An Improved MSCNN and GRU Model for Rolling Bearing Fault Diagnosis

Teng Wang* – Youfu Tang – Tao Wang – Na Lei

Northeast Petroleum University, Mechanical Science and Engineering Institute, China

In this paper, a novel fault diagnosis method based on the fusion of squeeze and excitation-multiscale convolutional neural networks (SENet-MSCNN) and gate recurrent unit (GRU) is proposed to address the problem of low diagnosis rate caused by the fact that normal samples are much larger than fault samples in the vibration big data. The method takes the time-domain vibration signal as input and fuses the spatial features extracted by SENet-MSCNN. The temporal features extracted by GRU in order to bring them into the fully connected layer for identification so as to realize the intelligent diagnosis of rolling bearing adaptive feature extraction. Finally, the method is applied to the simulated signal and experimental data for testing and analysis. The results reveal that the model can reach 98.98 % and 76.44 % migration diagnostic accuracy in bearing and gearbox datasets. At the same time, it has strong noise immunity, adaptivity, and robustness, providing an effective way for intelligent diagnosis of rolling bearing vibration big data.
Keywords: SENet, multiscale convolutional neural networks, gate recurrent unit, rolling bearing, fault diagnosis

Highlights
- A novel integration method of SENet-MSCNN and GRU is proposed, which can more effectively and adaptively extract the fault features of rolling bearing and fault diagnosis.
- Based on the existing problems of convolutional neural networks (CNN), we developed MSCNN. The multiscale convolution kernel in MSCNN not only considers the global basic features of the signal but also extracts local detail features.
- The SENet is added to MSCNN to recalibrate multiscale features, which can reduce attention to irrelevant information and pay more attention to motivate important information.
- SENet-MSCNN is good at reducing frequency variance and extracting spatial features, and GRU is good at extracting long sequence time-series features. The integration model has better robustness in real working conditions and can reach 98.98 % accuracy under variable load conditions.

0 INTRODUCTION

As a key component in rotating machinery, rolling bearings often work in severe environments of high load and high speed, so they are highly prone to failure. Research shows that bearing faults account for the majority of the total number of faults [1]. Therefore, it is of great theoretical significance and engineering application value to research rolling bearing fault diagnosis methods to ensure the continuous and safe operation of equipment and reduce the economic loss of downtime [2].

However, the rolling bearing vibration big data caused by variable working conditions and shock excitation have typical nonlinear non-stationary complex characteristics [3] and [4], which makes the existing signal processing techniques, such as time-domain statistical analysis [5] and [6], frequency-domain spectral analysis [7] and [8], short-time Fourier transform [9], wavelet analysis [10] and [11], Hilbert-Huang transform (HHT) [12], variational mode decomposition (VMD) [13], difficult to extract fault features adaptively.

In contrast, the normal samples are much larger than the fault samples in the massive rolling bearing vibration data collected in the field, which makes the diagnostic efficiency and recognition rate of the existing artificial intelligence diagnosis methods, such as support vector machine (SVM) [14], decision tree (DT) [15], and random forest (RF) [16], artificial neural network (ANN) [17], CNN [18] and [19], deep autoencoder (DAE) [20], deep belief network (DBN) [21], recurrent neural network (RNN) [22] and artificial immune algorithm (AIN) [23] difficult to apply in industrial contexts.

At present, researchers at home and abroad have fused signal-processing techniques with artificial intelligence diagnosis methods, such as wavelet packet decomposition and empirical mode decomposition (EMD) with back propagation (BP) network fusion [24], VMD and probabilistic neural network (PNN) [25], wavelet and CNN network fusion [26], HHT, and CNN network fusion [27]. These methods have been effective in improving the diagnostic performance of the large sample and various fault vibration data. However, vibration data is affected by different working conditions, structural parameters, fault types, fault degrees, and the number of faults. The above fusion methods have their applicability conditions

*Corr. Author's Address: Northeast Petroleum University, Mechanical Science and Engineering Institute, China, wangteng_1009@163.com

261

and need to be artificially selected based on experts' empirical knowledge, which has greater limitations.

In addition, some researchers improve fault diagnosis accuracy and robustness by fusing different artificial intelligence methods. These diagnosis methods mainly contain two parts: feature extraction and pattern recognition. A convolutional discriminative feature learning approach and support vector machine fusion method was proposed by Sun et al. [28] for fault diagnosis of induction motors; the network performance is improved. Chen et al. [29] proposed a mechanical fault diagnosis method based on CNN and extreme learning machine (ELM) by using ELM as a classifier of CNN with the advantages of fast learning speed and high generalization ability. The network performance is further improved, and the model generalization ability and convergence speed are also enhanced. Wang et al. [30] proposed a CNN-based hidden Markov model for rolling bearing fault identification by fusing the strong feature extraction capability of CNN and the excellent pattern recognition performance of the hidden Markov model. Compared with the CNN model alone, it has higher classification accuracy and robustness. Based on the excellent network performance of CNN, the above methods have achieved better performance by combining CNN with various artificial intelligence methods. However, the CNN, as a feature extraction layer, extracts high-dimensional features and contains a large amount of spatial information and sequence information. Another approach, as a classifier, classifies the spatial features extracted by the CNN without considering the connection between the features. Based on this idea, taking advantage of RNN for extracting temporal features and CNN for extracting spatial features, some researchers have combined CNN and RNN [31] for fault diagnosis and achieved better accuracy. However, RNNs are prone to gradient explosion and gradient disappearance. Gate recurrent unit (GRU) [32] and long short-term memory (LSTM) [33], as a variant of RNN, can solve gradient vanishing and gradient explosion problems. A planetary gearbox diagnosis method based on CNN and LSTM is proposed by Shi et al. [34]; the network is able to detect the type, location, and direction of gearbox faults with greater accuracy and a higher recognition rate than traditional a single CNN. Chen et al. [35] found that the features extracted by a size convolutional kernel are more singular; a multi-scale convolutional neural network and long shor-term memory (MSCNN-LSTM) fault diagnosis model was proposed. The average accuracy in the experimental data reached 98.46 % and has strong noise immunity.

Li et al. [36] proposed a method that combines CNN and GRU models with vibration and acoustic emission signals for gear-pitting fault diagnosis. The method can achieve a diagnosis rate of more than 98 % and exhibits stronger robustness compared with a single CNN and GRU for different loads and learning rates. The fusion of two different deep learning methods is to take the advantage of both models and make the model representation more powerful. However, it is easy to deepen the network layer depth, leading to model overfitting.

To address these problems, a fault diagnosis method based on the fusion of SENet-MSCNN and GRU is proposed. The width of the convolutional layers of the network is increased by adding convolutional kernels of different scales to form MSCNN layers, without increasing the depth of the network structure. Convolutional kernels of different sizes can capture different perceptual field features to obtain global and local information. In addition, the features extracted by MSCNN are not all important, which can easily cause redundant information and irrelevant information to influence the classification results. Thus, the SE-Net block [37] was introduced into MSCNN to recalibrate multiscale features to reduce attention to irrelevant information and pay more attention to motivate important information. Then, the spatial features extracted by SENet-MSCNN were input to GRU to extract time-series features. Compared with other fault diagnosis methods, the method has broad application prospects in improving the accuracy of rolling bearing fault diagnosis. In addition, the method is attractive in reducing failure rates, reducing maintenance and repair costs of machinery and equipment, and preventing accidents.

1 FAULT DIAGNOSIS MODEL BASED ON SENet-MSCNN AND GRU METHOD

The MSCNN extracts the fault features through several convolutional kernels of different sizes and fuses the multiscale features. Then the fused features are fed into the GRU network to extract the time-series features and classify them while adding the SE-Net into the MSCNN to enable the network model to recalibrate multiscale features, which can further improve the diagnosis rate and robustness of the fault model.

