IMDRSN-BiLSTM for Rolling Bearing Fault Diagnosis

Yuan Xu[1,2], Heng-Wei Liao[1,2], Wei Ke[3], Yan-Lin He[1,2], Qun-Xiong Zhu[1,2], Yang Zhang[1,2], Ming-Qing Zhang[1,2*]

[1] College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China

[2] Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing 100029, China

[3] Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR 999078, P.R. China

xuyuan@mail.buct.edu.cn, mqzhang@buct.edu.cn

Abstract. To tackle the problem of decreased accuracy of deep residual shrinkage networks (DRSN) in the presence of strong noise, this paper proposes an improved multi-scale deep residual shrinkage network (IMDRSN) combined with bidirectional long short-term memory (BiLSTM) for rolling bearing fault diagnosis. Firstly, raw fault data is transformed into time-frequency images, and the Xception module captures multi-scale information in the images. Secondly, multiple scales of residual shrinkage building units (RSBU) are used to denoise the captured image information. Thirdly, introduce a Xception module into each RSBU to enhance the model's information retrieval capabilities. Incorporate a convolutional block attention module (CBAM) into each RSBU to strengthen the model's focus on key features, and introduce an adaptive module to reduce the constant bias impact of soft thresholding between input and output. Finally, the BiLSTM module is employed to capture the dependencies within the time series data, and to perform the task of fault classification. The IMDRSN-BiLSTM model is applied to the rolling bearing fault diagnosis task on the case western reserve university (CWRU) dataset in noisy environments, and experimental outcomes demonstrate that the IMDRSN-BiLSTM model delivers higher precision and robustness in identifying bearing malfunctions.

Keywords: DRSN, Xception Module, Adaptive Module, BiLSTM.

1 Introduction

Rolling bearings are crucial components in mechanical devices, yet they are also prone to damage. Bearing faults can lead to equipment downtime and damage, significantly impacting productivity and equipment lifespan. Consequently, the research and application of rolling bearing fault diagnosis technology are of great importance.

There are typically two main types of techniques for diagnosing faults in mechanical transmissions: signal analysis methods and those based on machine learning. Signal analysis methods [1] utilize pertinent mathematical models to directly analyze signals,

* Corresponding authors:
E-mail addresses: mqzhang@buct.edu.cn (M. Zhang)

subsequently extracting diagnostic features from the signal for fault diagnosis. Nevertheless, this category of signal analysis methods encounters difficulties in resolving issues with a multitude of overlapping signals. Deep learning methods [2], however, have the capability to learn relevant features from raw bearing one-dimensional vibration signals, thereby supplanting traditional statistical features and attaining superior diagnostic accuracy. At present, deep learning techniques have been extensively utilized in the domain of mechanical fault detection, yielding remarkable achievements [3,4].

Traditional deep learning faces challenges in parameter optimization. The derivatives of the loss function are required to be back-propagated layer-wise, and following multiple stages of reverse propagation, the gradients exhibit instability [5]. This also makes it difficult for input layer parameters to be effectively updated. The deep residual network (ResNet) represents an improvement over convolutional neural network (CNN). The ResNet model introduces an identity path on the basis of the CNN model. The identity path can alleviate gradient propagation issues, provide feature reuse mechanisms, and optimize the model training objectives, reducing the difficulty of model training [6]. Due to its excellent parameter adjustment capabilities, the ResNet is widely used in bearing fault diagnosis [7,8]. Indeed, despite the superior characteristics of the ResNet architecture compared to traditional CNN, it fails to adequately mitigate the issue of noise contamination in the fault detection task, particularly in the context of mechanical bearing fault diagnosis. The presence of noise is unavoidable in such scenarios, and its excessive presence can lead to a decline in the model's precision [9]. To address the issues of noise interference and feature extraction in the process of bearing fault detection, literature [10] proposes two deep learning network structures. Both network structures combine residual learning and shrinkage methods to effectively extract useful features from fault signals and reduce the impact of noise. Among them, the network with channel-shared thresholds learns features by sharing thresholds between channels to reduce noise effects. In contrast, the network with channel-wise independent thresholds uses different thresholds to handle features in each channel, better adapting to the differences between different channels and having better noise reduction capabilities. However, under severe noise conditions, the fault detection accuracy of the deep residual shrinkage network (DRSN) might decline. By incorporating attention mechanisms into deep learning architectures, the network can prioritize crucial features [11], improving fault diagnosis accuracy. The literature [12] introduces attention mechanisms into DRSN, weighting different regions through an attention branch, allowing the network to focus more on important areas, and also achieving significant performance improvements. The literature [13] introduces bidirectional long short-term memory (BiLSTM) network into DRSN to better obtain the dependencies in time series data, improving fault diagnosis accuracy. The literature [14] replaces the soft threshold function in the DRSN model with a semi-soft threshold function and introduces an adaptive slope module to eliminate the constant bias between input and output signals, thereby improving fault diagnosis accuracy.

