



# Multiscale Redundant Second Generation Wavelet Kernel-Driven Convolutional Neural Network for Rolling Bearing Fault Diagnosis

Fengxian Su<sup>(✉)</sup>, Shuwei Cao, Tianheng Hai, and Jing Yuan

School of Mechanical Engineering, University of Shanghai for Science and Technology,  
516 JunGong Road, Shanghai 200093, China  
sfx202103@163.com

**Abstract.** Fault diagnosis of rolling bearings is crucial in areas related to rotating machinery and equipment applications. If faults are detected in time at an early stage, it can guarantee the safe and effective operation of the equipment, saving valuable time and high maintenance costs. Traditional fault diagnosis techniques have achieved remarkable results in rolling bearing fault detection, but they rely heavily on expert knowledge to extract fault features. Manual extraction of features in the face of massive industrial data exhibits poor timeliness. In recent years, with the development and wide application of deep learning, data-driven mechanical fault diagnosis methods are becoming a hot topic of discussion among related researchers. Among them, Convolutional Neural Network (CNN) is an effective deep learning method. In this study, a new method of multiscale redundant second generation wavelet kernel-driven convolutional neural network for rolling bearing fault diagnosis is proposed, called RW-Net. By performing two layers of redundant second generation wavelet decomposition on the input time-domain signal in the shallow layer of the network, the network can automatically extract fault features with rich information. The proposed method is validated by Case Western Reserve University (CWRU) bearing test data, and the average fault identification accuracy is 99.4%, which verifies the feasibility and effectiveness of the proposed method.

**Keywords:** Deep learning · CNN · Redundant second generation wavelet · Fault diagnosis · Rolling bearing

## 1 Introduction

With the rapid development of modern industry, mechanical equipment has become the cornerstone of promoting productivity and facilitating economic growth [1, 2]. Among them, rolling bearings are widely used in various mechanical equipment, as key components of rotating mechanical equipment, in the mechanical transmission process has the role of load-bearing weight and reducing friction. Rolling bearings are prone to pitting, spalling, cracks, and other local failures under harsh working conditions such as high load, strong impact, and high temperature. The local failure of rolling bearings is one of the main causes of rotating machinery faults. If not found in time, it will have an impact

on the safe operation of mechanical equipment, and even cause serious economic losses and personal casualties [3]. Therefore, the research of accurate and effective rolling bearing fault diagnosis and health monitoring methods is of great significance to reducing downtime, preventing safety accidents, and guaranteeing the safe and effective operation of equipment.

Generally, mechanical fault diagnosis techniques are divided into three parts: signal acquisition, feature extraction, and fault identification. Among them, the latter two of these steps are very important and greatly influence the accuracy of the final diagnosis [4]. Usually, the vibration signals of rolling bearings are complex and non-smooth with high background noise. It is a great challenge to effectively extract the representative fault features in the vibration signal for fault diagnosis [5]. Several researchers have achieved notable results in the field of rotating machinery fault diagnosis using signal processing-related techniques. Li et al. use the variational modal decomposition (EMD) method for feature extraction of fault signals. The authors solved the problem of information loss and over-decomposition and verified the effectiveness of the proposed method using high-speed locomotive wheelset bearings [6]. Chen et al. reveal the essence of wavelet transform inner product matching in rotating machinery fault diagnosis through simulation and field test experiments [7]. Ming et al. propose the spectral auto-correlation analysis method and applied it to the feature extraction of early faint faults of rolling bearings [8]. However, these traditional methods rely heavily on a priori knowledge and expert knowledge. This limits the wide usage of traditional fault diagnosis methods. Compared to traditional signal processing techniques, intelligent fault diagnosis is a new development in mechanical fault detection technology [9].

The development of intelligent machinery fault diagnosis has benefited from the rapid development of sensing technology, computer technology, and data storage technology in recent years. These technologies provide the technology for data acquisition, transmission, and storage in manufacturing systems [10, 11]. For the intelligent diagnosis model based on a neural network, its network structure contains large amount of learnable parameters. In order to fully training the network parameters, it requires a lot of engineering data. Most of the actual engineering data are generated during the fault-free period. Only a small fraction of the fault data is generated when a machine breaks down, and the data used for neural network training is artificially processed, and it contains a dataset corresponding to the real labels. Manual processing of data is time-consuming and laborious. Based on the above, how to obtain higher accuracy with relatively less training data is a hot topic worthy of study.

When a rolling bearing malfunctions during operation, it can cause the dynamic signal to contain non-stationary components. Wavelet transforms with good time-frequency multi-resolution characteristics is a powerful tool for dynamic signal analysis. However, in engineering practice, how to select the appropriate wavelet basis function from the library of wavelet basis functions to match the signal to be analyzed becomes a difficult problem. Hence, the second generation wavelet transform was born, which constructs predictors and updaters to match the signal to be analyzed adaptively by lifting methods. Nevertheless, the second generation wavelet transform has a splitting operation, which makes the number of points of the approximation signal decrease to half for each second

generation wavelet decomposition performed on the signal. As the number of decompositions increases, the approximation signal contains less and less information, the signal is also prone to distortion. The redundant second generation wavelet (i.e. RSGW) transform overcomes this shortcoming. There is no sampling operation in the signal decomposition process. The length of the approximation and detail signals are the same as the original signal, and the information in the decomposed signal is redundant [12, 13]. Gao et al. use RSGW for signal noise reduction to improve the SNR [14]. Jiang et al. interpolate the initial prediction operator and updating operator of the RSGW to obtain the redundant prediction operator and updating operator corresponding to the number of decomposition layers, and the experiment shows that the method can accurately extract the signal features [15].