1.1 Architecture of MSCNN

Since the rolling bearing vibration signal presents nonlinear and nonstationary characteristics, the high-

frequency features cannot be extracted by larger convolutional kernels and the low-frequency features cannot be extracted by smaller convolutional kernels. Therefore, to solve this problem, this paper uses the MSCNN structure, which includes a multiscale convolutional layer by connecting convolutional kernels of different sizes [1×1, 3×1, 5×1], whose structure is shown in Fig. 1. Convolutional kernels [3×1] extract high-frequency fault features; convolutional kernels [5×1] extract low-frequency fault features. The features extracted from several different receptive fields possess both global and local information [38] and [39]. A convolutional kernel of size [1×1] is added to each of the four branches of MSCNN, which has two advantages: first, although a [1×1] convolutional kernel cannot extract spatial features, it can extract features along the depth dimension to achieve a nonlinear feature map. Second, several [1×1] convolutional kernels are embedded in the front of [3×1, 5×1] convolutional kernels and can reduce dimensionality to reduce the computational cost. It can accelerate training and improve generalization. The calculation process of the convolution is as follows:

$$[x * g](n) = \sum_{\tau=-\infty}^{\infty} x(\tau) g(n-\tau), \qquad (1)$$

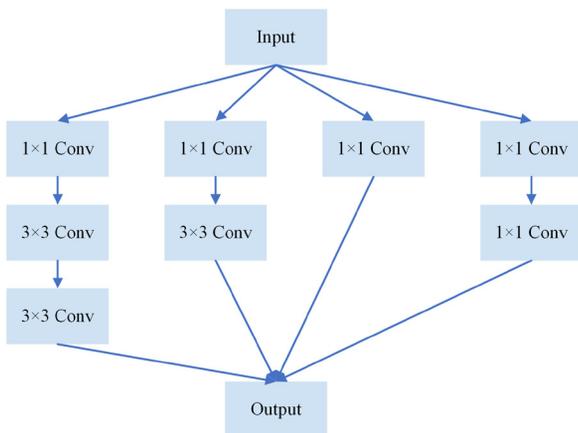where x denotes the amplitude, and g denotes the multiscale convolution kernel.

Fig. 1. *The structure diagram of MSCNN, which includes a multiscale convolutional layer by connecting convolutional kernels of different sizes [1×1, 3×1, 5×1]*

By using the scaled exponential linear units (SELU) activation function [40], the data distribution is self-normalized to satisfy a normal distribution with mean 0 and variance 1. Moreover, the SELU activation function is a non-saturated function, which can solve the vanishing gradient and exploding gradient problem. Its function expression is as follows:

$$selu(x) = \lambda \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases}, \qquad (2)$$

where α and λ denote constants.

1.2 Architecture of SENet

A new module is introduced in the MSCNN model: SENet, whose detailed structure is shown in Fig. 2. The biggest advantage of the SENet block is that it can construct interdependencies between channels [41]. SENet adopts a feature recalibration mechanism, which can obtain the dependency degree of each channel feature through global information. Then, by the dependency degree, the important information is selectively enhanced and the irrelevant information is squeezed to recalibrate the relationship between channel-wise features. Thus, it aids in strengthening the convolutional kernel learning capability and improve the feature representation capability of MSCNN. The formulas are as follows:

$$z_c = GAP(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j), \qquad (3)$$

$$s_c = \sigma\left(\mathbf{W}_2 * \delta\left(\mathbf{W}_1 * z_c\right)\right), \qquad (4)$$

$$\mathbf{M} = [m_1, m_2, m_3, ..., m_c] = F_{scale}\left(s_c \bullet u_c\right), \qquad (5)$$

where $u_c \in R^{H \times W}$ is input feature, $z_c = R^{1 \times c}$ the channel-wise feature vector, $s_c = R^{H \times W}$ recalibration vectors, and $\mathbf{M} \in R^{H \times W}$ reconstructing feature vector. $\mathbf{W}_1 \in R^{D/r \times D}$ and $\mathbf{W}_2 \in R^{D \times D/r}$ $z_c = R^{1 \times c}$ are weights $*$ convolution operator, σ Sigmoid function, δ Relu function, r reduction ratio, and $F_{scale}\left(\bullet\right)$ is scalar multiplication.
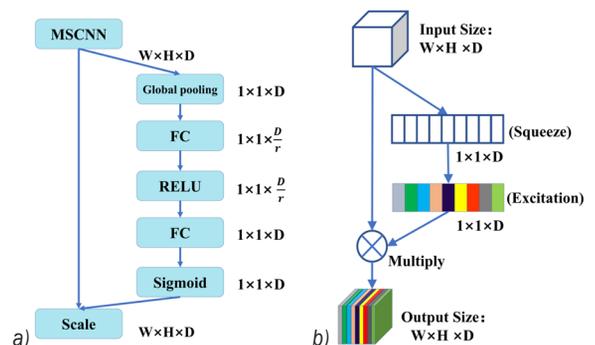

Fig. 2. *The structure diagram of SENet; a) SENet module, and b) SENet block*

1.3 Architecture of GRU

The GRU network is a simplified version of the LSTM, which has a simpler structure, with lower computational cost, faster iterations, and no reduction in network performance compared to the LSTM. The GRU network has only two gated units: the update gate and the reset gate. With these two gated units, it can learn, discard, and retain information in a long-term sequence and influence the output of the next iteration. As shown in Fig. 3, the input vector $\mathbf{x}_{(t)}$ and the previous state vector $\mathbf{h}_{(t-1)}$ are connected to two fully connected layers, through the Sigmoid function mapping the result $z_{(t)}$ and $r_{(t)}$ between 0 and 1.

Fig. 3. *The structure diagram of GRU; the GRU network has only two gated units: the update gate and the reset gate*

The formulas can be given as follows:

$$z_{(t)} = \sigma\left(\mathbf{W}_{xz}^{\mathrm{T}} + \mathbf{W}_{hz}^{\mathrm{T}}\mathbf{h}_{(t-1)} + b_z\right), \quad (6)$$

$$r_{(t)} = \sigma\left(\mathbf{W}_{xr}^{\mathrm{T}}x_{(t)} + \mathbf{W}_{hr}^{\mathrm{T}}\mathbf{h}_{(t-1)} + b_r\right), \quad (7)$$

$$g_{(t)} = tanh\left(\mathbf{W}_{xg}^{\mathrm{T}}x_{(t)} + \mathbf{W}_{hg}^{\mathrm{T}}\left(r_{(t)} \otimes \mathbf{h}_{(t-1)}\right) + b_g\right), \quad (8)$$

$$h_{(t)} = z_{(t)} \otimes \mathbf{h}_{(t-1)} + \left(1 - z_{(t)}\right) \otimes g_{(t)}, \quad (9)$$

where σ represents Sigmoid function. $\mathbf{W}_{xz}$, $\mathbf{W}_{xr}$, and $\mathbf{W}_{xg}$ represent the weight matrices of for their connection to the input vector $\mathbf{x}_{(t)}$. $\mathbf{W}_{hz}$, $\mathbf{W}_{hr}$, and $\mathbf{W}_{hg}$ represent the weight matrices of for their connection to the vector $\mathbf{h}_{(t-1)}$. b_z, b_r, and b_g are the bias. $\otimes$ is scalar multiplication.

1.4 Intelligent Fault Diagnosis Methods Based on SENet-MSCNN and GRU Model

As shown in Fig. 4, rolling bearing fault diagnosis consists of three major parts: the SENet-MSCNN layer, GRU layer, and Dense layer. The method is based on the improvement of the integration method of the CNN and LSTM, with which CNN is good at reducing the vibration frequency variance and GRU is good at extracting time-series features. Combining and improving the CNN and GRU fusion model, next the SENet-MSCNN and GRU fault diagnosis model is proposed. First, the time domain signal of bearing fault vibration is directly served as the input of SENet-MSCNN to extract multiscale features through a multiscale convolution kernel. The multiscale features are input to the SENet to recalibrate features. Then, the high dimensional multiscale features are Global Average Pooling to reduce dimensions. The low dimensional features are input to the GRU layer to extract time-series features. Finally, the features are input to the fully connected layer for classification by the Softmax function.
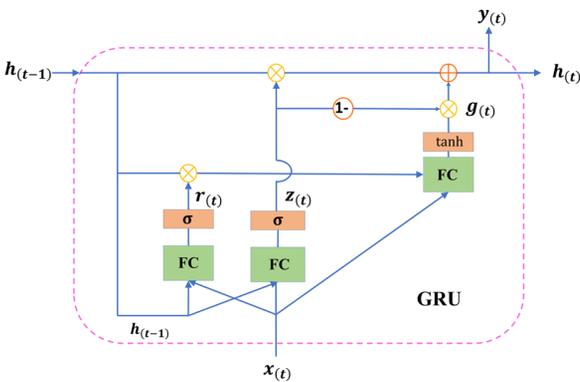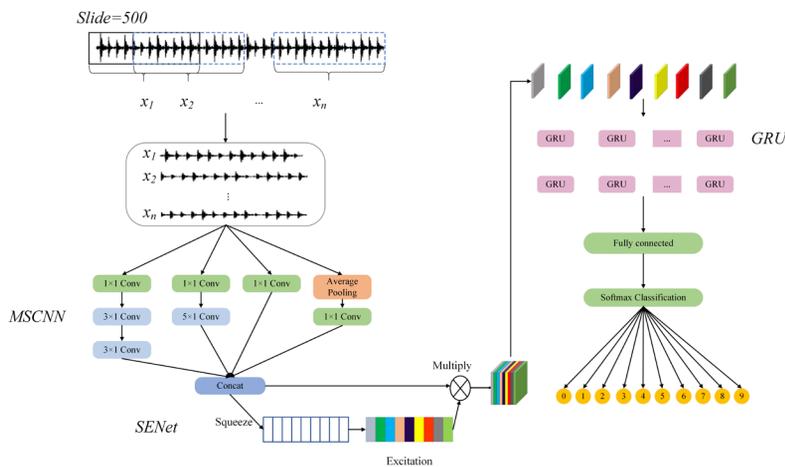

