



A convolutional neural-network-based diagnostic framework for industrial bearing

Bowen Yu and Chunli Xie

College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China

Correspondence: Chunli Xie (xcl08@126.com)

Received: 1 October 2022 – Revised: 11 November 2023 – Accepted: 22 December 2023 – Published: 19 February 2024

Abstract. The problem of industrial bearing health monitoring and fault diagnosis has recently been a popular research topic. Extracting sufficient features from the input raw vibration signals and mapping them to the most likely fault labels is the essence of bearing fault diagnosis. This study proposes a novel framework for bearing defect diagnostics by merging dilated residual convolutional neural networks and attention mechanisms. In this framework, multiple parallel dilated convolutional networks can automatically learn rich fault features at each scale from vibration signals. Simultaneously, the attention approach boosts fault-related features and suppresses irrelevant ones, improving fault detection performance and generalization. According to the experimental results of two different bearing datasets, the framework achieves a higher accuracy and can accurately identify various types of faults.

1 Introduction

With the advent of the era of big industrial data, mechanical equipment is constantly developing toward complexity and intelligence. As a basic component of mechanical equipment, industrial bearings are also components with a high incidence of failure. Industrial bearing failure can directly lead to the deterioration of the operating condition of mechanical equipment and pose significant safety issues. Based on statistics, bearing failures account for 40%–70% of the electro-mechanical drive system, resulting in substantial losses (Lessmeier et al., 2016). Therefore, real-time and accurate diagnostics of industrial bearings are critical for ensuring smooth operation and extending the equipment's life.

The commonly used methods for industrial bearing diagnosis include oil pressure, infrared thermal imaging, vibroacoustic measurements, electric current, etc. (Thoppil et al., 2021). Since low-cost vibration sensors can conveniently collect a wide range of vibration fault information, vibration-signal-based diagnostic methods are most widely adopted in health condition monitoring. The fault signals of industrial bearings are non-smooth and contain a lot of background noise (Lin et al., 2004), and the defects are rarely single. It is a great challenge for the diagnosis model to extract the effective fault information from the complex vibration signal and

ensure classification accuracy. Bearing fault diagnosis models usually have two major parts: feature extraction and fault classification (Chen et al., 2021). Feature extraction refers to the extraction of representative fault-related information from the raw data based on the technicians' signal processing knowledge and practical engineering experience, usually divided into time and frequency domain information.

A single time-domain signal often cannot accurately express bearing fault information. It is common to transform it into the frequency domain or time–frequency domain, such as wavelet packet (Yen and Lin, 2000), envelope analysis (Tsao et al., 2012), and empirical mode decomposition distribution (Yu and Junsheng, 2006). Various machine learning algorithms are used as classifiers on the extracted fault features, such as support vector machines (SVMs) (Yang et al., 2007), random forest methods (Roy et al., 2020), and K nearest neighbor (KNN) (Tian et al., 2015).

Fault diagnosis models based on shallow machine learning techniques and manual feature extraction methods have shown excellent recognition accuracy. However, there are several clear drawbacks: (1) it needs to manually extract features from the original vibration signal based on experience and signal processing knowledge, which requires intensive computation. What's more, the effect of diagnosis mainly de-

depends on the quality of feature extraction. (2) Feature extraction and classification are two independent processes, and the unsynchronized extraction and classification cannot meet the requirements of real-time diagnosis in the production system. (3) The diagnostic capability of shallow machine learning is slightly insufficient in the face of massive and strongly dynamic data, making it difficult to adapt to complex working conditions (Jiao et al., 2020).

The rapid development of deep learning has brought a new approach to bearing fault diagnosis. As a common deep learning model, the convolutional neural network (CNN) has achieved remarkable results in object recognition, image processing, and audio classification (Khan et al., 2020). Due to its multi-layer network structure, CNN has a strong adaptive feature learning ability, does not require any complicated manual extraction process, and can automatically learn fault feature representations from raw signal data.

Zhu et al. (2019) transformed the original one-dimensional time-domain signal into a time–frequency map through a short-time Fourier transform and input it into a convolutional neural network to identify fault features. Wang et al. (2019) compared the classification performance of eight different time–frequency analysis methods on the AlexNet model.

Mechanical equipment conditions are complicated and varied throughout the operation, and bearing failures come in various shapes and locations. As a result, the signal contains multiple scales of characteristics. Jiang et al. (2018) introduced multi-scale coarse-grained layers in CNNs to capture different granularity features by smooth shifting. Peng et al. (2020) offered a multi-branch, multi-scale CNN for learning rich and complementary defect information from wheel set bearings. Multi-scale convolutional networks have more layers and are susceptible to degradation. Liu et al. (2019) incorporated residual learning into CNNs to improve model training and prevent performance deterioration. Surendran et al. (2022) utilized a residual multi-scale CNN model (inception-resnet v2) to extract high-level fault characteristics and optimize the parameters using the sailfish algorithm.

Motivated by the prior studies, we have developed a novel fault diagnostic framework, which incorporates multi-filter dilated CNNs, a residual convolutional neural network and attentional mechanisms. The framework enables automatic feature extraction and end-to-end detection for fault identification, enhancing CNN stability and generalization in complicated conditions. The following is a list of the paper's major contributions:

1. A multi-scale extraction module based on dilated CNNs is proposed. Dilated CNNs produce diverse receptive fields to capture different fault features by adjusting dilation rates.
2. We design the residual connection module to transfer feature information between different layers, enabling data from shallow levels to flow into deeper layers and reducing information loss during transmission. Addi-

tionally, a wide convolutional kernel is used to capture long-term dependencies and mitigate noise interference.

3. Multiple attention mechanisms are applied in the module. The attention module assigns different weights to captured fault features, thus enhancing representative features while suppressing irrelevant ones.
4. The diagnostic framework is proposed. It was validated under various scenarios with bearing vibration signals, and the effect of various dilation rates and reduction ratios on the model extraction capability was examined. It is experimentally demonstrated that the framework performs well in complex situations.