Inspired by this, this paper combines RSGW with convolution layers to form a deep CNN (i.e., RW-Net) driven by multiscale RSGW convolution kernels. The first layer of the network is the Conv1, in which multiscale RSGW transform is performed. The more layers of RSGW decomposition, the clearer the feature extraction and the less noise. The kernel of this convolution layer is the RSGW kernel, which is obtained by interpolating the second generation wavelets to complement the zeros. Therefore, the deep CNN driven by multiscale RSGW convolution kernels has the following advantages: 1) Multiple RSGW transform can be performed within the Conv1, and the decomposition result is always the same length as the original signal, which is suitable for engineering data with small sample lengths. 2) The RSGW transform realizes the translational invariance of the signal and can extract and retain richer dynamic fault features, which can effectively enhance the overall fault diagnosis performance of RW-Net. The contributions of this paper are as follows:

- 1) A new rolling bearing fault diagnosis method is proposed.
- 2) RW-Net uses the collected time series as input to achieve end-to-end fault identification without expert knowledge, reducing the complexity and timeliness of fault diagnosis.
- 3) The RW-Net model is developed and applied to the engineering case to verify the validity of the model in comparison with classical and popular networks. The RSGW layers proposed in this model have universal applicability and can be applied to almost any network. In addition, compared with the traditional convolutional neural network, RW-Net has only two training parameters, which saves computing storage space and improves the convergence speed of the network training.

The rest of the paper is organized as follows. Section 2 introduces the theoretical foundation knowledge of RSGW transform and convolutional neural networks. Section 3 explains the design of wavelet kernels and the construction of RSGW layers. Section 4 introduces the effectiveness and result analysis of RW-Net in the experiment. Section 5 is the conclusion.

## 2 Theoretical Foundation

### 2.1 RSGW Transform

In the RSGW transform algorithm, there are only two steps, prediction and updating, and the splitting operation is removed compared to the second generation wavelet transform. The RSGW performs a predict-and-update operation on the input signal  $\hat{s}^{(k)}$ . The RSGW transform is schematically shown in Fig. 1.

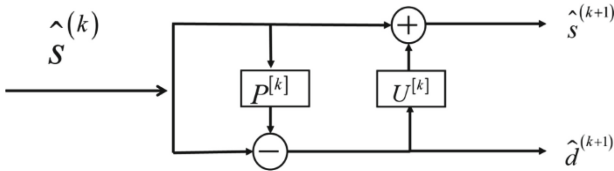


Fig. 1. Schematic diagram of RSGW transform

**Prediction:** The redundant predictor after interpolation zero padding is used to predict the signal, and the prediction error is defined as the detail signal of the RSGW transforms:

$$\hat{d}_i^{(k+1)} = \hat{s}_i^{(k)} - \sum_i p_r^{[k]} \hat{s}_{r+i-2^{k-1}N}^{(k)}. \tag{1}$$

**Updating:** Based on the detail signal, the redundant updater  $U^{[k]}$  after interpolation zero padding is used to update the detail signal, and the updated signal  $\hat{s}_i^{(k+1)}$  is defined as the approximation signal of the RSGW:

$$\hat{s}_i^{(k+1)} = \hat{s}_i^{(k)} + \sum_i u_l^{[k]} \hat{d}_{l+i-2^{k-1}\tilde{N}}^{(k)}. \tag{2}$$

### 2.2 Fundamental Theory of CNN

Convolutional neural networks contain two parts: feature extraction and feature selection, where the convolution, activation, and pooling layers perform feature extraction on the input signal and the fully connected layer filter the extracted features. The back-propagation algorithm calculates the gradient values of the variable parameters of the network and the adaptive optimization function to update the mode parameters so that the output of the network corresponds to the true labels of the input signals. As the input signal passes through the convolution layer, the filters (also called convolution kernels) in the convolution layer progressively scan the input matrix in specific steps to obtain a matrix of smaller size and containing fault features. The value of the convolution kernels determines the feature type extracted by the convolution layers. During the convolution operation, the convolution kernels are always the same batch, which is what makes the convolution layer different from the normal network layers. Therefore, one of the most

important features of the convolution layer is weight sharing, which reduces the trainable parameters. In addition, the output channels of the redundant second generation wavelet convolution layer mentioned in this paper are equal to the number of redundant second generation wavelet convolution kernels. The calculation of the convolution layer is as follows [16]:

$$Y^{l(i,j)} = \sigma_i^l * X^{l(j)} = \sum_{j'=0}^{W-1} \sigma_i^{l(j')} X^{l(j+j')}. \quad (3)$$

where  $\sigma_i^{l(j')}$  is the  $j'$ th weight on the  $i$ th convolution kernel of the  $l$ th layer.  $X^{l(j')}$  denotes the  $j'$ th convolved region of the  $l$ th layer.  $W$  indicates the width of the convolution kernel.  $Y^{l(i,j)}$  denotes the result of the  $i$ th convolution kernel in the layer  $l$  with the convolved region on the input signal  $X$ .