Fig. 4. *The framework of the SENet-MSCNN and GRU model, which consist of three major parts SENet-MSCNN layer, GRU layer, and Dense layer*

1.5 Signals Application of SENet-MSCNN and GRU Methods in Simulated Signals

1.5.1 Construction of Simulated Signal Data Sets

The raw signal of the rolling bearing is simulated using three simplified models from the literature [42]: $x_1(t)$, $x_2(t)$, $x_3(t)$, which are:

$$x_1(t) = 0.4\cos(2\pi f_1 t + 10), \qquad (10)$$

$$x_2(t) = 0.6\cos(2\pi f_2 t - 15), \qquad (11)$$

$$x_3(t) = \sin(2\pi f_b t)\left[1 + \sin(2\pi f_r t)\right], \qquad (12)$$

where $f_1 = 20$ Hz, $f_2 = 45$ Hz, $f_3 = 100$ Hz, $f_r = 10$ Hz, $N = 2048$, sampling frequency $f_s = 10$ Hz.

A random matrix **A** is used to construct the simulated signal and a white noise signal is added to form the simulated signal: $x_4(t)$, $x_5(t)$ and $x_6(t)$. As shown in Fig. 5, $x_4(t)$, $x_5(t)$ and $x_6(t)$ are added to the Gaussian white noise to obtain the simulated signal plot with a signal-noise ratio of 2 dB; the signal-noise ratio equation is as follows:

$$SNR = 10\log_{10}\frac{p_s}{p_n}, \qquad (13)$$

$$\mathbf{A} = \begin{pmatrix} 0.5214 & 0.4555 & 0.2013 \\ 0.4278 & 0.2002 & 0.4416 \\ 0.3122 & 0.3818 & 0.6058 \end{pmatrix}, \qquad (14)$$

$$\mathbf{A}\begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix} + \begin{pmatrix} S_{noise} \\ S_{noise} \\ S_{noise} \end{pmatrix} = \begin{pmatrix} x_4(t) \\ x_5(t) \\ x_6(t) \end{pmatrix}, \qquad (15)$$

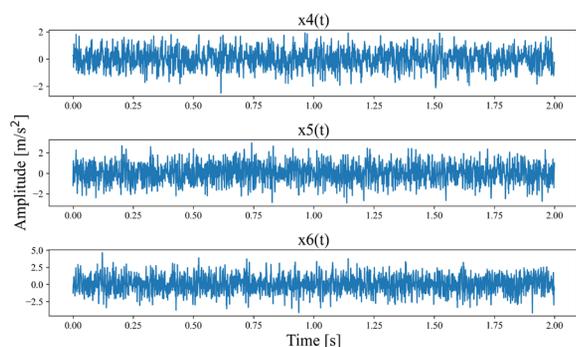where p_s is to input signal energy, p_n noise energy, and S_{noise} Gaussian white noise.

Fig. 5. *Time-domain plots of the simulated signal,
a) $x_4(t)$, b) $x_5(t)$, and c) $x_6(t)$ are added to Gaussian
white noise signals with a signal-noise ratio of -2*

$x_4(t)$, $x_5(t)$ and $x_6(t)$ are added to Gaussian white noise signals with a different signal-noise ratio in the range of [-4, 8]. The numbers of training samples and test samples for each signal-noise ratio of each fault are 30 and 10, respectively, which are given in Table 1.

Table 1. *The information of the simulated signal dataset*

Signals		SNR [dB]						
		-4	-2	0	2	4	6	8
$x_4(t)$	Train	30	30	30	30	30	30	30
	Test	10	10	10	10	10	10	10
$x_5(t)$	Train	30	30	30	30	30	30	30
	Test	10	10	10	10	10	10	10
$x_6(t)$	Train	30	30	30	30	30	30	30
	Test	10	10	10	10	10	10	10

### 1.5.2 Network Structure and Parameters

The detailed structure and parameter settings of the SENet-MSCNN and GRU networks models are shown in Table 2. Firstly, the simulation signal is input to the SENet-MSCNN network. The number of convolutional kernels in the SENet-MSCNN network is 128, and the activation function is Selu, which induces self-normalizing data. Then the output features of SENet-MSCNN are input to global average pooling and the remaining spatial features are discarded to reduce feature dimensionality. Twenty cell numbers are set for the GRU network and the Relu function is used. The last layer is the fully connected layer and the Softmax function is applied to classify the output results into three classifications. The hyperparameters are set: the learning rate is 0.001, the batch size is 32, the number of iterations is 196, and the loss function is the cross-entropy loss function.

1.5.3 Simulation Analysis

Fig. 6 is a comparison of the recognition rates in the test set between GRU, 1D-CNN, MSCNN, BP networks, and proposed methods. It can be seen that recognition rates of the SENet-MSCNN and GRU networks have stabilized at 100 % at the second iteration. The recognition rate of the GRU network has stabilized tending to 99.3 %. The BP network and MSCNN network recognition rate are stable below 99.8 %, and the 1DCNN network fluctuates more. Therefore, the recognition rate, convergence speed, and anti-noise performance of the SENet-MSCNN and GRU networks perform better in the simulation data.

Table 2. *The parameters of the SENet-MSCNN and GRU models used in the simulated signal dataset*

| Layer	Kernel size/step	Kernel num	Unit	Input size	Output size	Activation
Conv_1	3×1/2	128		32×2048×1	32×1024×128	Selu
Conv_a	1×1/2	128		32×1024×128	32×512×128	Selu
Conv_b	1×1/1	128		32×1024×128	32×1024×128	Selu
Conv_b	3×1/2	128		32×1024×128	32×512×128	Selu
AveragePooling c	3×1/2			32×1024×128	32×512×128	Selu
Conv_c	5×1/1	128		32×512×128	32×512×128	Selu
Conv_d	1×1/1	128		32×1024×128	32×1024×128	Selu
Conv_d	3×1/1	128		32×1024×128	32×1024×128	Selu
Conv_d	3×1/2	128		32×1024×128	32×512×128	Selu
Concatenate				(32×512×128)×4	32×512×512	
SENet				32×512×512	32×512×512	
GlobalAverage Pooling				32×1024×512	32×512	
ExpandDim				32×512	32×512×1	
Gru_1			20	32×512×1	32×512×20	Relu
Gru_2			20	32×512×20	32×512×20	Relu
MaxPooling	3×1/2			32×512×20	32×256×20	
Flatten				32×256×20	32×5210	
Dense			3	32×5210	32×3	Softmax
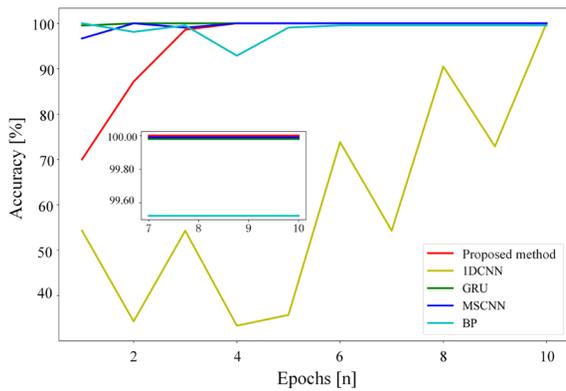
Fig. 6. *The diagnosis rate of different methods in simulation data, several methods including SENet-MSCNN and GRU, 1DCNN, GRU, MSCNN, and BP are compared and analysed*

2 APPLICATION OF SENet-MSCNN AND GRU METHODS IN ROLLING BEARING FAULT DIAGNOSIS

In this section, we first discuss the Case Western Reserve University, Cleveland, USA, bearing datasets [43] and the gearbox datasets, and our implementation details. Subsequently, the proposed method in this paper is applied to the comparative analysis of several typical methods in the two datasets. Meanwhile, we did ablation experiments to examine the effect of each model component. Finally, we design the variable working conditions experiment to analyse the migration performance of the diagnosis.