The following is the structure of this paper: Sect. 2 introduces the background knowledge of convolutional neural networks, Sect. 3 explains the diagnostic framework's components and its procedure, and Sect. 4 introduces the experimental datasets of Case Western Reserve University (CWRU) and Jiangnan University (JNU). Section 5 in this paper presents the experimental results under different tasks and their corresponding analyses. Conclusions are presented in Sect. 6.

2 Convolutional neural networks

Convolutional neural networks encompass three key concepts: sparse interaction, parameter sharing, and equi-variant representation (Goodfellow et al., 2016). As illustrated in Fig. 1, a typical convolutional neural network comprises convolutional, pooling, and fully connected layers.

2.1 Feature extraction

The convolutional layer in CNNs stands as a cornerstone of its architecture. In this layer, input data are meticulously scanned using convolutional kernels. These kernels, akin to filters, possess a strong ability to discern intricate patterns within the data. Through the convolutional operation, these filters capture hierarchical and abstract representations of the input. Activation functions in CNNs map the inputs of neurons to their respective outputs. This transformation commonly employs nonlinear operations, enabling the network to learn the complex nonlinear relationships within the data. Pooling layers play a crucial role in simplifying computations within neural networks, which condense input dimensions by synthesizing local regions of the feature map into a single outcome. A CNN systematically extracts discriminative features related to faults from vibration signals by executing convolution and pooling operations.

2.2 Classification

After the feature extraction phase, the acquired features, refined through layers of convolutions and pooling, flow into

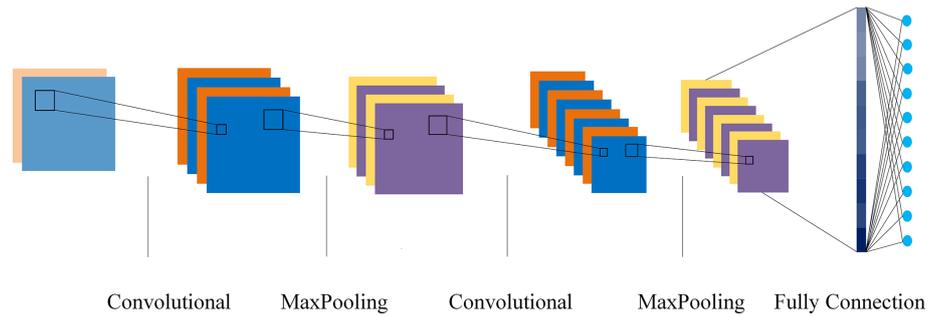


Figure 1. Convolutional neural network architecture.

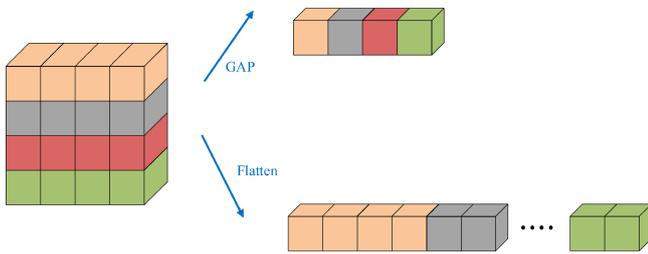


Figure 2. Comparison of fully connected layer and global average pooling.

densely connected layers, where intricate connections form, capturing abstract concepts. Each neuron in these fully connected layers acts as a learned feature detector, discerning complex combinations of visual elements. The final layer, often adorned with the softmax activation function, transforms these intricate features into probabilities, quantifying the network’s confidence in each potential class.

Traditional models often incorporate multiple fully connected layers to capture intricate dependencies within the data. However, as convolutional neural networks progress to depth and complexity, fully connected layers lead to a surge in parameters, which amplifies the risk of overfitting and puts forward higher requirements for the performance of diagnostic equipment. In response to the problem, Hinton et al. (2012) proposed the dropout method to randomly abandon the connection, reduce the co-adaptation between nodes, and empower the network to acquire robust features that generalize better to unseen data. To further improve the anti-fitting ability of the model and reduce the parameters in the training process, Lin et al. (2013) proposed a global average pooling method, which is different from the traditional fully connected layer by performing global average pooling on each feature map. Figure 2 compares the operational mechanisms of global average pooling and fully connected layers.

3 Methods

3.1 Multi-filter dilated convolutional neural network

The vibration signal gathered by the accelerometer is typically non-stationary, meaning that the signal’s frequency component fluctuates with time and has a significant degree of uncertainty. It comprises complex feature information of various timescales and presents typical multi-scale characteristics. Meanwhile, bearing faults come in various shapes and sizes, and different types of faults produce distinct characteristic frequencies. Due to these factors, traditional convolutional neural networks with a fixed filter do not extract enough information for accurate fault diagnosis.

We propose a multi-filter dilated convolution module to mine multi-scale information from the bearing vibration signals to perform the feature extraction work. The module’s structure is depicted in Fig. 3. The module uses four parallel dilated convolution structures with a filter w , sliding in the vibration signal for convolutions to obtain multiple feature maps (Wang and Ji, 2018). It can get several receptive fields in the vibration signal sequence, allowing each output of the local convolution stage to catch different scale features, the formula being defined as

$$o = \sum_{s=1}^S f[i + d \cdot s] w[i], \tag{1}$$

where d is the dilation rate and s denotes kernel size; the operation process of dilated convolution is depicted in Fig. 4a.