The activation layer is a nonlinear mapping of the values output from the convolution layer to a specific interval using an activation function. When the excitation input reaches a certain strength, the neuron is activated. Since the convolution operations in the convolution layer are linear, the complexity of the neural network and its ability to fit the target are greatly reduced if there is no activation function for nonlinear operations.

The role of the pooling layer is to extract higher dimensional fault features through pooling operations, thus reducing the computation and making the data representation more obvious. The common pooling operations are maximum pooling and average pooling [17]. Maximum pooling is defined as representing the maximum value of the pooled region in the data as this region, while average pooling is defined as representing the average value of the pooled region in the data as this region.

Fully connected layers are structures in which neurons in this layer are connected two by two with neurons in the upper layer, but not between neurons in this layer. The role of the fully connected layer is to enhance the nonlinear mapping capability of the network, to limit the size of the network, and to classify the features extracted by the convolutional and pooling layers. Converts the pooling layer output data into a one-dimensional vector, which is then used as input to the fully connected layers. The output length of the fully connected layer is the number of labels recognized by the neural network.

### 3 The Proposed Method

As shown in Fig. 2, the RSGW convolution layer will be explained according to the convolution calculation method and the construction of the convolution kernel. The convolution calculation method is the RSGW transform in signal processing theory. The construction of the convolution kernel will be constrained and designed according to the vanishing wavelet theory. The construction of the convolution kernel will be constrained and designed according to the vanishing moment in wavelet theory. The purpose of all these designs is to combine RSGW with convolutional neural networks to form a new deep CNN driven by RSGW.

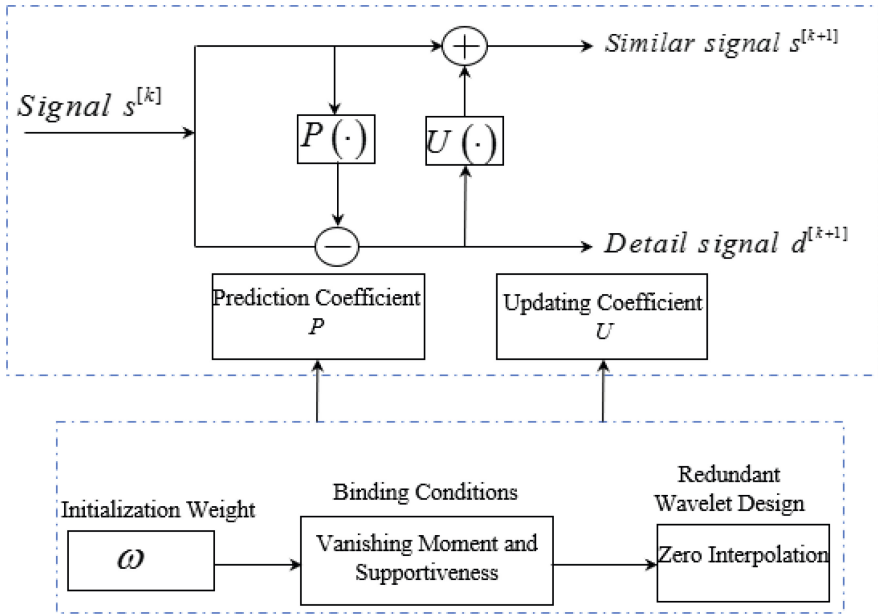


Fig. 2. Design of RSGW convolution layer

### 3.1 Design of RSGW Convolution Kernel

#### The Construction of the Initial Predictor $P$ and Updater $U$

Suppose that the coefficient of the predictor  $P$  with length  $N$  is  $P = [p_1, p_2, \dots, p_{N/2}, p_{N/2+1}, \dots, p_N]$ , and the coefficient of the updater  $U$  with length  $\tilde{N}$  is  $U = [u_1, u_2, \dots, u_{\tilde{N}}]$ . Claypoole uses the equivalent filter method to obtain the prediction operator  $P$  and updater  $U$ , that is, the specific coefficients of  $P$  and  $U$  can be obtained by solving the linear equations [18]. If the order of the prediction polynomial of constraint  $P$  is  $M$ . For a linear predictor  $P$  with a length of  $N$ , only the polynomial  $M < N$  order of predictor  $P$  is required to be constrained, and the remaining  $N - M$  degrees of freedom will be updated by the back-propagation algorithm to adapt to the signal characteristics. The relationship between RSGW equivalent high-pass filter  $\tilde{H}$  and prediction operator  $P$  is

$$\tilde{H} = [-p_1 \ 0 \ -p_2 \ 0 \ \dots \ -p_{N/2} \ 1 \ -p_{N/2+1} \ 0 \ \dots \ 0 \ -p_N]. \tag{4}$$