2.1 Data Preprocessing

This paper increases the data sample set through data augmentation techniques (by sliding windows and cuts) to create three different sizes of comparison test datasets of 500, 1500, and 4500 samples. As shown in Fig. 7, each sample slide window size is 2048 points, and a slide step size is 500 points, which consists of datasets.
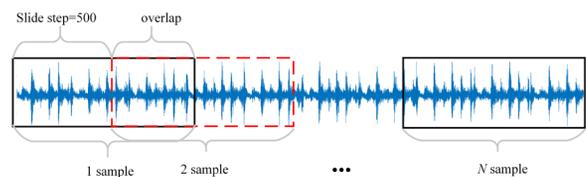
Fig. 7. *Data augmentation with overlap where. the sample length is 2048 and the sliding step is 500*

2.2 Bearing Dataset

As shown in Fig. 8, the experimental platform consists of four parts: motor, torque transducer/encoder, dynamometer, and electronic control. The sampling frequency is 12 kHz, and the data ARE collected from the vibration data of the drive end (DE). There are three fault types: the inner fault, the ball fault, and the outer fault with three fault diameters (0.1778 mm, 0.3556 mm, and 0.5334 mm) and a normal state. Three fault types are shown in Fig. 9. Therefore,
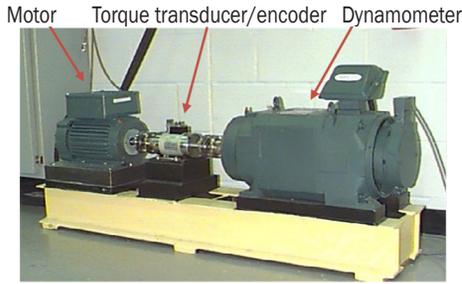
Fig. 8. *Experiment platform for rolling bearing fault used by CWRU which consists of four parts: motor, torque transducer/encoder, dynamometer, and electronic control*

there are 10 states in total and their fault time-domain waveforms are shown in Fig. 10. The rolling bearings worked at three motor loads (746 W, 1492 W, 2238 W with three motor speeds (1772 r/min, 1750 r/min, and 1730 r/min). The 10 states under the three working conditions (1 hp, 2 hp, 3 hp) are respectively denoted as A, B, and C. Dataset D consists of three working conditions. Details of the datasets are shown in Table 3; 80 % of the samples are extracted from the 10 states datasets under the three working conditions to form the training set and 20 % to form the test set.

2.3 Gearbox Dataset

The experiment in this paper used the HFXZ-1 planetary gearbox fault diagnosis experimental platform, as shown in Fig. 11. The experimental platform consists of seven parts: motor, gearbox, flexible coupling, planetary gearbox, helical gearbox, torque sensor, and magnetic powder brake. Three fault states were in the experiment (gear tooth breakage,
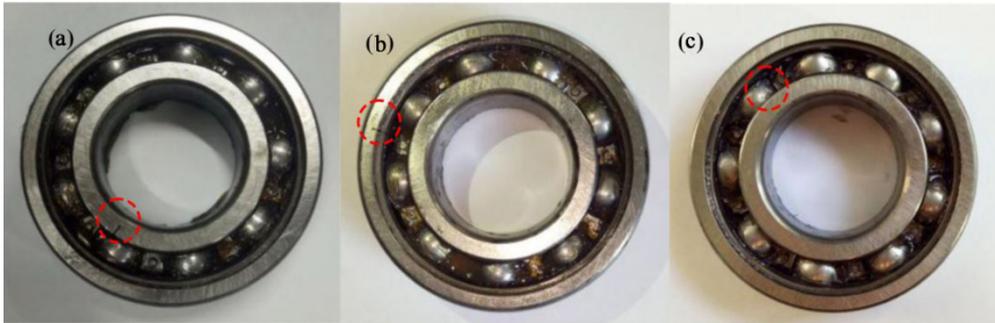

Fig. 9. *Different faults of the rolling bearings: a) inner fault, b) outer fault, and c) ball fault*
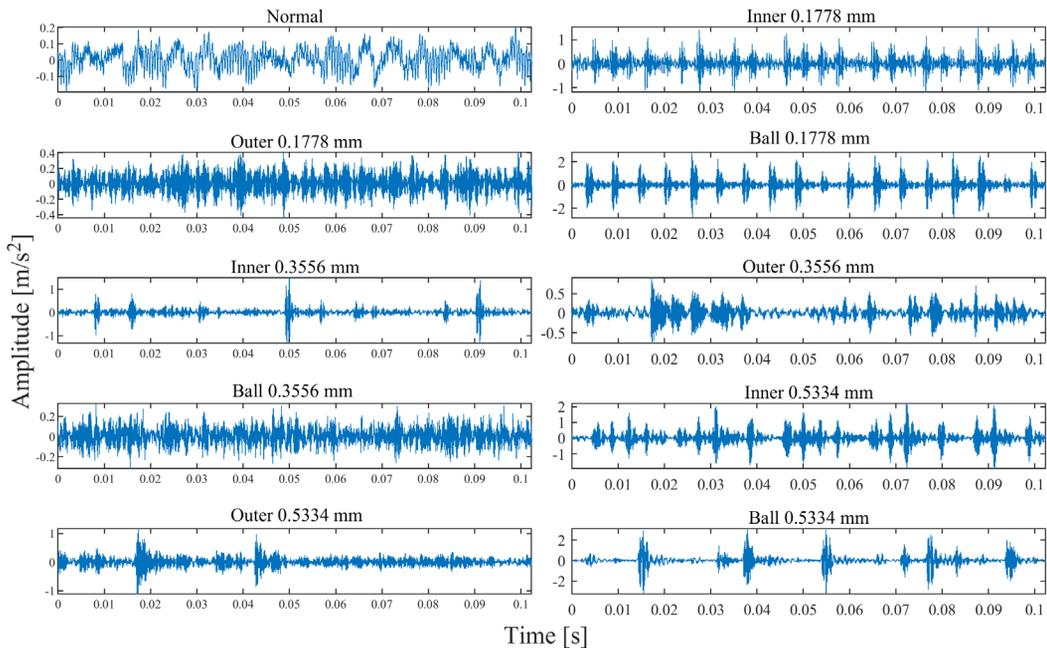
Fig. 10. *Raw vibration signals for 10 states; rolling bearings include 10 states: Normal, Inner 0.1778 mm , Outer 0.1778 mm , Ball 0.1778 mm , Inner 0.3556 mm , Outer 0.3556 mm , Ball 0.3556 mm , Inner 0.5334 mm , Outer 0.5334 mm , and Ball 0.5334 mm*

