitting, changes all the activation functions in the model to LeakyRelu, and uses the RMSprop algorithm to optimize the cross-entropy function. At the end of this article, we use ablation experiments and compare with other deep learning methods. The results show that the improved ShufflenetV2-LSTM system performs very well in terms of diagnostic efficiency and accuracy.

In summary, the main contributions of this study are shown as follows:

1. The original vibration signal was decomposed by VMD to obtain several IMF components. The first three IMF components with the minimum envelope entropy were selected to reconstruct the signal, and the one-dimensional reconstructed signal was used to construct a 32 by 32 two-dimensional characteristic matrix by transverse interpolation.

2. In order to make use of LSTM’s sensitivity to timing signals, the ShufflenetV2 network is selected as the main trunk, and LSTM modules are embedded in series after ShufflenetV2’s global pooling layer. In order to reduce the degree of overfitting, an improved ShufflenetV2-LSTM intelligent fault diagnosis model for rolling bearings is proposed by connecting the Dropout layer after the LSTM.

3. The addition of Dropout layer to the ShufflenetV2-LSTM model brings additional time cost and reduces the overall robustness of the system. Therefore, the activation function and optimizer are optimized in this paper to solve this problem and improve also the diagnostic accuracy of the model.

4. Compared with other models the experimental results show that the proposed method can effectively diagnose faults. The process of LSTM learning fault features is shown by visualization technology.

This paper is organized as follows: the theoretical background is given in Section 2. Section 3 presents the improved ShufflenetV2-LSTM fault diagnosis method. The implementation of the proposed method to deal with the fault diagnosis of data in Western Reserve University is illustrated in Section 4. Conclusions and future work are provided in Section 5.

2. The Basic Theory of the Model

2.1. ShufflenetV2. Traditional CNN models include convolutional layers, pooled layers, and fully connected layers. The existence of large convolution kernels and pooling layers makes the model computationally large. The model depth and size are increasing to improve the accuracy of the model. For some specific application scenarios such as mobile devices, because of their limited performance, the model requires high precision and small size. ShufflenetV1 solves the above problems. Based on the “Four Principles of Efficient Networks,” ShufflenetV2 is modified from the ShufflenetV1 network and verified the superiority of ShufflenetV2 [22]. As a lightweight network, the structure of ShufflenetV2 is shown in Table 1. Each stage unit in the table is stacked by the unit structure shown in Figures 1(a) and 1(b).

When Stride = 1, the structure shown in Figure 1(a) is adopted. In order to meet the same number of input and output feature matrix channels in the G1 criterion, firstly, the unit uses “Chanel Split” to divide the input feature channel into two parts c-c’ and c”. Secondly, one of the branches
2.2. Long Short-Term Memory Network (LSTM). LSTM is a variant of recurrent neural network (RNN) and solves the problem of gradient disappearance and gradient explosion [23]. The core design of LSTM is the cell state and the “gate” structure. The cellular state acts as a pathway for information to travel through a sequence. In theory, cell states can be carried into the cells of later time steps, which overcomes the effect of short-term memory. Information is added and removed through a “gate” structure, which learns what information to save or forget during training. LSTM cleverly combines short- and long-term memory through gate control, which can effectively process time series signals, as shown in Figure 2 [24]. Therefore, it is more reasonable to use LSTM to process vibration signals with temporal characteristics on the basis of extracting features by CNN.

LSTM is mainly composed of input gate, forget gate, and output gate. In part one, the old information to be discarded is determined through the forgetting gate, and the input \( x_t \) at time \( t \) and the output \( h_{t-1} \) of the hidden layer at the previous moment are used as the input of the forgetting gate. The calculation formula is as follows:

\[
f_t = \text{sigmoid}(W_f \times [h_{t-1}, x_t] + b_f),
\]

where sigmoid is the sigmoid activation function, \( W_f \) is the weight of the forgetting gate, \( b_f \) is the bias of the forgetting gate, and \( f_t \) is a measure of the importance of past memories.