Given the diverse output sizes stemming from different dilation rates, we employ padding techniques to ensure uniform output lengths. Subsequently, feature maps from distinct levels are connected along the channel dimensions (Chen and Shi, 2021). Each structure is equipped with 64 filters and a kernel size of 5, so we can get the output in Eq. (2).

$$o = [o_1, o_2, \dots, o_{n-h+1}] \in R^{L \times C} \tag{2}$$

Because the feature maps formed by convolution are substantially different in recognizing bearing fault features, we employ attention mechanism approaches to learn discriminative

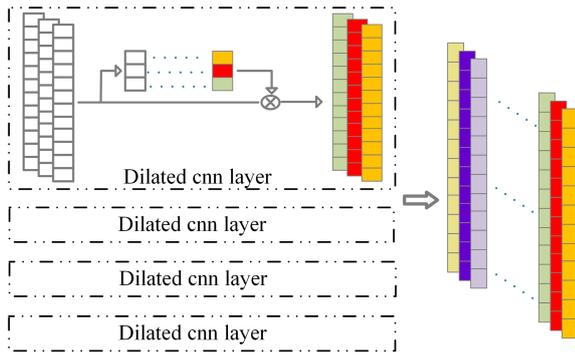


Figure 3. Multi-filter dilated convolutional neural network.

features and disregard valueless data. First, the global temporal information in the feature map of L length is compressed into a channel descriptor using the global average pooling layer (Hu et al., 2018).

$$z_c = \frac{1}{L} \sum_{i=1}^L o_c(i), \quad (3)$$

where z_c is channel-wise statistics, reflecting the global information of the c feature map.

To properly capture the correlation of the channels in the channel-wise statistics, the next step is to fuse the feature map information of each channel across the fully connection layer (Ye and Yu, 2021).

$$s = \sigma(F_2 \delta(F_1 z)), \quad (4)$$

where δ represents the ReLU activation function, F_1 and F_2 denote the fully connected layers, and σ is the sigmoid that compresses the dynamic range of the vector between $[0, 1]$. After that, channel-wise multiplication in Eq. (5) is performed to complete the rescaling of the original features in the channel dimension with the learned weights.

$$v \cdot s = [v_1 s_1, v_2 s_2, \dots, v_c s_c] \quad (5)$$

After extracting characteristics from each structure, they are combined across channels using concatenation.

3.2 Residual convolutional neural network

The residual network provides the shortcut connection approach (He et al., 2016) connecting earlier layers to later layers via shortcuts to allow the flow of information across distinct layers. Shortcut connections encompass identity and projection shortcuts, as illustrated in Eqs. (6)–(7).

$$y = F(x, W_i) + x, \quad (6)$$

$$y = F(x, W_i) + W_s x, \quad (7)$$

where y and x refer to input and output, F is residual mapping, and W_i indicates weight. The schematic diagram can be found in Fig. 4b.

We create a residual convolutional neural network that learns to integrate feature information from the 1D signal, which comprises convolutional layers with residual connections and a pooling layer. In designing our convolutional layers, we utilize a wider convolutional kernel of 32 for the initial filters to mitigate noise interference and capture global trend information more effectively (Liang and Zhao, 2021). Subsequently, the second layer employs filters with a kernel size of 7. The chosen number of filters for these convolutional layers are 128 and 256 to optimize the feature extraction capabilities of our model.

Due to the alteration in the number of channels across different layers, it is necessary to perform dimensional matching. This is accomplished via projection shortcuts as defined in Eq. (7), which employ a convolutional layer with a 1×1 kernel size to facilitate this transition. Finally, applying the ReLU activation function introduces nonlinearity into the model, enhancing its capacity to model complex functions.

3.3 Discriminative enhanced module

After the features are extracted by the multi-filter dilated and residual convolutional neural network, they are fused by element-wise addition. A discriminative enhanced module (DEM) is then applied to the extracted multi-level features to further deepen the model's ability to screen for critical features before entering the classification layer. The DEM module is shown in Fig. 5. It includes two branches: the spatial channel attention module and the channel attention module (Woo et al., 2018).

As for spatial information, we apply the convolutional operation to aggregate the compressed information from the channel dimensions (Roy et al., 2018).

The channel attention technique produces two feature maps with complementing global information through average and maximum pooling, respectively. Then, these two feature maps are then subjected to two separate convolution operations.

$$C_{\text{out}} = \sigma(W_2 \otimes (\delta(W_1 \otimes C_{\text{max}})) + W_2 \otimes (\delta(W_1 \otimes C_{\text{avg}}))), \quad (8)$$

where C_{max} and C_{avg} are the global maximum pooling and global average pooling feature, W_1 and W_2 are the weights, δ is ReLU activation, σ denotes sigmoid activation, and \otimes is convolution operation.

Input vectors can be optimized adaptively by DEM, to score the characteristics adaptively learned at various scales to enhance key information.

3.4 Fault detection framework

We present a framework for fault detection in Fig. 6. The framework adopts an end-to-end learning approach comprising four steps: signal acquisition and segmentation, model

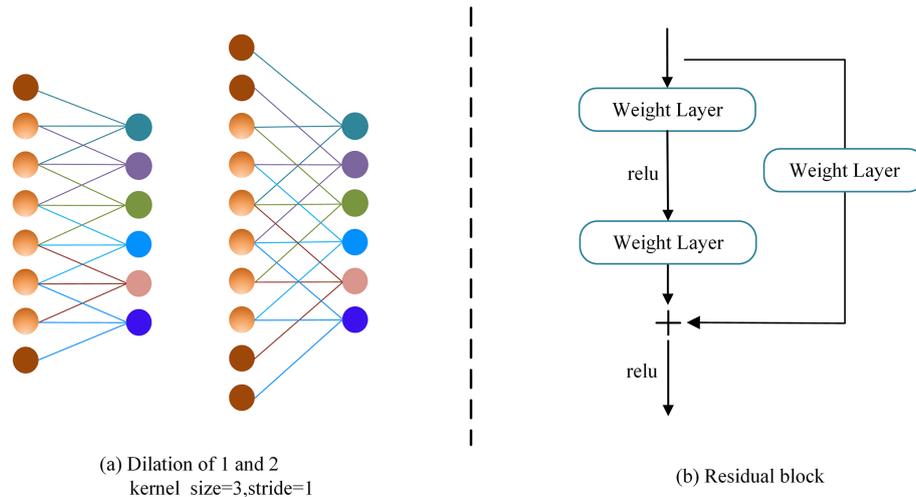


Figure 4. Dilated convolutional neural network and residual block.

