

# Lightweight rolling bearing fault diagnosis via dense connectivity and adaptive soft thresholding

Yanqi Wang<sup>1</sup>, Songlin Zhang<sup>2</sup>, Ruming Ding<sup>3</sup>, Cheng Luo<sup>4</sup>, Letian Zhong<sup>5</sup>

Department of Anhui Institute of Information Technology, Wuhu Anhui, 241000, China

<sup>1</sup>Corresponding author

E-mail: <sup>1</sup>yanqi\_wang@foxmail.com, <sup>2</sup>149721096@qq.com, <sup>3</sup>398029523@qq.com, <sup>4</sup>1466899043@qq.com, <sup>5</sup>a2371725573@gmail.com

Received 26 August 2025; accepted 1 January 2026; published online 24 February 2026

DOI <https://doi.org/10.21595/jve.2026.25314>



Copyright © 2026 Yanqi Wang, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract.** Deep learning-based fault diagnosis of rolling bearings is frequently challenged by strong ambient noise in vibration signals and the high computational cost of deployable models. While deeper networks can enhance performance, they often lead to parameter redundancy and information loss in deep layers, hindering industrial application. To achieve a balance between noise robustness and model lightness, this paper proposes SE-SDCTNet, a novel architecture built upon Sparse Dense Compact Thresholding (SDCT) blocks and Squeeze-and-Excitation (SE) blocks. The SDCT blocks employ dense connections for efficient feature reuse, while incorporating sparsity constraints and an integrated soft-thresholding mechanism to actively suppress noise and reduce parameters. Subsequently, SE blocks adaptively recalibrate channel-wise features to compensate for potential information loss due to sparsity and to enhance discriminative power. Furthermore, dilated convolutions are embedded to preserve multi-scale contextual information throughout the network. Evaluated on the Case Western Reserve University (CWRU) bearing dataset, SE-SDCTNet demonstrates superior diagnostic accuracy (e.g., 93.1 % under severe 2 dB noise) and robustness across various signal-to-noise ratios, while containing only 0.32 million parameters, merely about 3 % of ResNet18. In summary, this work provides a lightweight, accurate, and robust solution that facilitates the transition of data-driven fault diagnosis from theoretical research to practical industrial deployment.

**Keywords:** bearing fault diagnosis, SE-SDCTNet, sparse dense compact thresholding block, squeeze-and-excitation block, dilated convolution, lightweight model.

## 1. Introduction

Rotating machinery, such as bearings and gears, is widely used in industrial fields. The health state of these crucial mechanical components greatly influences the performance, stability, and lifespan of the machine. Reliable monitoring and fault diagnosis of rolling bearings are essential for the safe, stable operation of rotating machinery [1]. The development of computer technologies and big data has ushered in a new era of data-driven fault diagnosis for rolling bearings. Machine learning techniques, leveraging feature extraction and subsequent pattern recognition, have demonstrated significant potential in bearing fault diagnosis through their self-learning optimization capabilities. Notable examples include support vector machines [2], Bayesian networks [3], and deep learning networks [4].

Notably, deep networks have proven highly effective for fault diagnosis by extracting semantic features [5-7]. For instance, Xu et al. developed a model combining convolutional neural networks with deep forests, utilizing time-frequency images derived from vibration signals [8]. Liu et al. introduced a feature-integration-boosting-based discriminative stacked autoencoder (DSAE) network that leverages feature-integration boosting for bearing fault diagnosis. In their approach, multiple DSAEs serve as weak classifiers to extract fault-related features directly from raw vibration signals [9]. These advancements highlight the continuous evolution and refinement of deep learning techniques in the field of rolling bearing fault diagnosis. Among the available architectures, convolutional neural networks (CNNs) are particularly suitable for signal processing

due to the computational efficiency of their convolutional and pooling functions, while effectively capturing the periodic characteristics of signals [10]. As the architectures of CNNs become increasingly deep to capture more complex information, the vanishing gradient problem often arises in traditional network topologies. To address this issue, residual networks (ResNets) that utilize skip connections have brought a significant breakthrough in CNNs and have garnered considerable interest in the field of machinery fault diagnosis. Skip connections help reduce training difficulty, improve information flow, and facilitate parameter optimization, enabling deep and complex networks to be trained effectively through residual learning [11]. However, collected vibration signals are typically corrupted by heavy noise that obscures fault related signatures and severely erodes diagnostic accuracy [12]. Consequently, the development of effective and reliable fault diagnosis methods under noisy conditions remains a critical challenge.

To address this limitation, researchers have explored various strategies to enhance model noise robustness. Zhen et al. [13] introduced a VMD-SG denoising algorithm that combines two-dimensional grayscale images with ResNet for wind-turbine bearing fault diagnosis. Yu et al. [14] devised a convolutional autoencoder-based model that leverages residual learning to enhance the deep network's feature-extraction capability. However, these approaches tend to attenuate signals indiscriminately during pre-training. To tackle this limitation, the attention mechanism has been increasingly applied in intelligent fault diagnosis networks, particularly within a multi-scale structure [15]. Liang et al. [16] furthered this by integrating a squeeze-and-excitation attention into a parallel residual network structure, utilizing dilated convolutions to create multi-level connections. Zhang et al. [17] also presented a novel global multi-attention deep residual shrinkage network to mitigate noise interference. The attention mechanism selectively preserves prominent features, proving to be a viable approach for noise suppression. The field of fault diagnosis has witnessed considerable advancements through deep learning models, many of which exploit channel relationships and attention mechanisms to boost performance. Nevertheless, their practical deployment is often hindered by substantial computational overhead and parameter volume. In response, efficient architectures such as MobileNetV2 [18] and LiNet [19] have been developed, employing inverted residuals and linear bottlenecks to achieve a favorable balance between accuracy and efficiency. A pronounced susceptibility to performance degradation in noisy settings, however, persists as a major limitation.

Hence, designing models that are simultaneously parameter-efficient and noise-robust remains a vital and immediate research objective for industrial fault diagnosis. To address this dual challenge of noise interference and parameter redundancy, this paper proposes SE-SDCTNet, a lightweight, noise-robust model tailored for industrial deployment. Its core advantages lie in three synergistic innovations: first, stacked Sparse-Dense Compact Thresholding (SDCT) blocks integrate dense connectivity (for efficient feature reuse) with sparsity constraints and adaptive soft-thresholding (for targeted noise suppression), achieving high accuracy without excessive parameters; second, SE blocks calibrate channel-wise feature responses to emphasize informative channels and eliminate redundancy, compensating for the limitation of channel independence in conventional CNNs; third, dilated convolutional layers extract multi-scale context information to preserve global fault features as the network deepens. Together, these components enable SE-SDCTNet to deliver a lightweight, high-precision, and robust solution for rolling bearing fault diagnosis in noisy industrial environments.

The remainder of this paper is organized as follows: Section II summarizes related works on residual and deep residual shrinkage networks. Section III details the proposed SE-SDCTNet architecture, including the SDCT block, SE attention, and their synergy. Section IV presents the experimental setup, dataset description, and analysis of diagnostic results under varying noise levels. Finally, Section V concludes the study and identifies directions for future research.