Table 3. *The information of the rolling bearing datasets*

		Normal	Inner			Ball			Outer		
Label		0	1	2	3	4	5	6	7	8	9
Fault diameter [inches]		0	0.007	0.014	0.021	0.007	0.014	0.021	0.007	0.014	0.021
A (1 hp)	Train	160	160	160	160	160	160	160	160	160	160
	Test	40	40	40	40	40	40	40	40	40	40
B (2 hp)	Train	160	160	160	160	160	160	160	160	160	160
	Test	40	40	40	40	40	40	40	40	40	40
C (3 hp)	Train	160	160	160	160	160	160	160	160	160	160
	Test	40	40	40	40	40	40	40	40	40	40
D (1 hp 2 hp 3 hp)	Train	480	480	480	480	480	480	480	480	480	480
	Test	120	120	120	120	120	120	120	120	120	120
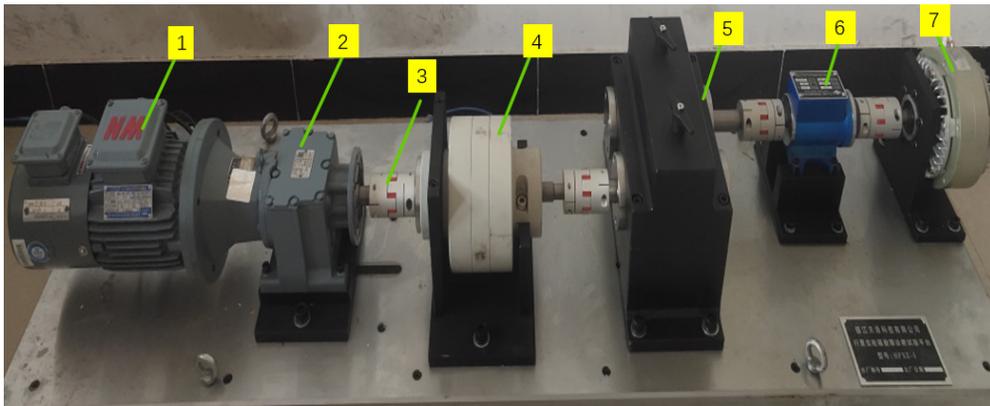
Fig. 11. *Planetary gearbox fault diagnosis experiment platform: 1 motor, 2 gearbox, 3 flexible coupling, 4 planetary gearbox, 5 helical gearbox, 6 torque sensor, and 7 magnetic powder brake*
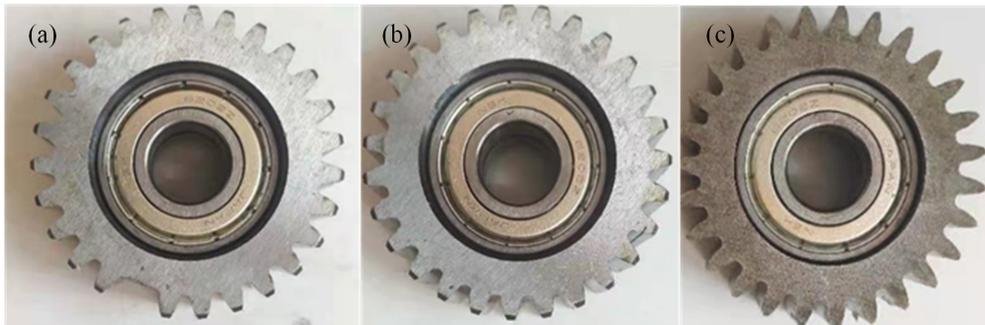
Fig. 12. *Different faults of the gearbox: a gear tooth breakage, b gear wear, and c gear crack*

gear wear, and gear crack), as shown in Fig. 12, and one normal state.

The motor speed was set to 600 rad/min and the sampling frequency was 5120. Three loads were set: 0.1 A, 0.05 A, and 0 A corresponding to datasets A, B, and C. The acceleration sensor was installed outside the planetary gearbox to detect the vibration signal, and the time domain waveform of the original vibration signal is shown in Fig. 13.

2.4 Analysis of Experimental Results

The detailed structure and parameter settings of the SENet-MSCNN and GRU network models in the experimental data are shown in Table 2, which are the same as the detailed parameters under the simulated signals. The number of units in the dense layer being 10 and divided into 10 classifications. The data in A, B, C, and D working loads are used as the model input. To evaluate the superiority of the proposed method, it was compared with 1D-CNN, CNN-GRU, MSCNN,

GRU, BP, SVM, RF, and DT. The results are shown in Table 4. The proposed method outperforms the other models in various working loads. The average accuracy is improved by 1.75 %, 1.42 %, and 3.58 % compared with CNN-GRU, MSCNN, and 1DCNN, respectively. The traditional machine learning methods (SVM, RF, and DT) perform poorly, with average accuracy below 80%, which demonstrates that deep learning methods have better performance in massive vibration data. Hence, the recognition rate of the SENet-MSCNN and GRU networks is remarkably better than the other methods.
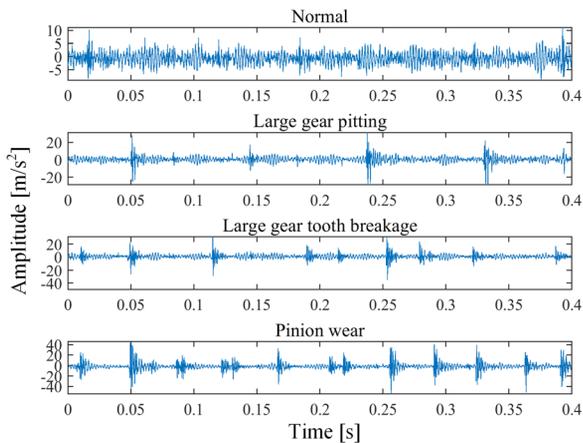

Fig. 13. *Time-domain plots of vibration signals under four state modes*

Table 4. *The diagnosis rate of different methods in experimental data*

Methods	Different working loads				Average [%]
	A	B	C	D	
Proposed method	100.00	100.00	100.00	100.00	100.00
1DCNN	98.00	94.67	93.00	100.00	96.42
CNN-GRU	97.67	99.00	97.00	99.33	98.25
MSCNN	97.33	99.00	99.00	99.00	98.58
GRU	96.33	84.67	95.66	94.00	92.64
BP	71.33	58.00	66.00	61.00	64.08
SVM	64.67	75.00	81.67	80.78	75.53
RF	47.67	60.33	49.33	47.33	51.16
DT	36.33	40.33	40.33	40.22	39.30

Considering the different performance of network models in different sizes of datasets, comparison experiments of three datasets (500 samples, 1500 samples, and 4500 samples) are established. The datasets are divided into training sets and test sets in the ratio of 4:1. The batch sizes of the three datasets are set to 16, 32, and 64, respectively. Compared with other methods,

the SENet-MSCNN and GRU model have the highest accuracy in three datasets with 100 %, 99.67 %, and 100 % respectively, which can be seen in Fig. 14. It also shows that the method has good performance in small datasets. Therefore, the accuracy and robustness of the proposed method in this paper are significantly better than others. Fig. 15 shows the time plots. The time consumed by the GRU network is the longest and the MSCNN is the shortest for each sample; the time consumed by the SENet-MSCNN and GRU networks is second. The MSCNN reduces the complexity of GRU model parameter computation and reduces the model training time. Furthermore, with the batch size increasing, the time consumed by every sample is further reduced. The consuming time of the proposed method is reduced to 19 ms/sample at 4500 samples, which further speeds up the model iteration. The SENet-MSCNN and GRU fault diagnosis models are significantly improved in terms of diagnostic accuracy, robustness, and diagnostic speed.
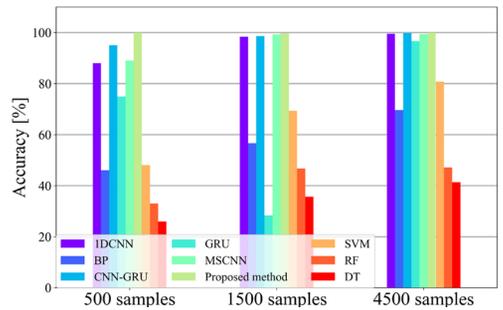
	500 samples	1500 samples	4500 samples
1DCNN	88.00 %	98.33 %	99.56 %
BP	46.00 %	56.67 %	69.56 %
CNN-GRU	95.00 %	98.67 %	99.78 %
GRU	75.00 %	28.33 %	96.56 %
MSCNN	89.00 %	99.33 %	99.40 %
Proposed method	100.00 %	99.67 %	100 %
SVM	48.00 %	69.33 %	80.78 %
RF	33.00 %	46.67 %	47.11 %
DT	26.00 %	35.67 %	41.33 %

Fig. 14. *Diagnosis accuracy of different methods in three fault sample sets*

2.5 Analysis of Ablation Experiments

In this section, we will do ablation experiments to compare the performance of the proposed method with serval baseline methods (MSCNN, GRU, MSCNN-GRU).
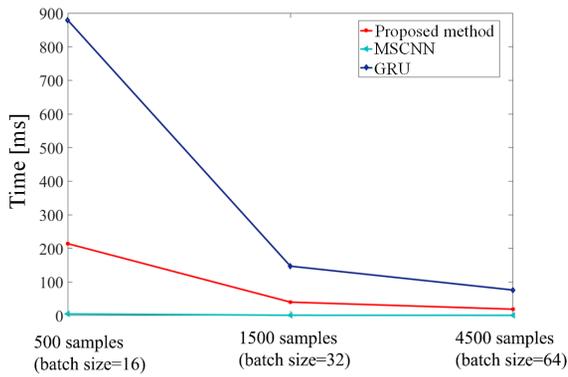
Fig. 15. *Iteration speed of different methods in three datasets*

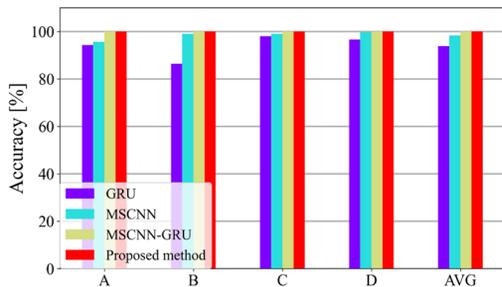|	A	B	C	D	AVG
GRU	94.33 %	86.33 %	98.00 %	96.67 %	93.83 %
MSCNN	95.67 %	99.00 %	99.00 %	99.67 %	98.33 %
MSCNN-GRU	99.97 %	100.00 %	100.00 %	100.00 %	99.99 %
Proposed method	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %

Fig. 16. *Ablation study results*

As Fig. 16 shows, the average accuracy of MSCNN and GRU is 93.83 % and 98.33 %, respectively. After the fusion of these two methods, MSCNN-GRU achieves 100 % accuracy in B, C, and D working conditions, with an average accuracy of 99.99 %. The proposed method adds the SENet to MSCNN-CNN, and the accuracy is further improved to reach 100 % in all four conditions. Thus, it is also

proved that the fusion of GRU, MSCNN, and SENet has a significant improvement in the recognition rate.