In part two, the new information to be remembered is determined through the input gate. The input \( x_t \) at time \( t \) and the output of the hidden layer at the previous time are again calculated through formulas (2) and (3) to obtain the new information importance measurement factor \( i_t \) and candidate vector \( C_h \). Finally, the cell state is updated by formula (4).

\[
i_t = \text{sigmoid}(W_i \times [h_{t-1}, x_t] + b_i),
\]

\[
C_h = \tanh(W_c \times [h_{t-1}, x_t] + b_c),
\]

\[
C_t = f_t \times C_{t-1} + i_t \times C_h,
\]

where \( W_i \) and \( W_c \) are the input gate weight and \( b_i \) and \( b_c \) are the input gate bias. \( C_{t-1} \) is the cell state at the previous moment, and \( C_t \) is the cell state of the updated input gate.

In part three, the output gate combines the results of the forget gate, and the final information to be output is determined. Firstly, the state of the cell at time \( t \) is calculated by formula (5) to obtain part of the output \( O_t \). What is more, the cell state at time \( t \) is normalized through the tanh function. Finally, multiply the two by formula (6) as the final output result.

\[
o_t = \text{sigmoid}(W_o \times [h_{t-1}, x_t] + b_o),
\]

\[
h_t = o_t \times \tanh(C_t),
\]

where \( W_o \) is the output gate weight and \( b_o \) is the output gate bias.

This study uses the sensitivity of LSTM to time series data and embeds LSTM in series after the global pool of ShufflenetV2 to strengthen the timing correlation of signals and suppress irrelevant features to improve training accuracy.

2.3. Dropout. Compared with the training sample, there are too many parameters in the network model, which makes the training model prone to overfitting—the accuracy of the model on the training data is higher, but the accuracy on the test data is lower [25].

Stringing the Dropout layer in the network model can effectively improve the overfitting phenomenon [26]. In each training batch, the working principle of Dropout is to randomly shield some neurons each time by ignoring the connections between some neurons with a certain probability, as shown in Figure 3.

If \( N \) is the number of hidden cells in any layer, \( P \) is the probability of retaining a cell. In the expected case, only \( P \)
Figure 1: Blocks of ShufflenetV2 2.2 long short-term memory network (LSTM).

(a) Stride = 1: basic unit

(b) Stride = 2: unit for spatial down sampling

Figure 2: LSTM structure.
N units will appear after the Dropout layer, instead of N hidden cells. This group of P x N units varied each time and did not allow the units to build coadaptations freely. Therefore, if a layer of size N is optimal for a standard neural network for any given task, then a good exit network should have at least N/P units.

Applying Dropout layer to a neural network is equivalent to extracting a "refined" network from it. The refined network consists of all the units that have survived from Dropout (Figure 3 (b)). Dropout introduces an additional hyperparameter, the probability of retaining unit P. This hyperparameter controls the intensity of Dropout. P = 1 means no Dropout, and a low P value means more Dropout. Typical P values for hidden units range from 0.5 to 0.8. A small P will slow down the training speed and lead to underfitting. Large P may not produce enough Dropout to prevent overfitting. According to experience, P is 0.5 in this study.

In this way, each neuron is not sensitive to another specific neuron. The parameters will not rely too much on the training data, which greatly enhances the generalization ability of the model. The Dropout layer connects to the back of the network which is more effective and needs to work with the fully connected layer. Therefore, a fully connected layer (FC1) and a layer of Dropout are serially connected after the LSTM, as shown in Figure 4.

3. The Improved ShufflenetV2-LSTM Rolling Bearing Fault Diagnosis System

The single ShufflenetV2 network cannot diagnose the bearing faults efficiently for multiposition and multiworking conditions. In order to introduce timing characteristics, improve training accuracy, and reduce overfitting degree, the ShufflenetV2 network is selected as the main stem in this paper. LSTM and Dropout modules are embedded in series after ShufflenetV2’s global pool layer. An improved ShufflenetV2-LSTM fault diagnosis method is proposed. Combined with the sensitivity of LSTM to time series data, the system strengthens the time series correlation of signals and restraints irrelevant features to improve the training accuracy of model. In addition, the generalization ability of the model is increased by using Dropout random shielding neurons. However, during the experiment, it was found that the addition of Dropout module would reduce the overall robustness of the system and bring additional computational costs. In this study, we also improved the activation function and optimizer of the model to improve this problem and improve the diagnostic accuracy of the model.