Based on the previous discussion, this paper presents an improved multi-scale deep residual shrinkage network integrated with a BiLSTM network (IMDRSN-BiLSTM) for the purpose of fault detection in rolling bearings diagnosis in noisy con-

ditions. Firstly, the approach employs continuous wavelet transform (CWT) to represent the fault data in the form of time-frequency representations, subsequently drawing upon the Xception module to gather information from these representations across various scales. Secondly, multiple scales of residual shrinkage building units (RSBU) are used to denoise the captured image information. Thirdly, each RSBU is integrated with Xception modules to boost the model's information acquisition capabilities. Incorporating a convolutional block attention modules (CBAM) into each RSBU to enhance the model's focus on key features, while an adaptive module is employed to reduce the constant bias impact of soft thresholding between input and output. Finally, the BiLSTM module is employed to capture the dependencies within the time series data, and to perform the task of fault classification. The IMDRSN-BiLSTM model is utilized for fault diagnosis tasks in noisy environments using the case western reserve university (CWRU) rolling bearing fault dataset. The proposed IMDRSN-BiLSTM model has demonstrated superior performance in terms of accuracy and robustness, as evidenced by experimental results.

2. Preliminaries

2.1 Deep Residual Shrinkage Network (DRSN)

DRSN primarily employs feature extraction, residual learning, and shrinkage denoising to reduce noise in the input signal [15]. DRSN is constructed by stacking convolutional layers, a certain number of residual shrinkage building units (RSBU), batch normalization (BN) layers, global average pooling (GAP) layers, fully connected layers, and an output layer. The principle of RSBU involves the convolutional layer transforming the relatively important features into differences significantly larger than a threshold, while the features corresponding to redundant information are transformed into differences close to the threshold. Subsequently, a subnetwork within the module autonomously learns a set of thresholds between important and redundant features, ensuring that thresholds remain independent across different channels. Finally, soft thresholding is applied to eliminate redundant features while preserving important ones.

The subnetwork of the RSBU [16,17] consists of a GAP layer, fully connected layers, BN layers, rectified linear unit (ReLU) activation functions, and a fully connected layer with Sigmoid activation function. The subnetwork autonomously learns the thresholds for each channel. The learning process is as follows: the subnetwork begins by determining the magnitude of all signals present in the input feature map. It then incorporates a GAP layer to consolidate the features. Following this, the pooled feature map is inputted into a fully connected layer. Lastly, the fully connected layer employs the Sigmoid activation function to restrict the output range to between 0 and 1, represented by the variable 'a'. The formula for the Sigmoid function is shown in Eq (1).

$$y = 1/(1 + e^{-x}) \tag{1}$$

Where x denotes the input signal, and y signifies the output signal. By employing

GAP on the feature maps, a group of features is acquired, represented as A. Later, the ultimate threshold is established by taking the product of A and a. The structure of RSBU is presented in Fig. 1.

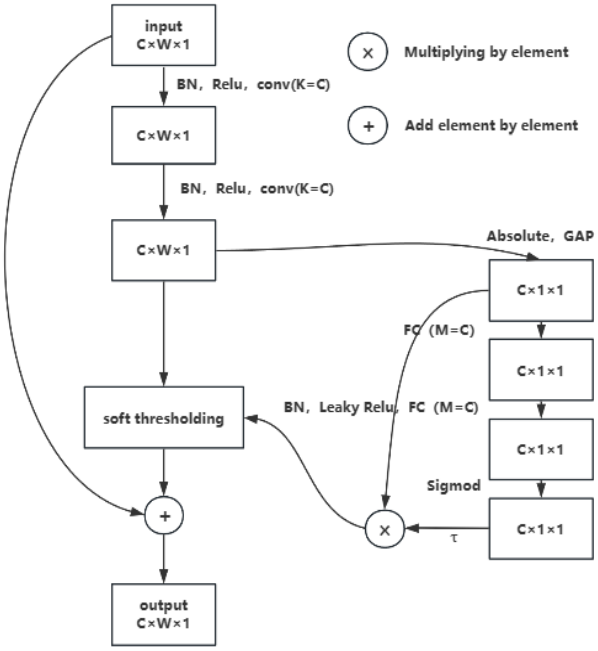

Fig. 1. The structure of RSBU.