## 2. Related works

### 2.1. Residual network

Residual network is composed of multiple serially connected residual blocks, specifically designed to facilitate deep feature learning while maintaining training stability. As illustrated in Fig. 1, each residual block mainly consists of two parallel functional subblocks. A shortcut connection links the input directly to the output feature map, effectively addressing the problems of gradient vanishing and network degradation that often occur during the training of deep networks. This architecture is particularly beneficial for bearing fault diagnosis, where accurate ex-traction of subtle fault features from complex vibration signals is crucial. The residual formulation introduces an auxiliary mapping:

$$F(x) = H(x) - x, \tag{1}$$

where  $x$  denotes the input of network, and let  $H(x)$  be the desired underlying mapping. Then, the traditional training algorithm through fitting  $H(x)$  is transformed equivalently to fitting the residual function  $F(x)$ :

$$F(x) \rightarrow 0 \Leftrightarrow H(x) \rightarrow x, \tag{2}$$

when  $F(x)$  reaches 0,  $H(x) = x$ .  $H(x) = x$  represents the best solution mapping for a set of stacked network layers. If the network becomes deeper, it is hard for the model to directly fit the actual mapping  $H(x)$ . The residual structure starts to fit the residual mapping  $F(x)$  by introducing shortcut connections, and the actual mapping  $H(x)$  is expressed as  $H(x) = F(x) + x$ . When  $F(x) = 0$ , an identity mapping  $H(x) = x$  is obtained. The approximation of the actual mapping is achieved by minimizing the residual function  $F(x) = H(x) - x$ , which solves the performance degradation problem with network layer are stacked. So, compared with traditional training methods, it provides a more convenient and efficient approaches [20].

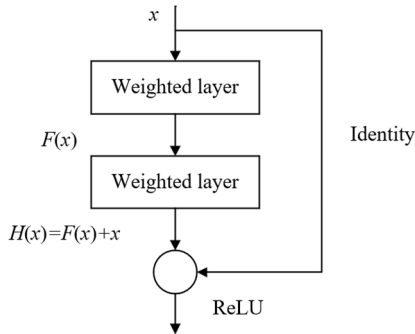


Fig. 1. The structure of the residual block

### 2.2. Deep residual shrinkage network

To enhance robustness against noise, the deep residual shrinkage network (DRSN) was proposed as an extension of ResNet. DRSN integrates soft thresholding with channel-wise attention to selectively suppress noise-dominated features while preserving discriminative fault information. Specifically, within each residual shrinkage module, a subnetwork first estimates adaptive thresholds based on global feature statistics via a squeeze-and-excitation mechanism. These thresholds are then applied through a soft thresholding operator:

$$y = \text{sign}(x) \cdot \max(|x| - \tau, 0), \tag{3}$$

where  $x$  and  $y$  denote the input and output features, respectively, and  $\tau > 0$  is a learnable threshold. This design enables DRSN to effectively attenuate small-magnitude coefficients, while retaining salient fault-induced transients, which are easily masked in early-stage failures [20].

Despite its improved noise robustness, DRSN inherits the parameter-intensive nature of deep ResNet variants. The inclusion of additional attention subnetworks and multiple convolutional layers lead to a substantial increase in model complexity. In practical industrial applications, where computational resources and deployment costs are constrained, such high-parameter models are often impractical. Therefore, there remains a critical need for a lightweight yet robust fault diagnosis architecture that simultaneously achieves high accuracy under noisy conditions and maintains low computational overhead. This gap motivates our development of SE-SDCTNet, a compact model that synergistically combines sparse dense connectivity, adaptive soft thresholding, and channel attention to deliver both strong noise immunity and parameter efficiency.

### 3. Proposed method

Increasing the number of layers in a deep neural network can enhance the system's ability to resist noise; however, networks with too many layers are more difficult to be trained. Additionally, a greater number of parameters requires more powerful hardware, making such networks challenging to deploy in practical applications. The computational cost of a convolutional neural network primarily depends on the number of convolutional layers and can be expressed as:

$$N = \sum_{k=1}^Q g(x, y), \quad (4)$$

where  $Q$  is the total number of convolutional layers,  $x$  and  $y$  represent the input and output of the convolutional layer, and  $g(x, y)$  represents the parameter quantity of the convolution operation, which is specifically expressed as:

$$g(x_i, y_j) = \sum_{i=1}^P \sum_{j=1}^c f(H, W, m, n, s), \quad (5)$$

$$f(H, W, m, n, s) = (H - s + 1) \times (W - s + 1) \times m \times n, \quad (6)$$

where  $H$  and  $W$  represent the height and width of the input feature map,  $m$  and  $n$  represent the size of the convolution kernel,  $s$  represents the step size,  $p$  represents the number of convolution layers, and  $c$  represents the number of channels. The number of channels is directly related to the amount of computation required for each convolution. Additionally, feature extraction and feature fusion operations during the convolution process significantly impact computational requirements. Therefore, it is crucial to reasonably reduce the number of channels and layers in the deep network model while maintaining the accuracy of fault diagnosis.

#### 3.1. Sparse dense compact thresholding block

To enhance feature utilization efficiency and reduce the number of network parameters, we employ dense blocks. Unlike residual blocks, dense blocks promote the propagation of features and make more efficient use of features at different levels within each block. Each dense block is composed of multiple bottleneck layers that utilize dense connections between layers, maintaining consistent output sizes throughout the block. As shown in Fig. 2, this architecture ensures that the output of each layer contains the characteristic information from all preceding layers. In a dense connection setup, every layer receives feature maps from all previous layers as inputs and passes

its own feature maps to all subsequent layers. This design not only increases the flow of information and gradients during training but also encourages feature reuse. The formula is as follows:

$$x_l = H_l(x_0 \oplus x_1 \oplus \dots \oplus x_{l-1}), \quad (7)$$

where  $l$  corresponds to the input vector of the  $l$ -th layer; the operator  $\oplus$  represents the concatenation operation of the feature vector; the function  $H_l$  represents the nonlinear conversion function, which is a combination of Conv ( $1 \times 1$ )-BN-ReLU. Here, BN refers to Batch Normalization, ReLU represents the activation function, Conv ( $1 \times 1$ ) indicates a convolution layer with a kernel size of  $1 \times 1$ , and Conv ( $3 \times 3$ ) indicates a convolution layer with a kernel size of  $3 \times 3$ . This mechanism ensures that each layer can directly access the final error signal, thereby improving the accuracy of the model in bearing fault diagnosis.

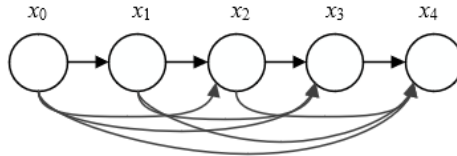


Fig. 2. The structure of the dense block

Inspired by sparse autoencoder [21], sparsity constraints are added to dense blocks to alleviate overfitting and suppress noise. By forcing the neural network to focus on the primary features of the data and ignore minor signals, this approach achieves stronger feature extraction and noise reduction capabilities. As illustrated in Fig. 3, the sparsity constraint ensures that only a portion of the neurons in the network are active. Through this sparse structure, key features in the data can be effectively extracted. It is implemented by L1 regularization.

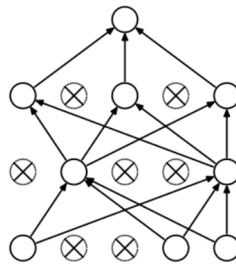


Fig. 3. The neural networks with sparsity constraints

In this work, the SDCT block is proposed. As illustrated in Fig. 4, this module innovatively integrates a soft thresholding mechanism into a dense connection architecture, which is inspired by DRSN for capturing weak fault features in bearing vibration signals. Within the SDCT Block architecture, dense connections are implemented across all layers except the last: this design ensures full reuse of multilevel fault features from preceding layers, while feature concatenation is intentionally omitted at the final layer to reduce parameter complexity, addressing the industrial deployment barrier of large-parameter models. Additionally, sparse regularization is incorporated into the dense block to enforce sparsity constraints, further suppressing overfitting by limiting the activation of redundant neurons that may learn noise-related patterns, and guiding the network to focus on salient bearing fault features. The SDCT Block incorporates a dedicated subnetwork to estimate adaptive thresholds for soft thresholding, enabling targeted suppression of noise-related features in bearing vibration data. Specifically, Global Average Pooling (GAP) is first applied to the absolute values of the input feature map  $x$  (which encodes spatial-frequency information of bearing vibrations), producing a one-dimensional channel-wise descriptor vector that aggregates

global feature distribution across channels.