3.1. Optimization Model

3.1.1. Activation Function. When improving the ShufflenetV2-LSTM system to identify bearing faults, the mapping relationship between fault types and labels is complicated. The activation function introduces nonlinear characteristics and transforms the current feature space to another space through a certain mapping [27]. The mapping ability of nonlinear equations for complex data learning is stronger than linear. In order for the data to be better classified, the selection of the activation function is crucial [28].

The activation function used by the original ShufflenetV2 is Relu, and its mathematical expression is as follows [29].

\[
\text{ReLU}(x) = \begin{cases} 
  x & x > 0, \\
  0 & x \leq 0.
\end{cases}
\]  

(7)
It can be seen from the expression that when \( x \leq 0 \), the output is 0 and its gradient is also 0. The neuron cannot update the parameters, which will cause the network to "necrotize." To solve this problem, the leakage value is introduced in the negative interval to obtain the LeakyReLU function \[ \text{LeakyReLU}(x) = \begin{cases} x & x > 0, \\ x/a & x \leq 0, \end{cases} \quad (8) \]

where \( a \) is the fixed parameter in the \((1, +\infty)\) and \( x \) is the input. According to reference [30], the empirical \( a \) value is 100 in the study. By introducing leakage value, unilateral suppression is achieved, and part of negative gradient information is retained so that it is not completely lost. The LeakyReLU function ranges from minus infinity to infinity. This expands the range of values compared to Relu.

In order to accelerate the network convergence and enhance the robustness of the model, this research changes all the Relu in the original model to the LeakyReLU activation function.

3.1.2. Optimizer. During the training period of the model, the optimizer continuously optimizes the value of loss function by updating and calculating the network parameters of...
the model, so that the model achieves the global optimal point [31]. In practical application, the selection of loss function and optimizer determines the convergence speed and effect of the model. The inappropriate loss function and the optimizer will cause the model to fall into the local optimal point, which is the value of loss function hovers around the local optimal point, unable to reach the global optimal point, resulting in poor accuracy of the final model.

In this paper, the cross-entropy cost function is used to calculate the difference between the current model probability distribution and the real distribution to obtain the loss function value. Equation (9) is a formula for the cross-entropy cost function, where \( y \) is the output value of the neuron activation function and \( y \) is the desired output value.

\[
C = -\frac{1}{n} \sum y \ln a + (1 - y) \ln (1 - a).
\]  

(9)

Therefore, the choice of the optimizer plays a crucial role in the training process [32]. RMSprop combines exponential moving average of gradient square to adjust the change of learning rate. This converges well in the case of nonstationary objective functions. The RMSprop optimizer introduces an attenuation factor \( \beta \) on the basis of AdaGrad and avoids the continuous accumulation of second-order momentum by focusing only on the descending gradient of the past period of time, thereby solving the problem of premature end of training [33]. The core part of RMSprop is the second-order momentum, which is calculated as follows:

\[
V_t = \beta \times V_{t-1} + (1 - \beta)dw^2,
\]  

(10)

where \( \beta \) is the attenuation factor, \( dw \) is the current gradient, and \( V_t \) is the second-order momentum at time \( t \). Adam is an algorithm that combines the moment and RMSprop [34].

\[
m_t = \beta_1 \times m_{t-1} + (1 - \beta_1)dw,
\]  

(11)

\[
V_t = \beta_2 \times V_{t-1} + (1 - \beta_2)dw^2,
\]  

(12)

where \( m_t \) is the first-order momentum, \( \beta_1 \) is the first-order momentum decay factor for formula (11), and other parameters in formula (12) have the same meaning as in formula (10).