The specific expression of the constraint predictor coefficient polynomial of order  $M$  is

$$\sum_{k=-N+1}^{N-1} k^q \tilde{H}_k = 0, 0 \leq q < M. \tag{5}$$

The relationship between RSGW reconstruction equivalent high-pass filter  $H$  and predictor  $P$  and updater  $U$  is

$$H_{2l-1} = \begin{cases} 1 - \sum_{m=1}^N P_{S_m} U_{S_{l-m+1}} & l = (N + \tilde{N})/2 \\ \sum_{m=1}^N P_{S_m} U_{S_{l-m+1}} & l \neq (N + \tilde{N})/2 \end{cases} \quad (6)$$

$$H_{2l+N-2} = U_{S_l} \quad l = 1, 2, \dots, \tilde{N}$$

Equation (6) can be simplified as follows:

$$\tilde{V}H = 0. \quad (7)$$

where  $\tilde{V}$  is a matrix of size  $\tilde{N} \times [2 \times (N + \tilde{N}) - 1]$ , whose elements are represented as follows:

$$[\tilde{V}]_{m,n} = n^m. \quad (8)$$

where  $n = -N - \tilde{N} + 2, -N - \tilde{N} + 3, \dots, N + \tilde{N} - 3, N + \tilde{N} - 2, m = 0, 1, \dots, \tilde{N} - 1$ . Since the coefficient  $P$  is determined by the order of the constraint coefficient polynomial and the adaptive signal, Eq. (7) is a linear system of equations containing only the coefficients of  $U$ , and then  $U$  can be determined by the least square method. Duan proved through experiments that when  $N - M \leq 2$  can make the predictor adapts to the signal best [12]. Therefore, the order of the predictor coefficient polynomial of the RSGW selected in this paper is  $N - 2$ , and the remaining degrees of freedom are determined by the neural network by fitting the input signal.

### The Construction of Redundant Predictor $P^{[k]}$ and Updater $U^{[k]}$

Based on the initial predictor, the coefficient  $p_r^{[k]}$  of the  $k$  th RSGW decomposition predictor is calculated as follows 15:

When  $r - 1$  can be divisible by  $2^k$ ,

$$p_r^{[k]} = p_{(r-1)/2^k}. \quad (9)$$

When  $r - 1$  can't be divisible by  $2^k$ ,

$$p_r^{[k]} = 0. \quad (10)$$

Then we can get the redundant predictor  $P^{[k]} = \{p_r^k, r = 1, 2, \dots, 2^r N\}$  in the decomposition of the layer  $k$ .

Based on the initial updater  $U$ , the coefficient  $u_l^{[k]}$  of the redundant updater decomposed by layer  $k$  is designed as follows:

When  $l - 1$  can be divisible by  $2^k$ ,

$$u_l^{[k]} = u_{(l-1)/2^k}. \quad (11)$$

When  $l - 1$  can't be divisible by  $2^k$ ,

$$u_l^{[k]} = 0. \tag{12}$$

Then the redundant updater  $U^{[k]} = \{u_l^k, l = 1, 2, \dots, 2^l \tilde{N}\}$  of the  $k$  th layer decomposition can be obtained.

### 3.2 Deep CNN Driven by Multiscale RSGW Kernels

The main structure of RW-Net includes an RSGW convolution layer (Conv1), 1D convolution layers, adaptive maximum pooling layers, and fully connected layers, as shown in Fig. 3. Two RSGW decompositions are performed in the Conv1, which can better extract the input signal features without losing the useful information in the signal. The RSGW convolution kernel is  $a$ , and then the initial prediction operator  $b$  and the updating operator  $c$  are obtained according to the wavelet vanishing moment and the equivalent filter method, and their lengths are 10. The longer the length of  $P$  and  $U$ , the more the waveform changes of the RSGW, and the stronger the ability of the RSGW to adapt to the signal. However, the longer the length  $P$  and  $U$  will increase the training time of the network, which is considered as 10. When performing RSGW transform, the initial  $P$  and  $U$  will be interpolated according to the number of scale transformations to increase the length. The length relationship between the redundant prediction operator  $P$  and the updating operator  $U$  of the  $k$  th scale transformation is shown in Eq. (13). Therefore, the length of the redundant prediction operator  $P^{[1]}$  and updating operator  $U^{[1]}$  of the first RSGW transform is 20, and the second is 40. Subsequently, too many RSGW convolution kernels will increase the training time of the network, and too few will affect