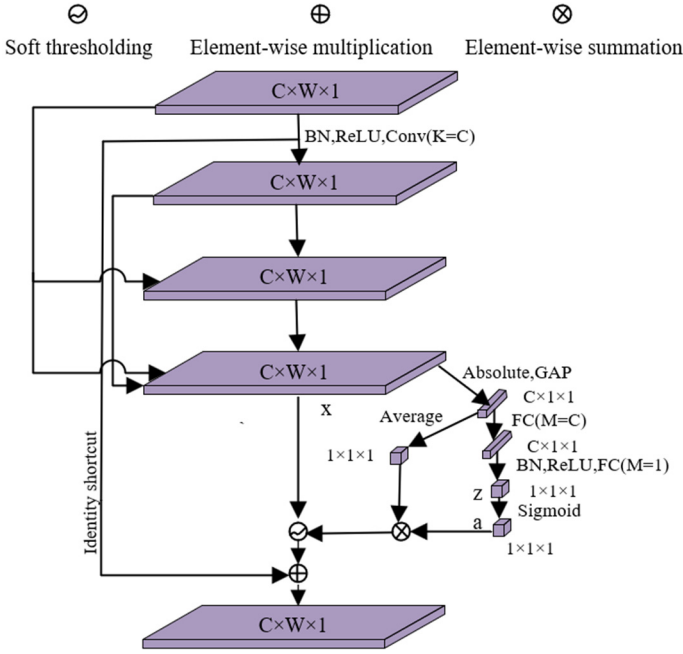


Fig. 4. The structure of the sparse dense compact thresholding block

### 3.2. SE block

Sparsity ensures that certain channels within the SDCT blocks remain inactive, effectively preventing overfitting caused by redundant noise-related features in bearing vibration data. To further enhance the model’s ability to capture discriminative bearing fault features, multiple SDCT blocks are stacked together. However, in rolling bearing fault diagnosis scenarios. To address these issues, an SE block is introduced immediately after each SDCT block to dynamically amplify the weights of informative channels that encode key fault characteristics, while downplaying channels dominated by noise. The channel attention mechanism embedded in the SE block enables the model to identify and emphasize task-relevant features with only a marginal increase in model complexity and computational overhead. Implemented through the SE block as illustrated in Fig. 5, it comprises three sequential processes: Squeeze, Excitation, and Reweight.

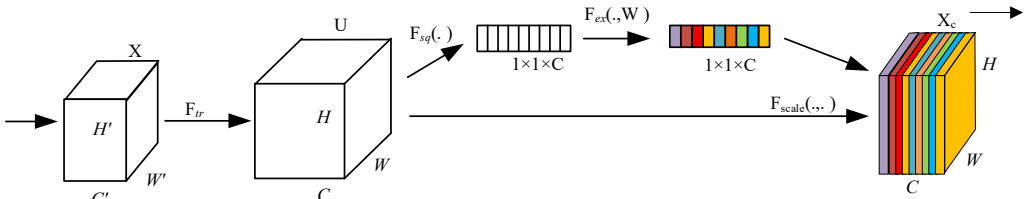


Fig. 5. The structure of the SE block

The globalization on feature map  $F_{tr}$  is employed as the generalized function, and it yields feature map  $G$ , with a size of  $W \times H \times C$ , as the input of the attention mechanism:

$$F_{tr}: X \rightarrow U, \quad X \in R^{H' \times W' \times D' \times C'}, \quad U \in R^{H \times W \times D \times C}, \quad (8)$$

$$F_{tr}: u_c = v_c * x = \sum_{s=1}^{c'} v_c^s * x^s, \tag{9}$$

where  $F_{tr}$  represents the convolution operation,  $U$  is the feature map,  $H$  is the height,  $W$  is the width,  $C$  is the number of channels, and  $D$  is the temporal depth.  $v_c$  represents the  $c$ -th convolution kernel, and  $u_c$  represents the  $c$ -th feature map. This process involves three distinct stages. Initially, the squeeze operation compresses features along the spatial dimension using global average pooling. Here, each spatial domain of size  $H \times W$  is reduced to a single real number, which not only possesses a global receptive field but also aligns with the input feature's channel count. These real numbers encapsulate the global distribution of responses across feature channels, enabling the network to gather global information even in early processing layers. The resulting output is a  $1 \times 1 \times C$  vector, formulated as:

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H \times D} \sum_{i=1}^W \sum_{j=1}^H \sum_{r=1}^D u_c(i, j, z). \tag{10}$$

The next step is the Excitation operation, which is similar to the gate mechanism in a recurrent neural network. Each feature channel generates a corresponding weight based on the parameters  $w$ . These weights are obtained through learning, aiming to clearly model the correlation between feature channels. The operation can be expressed as:

$$s = F_{ex}(z, w) = \sigma(g(z, w)) = \sigma(w_2 \delta(w_1 z)), \tag{11}$$

where  $w_1$  and  $w_2$  denote fully-connected operations,  $\delta$  is the ReLU function, and  $\sigma$  is the sigmoid function. The last stage is the reweight stage, which regards the weight  $s_c$  output by the Excitation stage as the importance of each feature channel after feature selection. Subsequently, these weights are weighted to the previous features  $u_c$  channel by channel through multiplication operations, thereby recalibrating the original features in the channel dimension. The final output feature map can be expressed as:

$$x_c = F_{scale}(u_c, s_c) = s_c \cdot u_c. \tag{12}$$

Stacking SDCT blocks and SE blocks enhances the feature extraction capability of the bearing fault diagnosis model, making the network both lightweight and efficient while improving its noise resistance.

### 3.3. The proposed model

The fault diagnosis model proposed in this paper, named SE-SDCTNet, is composed of sequentially stacked Sparse-Dense Compact Thresholding (SDCT) blocks and Squeeze-and-Excitation (SE) blocks, with its overall architecture illustrated in Fig. 6. The model takes a 2D grayscale image (converted from 1D bearing vibration signals to preserve raw feature information) as input, and its structure follows a data-driven flow: standard convolution layers, a max-pooling layer, stacked SDCT blocks (each followed by a serial SE block), transition layers, a Global Average Pooling (GAP) layer, and Fully Connected (FC) layers.

The architecture begins with a standard  $7 \times 7$  convolutional layer, processing the  $32 \times 32$  input image to produce initial feature maps. To reduce parameter count and computational overhead, thereby accelerating training and inference, a  $3 \times 3$  max-pooling layer with stride 2 is applied immediately afterward. All subsequent convolutional layers use  $3 \times 3$  kernels to balance feature extraction capability with computational efficiency. Following the initial pooling layer, two SDCT

blocks are sequentially stacked, each followed by an SE block and a transition layer. The SDCT block employs a three-layer convolutional structure (Fig. 4) to extract deep features: its dense connections enhance feature reuse by fully leveraging multi-level fault-related information from vibration signals, while its sparsity constraints suppress irrelevant noise and mitigate overfitting. The succeeding SE block adaptively recalibrates channel-wise feature weights, emphasizing inter-channel dependencies to further suppress noise and compensate for the channel-independent nature of standard convolutions. Each transition layer then performs two key functions: first, a  $1 \times 1$  convolution compresses the channel dimension to reduce model complexity and avoid feature redundancy; second, a max-pooling layer with stride 2 downsamples the spatial size of the feature maps, enabling multi-scale feature extraction in deeper layers. After processing through the stacked SDCT-SE-transition modules, the final feature map undergoes Global Average Pooling (GAP) to produce a one-dimensional feature vector. This vector is mapped to output units corresponding to each fault category (including the normal state). For this multi-classification task, a softmax activation is applied to generate class probabilities, and the category with the highest probability is selected as the final diagnostic label.