Theoretically speaking, the optimization effect of Adam should be better than RMSprop [35], but in the experiment process of this study, it is found that the network model using RMSprop optimizer can converge stably within a certain period of time, but the model using Adam is still oscillating near the optimal value. The reason for the analysis may be that the oscillation of the acceleration learning rate of the first-order momentum in the later training period makes the model difficult to converge.

The improved ShufflenetV2-LSTM fault diagnosis system uses the RMSprop algorithm to optimize the cross-entropy loss function instead of Adam, which makes the network better converge to compensate for the additional time cost of Dropout.

### 4. Experiments and Results

#### 4.1. Data Set

This article uses a data set from the Rolling Bearing Experimental Center of Case Western Reserve University (CWRU) for experimental verification [36]. The bearing designation is 6205-2RS JEM SKF deep groove ball bearing, and the sampling frequency is 12 kHz. Bearing damage is a single-point damage made by EDM, which can

<table>
<thead>
<tr>
<th>Service load</th>
<th>Normal or fault (degree of damage)</th>
<th>Number of samples (training set/testing set)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>IR (7/14/21)</td>
<td>100 (80/20) × 3</td>
</tr>
<tr>
<td>0/1/2/3</td>
<td>B (7/14/21)</td>
<td>100 (80/20) × 3</td>
</tr>
<tr>
<td></td>
<td>OR (7/14/21)</td>
<td>100 (80/20) × 3</td>
</tr>
</tbody>
</table>

#### 3.2. Improved ShufflenetV2-LSTM Basic Flow

An improved ShufflenetV2-LSTM fault diagnosis method for rolling bearings is proposed in this paper. The basic structure is shown in Figure 4, and the specific steps are as follows:

1. Collect vibration signals of rolling bearing running state
2. A series of IMF components were decomposed by VMD algorithm. The first three IMF components were selected to reconstruct the signal according to the envelope entropy value, and a 32 × 32 two-dimensional characteristic matrix was constructed by transverse interpolation
3. Sample training set and test set are divided according to 8:2 ratio
4. Set up network model and initialize parameters
5. The training set trains the network model, and the model learns the fault characteristics, uses RMSprop algorithm to calculate the gradient of the loss function, and then updates the parameters
6. Judge whether the training iteration number \( M \) reaches the preset value \( N \). If so, proceed to the next step; otherwise, repeat the training
7. Test set tests the trained model, outputs evaluation indicators, and finishes the training
be divided into three types of failure locations: inner ring, rolling body, and outer ring failure. Each position is divided into 3 different levels of failure: minor failure (7 mils), moderate failure (14 mils), and serious failure (21 mils), plus a group of healthy rolling bearing operating status, a total of ten different bearing operating status. In order to verify that the improved model can be applied to different types of faults in different working conditions, each operating state distinguishes four workloads of 0, 1, 2, and 3 horsepower. Therefore, the data set in this article contains a total of 40 types of rolling bearing operating states, each of which has 100 small samples of data, and each small sample contains 1024 data points. According to reference [36], randomly select 80% of the samples from each type of running state for training set and the remaining 20% for testing set. The samples are arranged out of order. For the convenience of presentation, the inner ring, rolling body, and outer ring are represented by IR, B, and OR, respectively, and the normal state is represented by N. Details of the data set are shown in Table 2.

Before training the model, the data need to be preprocessed. For clarity, take the sample data of 0 horsepower slight inner ring failure as an example. Firstly, the sample of 1024 data points is decomposed by VMD to obtain 7 IMF components. Then the envelope entropy is used as the evaluation criterion [37], and the first three IMFs with the smallest entropy value are selected to reconstruct the signal. In order to meet the instantaneity, the vibration signal is directly used as the training set to maintain the timing characteristics of the vibration signal. Finally, the one-dimensional reconstruction signal is sequentially constructed in a horizontal order to construct a two-dimensional matrix of 32 by 32, which is used as the model input layer.