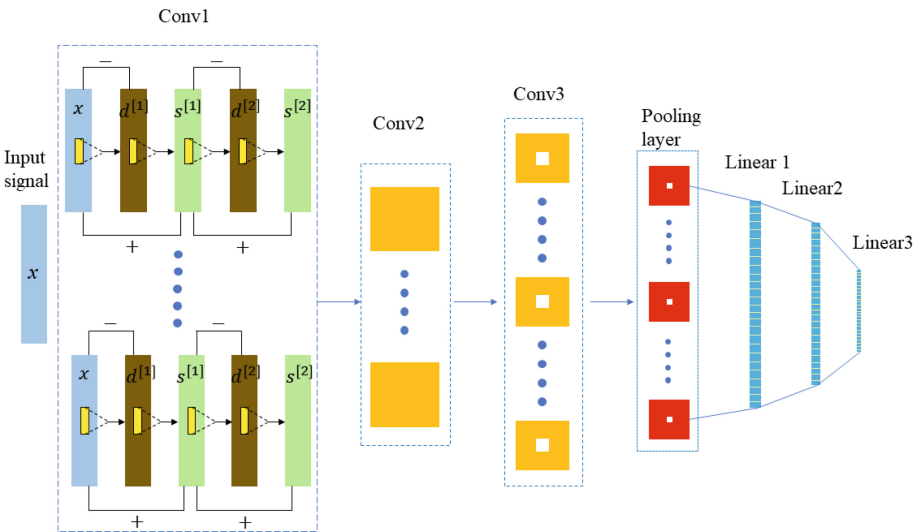


Fig. 3. RW-Net network structure diagram



the network performance. The number of kernels in Conv1 is set to 6.

$$\begin{aligned} P^{[k]} &= 2^k P \\ U^{[k]} &= 2^k U \end{aligned} \quad (13)$$

The specific parameters of RW-Net are shown in Table 1.

**Table 1.** Specific parameters of RW-Net

Network layers	Kernel size	Channels	Output	Padding	Activation function
Conv1	2	6	1*6*1024	Yes	Tanh
Conv2	5	16	1*16*1020	No	Tanh
Conv3	25	32	1*32*996	No	Tanh
Pooling	–	32	1*32*32	–	–
Linear1	–	–	1*1*120	–	Tanh
Linear2	–	–	1*1*72	–	Tanh
Linear3	–	–	m	–	Softmax

## 4 Experimental Verification

### 4.1 Case1: CWRU

#### 4.1.1 Dataset Description

In this paper, the bearing fault dataset of CWRU is selected as the experimental object [19]. The motor speed is 1730 RPM and the sampling frequency is 12000 Hz. The bearing label categories to be identified are shown in Table 2. There are 10 kinds of labels, including 0.007, 0.014, and 0.021 inch fault diameter bearing with inner ring, roller, and outer ring fault and healthy bearing. The total number of samples in the bearing signal dataset is 2000, and the training and test samples are divided in a ratio of 3:1. That is, the training data sample is 1500, and the test data sample is 500. In order to reflect the advantages of this method for short sample length (fewer sample points), the input signal length is 512.

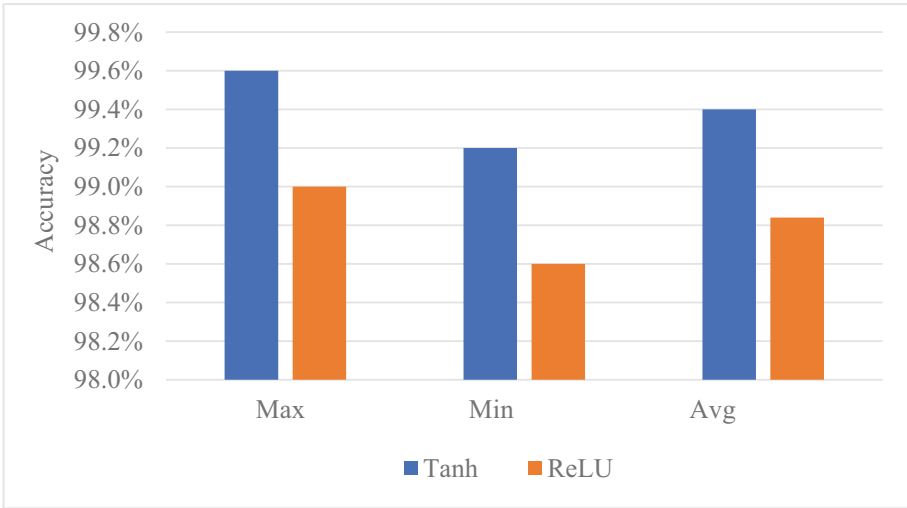
#### 4.1.2 Selection of Activation Function

The activation function is very important for neural network nonlinearity and diagnostic ability. The ReLU activation function allows eigenvalues greater than zero in the signal to pass through, and values less than zero are treated as zero, while the Tanh activation function can map eigenvalues to the interval  $[-1, 1]$ . In order to explore the influence of ReLU and Tanh activation functions on the diagnostic performance of RW-Net, RW-Net with different activation functions is utilized to do comparative experiments on CWRU

**Table 2.** Label information

	0 Inch	0.007 Inch	0.014 Inch	0.021 Inch	Motor speed
Normal	Label1	–	–	–	1797 RPM
Inner ring	–	Label2	Label3	Label4	
Roller	–	Label5	Label6	Label7	
Outer ring	–	Label8	Label9	Label	

bearing fault dataset. In order to guarantee the objectivity of the results, each network is trained 5 times. Figure 4 and Table 3 show the influence of ReLU and Tanh activation functions on the accuracy of the network. It can be seen from Fig. 4 that the Tanh activation function can make the RW-Net diagnostic performance better than the ReLU activation function, and its maximum, minimum, and average accuracy are the highest. Table 3 records the specific values of the correct rate 99.6%, 99.2% and 99.4%. In this paper, Tanh is selected as the activation function of RW-Net.