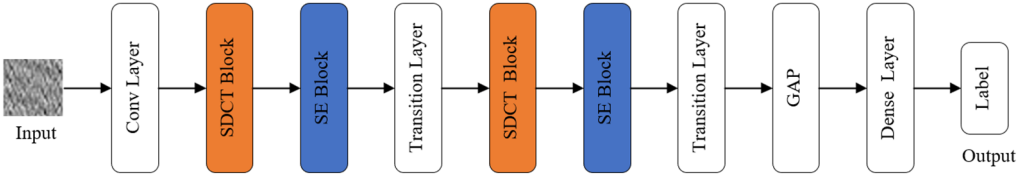


Fig. 6. The structure of the proposed model

The module of multiscale feature extraction is built using dilated convolutions with diverse receptive fields [22]. This approach allows the network to capture information at different scales while preserving feature details. Dilated convolutions achieve a larger receptive field by introducing gaps into standard convolution kernels. These inserted holes are treated as zero-weight elements during the convolution operation on input feature maps. As a result, dilated convolutions can expand the receptive field without elevating computational expenses. By introducing the dilation rate  $d$ , the effective size of the convolution kernel in dilated convolutions is adjusted as follows:

$$k' = k + (k - 1) \times (d - 1), \quad (13)$$

where  $k$  represents the size of the standard convolution kernel. In the SE-SDCTNet fault diagnosis model, different receptive fields are employed to acquire multi-scale feature information from vibration signals by configuring varying dilation rates for dilated convolutions. It is applied to a second SDCT block and  $d$  is set to 2.

### 3.4. Bearing image preparation from the vibration signal

To leverage the powerful pattern recognition capabilities of CNNs which are inherently well-suited for processing two-dimensional data, one-dimensional vibration signals are converted into grayscale images. While time-frequency analysis is a common approach, it often relies on hand-crafted transformations that may discard valuable information or introduce complexity. Instead, we employ a direct and computationally efficient reshaping method, which provides the CNN with a structural view of the signal segment conducive to learning local temporal patterns [23].

Grayscale images of size  $D \times D$  are generated by segmenting the raw vibration signal and reshaping the segment into a matrix. A length- $D^2$  segment – containing  $D^2$  consecutive one-dimensional samples is first extracted; this length is selected so that the segment spans at least one complete revolution of the bearing. The samples are then arranged row-wise into a  $D \times D$

matrix and normalized to the  $[0, 255]$  range, after which they are rounded to obtain integer pixel values. The resulting grayscale image is thus defined as:

$$P(h, i) = \text{Round} \left\{ \frac{L((h - 1) \times D + i) - \min(L)}{\max(L) - \min(L)} \times 255 \right\}, \quad (14)$$

where  $L(h)$ ,  $h = 1, 2, \dots, D^2$  represents the  $h$ -th the signal segment's one-dimensional data point.  $P(h, i)$  represents the value of the  $h$ -th row and the  $i$ -th column in the two-dimensional gray image with  $h = 1, 2, \dots, D$ ,  $i = 1, 2, \dots, D$ . The minimal and maximal values of the intercepted signal are denoted by  $\min(L)$  and  $\max(L)$ , respectively.  $\text{Round}(\cdot)$  is the rounding operator.

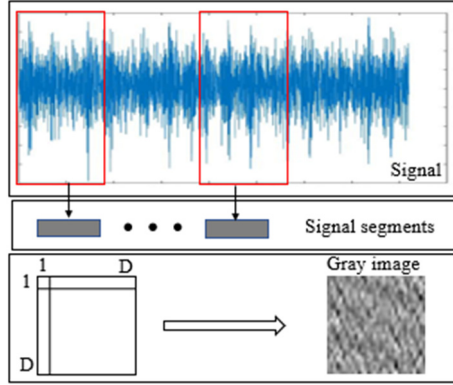


Fig. 7. The process of signal-to-image conversion

### 3.5. Loss function formulation

Rolling bearing fault diagnosis is a multiclass classification problem. In the proposed SE-SDCTNet model, the softmax function is employed to estimate the probability distribution over fault classes. The softmax function transforms the raw output scores (logits) from the neurons into values within the range  $(0, 1)$ , ensuring that their sum equals 1, thus forming a valid probability distribution. The model's loss function is defined as the categorical cross-entropy between the predicted probability distribution and the true label distribution, which effectively measures the divergence between the model's predictions and the ground truth, guiding the optimization process during training. It can be described as:

$$L_\varphi = H(p, q) = - \sum_{i=1}^c y(i) \log(\hat{y}_i), \quad (15)$$

where  $p(x)$  is the true distribution and  $q(x)$  is the estimated distribution.  $c$  is the number of categories.  $y(i)$  is the true label and  $\hat{y}_i$  is predicted label of the model. In addition, L1 regularization is added to the loss function in the SDCT blocks. The final loss function is described as:

$$L_{total} = - \sum_{i=1}^c y(i) \log(\hat{y}_i) + \lambda \sum_{j=1}^n |w_j|, \quad (16)$$

where  $w_j$  is the weight parameter of the model, and  $n$  is the total number of weight parameters, and  $\lambda$  is the regularization strength coefficient, with a value of 0.001 in the proposed model. The model parameters undergo training by minimizing the loss function via the Adaptive Moment

Estimation (Adam) optimization algorithm.

## 4. Experimental validations and analysis

To assess the fault diagnosis performance of the proposed approach, experiments were carried out using the CWRU rolling bearing datasets.

### 4.1. Experimental configuration

The CWRU dataset is extensively utilized as a benchmark for rolling bearing fault diagnosis [15]. Its data acquisition testbed is depicted in Fig. 8. In this setup, a faulty bearing was mounted on the test motor, and acceleration sensors at both the motor drive end and fan end recorded vibration data under varying loads. The dataset encompasses four types of bearing conditions: normal, inner race fault (IF), ball fault (BF), and outer race fault (OF).

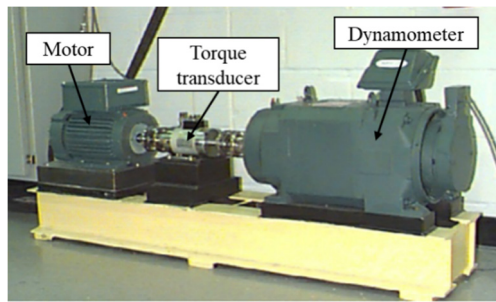


Fig. 8. The data acquisition testbed of the CWRU dataset

The experimental CWRU dataset comprised rolling bearing vibration signals under both normal and faulty conditions, with the data distribution summarized in Table 1. Specifically, six health states were examined at 1730, 1750, 1772, and 1797 RPM: normal state (label 0), inner race fault at 0.007 inches (label 1), inner race fault at 0.021 inches (label 2), ball fault at 0.021 inches (label 3), outer race fault at 0.007 inches (label 4), and outer race fault at 0.021 inches (label 5). Each sample contained 1024 sampling points. The input image samples for the SE-SDCTNet model were created through signal-to-image conversion, yielding  $32 \times 32$  images. A total of 2400 samples were utilized, with 300 training samples and 100 test samples for each type.

Table 1. Description of the CWRU's bearing dataset

Signal description	Fault diameter/ inch	Training sample	Test sample	Fault type	Label
Normal	–	300	100	Normal	0
IF	0.007	300	100	IF007	1
IF	0.021	300	100	IF021	2
BF	0.021	300	100	BF021	3
OF	0.007	300	100	OF007	4
OF	0.021	300	100	OF021	5

### 4.2. Ablation experiments on the CWRU dataset

To validate the effectiveness of each key component in the proposed SE-SDCTNet architecture, we conduct a comprehensive ablation study on the CWRU dataset under the four operating conditions (1730, 1750, 1772, and 1797 RPM). All variants share identical data splits, training procedures, and evaluation protocols. Specifically, models are optimized using the Adam optimizer with an initial learning rate of 0.0001. A learning rate decay strategy is applied, where the learning rate is multiplied by a factor of 0.96 every 5 epochs.