The soft thresholding in the DRSN is a widely employed signal denoising technique. It achieves this by shrinking the signal values towards zero, thereby eliminating redundant information and focusing on the relevant information linked to the labels. The soft thresholding function, as expressed in Eq (2). Where x is the input signal, y is the output signal, and t is a pre-defined threshold. The output signal exhibits a certain degree of contraction towards zero compared to the input, effectively performing a nonlinear filtering process.

$$y = \begin{cases} x - t & , \ x \geq t \\ 0 & , \ -t \leq x \leq t \\ x + t & , \ x \leq -t \end{cases} \tag{2}$$

2.2 Convolutional Block Attention Module (CBAM)

The CBAM is a compact architectural element that integrates both spatial and channel-wise attention mechanisms. Its foundational concept revolves around the introduction of an attention mechanism in the spatial and channel dimensions, which enables the network to concentrate on more salient features and enhance its representational strength [19].

The spatial attention module enhances the model's ability to focus on significant spatial regions, allowing it to concentrate more intensively on essential areas and improve its perception of local details. The mechanism of spatial attention operates as follows: initially, both maximum and average pooling operations are executed along the channel axis of the incoming feature map to emphasize salient features. These features are subsequently amalgamated along the channel axis, thereby creating a consolidated feature map. This map undergoes processing by a convolutional layer, resulting in the generation of spatial attention coefficients. A Sigmoid activation function is then utilized to confine these coefficients within the bounds of 0 and 1. Ultimately, the derived spatial attention coefficients are imposed upon the initial feature map, thereby regulating the significance of features at each spatial position.

The channel attention module emphasizes channels that are relevant to the current task while reducing the impact of unrelated ones, thereby enhancing the feature representation of the relevant channels. The process of channel attention is described as follows: for the input signal, there is a dual application of global maximum and global average pooling operations conducted individually on each channel, resulting in the isolation of global maximum and average feature vectors. These vectors are subsequently inputs into a shared fully connected layer, which computates the attention weights for each channel. The outcome of this layer is a concatenation of the two vectors, producing the final attention weight vector. An application of the Sigmoid activation function then follows, which calculates the channel attention weights. These weights are applied to each channel of the original feature map, generating an updated feature map that incorporates the computed attention weights. By integrating these two modules, CBAM can significantly amplify the network's concentration on crucial features, thereby improving overall network performance.

3. Fault Diagnosis Based on IMDRSN-BiLSTM

In response to the reduced diagnostic accuracy of DRSN in noisy conditions, this paper introduces a fault diagnosis approach for rolling bearings that utilizes IMDRSN-BiLSTM. The procedure commences with the extraction of features from time-frequency images through the application of the Xception module. Subsequently, multi-scale feature processing is conducted using improved residual shrinkage building units (IRSBU) at various scales. Finally, a BiLSTM network is employed to facilitate fault categorization, thereby accomplishing the fault diagnosis mission.

The key components of the IMDRSN-BiLSTM network include the Xception module, IRSBU, CBAM and BiLSTM. The structure of IMDRSN-BiLSTM model is shown in Fig. 2.