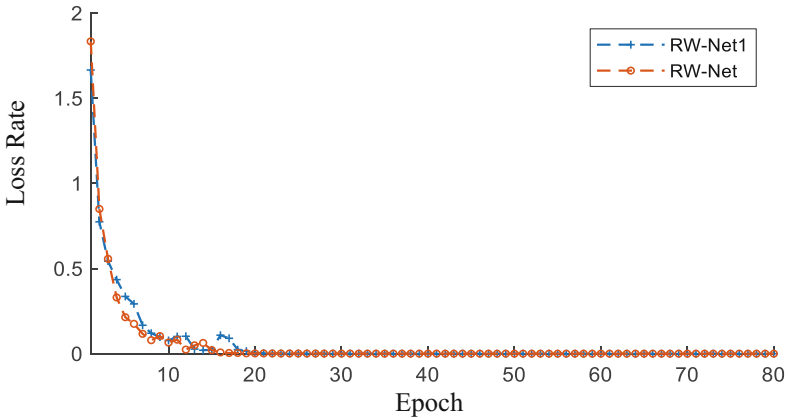
**Fig. 4.** The effect of ReLU and Tanh activation functions on the accuracy of RW-Net

**Table 3.** The statistical results of the test under different activation functions

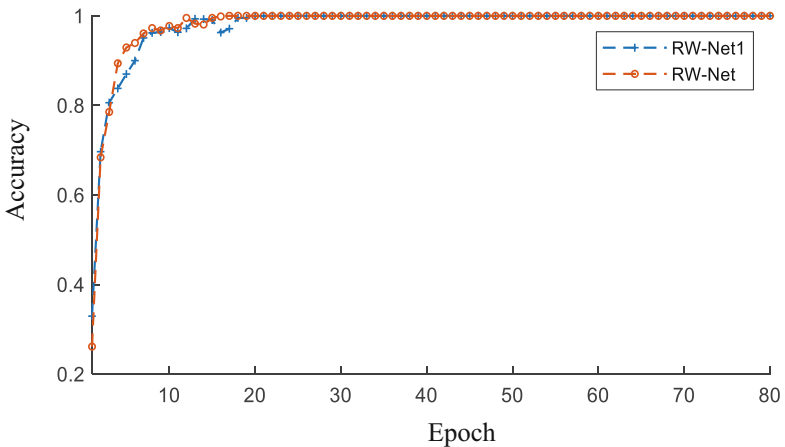
Activation function	Maximum accuracy	Minimum accuracy	Average accuracy
Tanh	<b>99.6%</b>	<b>99.2%</b>	<b>99.40%</b>
ReLU	99.0%	98.6%	98.84%

### 4.1.3 Experimental Contrast Analyses

In order to explore the influence of the number of RSGW transform in the Conv1 on the overall performance of RW-Net, this paper will perform an RSGW transform in the convolution layer, called RW-Net1. Figure 5 and Fig. 6 show the change in loss rate and accuracy rate of RW-Net and RW-Net1 during training on the CWRU dataset. It can be seen that the loss rate of RW-Net in training decreases faster than that of RW-Net1, and the loss has been close to 0 as the number of iterations increases. In terms of accuracy, RW-Net can also reach 100% quickly in training and remain stable. All these indicate that RW-Net is better than RW-Net1 in instability and convergence in training.



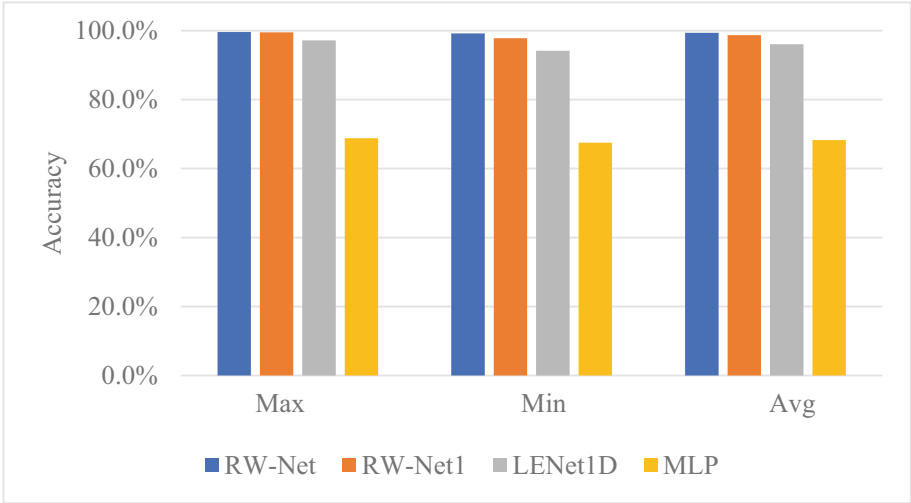
**Fig. 5.** The loss rate of RW-Net and RW-Net1 during training