The ablation experiment results of SE-SDCTNet on the CWRU dataset (tested in noise-free

environments) are presented in Fig. 9: under these ideal conditions, all models (DenseNet, SE-DenseNet, SDCTNet, and SE-SDCTNet) achieve 100 % diagnostic accuracy, reflecting strong fault classification performance across architectures. The mechanism behind this consistent high performance can be visualized via the t-distributed stochastic neighbor embedding (t-SNE) results of SE-SDCTNet under four operating conditions (1730/1750/1772/1797 RPM) in Fig. 10. In these plots, samples of different fault categories form distinct, non-overlapping clusters, this clear feature separation directly reflects SE-SDCTNet’s ability to capture unique fault characteristic patterns, which underpins the 100 % accuracy observed in Fig. 9. This alignment between high accuracy and well-separated feature clusters stems from ideal noise-free environments: bearing vibration signals carry unobscured fault-related spectral information, allowing both SE-SDCTNet and other baseline models to easily distinguish distinct fault patterns.

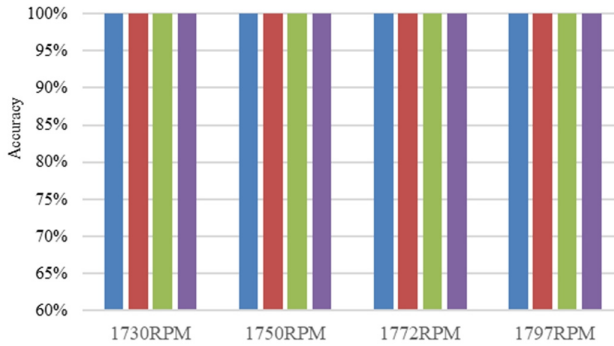


Fig. 9. Ablation experiment results of SE-SDCTNet on the CWRU dataset

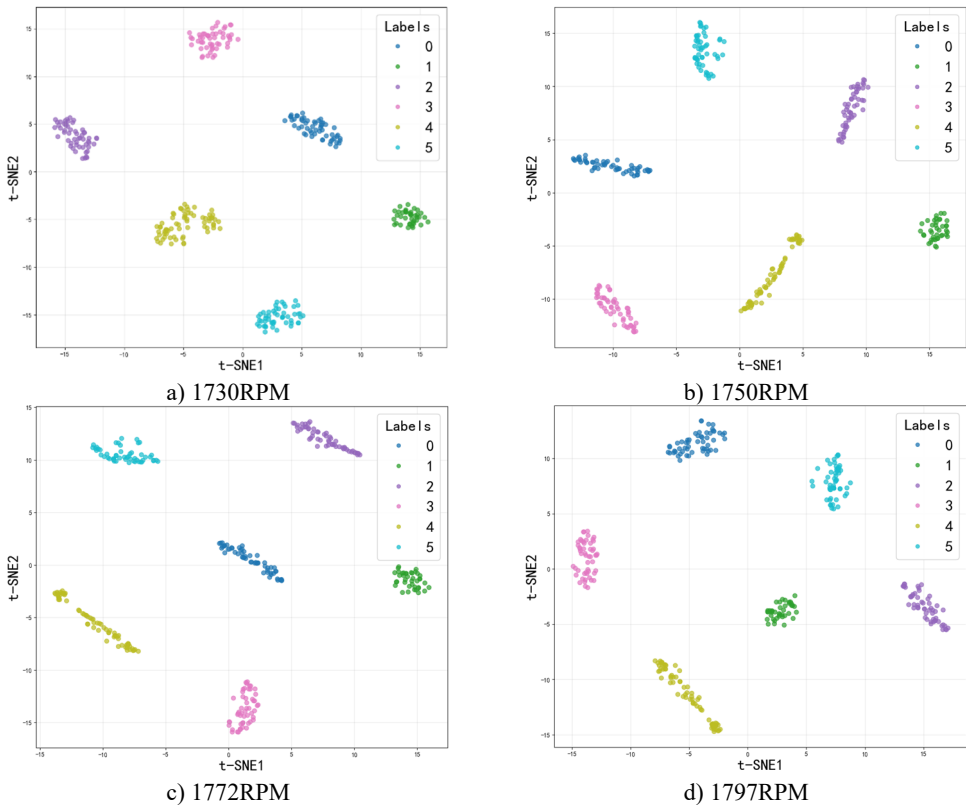


Fig. 10. The t-SNE visualization of the SE-SDCTNet under four operating conditions

However, real industrial scenarios involve significant background noise, which obscures fault features and degrades feature discriminability. To simulate this practical challenge, Gaussian noise was introduced into the test set’s vibration signals, with signal-to-noise ratio (SNR) quantifying noise intensity, this framework will enable systematic evaluation of whether SE-SDCTNet can maintain clustered, separable features (and high accuracy) under varying noise levels. The SNR was employed to quantify noise intensity, enabling a systematic assessment of each model’s robustness and its ability to preserve feature discriminability under varying levels of noise interference. The SNR is defined as [26]:

$$SNR = 10 \log_{10} \left( \frac{P_{signal}}{P_{noise}} \right), \quad (17)$$

where  $P_{signal}$  and  $P_{noise}$  denote the signal and noise powers, respectively.

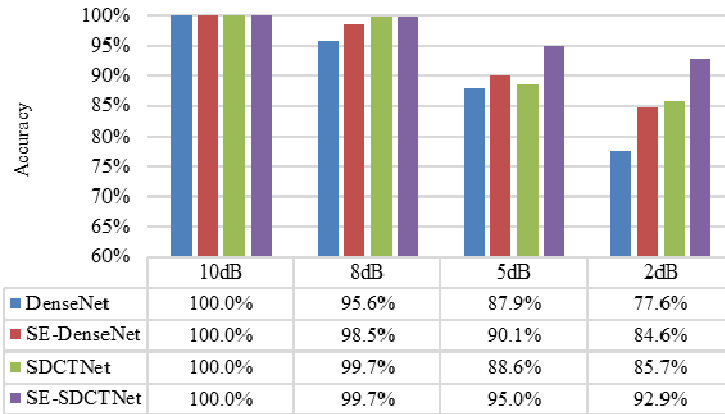


Fig. 11. Ablation experiment results of SE-SDCTNet with noise interference at 1797 RPM

Fig. 11 illustrates the diagnostic accuracy of various model variants under different noise levels (from 10 dB to 2 dB) when classifying six types of bearing faults, with 100 test samples per category. The results demonstrate a clear degradation in performance as noise intensity increases for all models. Specifically, DenseNet achieves high accuracy (100 %) under low-noise conditions but drops sharply to 77.6 % under 2 dB noise – a decline of 22.4 percentage points, indicating limited robustness in noisy environments. Incorporating SE blocks into DenseNet (SE-DenseNet) improves noise robustness, maintaining 84.6 % accuracy at 2 dB noise, representing only a 15.4 percentage point drop. This improvement arises from the SE module’s ability to adaptively recalibrate channel-wise feature responses, enhancing the representation of fault-relevant information while suppressing irrelevant activations. Replacing conventional dense blocks with SDCT blocks (SDCTNet) further enhances performance, achieving 85.7 % accuracy under 2 dB noise (a 14.3 percentage point drop), demonstrating the effectiveness of sparsity constraints in reducing redundant neuron activation and filtering out noise interference. When both SE and SDCT blocks are combined in the proposed SE-SDCTNet, the model achieves the highest overall accuracy across all noise levels. Notably, it maintains 92.9 % accuracy even under severe 2 dB noise, with only a 7.1 percentage point reduction from the noise-free case. This superior robustness stems from the synergistic interaction between the two mechanisms: the SDCT block enforces sparse representations to suppress noise, while the SE block selectively amplifies salient features through channel attention. Together, they enable SE-SDCTNet to achieve an optimal balance between lightweight design and strong noise resilience, significantly outperforming individual components and baseline architectures.