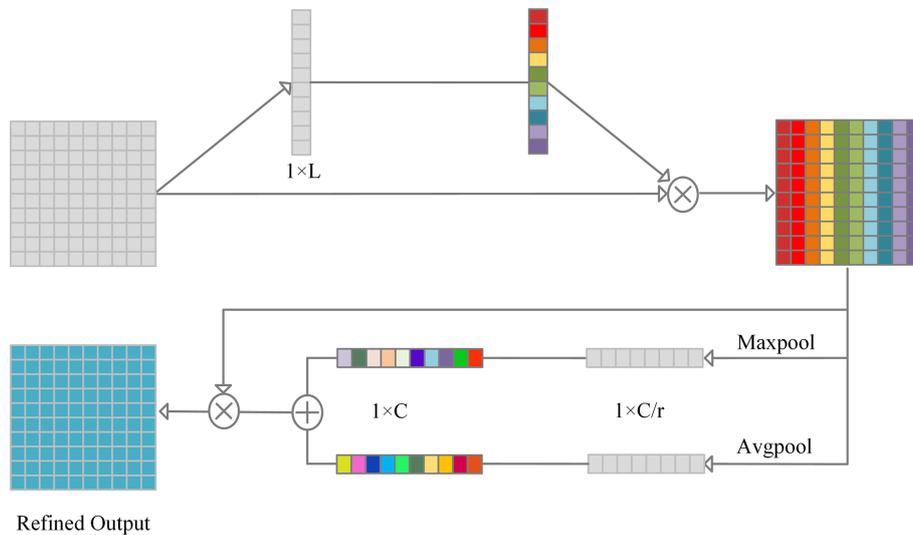


Figure 5. Discriminative enhanced module.

architecture development, model training workflow, and fault classification.

Step 1: signal acquisition and segmentation. The acquisition system collects the vibration data of mechanical components in varying states. Next, the original signal is divided into smaller units every 1024 data points, represented by

$$v = \{v_1, v_2, \dots, v_n\}. \tag{9}$$

Each segment of the vibration signal is tagged with one hot code during processing, so a separate bearing time-domain vibration dataset is represented as

$$\{(v_1, F_1), (v_1, F_1), \dots, (v_n, F_n)\}. \tag{10}$$

If the number of data obtained are insufficient, data augmentation techniques are used to increase the sample size; otherwise, they are not necessary.

Step 2: model architecture development. Fault characteristics are extracted using multi-filter dilated and residual convolutional neural networks. The discriminative enhanced module picks characteristics from the extracted information that enhance the discriminating ones. Finally, we use the global average pooling layer to generate feature vectors and transfer them into softmax to output several fault sorts.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \tag{11}$$

Step 3: model training workflow. We feed the model our training data and then iteratively move forward through each model layer to get the prediction. According to the loss function, the loss between the prediction and the target is determined. The error is then back-propagated while modifying the training parameters to minimize the difference.

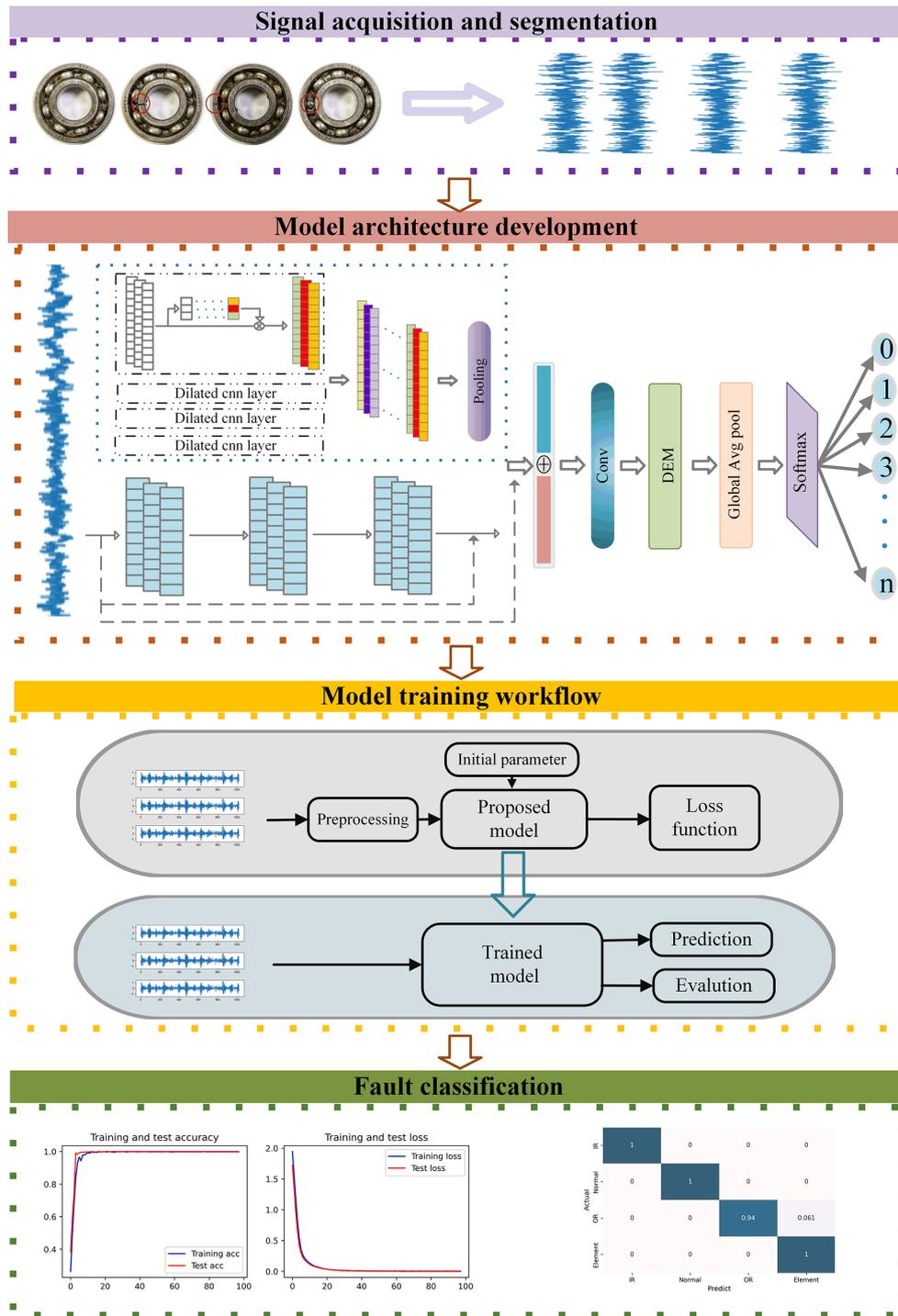


Figure 6. Fault detection framework.