**Fig. 6.** The accuracy of RW-Net and RW-Net1 during training

Compared with the current mainstream intelligent diagnosis models LeNet1D and MLP. Figure 7 and Table 4 show the test results of RW-Net and comparison methods

on the CWRU bearing dataset, respectively. It can be seen that RW-Net has achieved the best results. Although RW-Net1 also has good diagnostic ability on this dataset, the diagnostic accuracy is always less than RW-Net. By analyzing the experimental results, it can be concluded that: 1) RW-Net and RW-Net1 have better recognition ability for CWRU bearing dataset than comparison methods; 2) Two RSGW transforms in the Conv1 can improve the network diagnostic ability more than only once.



**Fig. 7.** The accuracy of RW-Net and comparison methods on the test set of CWRU

**Table 4.** The accuracy of RW-Net and comparison methods on the test set of CWRU

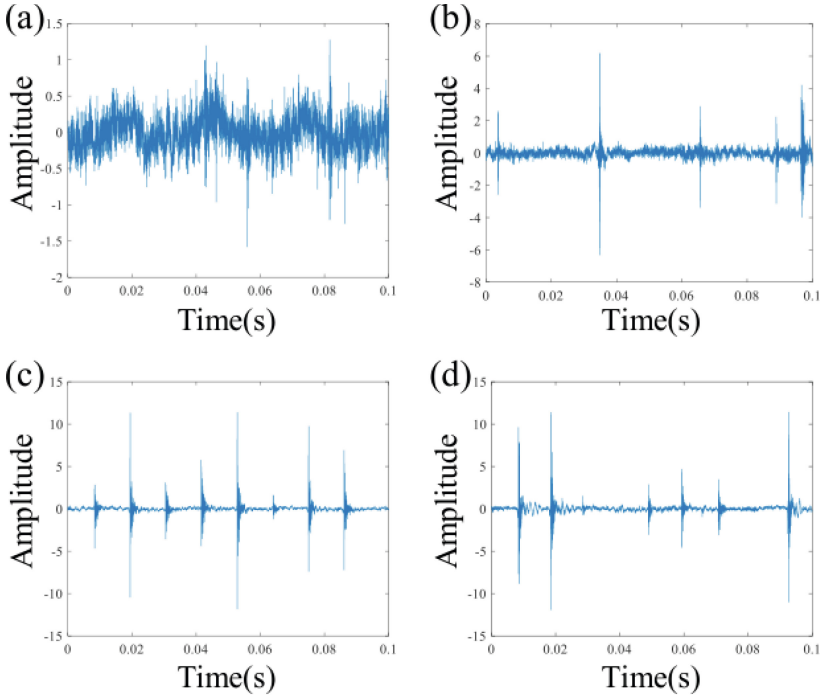
Network models	Maximum accuracy	Minimum accuracy	Average accuracy
RW-Net	<b>99.6%</b>	<b>99.2%</b>	<b>99.40%</b>
RW-Net1	99.5%	97.8%	98.70%
LENet1D	97.2%	94.2%	96.07%
MLP	68.8%	67.5%	68.27%

## 4.2 Case2: JNU

### 4.2.1 Dataset Description

Jiangnan University (JNU) bearing datasets were provided by Jiangnan University [20, 21] The JNU datasets are composed of three bearing vibration datasets with different rotating speeds, and the data acquisition frequency is 50 kHz. In this experiment, the 1000 RPM bearing fault dataset is adopted. Its vibration signals are shown in Fig. 8. This dataset contains one health state and three fault patterns, including inner ring fault,

outer ring fault and rolling element fault. Each category of health condition takes 400 samples, and each sample contains 512 data points. The dataset contains a total of 1200 samples. The samples corresponding to each label were assigned to the training and test sets in a ratio of 3:1, respectively.



**Fig. 8.** (a) Health state (b) Inner ring (c) Outer ring (d) Rolling element

#### 4.2.2 Experimental Contrast Analyses

Likewise, the experiment is used as a comparison with the current mainstream intelligent diagnostic models LeNet1D and MLP. Figure 9 and Table 5 show the test results of RW-Net and comparison methods on the CWRU bearing dataset, respectively. It can be seen that RW-Net has achieved the best results. The experiments demonstrate not only the superior fault identification capability of RW-Net, but also its robustness and generalization capability.