2.6 Variable Working Conditions Experiment

To further signify the migration characteristics of SENet-MSCNN and GRU models for different working conditions, the experiment datasets of A, B, C, and D working conditions are used as the input of the models. One of A, B, C, D working conditions is used as the source domain and another working condition dataset is used as the target domain, which constitutes 12 domains. Table 5 presents the accuracy of different methods in variable working conditions.

The average accuracy of BP, DT, RF, SVM models with shallow network structures is below 70 %, and the migration characteristics of these models are poor. The average accuracy of single deep network CNN, GRU, MSCNN is 91.12 %, 84.75 %, and 90.12 %, respectively. The network performance is improved. The average accuracy of the fused models, CNN-GRU, MSCNN-GRU, SENet-MSCNN and GRU are above 94.12 %. The network performance is further improved. The proposed method achieves the highest average accuracy of 98.98 %, which is 4.83 % and 2.81 % higher than CNN-GRU and MSCNN-GRU. Only in A-C, the MSCNN-GRU model performs better, with an improvement of 0.6 %. In the rest of the variable conditions, the proposed method has a significant improvement, especially in C-A, with an improvement of 11.2 %. The accuracy of all methods is reduced when C is used as the source domain or C is used as the target domain., which is explained by the fact that C enhances the cyclic shock response of rolling bearings, making the data more regular and the features learned by the network model simpler. This makes accuracy significantly reduce when testing low load data (especially 1hp working conditions), which

Table 5. *Experimental diagnosis results of various methods*

Methods	Variable working conditions test												Average [%]
	A-B	A-C	A-D	B-A	B-C	B-D	C-A	C-B	C-D	D-A	D-B	D-C	
1DCNN	99.60	91.40	96.60	90.00	74.80	89.00	79.40	86.40	87.00	99.20	100.00	100.00	91.12
BP	54.20	54.40	64.80	50.60	52.20	58.00	49.60	55.00	59.80	78.80	73.20	79.60	60.85
CNN-GRU	96.40	94.80	95.80	90.40	85.00	90.20	90.40	94.80	92.60	99.40	100.00	100.00	94.15
GRU	73.60	81.00	81.60	87.60	88.80	88.60	72.20	73.60	77.40	97.60	97.00	98.00	84.75
MSCNN	80.80	86.00	86.00	90.00	93.80	95.60	77.80	86.00	86.20	99.80	99.40	100.00	90.12
MSCNN-GRU	99.60	96.20	99.00	98.20	94.80	98.40	85.40	91.40	91.00	100.00	100.00	100.00	96.17
Proposed method	99.00	98.80	99.80	98.20	99.00	99.20	96.60	98.80	98.40	100.00	100.00	100.00	98.98
DT	33.60	35.80	49.60	32.00	34.20	45.20	35.20	36.80	44.80	63.80	64.20	62.80	44.83
RF	42.60	48.00	46.80	42.80	46.80	46.80	41.60	40.80	46.40	51.00	52.60	54.80	46.75
SVM	54.80	66.40	66.00	64.00	65.00	67.60	63.80	56.40	67.80	87.20	79.40	89.00	68.95

influences the migration characteristics of the model. Thus, the proposed method maintains a high accuracy under conditions with large differences, which proves that the feature extracted by the proposed method have stronger transfer characteristics.

Although the SENet-MSCNN and GRU model exhibit strong migration properties in the Western Reserve University bearing data, the bearing dataset is very transferable. Therefore, to further validate the migration properties of the method, the gearbox dataset was used to test. Table 6 presents the accuracy of different methods in variable working conditions. The SENet-MSCNN and GRU fusion model still exhibits the best performance on the gearbox dataset with an average accuracy of 76.44 %, an improvement of 4.61 %, 11.36 %, 9.77 %, and 2.69 % compared to the MSCNN, GRU, MSCNN-GRU, and CNN-GRU respectively. The average accuracy of traditional machine learning methods is below 56 %, with poor migration characteristics.

2.7 Visualization Results

(1) **Visualization of mid-layer activation**. A sample of the fourth fault type of rolling bearing (fault position is outer, fault diameter is 0.1778 mm) is used as input to the model; the feature map of each hidden layer output is visualized in a 2D image, which is shown in Fig. 17. The yellow parts represent the activated part and the blue parts represent the inactivated part. The features extracted by the first convolutional layer Conv 1, which is the yellow activated part, correspond to the shock component of the vibration signal. The global features extracted at different scales by the MSCNN layer are the same, while the local features are different. The convolutional kernels of 3×1 and 5×1 sizes are used in MSCNN_b and MSCNN_c. The branches are more sensitive to the shock signal and the extracted feature information has a higher resolution. As the number of network layers deepens, irrelevant information is filtered out and useful information is refined and

Table 6. *Experimental diagnosis results of various methods*

Methods	Variable working conditions test						Average [%]
	A-B	A-C	B-A	B-C	C-A	C-B	
1DCNN	69.00	53.00	61.50	100.00	48.00	89.50	70.17
BP	27.50	27.50	50.00	55.00	44.00	54.50	43.08
CNN-GRU	73.50	56.50	53.00	99.00	61.00	99.50	73.75
GRU	35.00	34.00	57.00	99.50	65.00	100.00	65.08
MSCNN	59.50	65.00	50.50	99.00	57.00	100.00	71.83
MSCNN-GRU	41.50	40.50	50.50	99.50	68.00	100.00	66.67
Proposed method	74.67	58.00	54.00	100.00	72.67	99.33	76.44
DT	28.50	26.00	33.00	42.00	34.50	28.00	32.00
RF	27.50	27.50	50.00	43.50	48.00	56.00	42.08
SVM	34.00	36.50	50.50	83.50	50.50	81.00	56.00
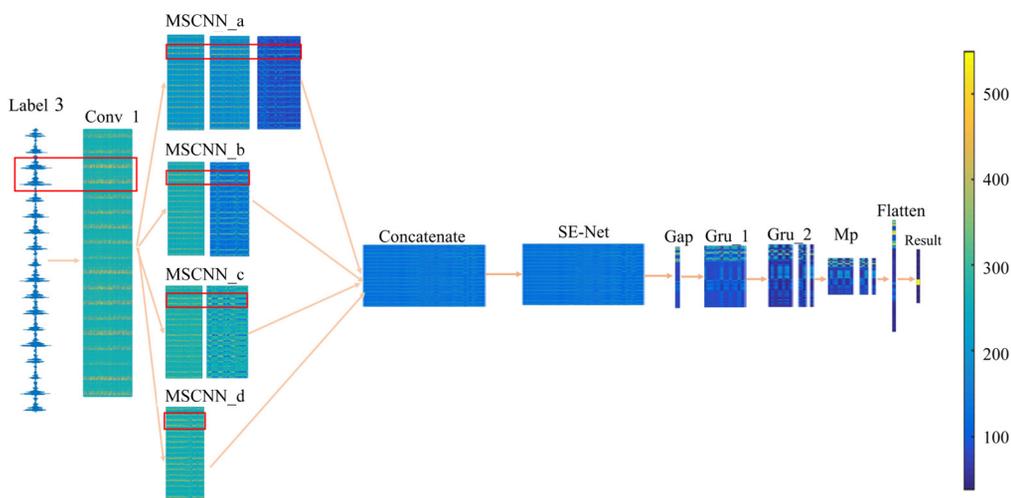
Fig. 17. *Visualization of the hidden layer activations of SENet-MSCNN and GRU, label 3 represents the fourth fault type of rolling bearing*

scaled up. The feature map becomes clearer, the extracted features are more abstract, and the source domain information is less relevant. The target domain information is gradually more relevant.
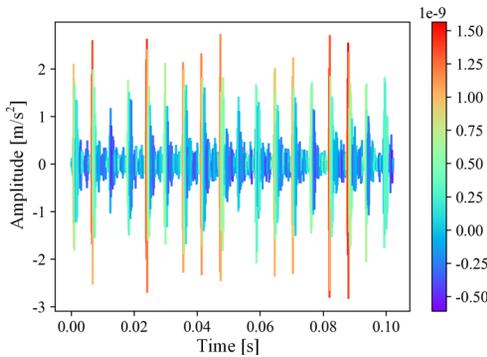
Fig. 18. *Visualization of time-domain waveforms for class activation; the red parts represent the activated part and the blue parts represent the inactivated part*

(2) **Visualization of the time-domain plot for class activation**. The time-domain waveform plot of the activation intensity of the vibration signal for the fourth fault states(fault location is the outer and the fault diameter is 0.1778 mm) is obtained by the method of class activation visualization [44]. As shown in Fig. 18, the medium and low-frequency shock signals in the red parts have a strong influence on the classification results of the fault diagnosis model, and the high-frequency signals in the blue parts have less influence on the classification results.