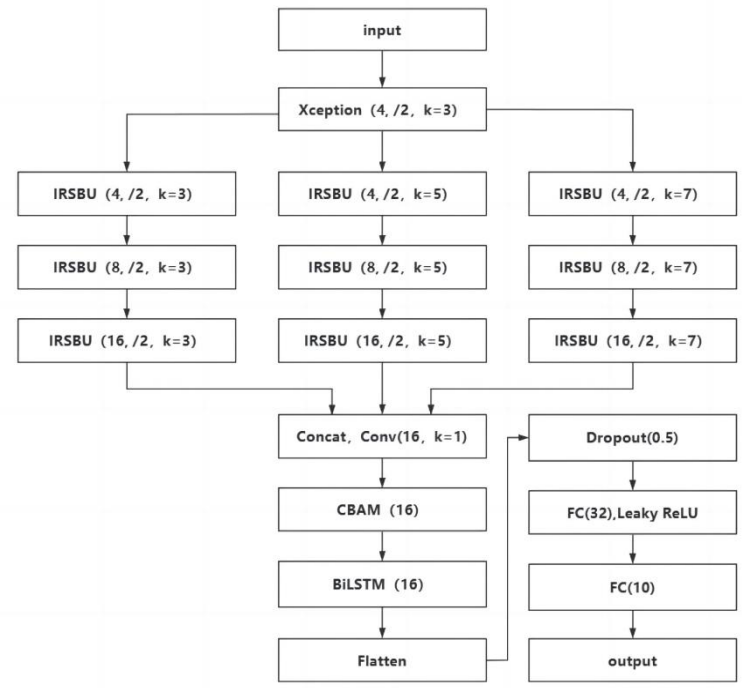

Fig. 2. The structure of IMDRSN-BiLSTM model.

3.1 Feature Extraction Based on Xception Model

The Xception network is a deep convolutional neural network that builds upon inception v3, offering improvements in accuracy, model size, and computational efficiency [18]. The key innovation of Xception is the replacement of traditional convolution layers in the inception model with depthwise separable convolutions, thereby maintaining performance while substantially reducing parameters and computational costs.

To improve feature extraction and decrease the dimensionality of time-frequency images, this paper simplifies the intermediate layers of the Xception network to function as a feature extraction module, thereby preventing data loss. Furthermore, a max pooling layer is incorporated to reduce the dimensionality of the time-frequency images. The intermediate layers of Xception primarily consist of basic blocks, residual connections, and bottleneck layers. The basic block is the core of the Xception network, utilizing depthwise separable convolutions to replace the multiple convolutional layers found in traditional inception modules. Residual connections connect the input signal to the output signal, mitigating the vanishing gradient problem. The constriction point is represented by a 1×1 convolutional layer, utilized for reestablishing the channel count. The Xception module developed in this study comprises fundamental blocks, residual connections, bottleneck stages, and a max pooling layer, as depicted in Fig. 3.

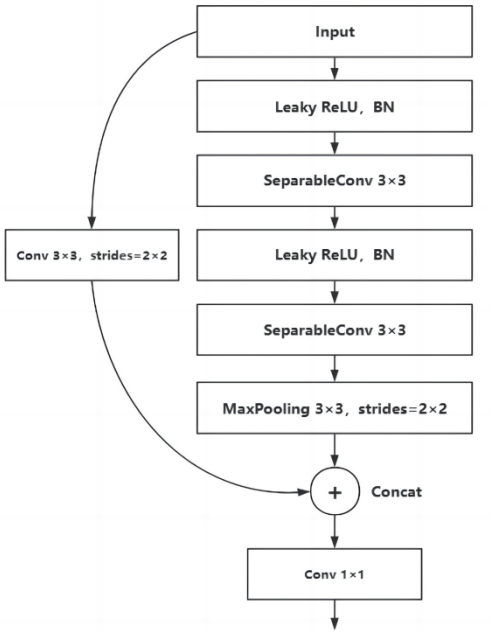

Fig. 3. The structure of Xception module.

The feature extraction process using the Xception module mainly involves using depthwise separable convolutional layers within the module to process the data. Depthwise convolution and pointwise convolution are the two distinct steps that make up depthwise separable convolutions. In depthwise convolution, each input channel is convolved independently, resulting in a significant reduction in the number of parameters. Subsequently, pointwise convolution, which has a kernel size of 1, combines the channels from the output of the depthwise convolution to produce the final feature map.In depthwise convolution, each channel of the input feature maps is processed separately. The computation process for the output ($\mathrm{F_{dw}}$) of depthwise convolution is depicted in Eq (3).

$$\mathrm{F_{dw}(i,j,m)} = \sum_{k}^{K} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} K(p,q,m) * F_{in}(i+p, j+q, m) \tag{3}$$

Where $\mathrm{F_{in}}$ refers to the input feature map, with i and j symbolizing the height and width of the resulting feature map, respectively, while m signifies the count of channels in the input feature map. The convolutional kernel utilized for the depthwise convolution is of size P×Q, represented by K. Following the depthwise convolution, the pointwise convolution operation consolidates the outputs. The computational procedure for the resulting output ($\mathrm{F_{pw}}$) of the pointwise convolution is delineated in Eq (4).