4.2. Ablation Experiment. In order to show that the various improvements made to the ShufflenetV2 network in this article contribute to the improvement of the model’s performance, an ablation experiment is carried out, where training accuracy, test accuracy, and training time are mainly used as evaluation indicators. During the experiment, the same sample data set was used, the learning rate was set to 0.001, the loss function used cross-entropy, the training iteration was 30 times, and the minimum batch was 64.

In order to prove that LSTM can improve the accuracy of the model, insert the LSTM after the global pooling layer of the ShufflenetV2 model. In order to prove that Dropout can prevent the model from overfitting, the fully connected layer and Dropout layer are connected in series after the LSTM layer. In order to prove the contribution of LeakyRelu activation function to the model, replace all activation functions in the network with LeakyRelu activation function. In order to prove that the RMSprop optimizer is more suitable for the model and accelerate the stable convergence of the model, a comparative experiment was designed and the Adam optimizer was changed to the RMSprop optimizer.

| Table 3: Results of ablation experiments with data from CWRU. |
|---------------------------------|-----------------|-----------------|-----------------|
| **Model** | **Factor Dropout Activation function** | **Training accuracy (%)** | **Test accuracy (%)** | **Time (min)** |
| ShufflenetV2 | 0 0 0 | 92.19 | 88.25 | 9'42 |
| | 1 0 0 | 98.44 | 95.63 | 9'57 |
| | 0 1 0 | 100 | 94 | 9'44 |
| | 0 0 1 | 91.88 | 90.75 | 20'39 |
| | 0 0 1 | 94.44 | 89.37 | 10'10 |
| This study | 1 1 1 | 100 | 97.38 | 9'23 |

| Table 4: Hyperparameter settings of each model. |
|---------------------------------|-----------------|-----------------|-----------------|
| **Model** | **LSTM Optimizer Activation function Dropout** |
| Improved Alexnet | Add | RMSprop | LeakyRelu | Delete the original |
| Improved MobilenetV2 | Add | RMSprop | LeakyRelu | Without |
| Improved Squeezenet | Add | RMSprop | LeakyRelu | Keep the original |
| Improved Resnet-18 | Add | Adam | LeakyRelu | Without |
| Improved Xception | Add | Adam | LeakyRelu | Without |

| Table 5: Diagnostic results of different models under CWRU bearing data set. |
|---------------------------------|-----------------|-----------------|-----------------|
| **Model** | **Training accuracy (%)** | **Test accuracy (%)** | **Time (min)** |
| Improved Alexnet | 40.63 | 38.75 | 25'16 |
| Improved MobilenetV2 | 98.44 | 94.63 | 12'59 |
| Improved Squeezenet | 39.04 | 32.37 | 7'49 |
| Improved Resnet-18 | 96.88 | 91.87 | 21'53 |
| Improved Xception | 93.76 | 93.63 | 31'53 |
| This study | 100 | 97.38 | 9'23 |
The experimental results are shown in Table 3. The temporal correlation between vibration data was enhanced by adding LSTM. The running time of the ShufflenetV2 network model is only increased by 15 s, but the accuracy rate can be increased by 7.38%. For the ShufflenetV2 model, the time cost of using RMSprop or Adam is the same basically. The accuracy rate of the model using the Adam optimizer can only oscillate around 92.19% at the end of the training. However, the model using the RMSprop optimizer can stably maintain 100% at the end of training, and the accuracy of the test set is 5.75% higher than that of Adam. This shows that the Adam optimizer will not always perform better than other optimizers, and the choice of optimizer cannot be generalized. Replace all model Relu activation functions with LeakyRelu activation functions to increase a small amount of time cost (28 s) in exchange for a certain classification accuracy rate. The improved model accuracy rate increased by 2.25%. For this model, adding, alone, the Dropout layer can reduce the degree of overfitting the difference between training accuracy, and test accuracy is reduced from 3.94% to 1.13%. However, the accuracy of the model will not be saturated until almost 2 times the time cost before the improvement as a cost. Therefore, it is not advisable to add alone the Dropout layer to avoid overfitting, and other improvements proposed in this paper are needed to compensate for the slow convergence rate and reduce the time cost.