Step 4: fault classification. Input test data to the trained model. The detection model returns the fault category corresponding to the input signal.

4 Data description

4.1 CWRU dataset

The rolling bearing dataset from Case Western Reserve University (CWRU; <http://csegroups.case.edu/bearingdatacenter>, last access: September 2022) has been

Table 1. Description of CWRU dataset.

Load (hp, horsepower)			Fault location	Defect size (inch)
A	B	C		
1	2	3	Normal	–
1	2	3	Inner race	0.007/0.014/0.021
1	2	3	Outer race	0.007/0.014/0.021
1	2	3	Ball	0.007/0.014/0.021

Table 2. Description of JNU dataset.

Speed (rpm)	Fault location	Defect size (mm)
800	Normal	–
800	Inner race	0.3×0.25
800	Outer race	0.3×0.25
800	Roller element	0.5×0.15

used extensively and has been considered a benchmark (Smith and Randall, 2015) in recent years.

The vibration signals were collected from an accelerometer mounted on the drive end (DE) with a sampling frequency of 12 kHz. Three locations of failure were considered: inner-ring failure, outer-ring failure, and ball failure. Each position had three different fault diameters: 0.007, 0.014, and 0.021 in. Consequently, there were three fault types times three different fault diameters, along with normal operating conditions, totaling 10 fault types.

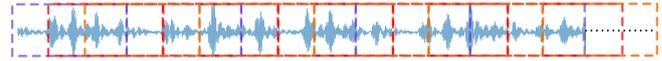
Datasets A, B, and C encompass vibration data from SKF 6205 bearings in three conditions: 1772 rpm (1 hp), 1750 rpm (2 hp), and 1730 rpm (3 hp), where bearing failures are seeded through electro-discharge machining, as shown in Table 1.

4.2 JNU dataset

Two types of roller bearings (N205 and NU205) were used in Jiangnan University bearing datasets (Li et al., 2013). The sampling frequency is 50 kHz, and the sampling duration is 20 s. The vertical vibration signals of the bearings were measured in four states: normal, defective inner ring, defective outer ring, and defective roller element. The measurements were conducted independently using an accelerometer, amplified through a signal conditioner, and recorded. We utilize data obtained at 800 rpm; detailed information is outlined in Table 2.

4.3 Data augmentation

Training the model with a substantial volume of vibration signal data is critical to ensuring robust model fit. We have implemented an overlapping data augmentation technique

**Figure 7.** Overlapping data augmentation.

(Zhang et al., 2017) to address the limited number of samples, as illustrated in Fig. 7. This method significantly increases the available training data, enhancing our model's performance.

5 Experimental validation and results

In this section, we validate various scenarios and analyze the corresponding outcomes.

5.1 Experimental setting

The training set of A, B, and C comprises 2000 samples, with a separate validation set and test set, each containing 300 samples.

We choose cross-entropy loss as the loss function to measure performance, and its mathematical formula is as follows:

$$H(p, q) = - \sum_x p(x) \log q(x), \quad (12)$$

where $p(x)$ indicates the true probability distribution and $q(x)$ is the predicted probability distribution.

In order to accelerate the training speed while avoiding falling into local optimal points, this paper uses the Adam stochastic optimization algorithm for training, which can dynamically adjust the learning rate of different parameters by iterating the weights according to the training data. The dropout rate during training is 0.3.

The framework used for the experiments is TensorFlow 2.6.0, running on a computer with an Intel i5 11400 CPU and an RTX3060 12 GB GPU. To better train the model in the TensorFlow framework, callback functions, early stop, and exponential decay learning rate scheduler are utilized to ensure optimum generalization performance.

5.1.1 The effect of dilation rates

If the kernel size of the filter is ks and the dilation rate is d , then the equivalent convolution kernel size ks' is

$$ks' = ks + (ks - 1) \times (d - 1). \quad (13)$$

When the dilation rate exceeds 1, the receptive field of the convolution kernel can be enlarged based on Eq. (13). By configuring a set of dilation rates to establish multiple receptive fields, the module can capture signal features across a broad range of scales.

To find the optimal combination, we attempted to validate sets with various rates. The standard convolution combination is (1, 1, 1, 1), while the other combinations tested include (1, 2, 3, 4), (1, 2, 4, 8), (1, 3, 5, 7), and (1, 4, 6, 9).

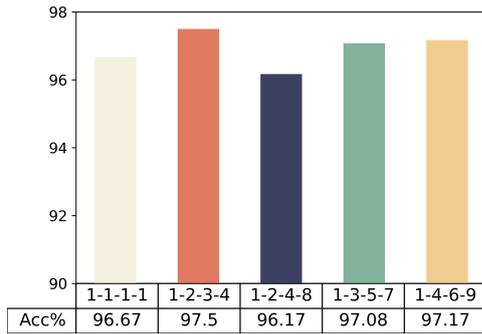


Figure 8. Results of different sets of dilation rates.

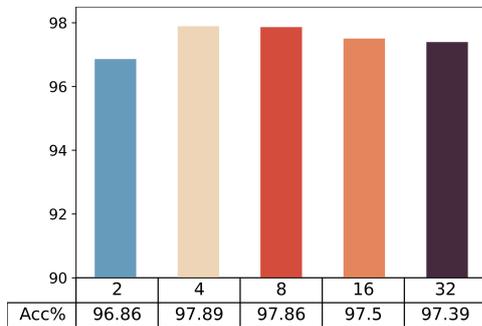


Figure 9. Results of different reduction ratios.