$$\mathrm{F_{pw}(i,j,m)} = \sum_{\mathrm{m=1}}^{\mathrm{M}} \sum_{\mathrm{p=0}}^{1} \sum_{\mathrm{q=0}}^{1} \mathrm{K(p,q,n,m)} * \mathrm{F_{dw}(i+p, j+q, m)} \tag{4}$$

Where P represents the pointwise convolutional kernel of dimensions 1×1, while N signifies the total number of output channels. Through these two processes, the

depthwise separable convolutional layer is able to generate new feature maps from the initial input and integrate them across the channel axis.

3.2 Feature Denoising Based on Multiple Scales of IRSBU.

The primary function of the RSBU is to process the acquired fault signals and reduce the interference of noise. The multi-scale feature processing based on the IRSBU proceeds as follows:

Firstly, RSBU of multiple scales utilize convolutional layers to downsample the data extracted by the Xception module. However, the downsampling process can lead to data loss, which might result in a decline in the precision of fault identification. In this paper, the second convolutional layer of the RSBU is replaced with the Xception module. The introduction of the Xception module allows each RSBU to read data in multiple scales while reducing the parameters and computational load introduced by multi-scale processing.

Secondly, the data distilled from the downsampled Xception component is reinforced by the CBAM module, emphasizing crucial features. The spatial attention mechanism discerns the interdependencies among features at distinct spatial locations, whereas the channel attention mechanism highlights connections within input feature channels. By integrating these dual attention mechanisms, CBAM efficiently boosts the network's attention on vital details, ultimately enhancing its overall efficacy.

Then, the soft thresholding function is used to directly subtract the threshold from intermediate signals that exceed the threshold, effectively removing noise interference. However, this process can also lead to the loss and distortion of some valid information due to the constant bias between the input and output signals. Therefore, this paper constructs and improves an adaptive module within the RSBU to analyze the input signals and generate parameters to correct the output signals. Within the adaptive module, the input signal is first transformed into a one-dimensional vector by the GAP layer, then two fully connected layers with different channel numbers generate scale parameters, and finally, a fully connected layer with a Sigmod activation function converts the scale parameters into a range of 0 to 1, yielding correction parameters. Through this process, the adaptive module can autonomously learn to generate correction parameters for different channels. Utilizing the parameters of the corresponding channels to individually correct the output signals can reduce the impact of the constant deviation between the input and output signals. The output signal correction formula is shown in Eq (5), where F_x is the input signal, F_y is the output signal, and t is the threshold determined by the model.

$$F_y = \begin{cases} \alpha\left(F_x - t\right) & , \ F_x \geq t \\ 0 & , \ -t \leq F_x \leq t \\ \alpha\left(F_x + t\right) & , \ F_x \leq -t \end{cases} \tag{5}$$

Finally, the feature processing outcomes from the IRSBU (see Fig. 4) at diverse scales are merged along the channel axis, and the channel count is reduced by employing dot product operations to reduce overall complexity. Subsequently, the CBAM is

employed to adjust the weights of the fault data both in the channel and spatial dimensions.