The part of the SENet-MSCNN and GRU model that is more sensitive to the vibration signal of the fourth fault state is similar to the characteristic frequency of the vibration signal of the fourth fault state, which further explains the fault diagnosis model to diagnose the input signal as the fourth fault state.

(3) **T-SNE Visualization**. T-SNE (T-distributed Stochastic Neighbor Embedding) is a common method used for data dimensionality reduction and visualization. In this paper, high dimension data is represented by low dimension distribution using the T-SNE method. Fig. 19 presents 1000 validation sets classified by the SENet-MSCNN and GRU models and the T-SNE visualization of their intermediate processes. The T-SNE visualization picture of raw signal through the Input Layer is confusing in the two-dimension space. The T-SNE visualization picture of raw signal through the SENet-MSCNN Layer has initial classification characteristics. GRU Layer already has obvious classification features, and the 10 states are remarkably separated from each other. The Dense Layer has even more obvious classification features, with the same state clustered at the same location and the distance between different states is larger.

3 CONCLUSIONS AND FUTURE WORK

In this paper, a rolling bearing fault diagnosis method based on SENet-MSCNN and GRU model is proposed. The method was applied to the comparative
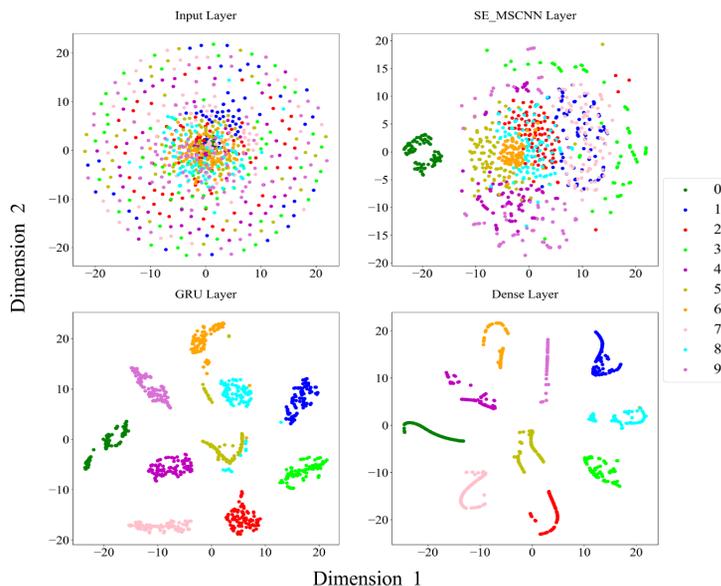
Fig. 19. *Visualization feature distribution map using T-SNE; feature visualization of the Input Layer, SENet-MSCNN Layer, GRU Layer, and Dense Layer structures of SENet-MSCNN and GRU*

analysis of the bearing data set and achieved a recognition rate of over 99.67 %. Compared with other representative fault diagnosis methods, the proposed method has significant advantages in terms of fault identification rate and robustness performance. In addition, we tested the migration capability of the model under variable working conditions in both the bearing dataset and the gearbox dataset. The method achieved recognition rates of 98.98 % and 76.44 % in the cross-service tests, respectively. The results show that the method exhibits better migration properties.

Therefore, the method is expected to provide a new method for rolling bearing fault diagnosis. In future research, we will apply the proposed fault diagnosis method to other mechanical fault types to determine its effectiveness in diagnosing a wider range of mechanical faults.

4 ACKNOWLEDGEMENTS

5 REFERENCES

[1] Zhao, C., Sun, H. (2019). Dynamic distributed monitoring strategy for large-scale nonstationary processes subject to frequently varying conditions under closed-loop control. *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, p. 4749-4758, DOI:10.1109/tie.2018.2864703.

[2] AlShorman, O., Alkahatni, F., Masadeh, M., Irfan, M., Glowacz, A., Althobiani, F., Kozik, J., Glowacz, W. (2021). Sounds and acoustic emission-based early fault diagnosis of induction motor: A review study. *Advances in Mechanical Engineering*, vol. 13, no. 2, DOI:10.1177/1687814021996915.

[3] Gong, T., Yang, J., Liu, S., Liu, H. (2022). Non-stationary feature extraction by the stochastic response of coupled oscillators and its application in bearing fault diagnosis under variable speed condition. *Nonlinear Dynamics*, vol. 108, no. 4, p. 3839-3857, DOI:10.1007/s11071-022-07373-y.

[4] Yun, K., Chong, Y., Enzhe, S., Liping, Y., Quan, D. (2021). Fault diagnosis method of diesel engine injector based on hierarchical weighted permutation entropy. *IEEE International Instrumentation and Measurement Technology Conference*, p. 1-6, DOI:10.1109/I2MTC50364.2021.9460083.

[5] Li, Y., Dai, W., Zhang, W. (2020). Bearing fault feature selection method based on weighted multidimensional feature fusion. *IEEE Access*, vol. 8, p. 19008-19025, DOI:10.1109/access.2020.2967537.

[6] Wrzochal, M., Adamczak, S., Piotrowicz, G., Wnuk, S. (2022). Industrial experimental research as a contribution to the development of an experimental model of rolling bearing vibrations. *Strojniški vestnik - Journal of Mechanical Engineering*, vol. 68, no. 9, p. 552-559, DOI:10.5545/sv-jme.2022.184.

[7] Yi, C., Wang, H., Ran, L., Zhou, L., Lin, J. (2022). Power spectral density-guided variational mode decomposition for the compound fault diagnosis of rolling bearings. *Measurement*, vol. 199, DOI:10.1016/j.measurement.2022.111494.

[8] Glowacz, A., Tadeusiewicz, R., Legutko, S., Caesarendra, W., Irfan, M., Liu, H., Brumercik, F., Gutten, M., Sulowicz, M., Antonino Daviu, J.A., Sarkodie-Gyan, T., Fracz, P., Kumar, A., Xiang, J. (2021). Fault diagnosis of angle grinders and electric impact drills using acoustic signals. *Applied Acoustics*, vol. 179, DOI:10.1016/j.apacoust.2021.108070.

[9] Ribeiro Junior, R.F., dos Santos Areias, I.A., Campos, M.M., Teixeira, C.E., da Silva, L.E.B., Gomes, G.F. (2022). Fault detection and diagnosis in electric motors using convolution neural network and short-time fourier transform. *Journal of Vibration Engineering & Technologies*, vol. 10, no. 7, p. 2531-2542, DOI:10.1007/s42417-022-00501-3.

[10] Li, Y., Cheng, G., Liu, C. (2021). Research on bearing fault diagnosis based on spectrum characteristics under strong noise interference. *Measurement*, vol. 169, DOI:10.1016/j.measurement.2020.108509.

[11] Peng, H., Zhang, H., Fan, Y., Shangguan, L., Yang, Y. (2022). A review of research on wind turbine bearings' failure analysis and fault diagnosis. *Lubricants*, vol. 11, no. 1, DOI:10.3390/lubricants11010014.

[12] Peng, Z., Zhike, P., Shiqian, C. (2020). Review of signal decomposition theory and its applications in machine fault diagnosis. *Journal of Mechanical Engineering*, vol. 56, no. 17, DOI:10.3901/jme.2020.17.091.