Validation was conducted using training set C, test set A, and test set B, while keeping all other variables constant and with an initial reduction ratio of 16. In the comparison of sets of dilation rates in Fig. 8, the combination of (1, 2, 3, 4) achieved the best results with an average diagnostic accuracy of 97.5 %, indicating it has the strongest feature extraction performance.

5.1.2 The effect of reduction ratios

The attention mechanism enhances the model’s ability to discern fault features by adjusting weights, and the reduction ratio stands out as a crucial parameter. This ratio reduces computational complexity by decreasing the number of input channels, aiding the model in efficiently capturing inner correlations. An appropriate reduction ratio helps regulate weight ranges, concentrating on the most pertinent elements in the input and enhancing the model’s expressiveness. Typically, the reduction ratio is a positive integer, with common values being 2, 4, 8, 16, etc. Comparative diagnostic results for the model with different reduction ratios are presented in Fig. 9. The optimal ratio for this diagnostic task is 4.

5.2 Performance under different workloads

Due to production requirements and unforeseen external environmental factors, machines often operate under varying conditions, including speed, load, and temperature. These

fluctuations impact the vibration frequency and amplitude of signals captured by accelerometers, leading to notable differences in signal characteristics. Consequently, these changes introduce interference, affecting the accuracy of classification. The variation in data distribution caused by these fluctuations significantly impacts the overall generalization performance of fault diagnosis models.

It requires good robustness of the diagnostic model to adapt to various changes in operating conditions. In the following section, the adaptability of the models, i.e., the proposed method, ConvNet, DRSN-CW (Zhao et al., 2019), ResNet, and TICNN (Zhang et al., 2018), is tested under different workloads. The validation method involves training the model on a specific workload and then applying it to test sets from a different workload, with results presented in Fig. 10.

TICNN, configured with a wide kernel size of 64, consistently achieves accuracy above 96 % for the initial four workloads. However, a substantial decline in diagnostic accuracy is observed for C–A and C–B, plummeting to 78.06 % and 86.72 %, respectively. In contrast to TICNN, ConvNet employs small kernels in each layer. Notably, ConvNet gains better performance exclusively in the A–B scenario. This observation demonstrates that a wide kernel is instrumental in extracting crucial vibrational features from the signal while suppressing spurious feature interference.

DRSN-CW embeds soft thresholding within the architecture, achieving an average diagnostic accuracy of 96.78 % for the initial four loads. Nevertheless, it incorporates a limited number of filters, potentially impeding the extraction of intricate feature representations and diminishing the capacity to distinguish various fault features effectively. DRSN-CW can just achieve more than 80 % diagnostic ability under C–A and C–B.

ResNet uses the same filter quantity as DRSN-CW in the B–C scenario but falls behind DRSN-CW by over 9 %, indicating that incorporating soft thresholding aids the DRSN-CW model in learning more robust feature representation. Our model achieves the best average diagnostic results in multi-load domain adaptation tasks, demonstrating that our model in this study can learn more discriminative features about these defects and enhance the model’s robustness in complex scenarios.

To intuitively observe the final classification results, we employ the t-SNE algorithm (Van Der Maaten and Hinton, 2008) to map abstract features under the A–C workload into a comprehensible geometric plane, as illustrated in Fig. 11. Different colors represent various bearing failures, visually demonstrating the degree of separation among failure characteristics. The model equipped with the DEM module exhibits higher separation when distinguishing between different classes of samples, underscoring its ability to prioritize fault-relevant information and enabling more accurate classification decisions.

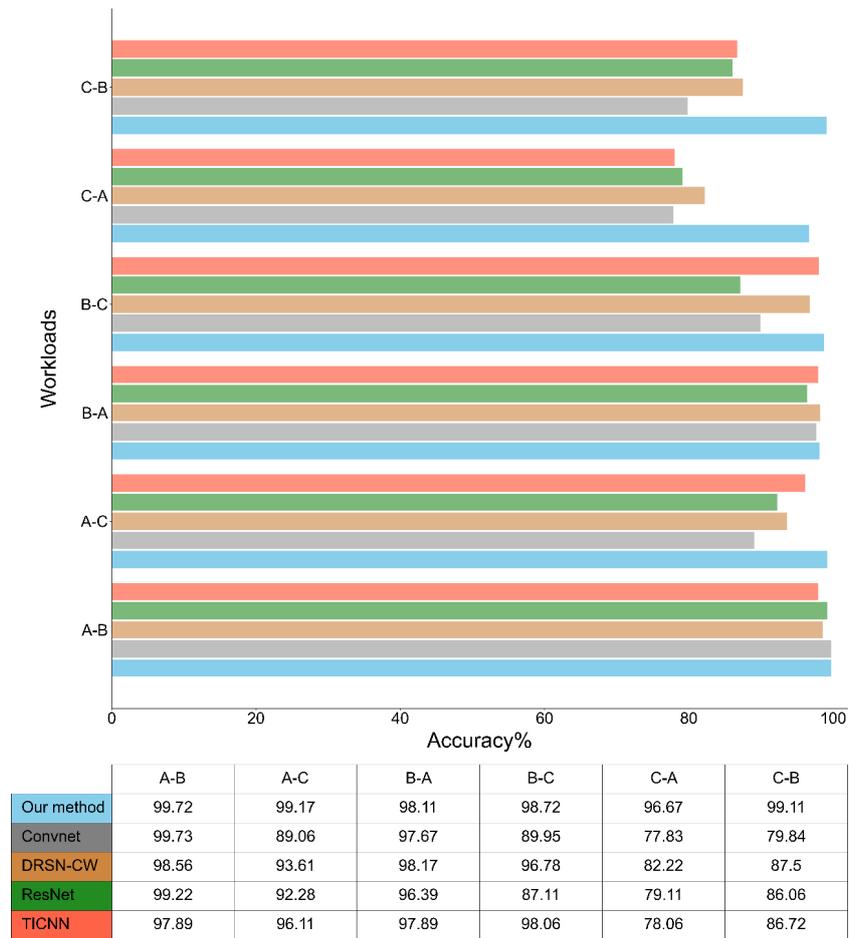


Figure 10. Results of different workloads.