### 4.3. Comparison of different diagnostic models on the CWRU dataset

For comparison, seven well-established deep learning baselines including LeNet5 [24], LiNet [19], Swin Transformer [25], ResNet18 [26], SENet18 [16], DenseNet [27], and MobileNetV2 [18] were also adopted for bearing fault diagnosis on the CWRU dataset. These networks span both classical CNNs and modern architectures about attention for bearing fault diagnosis. LeNet5 is one of the earliest and most representative lightweight convolutional networks. It employs 5×5 convolutions followed by 2×2 max-pooling. LiNet is a specialized lightweight network explicitly tailored for mechanical fault diagnosis. It is realized by one convolutional block with 3×1 convolution kernels, two light modules consisting of uneven-sized convolution blocks, residual structures, and one convolutional block with 1×1 convolution kernels. The Swin Transformer splits input images into patches for embedding, processes them via 4 Swin Transformer blocks, then outputs classifications through layer norm, global pooling, and a dense layer. The ResNet18 contains one 3×3 convolutional stem, eight basic residual blocks, GAP, and an FC classifier. The SENet18 augments ResNet18 by inserting SE blocks into every residual block. The DenseNet comprises an initial 3×3 convolution, 58 densely connected blocks, three transition layers, GAP, and an FC layer. MobileNetV2 is a highly efficient, lightweight convolutional architecture built upon inverted residuals and linear bottlenecks. Unlike MobileNetV1, which uses depthwise separable convolutions throughout, MobileNetV2 introduces an expansion layer before the depthwise convolution to increase representational capacity, followed by a linear projection layer that preserves information flow in the low-dimensional space.

**Table 2.** Parameters of the comparative models

Parameters	Batch size	Learning rate	Optimizer	Decay rate	Decay steps	Parameters / M
SE-SDCTNet	16	0.0001	ADAM	0.96	5	0.32
LeNet5	16	0.0001	ADAM	0.96	5	0.06
LiNet	16	0.0001	ADAM	0.96	5	0.02
Swin Transformer	16	0.0001	ADAM	0.96	5	0.80
ResNet18	16	0.0001	ADAM	0.96	5	11.17
SENet18	16	0.0001	ADAM	0.96	5	11.31
DenseNet	16	0.0001	ADAM	0.96	5	7.05
MobileNetV2	16	0.0001	ADAM	0.96	5	2.19

All subsequent experiments used identical data splits, training schedules, and evaluation protocols. The models tested include LeNet5, LiNet, Swin Transformer, ResNet18, SENet18, DenseNet, and MobileNetV2. These models were trained using the Adam optimizer. Decay steps refer to the number of iterations after one decay and decay rate refers to the learning rate after one decay. Parameters refers to the number of model parameters, in millions.

Fig. 12 compares the diagnostic accuracy of multiple models, including the proposed SE-SDCTNet, across a range of noise intensities (from noise-free to 2 dB) on six rolling bearing fault types, with 100 test samples per category. At 1730 RPM, basic lightweight models such as LeNet5 and LiNet achieve high accuracy (99.1 % and 99.7 %, respectively) under noise-free conditions, confirming their effectiveness in ideal, low-interference scenarios. However, their performance degrades sharply as noise increases: under the harsh 2 dB condition, LeNet5’s accuracy drops sharply to 40.7 % (a decline of 58.4 percentage points), and LiNet falls to 59.6 % (a 40.1-point reduction). This severe sensitivity to noise renders them unsuitable for real industrial environments, where noise is inevitable. Mainstream models with higher parameter counts, such as ResNet18, SENet18, and DenseNet, exhibit stronger noise robustness. For instance, ResNet18 maintains 78.3 % accuracy under 2 dB noise. Nevertheless, their considerable model complexity hinders deployment on resource-constrained edge devices, a key practical requirement for industrial bearing diagnosis. In contrast, SE-SDCTNet achieves 100 % accuracy in both noise-free and 10 dB noise at 1730 RPM, and retains 93.1% under severe 2 dB noise – a decrease of only 6.9 percentage points. This trend is consistent at 1797 RPM, where SE-SDCTNet maintains

leading accuracy across all noise levels, outperforming both lightweight and large-parameter counterparts. This robustness stems from the synergistic design of SE-SDCTNet’s core components: the SDCT block suppresses noise through sparsity constraints, limiting activation of redundant neurons; the SE block adaptively recalibrates channel-wise feature responses, preserving critical fault signatures even when sparse filtering removes less relevant channels; and dilated convolutions expand the receptive field without added computation, retaining global fault characteristics. Together, these mechanisms allow SE-SDCTNet to achieve an optimal balance – surpassing conventional lightweight models in noisy settings through superior robustness, while exceeding large-parameter networks in deployment efficiency due to its compact architecture.

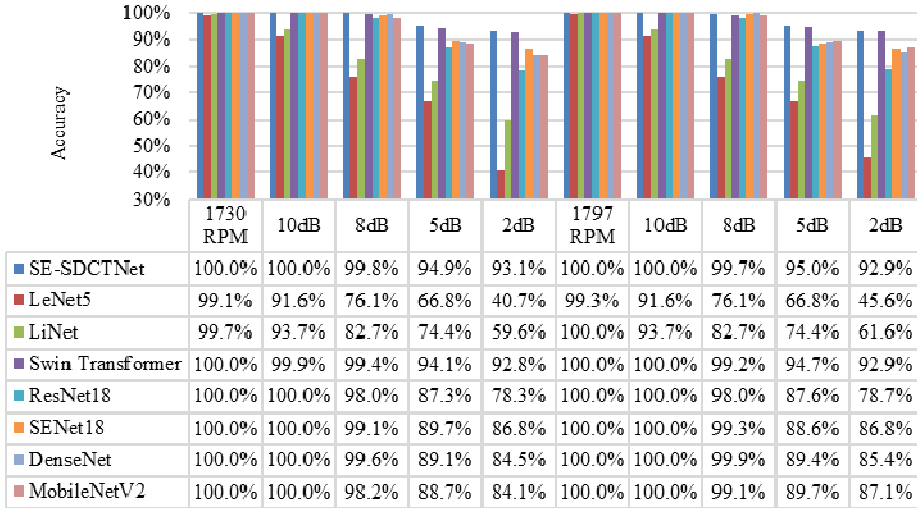


Fig. 12. Composed experiment results of SE-SDCTNet on the CWRU dataset

To further analyze the experimental results, additional evaluation metrics (precision, recall, and F1-score) were used. Table 3 presents the precision, recall, and F1-score of the SE-SDCTNet model when subjected to a noise level of 2 dB. These metrics serve to assess the overall diagnostic performance of the SE-SDCTNet model in the presence of noise interference and are defined as follows:

$$P = \frac{TP}{TP + FP'} \tag{18}$$

$$R = \frac{TP}{TP + FN'} \tag{19}$$

$$F = \frac{2TP}{2TP + FP + FN'} \tag{20}$$

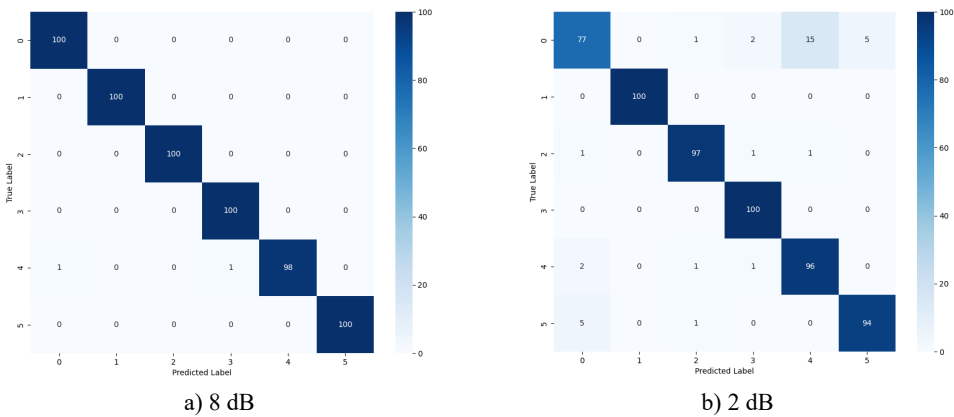
where TP denotes the count of positive samples accurately predicted as positive; FP represents the number of negative samples mistakenly predicted as positive; FN signifies the positive samples incorrectly predicted as negative; and TN indicates the negative samples correctly predicted as negative. Additionally, P, R, and F stand for precision, recall, and the F1-score, respectively.