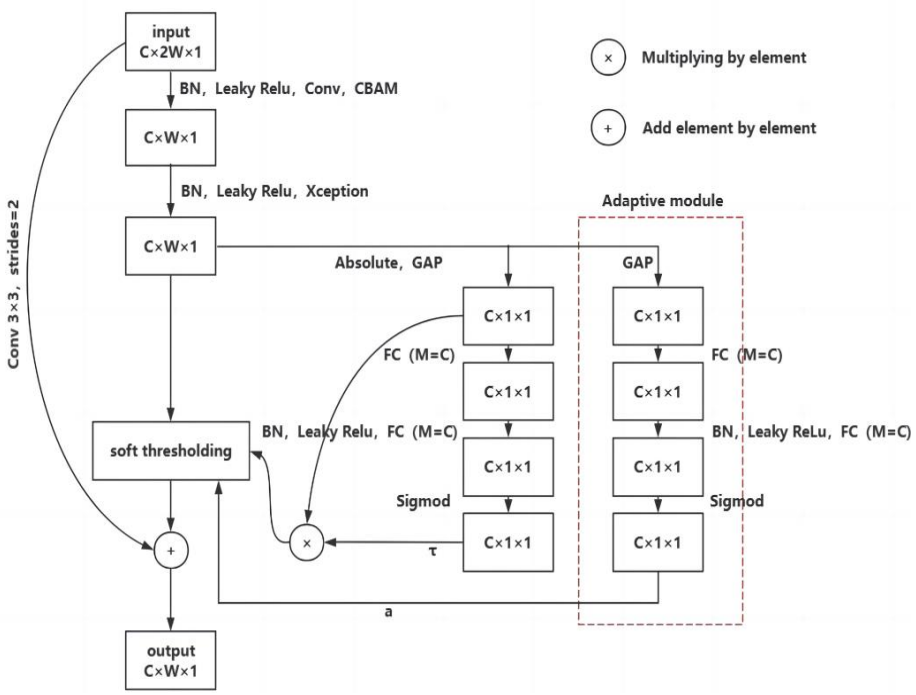

Fig. 4. The structure of IRSBU.

3.3 Fault Classification Based on BiLSTM

The BiLSTM is capable of remembering long-term dependencies within a sequence due to its built-in memory cells that learn contextual relationships, thereby enhancing the representation capability of the model [20]. The BiLSTM model's architecture primarily comprises the following elements: the input layer, the long short-term memory (LSTM) cells, the output layer, and a classification layer for tasks involving classification. The process of fault detection using BiLSTM operates as follows:

Firstly, the input layer of BiLSTM receives sequence data that has undergone feature-weighted processing. Next, within the hidden layer of the BiLSTM, two LSTM operating in opposite directions process the data sequentially and in reverse temporal order, respectively, to capture contextual information within the sequence. The process of BiLSTM data processing is illustrated in Fig. 5, where the symbols $\overrightarrow{h_n}$ and $\overleftarrow{h_n}$ represent the hidden layer outputs of the BiLSTM during forward and backward propagation, respectively.

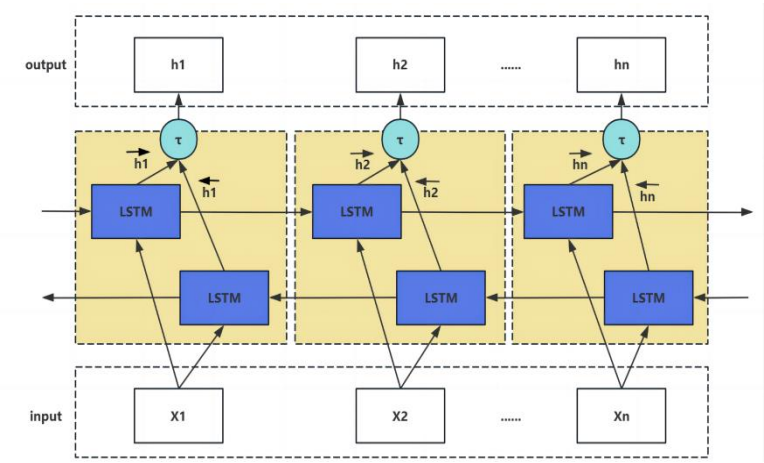

Fig. 5. Forward and backward data processing in BiLSTM model

Subsequently, the BiLSTM combines the outputs of the forward and backward LSTM layers by concatenation to obtain the output results, as shown in Eq (6). Finally, the classification layer of the BiLSTM performs diagnostic classification on the fault data.

$$h_n=[\overrightarrow{h_n},\overleftarrow{h_n}] \tag{6}$$

4. Experimental Validation And Result Analysis

4.1 Dataset Description

The CWRU rolling bearing dataset is a widely recognized fault diagnosis dataset for bearings [21]. The data is classified based on the location of the fault, which includes the inner race (IR), outer race (OR) and rolling body (B) and the size of the fault, which includes fault diameters of 7, 14 and 21 mils [22]. The data is also classified based on the faulty load, which includes 0HP, 1HP, 2HP and 3HP. The load affects the speed of the bearing, and the operating conditions are also different. The normal state is represented as NO.

For the experiment, the experiment is conducted using fault data from the drive end in the 0HP state, with the bearing rotational speed at 1797 RPM. The faults are classified into 10 categories based on their location and size, with each category having 500 samples, and each sample having a length of 1024 signals. The data division of the training set and test set follows a 90% and 10% ratio. The fault situations NO, B-07, B-14, B-21, IR-07, IR-14, IR-21, OR-07, OR-14, OR-21 correspond to the labels 0-9, respectively.