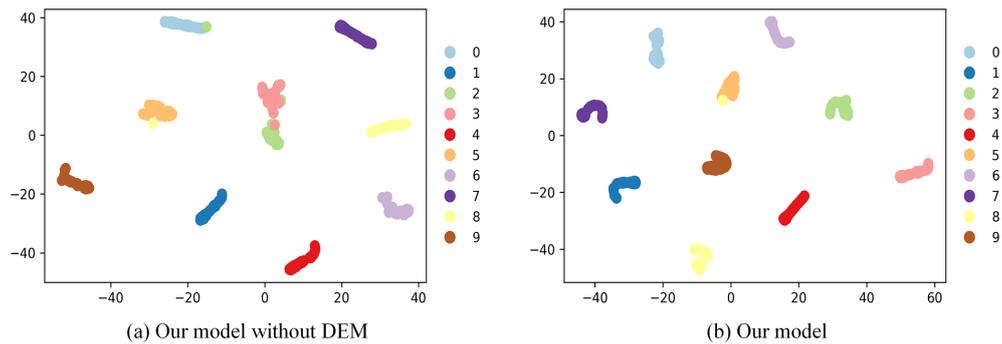


Figure 11. t-SNE visualization results.

5.3 Performance with the JNU dataset

To assess the generalizability of the proposed model, validation experiments were performed using the JNU dataset. In order to gain a deeper understanding of the model’s diagnostic capabilities, we introduced the confusion matrix to provide a detailed representation of the model’s performance within each fault category.

As illustrated in Fig. 12, the four confusion matrices represent the diagnostic results of our model, TICNN, DRSN-CW, and ResNet. In the case of outer-race fault detection, DRSN-CW exhibits a 12 % probability of erroneously classifying such faults as roller element faults. Regarding roller element damage diagnosis, this model encounters a 2 % chance of confusion with inner-race faults and an 8.2 % likelihood of incorrect classification as outer-race faults. TICNN accu-

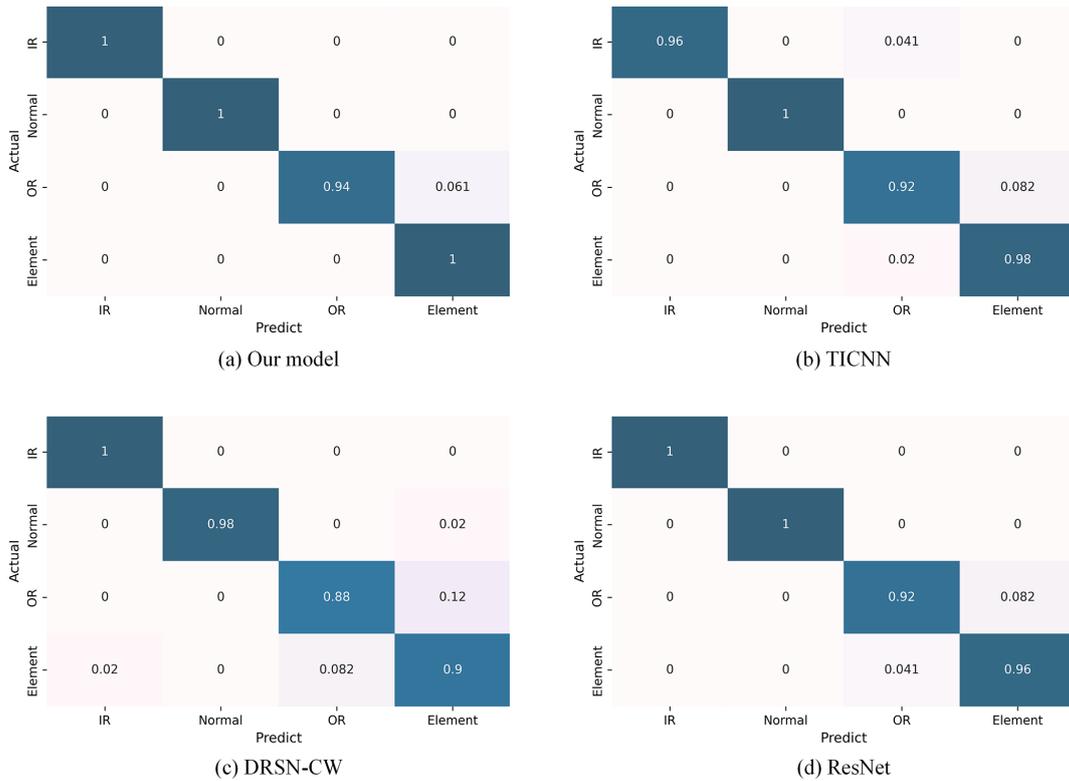


Figure 12. Comparison of confusion matrices.

rately predicts all samples in the healthy state; however, it trails our model by 4% in diagnosing inner-race faults and lags by a margin of 2% in the other two fault scenarios. ResNet shows commendable accuracy in diagnosing faults in inner-race and healthy-state samples, with a 100% correct prediction rate in these categories. Nonetheless, its diagnostic accuracy diminishes by 8% and 4% for the remaining fault types.

Our proposed model excels in superior diagnostic proficiency across a spectrum of bearing faults and degrees of damage, accurately detecting all instances of faults in three scenarios and attaining a 94% accuracy in identifying outer-race faults. The experimental outcomes demonstrate that our model achieves a convincing performance with the JNU dataset, substantiating its robust generalizability.

6 Conclusion

This study offers a novel framework based on convolutional neural networks for detecting industrial bearing faults. The framework incorporates dilated convolution, residual convolutional neural network, and attention mechanisms to gain rich fault feature representations and adaptively enhance key information. It is able to extract multi-scale features from nonlinear vibration signals to overcome the limitations of single-structure convolutional neural networks' weak flexi-

bility and extraction capacity. Multiple types of experimental validation are performed on bearing datasets. The experimental results show that the proposed model considerably exceeds traditional CNNs in feature learning and classification ability.