The classification results indicate that the model performs well overall in identifying various samples, with the Precision, Recall, and F1-score for different fault types and the normal state all remaining above 0.87. Among them, the fault type BF007 corresponding to label 1 achieves the optimal recognition performance, with Precision, Recall, and F1-score all reaching 1.00, realizing perfect classification. The fault types IF021 (label 2) and BF021 (label 3) also perform excellently, with F1-scores of 0.97 and 0.94 respectively, demonstrating that the model has high accuracy and completeness in identifying these two types of faults. In comparison, the OF007 (label 4) and

normal state (label 0) show slightly weaker performance, with F1-scores of 0.88 and 0.89 respectively, but they are still at a good level. Overall, the model can effectively distinguish between different fault types and the normal state.

**Table 3.** Performance metrics of the SE-SDCTNet model at SNR = 2 dB

Label	Fault type	P	R	F
0	Normal	0.90	0.89	0.89
1	IF007	1.00	1.00	1.00
2	IF021	1.00	0.95	0.97
3	BF021	0.91	0.97	0.94
4	OF007	0.90	0.87	0.88
5	OF021	0.89	0.91	0.90



**Fig. 13.** The confusion matrices of the SE-SDCTNet in 8dB and 2dB

To further visualize the fine-grained classification performance of SE-SDCTNet at 1730 RPM, the confusion matrices of SE-SDCTNet under 8 dB and 2 dB noise conditions are presented in Fig. 13, with fault labels corresponding to the definitions in Table 1 (0: Normal, 1: IF007, 2: IF021, 3: BF021, 4: OF007, 5: OF021). At 8 dB SNR (Fig. 13(a)), the model achieves near-perfect classification with an overall accuracy of 99.7 %. Most fault types (e.g., IF007, BF021, OF007, OF021) are correctly identified across all 100 samples. The only errors are two OF007 (Label 4) samples misclassified (one as Normal and one as BF021), demonstrating the model’s strong discriminative power in moderate-noise environments. Under the harsher 2 dB condition (Fig. 13(b)), the model maintains a high overall accuracy. Misclassifications are primarily concentrated between specific pairs: 15 Normal samples are misclassified as OF007, 5 OF021 sample is confused with Normal, and 3 IF021 samples are scattered among Normal, BF021, and OF007. This error pattern aligns with the inherent subtlety of the corresponding fault signatures. Normal samples, lacking fault-specific components, are prone to being mislabeled as low-amplitude faults (e.g., BF021) under high noise. Similarly, OF021 features are weaker than those of larger-magnitude faults, leading to confusion with the Normal state. Notably, the confusion matrix retains strong diagonal dominance, with 100 % correct identification for faults like IF007 and BF021. This indicates that SE-SDCTNet possesses reliable fault discrimination capability even in challenging noise conditions.

### 5. Conclusions

To enhance the noise robustness and feature extraction capabilities of deep learning models in rolling bearing fault diagnosis, this paper proposes an improved lightweight network architecture named SE-SDCTNet based on stacked sparse dense compact thresholding blocks, SE blocks, and dilated convolutions. This design balances lightweight properties and performance: dense

connections boost feature reuse efficiency, sparsity constraints mitigate overfitting to strengthen noise suppression, SE blocks adaptively recalibrate feature responses, and dilated convolutions expand the receptive field without additional computation to preserve global fault characteristics. Experimental results on the benchmark bearing fault dataset validate the model's superiority. Under different rotational speeds (1730 RPM and 1797 RPM) and noise levels (10 dB to 2 dB), the proposed SE-SDCTNet outperforms representative models (LeNet5, LiNet, Swin Transformer, etc.) in diagnostic accuracy and noise robustness. In high-noise scenarios (2 dB), SE-SDCTNet maintains 92.9 %-93.1 % accuracy, while lightweight counterparts like LeNet5 only reach 40.7 %-45.6 % (a gap of over 47 %), and even advanced models (e.g., Swin Transformer) achieve comparable but not superior performance (92.8 % at 1730 RPM 2 dB). Across all noise conditions (10dB to 5dB), SE-SDCTNet consistently achieves at least 94.9 % accuracy (e.g., 99.8 % at 8dB, 1730 RPM) surpassing heavier models (ResNet18, DenseNet) while using only 0.32 million parameters. These results confirm that this lightweight densely connected CNN significantly improves noise robustness and global information extraction for rolling bearing fault diagnosis.

However, the model relies on sufficient labeled samples for effective training, limiting applicability in data-scarce scenarios. Future work will explore GAN-based data augmentation: this will expand the dataset and enhance generalization, providing a more robust solution for rolling bearing fault diagnosis when labeled data is scarce.

## Acknowledgements

This work was supported by the Key Technologies Research Grant for Fault Diagnosis of New Energy Vehicle Drive Motors with Small Samples (Grant No. 2025kj035) and the Intelligent Control and Information Processing Team Grant (Grant No. 25kytdzd003).

## Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Author contributions

Yanqi Wang: writing-review and editing, writing-original draft, validation, methodology, investigation, formal analysis, data curation and conceptualization. Songlin Zhang: project administration and funding acquisition. Rumeng Ding: conception and design. Cheng Luo: draft manuscript preparation. Letian Zhong: analysis and interpretation of results.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- [1] Y. Li, Y. Wang, X. Zhao, and Z. Chen, "A deep reinforcement learning-based intelligent fault diagnosis framework for rolling bearings under imbalanced datasets," *Control Engineering Practice*, Vol. 145, p. 105845, Apr. 2024, <https://doi.org/10.1016/j.conengprac.2024.105845>
- [2] H. Liu, R. Ma, D. Li, L. Yan, and Z. Ma, "Machinery fault diagnosis based on deep learning for time series analysis and knowledge graphs," *Journal of Signal Processing Systems*, Vol. 93, No. 12, pp. 1433–1455, Nov. 2021, <https://doi.org/10.1007/s11265-021-01718-3>
- [3] Z. Li, S. Tong, and Z. Wang, "A method for quality fault diagnosis based on Bayesian network D-cut set," *Journal of Physics: Conference Series*, Vol. 2855, No. 1, p. 012002, Sep. 2024, <https://doi.org/10.1088/1742-6596/2855/1/012002>