4.2 Simulation Experiment

To guarantee the impartiality of experimental outcomes, the structures and parameters of all models are maintained uniform. All these models employ the Adam optimizer and the cross-entropy loss function. The batch size for training is set to 32, with a total of 40 training epochs being performed. The time-frequency images used as input have a size of 128×128×1, and the Leaky ReLU activation function has a coefficient of 0.01. The L2 regularization coefficient is 0.001. The BiLSTM model has a hidden layer size of 16, and the dropout layer has a coefficient of 0.5. Gaussian noise is added with a fixed size, and the signal-to-noise ratio is set to -2. The models used for comparison are: BiLSTM, ResNet-BiLSTM, DRSN, IMDRSN and IMDRSN-BiLSTM.

To minimize the effect of random variables on the experimental outcomes, each model underwent a quadruple replication process, averaging the results from four successive trials. The mean and standard deviation derived from these repetitions served as the decisive evaluation criteria. The outcomes of four replicate experiments along with their mean values are presented in Tables 1 and 2, respectively.

Table 1. Accuracy results of repeated fault diagnosis experiments for each model. (unit:%)

Diagnostic model	Experimental rounds			
	1	2	3	4
IMDRSN-BiLSTM	**99.67**	**99.50**	**99.83**	**99.67**
IMDRSN	99.60	99.20	99.60	99.40
DRSN	99.40	99.40	98.60	99.00
ResNet-BiLSTM	98.80	98.80	99.20	99.20
BiLSTM	98.40	98.80	98.60	97.80

Table 2. Average accuracy and standard deviation results of each model. (unit:%)

Model	Accuracy rate ± Standard deviation
IMDRSN -BiLSTM	**99.67±0.12**
IMDRSN	99.45±0.17
DRSN	99.10±0.33
ResNet-BiLSTM	99.00±0.20
BiLSTM	98.40±0.37

As indicated in Table 2, the diagnostic accuracy of the BiLSTM model is the lowest, at 98.40%, primarily because the BiLSTM model is designed for processing sequential and textual data. The diagnostic accuracy of the ResNet-BiLSTM model is 99.00%, a 0.60% improvement over the BiLSTM model. This can be attributed to the introduction of identity paths, which facilitate the preservation of crucial feature information throughout the network, thereby significantly boosting the model's representational power. The diagnostic accuracy of the DRSN is 99.10%, which is an improvement of 0.10% over the ResNet-BiLSTM model, effectively removing some noise interference. The IMDRSN shows a 0.35% improvement over the DRSN. This is because the Xception module expands the receptive field of the DRSN, the CBAM helps the DRSN focus more on important features, the adaptive component focuses on resolving the recurrent discrepancy between the output and the input data, and the multi-scale information processing of the DRSN ensures more complete data preservation. The IMDRSN-BiLSTM model demonstrates a 0.22% increase in accuracy compared to the IMDRSN. This is due to the fact that the IMDRSN architecture integrates the BiLSTM model, allowing the model to track the temporal relationships within time-series data. Experimental outcomes demonstrate that the IMDRSN-BiLSTM model delivers higher precision and robustness in identifying bearing malfunctions.

5. Conclusions

This paper presents a fault diagnosis approach based on the IMDRSN-BiLSTM model. Firstly, raw fault data is transformed into time-frequency images, and the Xception module captures multi-scale information in the images. Secondly, the RSBU module within the DRSN framework is replaced with a multi-scale RSBU to effectively denoise features extracted from the image data. Thirdly, each RSBU is integrated with the Xception module to enhance the model's capacity for information extraction. By incorporating the CBAM module with each RSBU, the model's attention to crucial features is improved. An adaptive module is introduced to mitigate the inherent bias between input and output during the soft thresholding process. Finally, the BiLSTM module is employed to capture the temporal dependencies within the sequence data, facilitating the fault classification task. Experimental results on the CWRU dataset, which is subjected to noise interference, demonstrate the superior performance of the proposed IMDRSN-BiLSTM model in terms of both accuracy and robustness.