[13] Kumar, A., Gandhi, C.P., Vashishtha, G., Kundu, P., Tang, H., Glowacz, A., Shukla, R.K., Xiang, J. (2021). Vmd based trigonometric entropy measure: A simple and effective tool for dynamic degradation monitoring of rolling element bearing. *Measurement Science and Technology*, vol. 33, no. 1, DOI:10.1088/1361-6501/ac2fe8.

[14] Zhou, J., Xiao, M., Niu, Y., Ji, G. (2022). Rolling bearing fault diagnosis based on WGWOA-VMD-SVM. *Sensors*, vol. 22, no. 16, DOI:10.3390/s22166281.

[15] Jiang, J., Liu, Y., Xu, C., Shen, H., Soni, M. (2022). Research on motor bearing fault diagnosis based on the adaboost algorithm and the ensemble learning with bayesian optimization in the industrial internet of things. *Security and Communication Networks*, vol. 2022, p. 1-12, DOI:10.1155/2022/4569954.

[16] Hosseinpour-Zarnaq, M., Omid, M., Biabani-Aghdam, E. (2022). Fault diagnosis of tractor auxiliary gearbox using vibration analysis and random forest classifier. *Information Processing in Agriculture*, vol. 9, no. 1, p. 60-67, DOI:10.1016/j.inpa.2021.01.002.

[17] Gunerkar, R.S., Jalan, A.K., Belgamwar, S.U. (2019). Fault diagnosis of rolling element bearing based on artificial neural network. *Journal of Mechanical Science and Technology*, vol. 33, no. 2, p. 505-511, DOI:10.1007/s12206-019-0103-x.

[18] Zhang, Z., Li, H., Chen, L., Han, P., Shi, H. (2021). Rolling bearing fault diagnosis using improved deep residual shrinkage networks. *Shock and Vibration*, vol. 2021, p. 1-11, DOI:10.1155/2021/9942249.

[19] Wang, H., Liu, Z., Peng, D., Cheng, Z. (2022). Attention-guided joint learning cnn with noise robustness for bearing fault

diagnosis and vibration signal denoising. *ISA Transactions*, vol. 128, p. 470-484, **DOI:10.1016/j.isatra.2021.11.028**.

[20] Mao, W., Feng, W., Liu, Y., Zhang, D., Liang, X. (2021). A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. *Mechanical Systems and Signal Processing*, vol. 150, **DOI:10.1016/j.ymssp.2020.107233**.

[21] Lv, D., Wang, H., Che, C. (2021). Fault diagnosis of rolling bearing based on multimodal data fusion and deep belief network. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 235, no. 22, p. 6577-6585, **DOI:10.1177/09544062211008464**.

[22] Liu, H., Zhou, J., Zheng, Y., Jiang, W., Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Trans*, vol. 77, p. 167-178, **DOI:10.1016/j.isatra.2018.04.005**.

[23] Li, D., Liu, S., Gao, F., Sun, X. (2021). Continual learning classification method for time-varying data space based on artificial immune system. *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 5, p. 8741-8754, **DOI:10.3233/jifs-200044**.

[24] Bin, G.F., Gao, J.J., Li, X.J., Dhillon, B.S. (2012). Early fault diagnosis of rotating machinery based on wavelet packets-empirical mode decomposition feature extraction and neural network. *Mechanical Systems and Signal Processing*, vol. 27, p. 696-711, **DOI:10.1016/j.ymssp.2011.08.002**.

[25] Lin, Y., Xiao, M., Liu, H., Li, Z., Zhou, S., Xu, X., Wang, D. (2022). Gear fault diagnosis based on cs-improved variational mode decomposition and probabilistic neural network. *Measurement*, vol. 192, **DOI:10.1016/j.measurement.2022.110913**.

[26] Guo, S., Yang, T., Gao, W., Zhang, C. (2018). A novel fault diagnosis method for rotating machinery based on a convolutional neural network. *Sensors*, vol. 18, no. 5, **DOI:10.3390/s18051429**.

[27] Guo, M.-F., Yang, N.-C., Chen, W.-F. (2019). Deep-learning-based fault classification using hilbert-huang transform and convolutional neural network in power distribution systems. *IEEE Sensors Journal*, vol. 19, no. 16, p. 6905-6913, **DOI:10.1109/jsen.2019.2913006**.

[28] Sun, W., Zhao, R., Yan, R., Shao, S., Chen, X. (2017). Convolutional discriminative feature learning for induction motor fault diagnosis. *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, p. 1350-1359, **DOI:10.1109/tii.2017.2672988**.

[29] Chen, Z., Gryllias, K., Li, W. (2019). Mechanical fault diagnosis using convolutional neural networks and extreme learning machine. *Mechanical Systems and Signal Processing*, vol. 133, **DOI:10.1016/j.ymssp.2019.106272**.

[30] Wang, S., Xiang, J., Zhong, Y., Zhou, Y. (2018). Convolutional neural network-based hidden markov models for rolling element bearing fault identification. *Knowledge-Based Systems*, vol. 144, p. 65-76, **DOI:10.1016/j.knosys.2017.12.027**.

[31] Hao, S., Ge, F.-X., Li, Y., Jiang, J. (2020). Multisensor bearing fault diagnosis based on one-dimensional convolutional long short-term memory networks. *Measurement*, vol. 159, **DOI:10.1016/j.measurement.2020.107802**.

[32] Zhang, P., Chen, C. (2022). Wind turbine planetary gearbox fault diagnosis using circular pitch cyclic vector and a bidirectional gated recurrent unit. *Measurement Science and Technology*, vol. 34, no. 1, **DOI:10.1088/1361-6501/ac95b2**.

[33] Zhu, Y., Zhu, C., Tan, J., Tan, Y., Rao, L. (2022). Anomaly detection and condition monitoring of wind turbine gearbox based on LSTM-FS and transfer learning. *Renewable Energy*, vol. 189, p. 90-103, **DOI:10.1016/j.renene.2022.02.061**.

[34] Shi, J., Peng, D., Peng, Z., Zhang, Z., Goebel, K., Wu, D. (2022). Planetary gearbox fault diagnosis using bidirectional-convolutional lstm networks. *Mechanical Systems and Signal Processing*, vol. 162, **DOI:10.1016/j.ymssp.2021.107996**.

[35] Chen, X., Zhang, B., Gao, D. (2020). Bearing fault diagnosis base on multi-scale cnn and lstm model. *Journal of Intelligent Manufacturing*, vol. 32, no. 4, p. 971-987, **DOI:10.1007/s10845-020-01600-2**.

[36] Li, X., Li, J., Qu, Y., He, D. (2019). Gear pitting fault diagnosis using integrated cnn and gru network with both vibration and acoustic emission signals. *Applied Sciences*, vol. 9, no. 4, **DOI:10.3390/app9040768**.

[37] Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E. (2020). Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell*, vol. 42, no. 8, p. 2011-2023, **DOI:10.1109/TPAMI.2019.2913372**.

[38] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June, p. 1-9, **DOI:10.1109/CVPR.2015.7298594**.

[39] Peng, D., Wang, H., Liu, Z., Zhang, W., Zuo, M.J., Chen, J. (2020). Multibranch and multiscale cnn for fault diagnosis of wheelset bearings under strong noise and variable load condition. *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, p. 4949-4960, **DOI:10.1109/tii.2020.2967557**.

[40] Gao, T., Yang, J., Jiang, S., Yan, G. (2020). A novel fault diagnosis method for analog circuits based on conditional variational neural networks. *Circuits, Systems, and Signal Processing*, vol. 40, no. 6, p. 2609-2633, **DOI:10.1007/s00034-020-01595-4**.

[41] Lv, H., Chen, J., Pan, T., Zhang, T., Feng, Y., Liu, S. (2022). Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application. *Measurement*, vol. 199, **DOI:10.1016/j.measurement.2022.111594**.

[42] Lv, Y., Yuan, R., Song, G. (2016). Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing. *Mechanical Systems and Signal Processing*, vol. 81, p. 219-234, **DOI:10.1016/j.ymssp.2016.03.010**.

[43] He, X., Zhou, X., Yu, W., Hou, Y., Mechefske, C.K. (2021). Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *ISA Trans*, vol. 111, p. 360-375, **DOI:10.1016/j.isatra.2020.10.060**.

[44] Chao, Q., Wei, X., Tao, J., Liu, C., Wang, Y. (2022). Cavitation recognition of axial piston pumps in noisy environment based on grad-cam visualization technique. *CAAI Transactions on Intelligence Technology*, **DOI:10.1049/cit2.12101**.