- [4] Y. Dong, Y. Li, H. Zheng, R. Wang, and M. Xu, "A new dynamic model and transfer learning based intelligent fault diagnosis framework for rolling element bearings race faults: Solving the small sample problem," *ISA Transactions*, Vol. 121, pp. 327–348, Feb. 2022, <https://doi.org/10.1016/j.isatra.2021.03.042>
- [5] X. Zhao et al., "Intelligent fault diagnosis of gearbox under variable working conditions with adaptive intra-class and inter-class convolutional neural network," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 34, No. 9, pp. 6339–6353, Sep. 2023, <https://doi.org/10.1109/tnnls.2021.3135877>
- [6] X. Zhang, Y. Ma, Z. Pan, and G. Wang, "A novel stochastic resonance based deep residual network for fault diagnosis of rolling bearing system," *ISA Transactions*, Vol. 148, pp. 279–284, May 2024, <https://doi.org/10.1016/j.isatra.2024.03.020>
- [7] A. Tian, Y. Zhang, C. Ma, H. Chen, W. Sheng, and S. Zhou, "Noise-robust machinery fault diagnosis based on self-attention mechanism in wavelet domain," *Measurement*, Vol. 207, p. 112327, Feb. 2023, <https://doi.org/10.1016/j.measurement.2022.112327>
- [8] Z. Xu et al., "A novel health indicator for intelligent prediction of rolling bearing remaining useful life based on unsupervised learning model," *Computers and Industrial Engineering*, Vol. 176, p. 108999, Feb. 2023, <https://doi.org/10.1016/j.cie.2023.108999>
- [9] S. Liu, J. He, Z. Chen, D. Chen, and Y. Chen, "Discriminative stacked autoencoder: feature-integration boosting for bearing fault diagnosis," *IEEE Sensors Journal*, Vol. 23, No. 22, pp. 27549–27558, Nov. 2023, <https://doi.org/10.1109/jsen.2023.3317873>
- [10] G. Wang, Y. Zhang, F. Zhang, and Z. Wu, "An ensemble method with DenseNet and evidential reasoning rule for machinery fault diagnosis under imbalanced condition," *Measurement*, Vol. 214, p. 112806, Jun. 2023, <https://doi.org/10.1016/j.measurement.2023.112806>
- [11] Z. Ye and J. Yu, "Deep morphological convolutional network for feature learning of vibration signals and its applications to gearbox fault diagnosis," *Mechanical Systems and Signal Processing*, Vol. 161, p. 107984, Dec. 2021, <https://doi.org/10.1016/j.ymsp.2021.107984>
- [12] P. Lyu, K. Zhang, W. Yu, B. Wang, and C. Liu, "A novel RSG-based intelligent bearing fault diagnosis method for motors in high-noise industrial environment," *Advanced Engineering Informatics*, Vol. 52, p. 101564, Apr. 2022, <https://doi.org/10.1016/j.aei.2022.101564>
- [13] C. Zhen and M. Liu, "Fault diagnosis of wind turbine rolling bearing based on VMD-SG-ResNet," *Academic Journal of Science and Technology*, Vol. 5, No. 1, pp. 233–238, Mar. 2023, <https://doi.org/10.54097/ajst.v5i1.5639>
- [14] J. Yu and X. Zhou, "One-dimensional residual convolutional autoencoder based feature learning for gearbox fault diagnosis," *IEEE Transactions on Industrial Informatics*, Vol. 16, No. 10, pp. 6347–6358, Oct. 2020, <https://doi.org/10.1109/tii.2020.2966326>
- [15] X. Zhang, Y. Wang, S. Wei, Y. Zhou, and L. Jia, "Multi-scale deep residual shrinkage networks with a hybrid attention mechanism for rolling bearing fault diagnosis," *Journal of Instrumentation*, Vol. 19, No. 5, p. P05015, May 2024, <https://doi.org/10.1088/1748-0221/19/05/p05015>
- [16] Y. Liang, B. Li, and B. Jiao, "A deep learning method for motor fault diagnosis based on a capsule network with gate-structure dilated convolutions," *Neural Computing and Applications*, Vol. 33, No. 5, pp. 1401–1418, May 2020, <https://doi.org/10.1007/s00521-020-04999-0>
- [17] Z. Zhang, L. Chen, C. Zhang, H. Shi, and H. Li, "GMA-DRSNs: A novel fault diagnosis method with global multi-attention deep residual shrinkage networks," *Measurement*, Vol. 196, p. 111203, Jun. 2022, <https://doi.org/10.1016/j.measurement.2022.111203>
- [18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4510–4520, Jun. 2018, <https://doi.org/10.1109/cvpr.2018.00474>
- [19] T. Jin, C. Yan, C. Chen, Z. Yang, H. Tian, and S. Wang, "Light neural network with fewer parameters based on CNN for fault diagnosis of rotating machinery," *Measurement*, Vol. 181, p. 109639, Aug. 2021, <https://doi.org/10.1016/j.measurement.2021.109639>
- [20] Z. He, Q. Shu, Y. Wang, and J. Wen, "A ReLU-based hard-thresholding algorithm for non-negative sparse signal recovery," *Signal Processing*, Vol. 215, p. 109260, Feb. 2024, <https://doi.org/10.1016/j.sigpro.2023.109260>
- [21] P. Guo, T. Shan, Z. Jie, H. Wang, Y. Shen, and J. Wang, "Research on the method of stress wave signal generation for aviation bearing based on SAE-DeGAN," in *2024 IEEE International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, Vol. 55, pp. 82–87, Jul. 2024, <https://doi.org/10.1109/sdpc62810.2024.10707733>

- [22] H. Liang and X. Zhao, "Rolling bearing fault diagnosis based on one-dimensional dilated convolution network with residual connection," *IEEE Access*, Vol. 9, pp. 31078–31091, Jan. 2021, <https://doi.org/10.1109/access.2021.3059761>
- [23] P. Liang, C. Deng, J. Wu, and Z. Yang, "Intelligent fault diagnosis of rotating machinery via wavelet transform, generative adversarial nets and convolutional neural network," *Measurement*, Vol. 159, p. 107768, Jul. 2020, <https://doi.org/10.1016/j.measurement.2020.107768>
- [24] S. Li, G. Xie, W. Ji, X. Hei, and W. Chen, "Fault diagnosis of rolling bearing based on improved LeNet-5 CNN," in *2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS)*, pp. 117–122, Nov. 2020, <https://doi.org/10.1109/ddcls49620.2020.9275100>
- [25] Z. Liu et al., "Swin transformer: hierarchical vision transformer using shifted windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9992–10002, Oct. 2021, <https://doi.org/10.1109/iccv48922.2021.00986>
- [26] J. Niu, J. Pan, Z. Qin, F. Huang, and H. Qin, "Small-sample bearings fault diagnosis based on resnet18 with pre-trained and fine-tuned method," *Applied Sciences*, Vol. 14, No. 12, p. 5360, Jun. 2024, <https://doi.org/10.3390/app14125360>
- [27] J. Yu, X. Zhou, L. Lu, and Z. Zhao, "Multiscale dynamic fusion global sparse network for gearbox fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, Vol. 70, pp. 1–11, Jan. 2021, <https://doi.org/10.1109/tim.2021.3076855>



**Yanqi Wang** received the M.Eng. degree from Henan Polytechnic University, Jiaozuo, China, in 2024. Now he works at the Anhui Institute of Information Technology. His current research interests include control, deep learning and fault diagnosis.



**Songlin Zhang** received the M.Eng. degree in control engineering from Jiangnan University, Wuxi, China, in 2014. Now he is currently an Associate Professor at the Anhui Institute of Information Technology. His current research interests include intelligent perception and control.



**Rumeng Ding** received her M.Eng. degree from Anhui Polytechnic University, Wuhu, China, in 2023. She is currently affiliated with the School of Electrical Engineering at Anhui Institute of Information Technology. Her current research interests include cluster analysis, image processing, and pattern recognition.



**Cheng Luo** received the M.Eng. degree from Nanchang Hangkong University Nanchang, China, in 2023. Now he works at the Anhui Institute of Information Technology. His current research interests include control, dynamics and fault diagnosis.



**Letian Zhong** received the M.S. degree from Hefei University of Technology, Hefei, China, in 2023. Now he works at the Anhui Institute of Information Technology. His current research interests include control, virtual simulation, and semantic segmentation.