Acknowledgments. This research is supported by the National Natural Science Foundation of China under Grant No.61973022, the Fundamental Research Fund for the Central Universities No.ZY2417, and Hubei Key Laboratory of Intelligent Robot (Wuhan Institute of Technology) No.ZHBIR202306.

References

1. Li H, Liu T, Wu X, et al. A bearing fault diagnosis method based on enhanced singular value decomposition[J]. IEEE Transactions on Industrial Informatics, 2020, 17(5): 3220-3230.

2. Shao H, Han S Y, et al. Novel joint transfer network for unsupervised bearing fault diagnosis from simulation domain to experimental domain[J]. IEEE Transactions on Mechatronics, 2022, 27(6): 5254-5263.
3. Hoang D T, Kang H J. A motor current signal-based bearing fault diagnosis using deep learning and information fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 69(6): 3325-3333.
4. Chen X, Zhang B, Gao D. Bearing fault diagnosis base on multi-scale CNN and LSTM model[J]. Journal of Intelligent Manufacturing, 2021, 32(4): 971-987.
5. Tao H, Qiu J, Chen Y, et al. Unsupervised cross-domain rolling bearing fault diagnosis based on time-frequency information fusion[J]. Journal of the Franklin Institute, 2023, 360(2): 1454-1477.
6. Wen L, Li X, Gao L. A transfer convolutional neural network for fault diagnosis based on ResNet-50[J]. Neural Computing and Applications, 2020, 32(10): 6111-6124.
7. Liu R, Wang F, Yang B, et al. Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions[J]. IEEE Transactions on Industrial Informatics, 2020, 16(6): 3797-3806.
8. Roy S K, Manna S, Song T, et al. Attention-based adaptive spectral–spatial kernel ResNet for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(9): 7831-7843.
9. Dandan P, Huan W, Zhiliang L, et al. Multibranch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition[J]. IEEE Transactions on Industrial Informatics, 2020, 16(7): 4949-4960.
10. Zhao M, Zhong S, Fu X, et al. Deep residual shrinkage networks for fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2019, 16(7): 4681-4690.
11. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.
12. Lu W, Li J, Wang J, et al. A CNN-BiLSTM-AM method for stock price prediction[J]. Neural Computing and Applications, 2021, 33(10): 4741-4753.
13. Li Tianhui, Xie Yuncheng, Che Ronghua, et al. Power information network intrusion detection model based on DRSN-BiLSTM[J]. Electric Power Automation Equipment, 2023, 21(09): 30-37.
14. Tang Shiyu, Tong Jin Yu, Zheng Jinde, et al. Improved fault diagnosis method of deep residual shrinkage network bearings[J]. Journal of Vibration and Shock, 2023, 42(18): 217-224+285.
15. Ehteram M, Afshari Nia M, Panahi F, et al. Gaussian mutation–orca predation algorithm–deep residual shrinkage network (DRSN)–temporal convolutional network (TCN)–random forest model: an advanced machine learning model for predicting monthly rainfall and filtering irrelevant data[J]. Environmental Sciences Europe, 2024, 36(1): 13.
16. Dai C, Hu S, Zhang Y, et al. Cavitation state identification of centrifugal pump based on CEEMD-DRSN[J]. Nuclear Engineering and Technology, 2023, 55(4): 1507-1517.
17. Yang J, Kong L, Ye H. Surface hardness determination of laser cladding using laser-induced breakdown spectroscopy and machine learning (PLSR, CNN, ResNet, and DRSN)[J]. Applied Optics, 2024, 63(10): 2509-2517.
18. Gülmez B. A novel deep neural network model based Xception and genetic algorithm for detection of COVID-19 from X-ray images[J]. Annals of Operations Research, 2023, 328(1): 617-641.
19. Chen B, Zhang Z, Liu N, et al. Spatiotemporal convolutional neural network with convolutional block attention module for micro-expression recognition[J]. Information, 2020, 11(8): 380.

20. Lu W, Li J, Wang J, et al. A CNN-BiLSTM-AM method for stock price prediction[J]. Neural Computing and Applications, 2021, 33(10): 4741-4753.
21. Liu C, Gryllias K. Simulation-driven domain adaptation for rolling element bearing fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2021, 18(9): 5760-5770.
22. Ayas S, Ayas M S. A novel bearing fault diagnosis method using deep residual learning network[J]. Multimedia Tools and Applications, 2022, 81(16): 22407-22423.