**Fig. 9.** The accuracy of RW-Net and comparison methods on the test set of JNU

**Table 5.** The accuracy of RW-Net and comparison methods on the test set of JNU

Network models	Maximum accuracy	Minimum accuracy	Average accuracy
RW-Net	98.5%	98.0%	98.32%
RW-Net1	97.5%	97.3%	97.42%
LENet1D	97.2%	94.2%	96.07%
MLP	68.8%	67.5%	68.27%

## 5 Conclusion

This paper proposes an improved CNN based on RSGW theories, called RW-Net. The shallow layer of RW-Net performs RSGW transform on time-domain signals. This layer inherits the advantages of RSGW in signal processing to extract signal features. RW-Net takes time domain signal as input. The RSGW layers are used as a multi-channel filter to simultaneously extract multiple fault features, and then the features extracted by the RSGW layers are fused as the input of the pooling layer. By enhancing the feature extraction ability of the shallow layer of the proposed method, the fault features can be accurately extracted by using small sample datasets, and the network training parameters are reduced. In this paper, the feasibility of RW-Net is verified by the CWRU and JNU bearing fault datasets. From the analysis of the experimental results, it can be seen that the average accuracy of RW-Net reaches 99.4% and 98.32%, which is better than other comparison methods. The feasibility and effectiveness of the proposed method are fully illustrated.

In this paper, we use the time-domain signal as the input signal of the network and each labeled data segment is independent of each other when segmenting the data set, i.e., the correlation between data segments is not considered. We think the correlation between data segments can be the next research direction. The graph convolutional neural network may be a good method.

## References

1. Huang, Z., Lei, Z., Huang, X., et al.: A multisource dense adaptation adversarial network for fault diagnosis of machinery. *IEEE Trans. Ind. Electron.* **69**(6), 6298–6307 (2021)
2. Jiang, Y., Yin, S.: Recursive total principle component regression based fault detection and its application to vehicular cyber-physical systems. *IEEE Trans. Ind. Inf.* **14**(4), 1415–1423 (2017)
3. Yuan, J., Cao, S., Ren, G., et al.: LW-Net: an interpretable network with smart lifting wavelet kernel for mechanical feature extraction and fault diagnosis. *Neural Comput. Appl.* **34**(18), 15661–15672 (2022)
4. Yuan, J., He, Z., Zi, Y., et al.: Adaptive multiwavelets via two-scale similarity transforms for rotating machinery fault diagnosis. *Mech. Syst. Sig. Process.* **23**(5), 1490–1508 (2009)
5. Shao, H., Jiang, H., Zhang, H., et al.: Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing. *Mech. Syst. Sig. Process.* **100**, 743–765 (2018)
6. Li, Z., Chen, J., Pan, J.: Independence-oriented VMD to identify fault feature for wheelset bearing fault diagnosis of high-speed locomotive. *Mech. Syst. Sig. Process.* **85**, 512–529 (2016)
7. Chen, J., Li, Z., Chen, G., et al.: Wavelet transform based on inner product in fault diagnosis of rotating machinery: a review. *Mech. Syst. Sig. Process.* **70**, 1–35 (2016)
8. Ming, A., Qin, Z., Zhang, W., et al.: Spectrum auto-correlation analysis and its application to fault diagnosis of rolling element bearings. *Mech. Syst. Sig. Process.* **41**(1–2), 141–154 (2013)
9. Wang, P., Song, L., Guo, X., et al.: A high-stability diagnosis model based on a multiscale feature fusion convolutional neural network. *IEEE Trans. Instrum. Meas.* **70**, 1–9 (2021)
10. Lund, D., MacGillivray, C., Turner, V., et al.: Worldwide and regional internet of things (IoT) 2014–2020 forecast: a virtuous circle of proven value and demand. International Data Corporation (IDC), Technical Report, vol. 1, no. 1, p. 9 (2014)
11. Chen, Z., Li, W.: Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Trans. Instrum. Meas.* **66**(7), 1693–1702 (2017)
12. Duan, C., Li, L., He, Z.: Application of second generation wavelet transform to fault diagnosis of rotating machinery. *Mech. Sci. Technol.* **23**(2), 224–226 (2004)
13. Duan, C., He, Z.: Second generation wavelet denoising and its application in machinery monitoring and diagnosis. *Mini-Micro Syst.* **25**(7), 1341–1343 (2004)
14. Gao, L., Tang, W., et al.: Noise reduction technology based on redundant second generation wavelet. *J. Beijing Univ. Technol.* **34**(12), 1233–1237 (2008)
15. Jiang, H., He, Z., Duan, C.: Construction of redundant second generation wavelet and mechanical signal feature extraction. *J. Xi’an Jiaotong Univ.* **38**(11), 1140–1142 (2004)
16. Zhang, W.: Study on Bearing Fault Diagnosis Algorithm Based on Convolutional Neural Network. Harbin University of Science and Technology (2017)
17. Chen, X., Xiang, S., Liu, C., et al.: Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **11**(10), 1797–1801 (2014)

18. Claypoole, R.L., Baraniuk, R.G., et al.: Adaptive wavelet transforms via lifting. In: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 3, pp. 1513–1516 (1998)
19. Bearing Data Center, Case Western Reserve University, Cleveland, OH, USA (2004). [http://csegroups.case.edu/bearing\\_datacenter/home](http://csegroups.case.edu/bearing_datacenter/home)
20. Li, K.: School of Mechanical Engineering, Jiangnan University (2019)
21. Li, K., Ping, X., Wang, H., Chen, P., Cao, Y.: Sequential fuzzy diagnosis method for motor roller bearing in variable operating conditions based on vibration analysis. *Sensors* **13**(6), 8013–8041 (2013)