Ultimately, the algorithm proposed in this research, introducing LSTM, selecting the RMSprop optimizer, using the LeakyRelu activation function, and adding the Dropout layer, reduces the degree of overfitting, improving the classification accuracy, and speeds up the model convergence speed to compensate for the additional time cost brought by the Dropout layer. The difference between the prediction accuracy and the training accuracy is reduced from 3.94% before the improvement to 2.62% in the rolling bearing data set. During the training process, the accuracy of the model can be stabilized at 100%, and the test accuracy can be as high as 97.38%, an increase of 9.13% compared to the model before the improvement.

4.3 Algorithm Comparison and Analysis. The current classic popular and lightweight neural network are selected for experimental comparison to compare with the method in this article in order to further verify the performance of this research method in the diagnosis of bearing faults. The experimental model selected Alexnet, MobilenetV2, Squeezenet, Resnet-18, and Xception. It is necessary to improve the selected network model in order to ensure the fairness of the comparison experiment. Since the optimal hyperparameter settings of each network model are different for different networks, the optimization method proposed in this paper is not necessarily suitable for all the above models. Therefore, each of the above-mentioned network needs to pass ablation experiments to find the best hyperparameter settings of their respective networks with accuracy and time cost as indicators, as shown in Table 4.

A large number of experimental results show that no matter which type of network, embedding the LSTM can effectively improve the original network for the time series vibration signal from Table 4. The choice of the optimizer should be selected appropriately according to different networks. The only model in the table is MobilenetV2, which is replaced with LeakyRelu by the clippedRelu activation function, and the rest are Relu instead of LeakyRelu. The above-mentioned improved network is compared with the research in this paper. The same sample data set is used during the experiment, and the other parameter settings are the same as those described above. The final experimental results are shown in Table 5 and Figure 5 (the data in Table 5 are plotted in Figure 5). It can be seen that the improved ShufflenetV2-LSTM model training set accuracy rate is the only one to reach 100% and the test set accuracy rate is as high as 97.38%, which is higher than the accuracy rate of other models. In addition, the loss function value is
the lowest, indicating that there is little difference between the actual output and predicted value of the model. The training time of this model is shorter compared with other models, which is only longer than the improved Squeezenet model. The improved Squeezenet model has an accuracy rate of only 32.37%. The improved Alexnet is the network with the least number of convolutional layers in the table, with only 8 layers, but the time is nearly 3 times longer than the improved ShufflenetV2-LSTM. The test accuracy rate exceeds 90%: the improved MobilenetV2, the improved Resnet-18, and the improved Xception. However, it is, respectively, 2.75%, 5.51%, and 3.75% lower than the improved ShufflenetV2-LSTM and the time cost is also high. In particular, the Xception training time is 3 times that of the improved ShufflenetV2-LSTM.

Therefore, the improved ShufflenetV2-LSTM proposed in this article stands out in the face of many classification models, with the best timeliness and the highest accuracy.
and can efficiently and reliably solve the problem of fault diagnosis of rolling bearings.

4.4. Data Visualization. In order to more intuitively study and analyze the model learning process and LSTM classification capabilities proposed, t-SNE (T-Stochastic Neighbor Embedding) technology is used to visualize the feature distribution and analyze the performance of the constructed model [37].

As shown in Figure 6(a), the input data is messy and difficult to classify. Though the features extraction of the multi-convolutional neural layers, some faults can be distinguished effectively. However, different working conditions of the same fault location cannot be effectively distinguished, for instance, the 2 and 3 horsepower conditions with moderate inner ring failure and the 0 and 1 horsepower conditions with serious outer ring failure. In addition, the minor and moderate rolling body failures cannot be effectively distinguished with the same 0 horsepower condition. It can be seen that LSTM strengthens the correlation of data timing characteristics, separating and gathering effectively the above-mentioned confusing faults comparing the two Figures 6(b) and 6(c). This shows that LSTM is sensitive to vibration signals and strengthens the ability to distinguish subtle faults. The model can extract and classify the forty types of fault features of rolling bearings under different working conditions and at different positions in the Figure 6(c). The model can effectively learn different fault features in order to realize diagnosis and classification.

4.5. Model Generalization Ability Experiment and Result Analysis. In order to further verify the generalization ability of the proposed method, a rolling bearing fault test bench was built, as shown in Figure 7. A rolling bearing of TYPE ER-16K was selected to set up inner ring fault, rolling body fault, and outer ring fault, plus a group of healthy bearings. There were four bearing state types, and the specific faults are shown in Figure 8. Twelve kinds of fault signals were measured under 150 kg, 300 kg, and 500 kg loads, respectively. The sampling frequency was 12.8 kHz. Each small sample contained 1024 data points, and each type of signal was composed of 1280 small samples. 80% of the samples in the same category were randomly selected as the training set and the rest as the test set from Table 6 for data details. The same pretreatment is performed in Section 3.1: sample signals are decomposed by VMD, the first three IMF components with the best entropy are reconstructed, and the two-dimensional characteristic matrix is constructed. The network model mentioned in the previous section is used for fault diagnosis. The results of diagnosis accuracy and training time are shown in Table 7.

It can be seen from Table 7 that when the fault categories are less, the diagnosis effect of the improved Alexnet model becomes better and the accuracy can reach 96.74%, but its accuracy and operation time are still inferior to the algorithm proposed in this paper. The test accuracy of this model is 98.86%, only lower than 99.09% of the improved Xception, but the training time of the improved Xception is up to 70 minutes. In the table, only the improved SqueezeNet has a shorter calculation time than the model in this paper, but its test accuracy is 81.93%, which is the lowest among all models.
It can be seen that when the fault category becomes smaller, the gap between the algorithms is narrowed. However, considering the diagnosis accuracy and training time comprehensively, the improved ShufflenetV2-LSTM model still performs best. In addition, combined with the experimental fault diagnosis results in the previous section, it is shown that the proposed algorithm performs better than other algorithms and has certain reliability and generalization ability, which is suitable for intelligent diagnosis of various bearing faults.

5. Conclusion

This study proposes an improved ShufflenetV2-LSTM diagnosis system for the problem of rolling bearing fault diagnosis. Through the experimental study of the bearing failure data set, the following conclusions can be obtained:

(1) This model strengthens the time series correlation between vibration data through LSTM based on the capability to extract fault features of ShufflenetV2. It can be found that the ability to distinguish rolling bearing faults is significantly enhanced from the visual classification results. The experimental results of the two groups also show that compared with other network models, the proposed algorithm model can diagnose rolling bearing faults quickly and effectively and has excellent performance in diagnosis accuracy and training time

(2) In view of the problem of overfitting caused by many model parameters and small sample data, embedding the Dropout layer alone can effectively alleviate the degree of overfitting, but it will cause model instability and bring additional training time costs. This study proposes an improved method of replacing the Relu with LeakyRelu activation function and using the RMSprop optimizer to optimize the cross-entropy function to achieve the purpose of making up for the time cost of the Dropout layer. Experimental verification shows that the proposed improved method can not only avoid overfitting but also meet the real-time requirements

(3) Breaking the convention, this research found that the optimization effect of RMSprop is better than Adam for the model in this paper. It can be seen that the choice of optimizer cannot be generalized. For specific target tasks, how to select a suitable optimizer for a specific network model needs to be further explored in the future

Data Availability

The data used to support the findings of this study have not been made available because they are confidential.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors’ Contributions

Shaohui Ning conceived the experiments; Yansong Wang, Yukun Wu, and Yonglei Ren performed the experiments; Yansong Wang and Zhenlin Zhang programed the algorithm; Kangning Du and Wenan Cai processed the data; Shaohui Ning wrote the paper. All authors have read and approved the final manuscript.

Acknowledgments

This work was supported in part by the Shanxi Provincial Natural Science Foundation of China under Grant 201901D111239 and in part by PhD Starting Fund of Taiyuan University of Science and Technology (20182010).

References