Data availability. The data in this study can be requested from the corresponding author.

Author contributions. BY conceptualized the work, decided on the methodology, and wrote the article. CX led the review and editing of the paper.

Competing interests. The contact author has declared that neither of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. The authors would like to express their thanks for the open source data from Case Western Reserve University and Jiangnan University.

Financial support. This research has been supported by the Natural Science Foundation of Heilongjiang Province (grant no. LH2021F002).

Review statement. This paper was edited by Jeong Hoon Ko and reviewed by three anonymous referees.

References

- Chen, W. and Shi, K.: Multi-scale Attention Convolutional Neural Network for time series classification, *Neural Networks*, 136, 126–140, <https://doi.org/10.1016/j.neunet.2021.01.001>, 2021.
- Chen, X., Zhang, B., and Gao, D.: Bearing fault diagnosis base on multi-scale CNN and LSTM model, *J. Intell. Manuf.*, 32, 971–987, <https://doi.org/10.1007/s10845-020-01600-2>, 2021.
- Goodfellow, I., Bengio, Y., and Courville, A.: *Deep learning*, MIT press, ISBN 9780262035613, 2016.
- He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, Nevada, USA, 26 June–1 July 2016, IEEE, 770–778, <https://doi.org/10.1109/cvpr.2016.90>, 2016.
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R.: Improving neural networks by preventing co-adaptation of feature detectors, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1207.0580>, 3 July 2012.
- Hu, J., Shen, L., and Sun, G.: Squeeze-and-excitation networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, Utah, USA, 18–22 June 2018, IEEE, 7132–7141, <https://doi.org/10.1109/cvpr.2018.00745>, 2018.
- Jiang, G., He, H., Yan, J., and Xie, P.: Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox, *IEEE T. Ind. Electron.*, 66, 3196–3207, 2018.
- Jiao, J., Zhao, M., Lin, J., and Liang, K.: A comprehensive review on convolutional neural network in machine fault diagnosis, *Neurocomputing*, 417, 36–63, <https://doi.org/10.1016/j.neucom.2020.07.088>, 2020.
- Khan, A., Sohail, A., Zahoora, U., and Qureshi, A. S.: A survey of the recent architectures of deep convolutional neural networks, *Artif. Intell. Rev.*, 53, 5455–5516, <https://doi.org/10.1007/s10462-020-09825-6>, 2020.
- Lessmeier, C., Kimotho, J. K., Zimmer, D., and Sextro, W.: Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification, *PHM Society European Conference*, Bilbao, Spain, 5–8 July 2016, 152–156, 2016.
- Li, K., Ping, X., Wang, H., Chen, P., and Cao, Y.: Sequential fuzzy diagnosis method for motor roller bearing in variable operating conditions based on vibration analysis, *Sensors*, 13, 8013–8041, <https://doi.org/10.3390/s130608013>, 2013.
- Liang, H. and Zhao, X.: Rolling bearing fault diagnosis based on one-dimensional dilated convolution network with residual connection, *IEEE Access*, 9, 31078–31091, <https://doi.org/10.1109/access.2021.3059761>, 2021.
- Lin, J., Zuo, M. J., and Fyfe, K. R.: Mechanical fault detection based on the wavelet de-noising technique, *J. Vib. Acoust.*, 126, 9–16, <https://doi.org/10.1115/1.1596552>, 2004.
- Lin, M., Chen, Q., and Yan, S.: Network in network, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1312.4400>, 16 December 2013.
- Liu, R., Wang, F., Yang, B., and Qin, S. J.: Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions, *IEEE T. Ind. Inform.*, 16, 3797–3806, <https://doi.org/10.1109/tii.2019.2941868>, 2019.
- Peng, D., Wang, H., Liu, Z., Zhang, W., Zuo, M. J., and Chen, J.: Multibranch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition, *IEEE T. Ind Inform.*, 16, 4949–4960, 2020.
- Roy, A. G., Navab, N., and Wachinger, C.: Recalibrating fully convolutional networks with spatial and channel “squeeze and excitation” blocks, *IEEE T. Med. Imaging*, 38, 540–549, 2018.
- Roy, S. S., Dey, S., and Chatterjee, S.: Autocorrelation aided random forest classifier-based bearing fault detection framework, *IEEE Sens. J.*, 20, 10792–10800, <https://doi.org/10.1109/jsen.2020.2995109>, 2020.
- Smith, W. A. and Randall, R. B.: Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study, *Mech. Syst. Signal Pr.*, 64, 100–131, <https://doi.org/10.1016/j.ymssp.2015.04.021>, 2015.
- Surendran, R., Khalaf, O. I., and Andres, C.: Deep learning based intelligent industrial fault diagnosis model, *CMC-Comput. Mater. Con.*, 70, 6323–6338, <https://doi.org/10.32604/cmc.2022.021716>, 2022.
- Thoppil, N. M., Vasu, V., and Rao, C.: Deep learning algorithms for machinery health prognostics using time-series data: a review, *J. Vib. Eng. Technol.*, 9, 1123–1145, <https://doi.org/10.1007/s42417-021-00286-x>, 2021.
- Tian, J., Morillo, C., Azarian, M. H., and Pecht, M.: Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K -nearest neighbor distance analysis, *IEEE T. Ind. Electron.*, 63, 1793–1803, <https://doi.org/10.1109/tie.2015.2509913>, 2015.
- Tsao, W.-C., Li, Y.-F., Du Le, D., and Pan, M.-C.: An insight concept to select appropriate IMFs for envelope analysis of bearing fault diagnosis, *Measurement*, 45, 1489–1498, <https://doi.org/10.1016/j.measurement.2012.02.030>, 2012.
- Van der Maaten, L. and Hinton, G.: Visualizing data using t-SNE, *J. Mach. Learn. Res.*, 9, 2579–2605, 2008.
- Wang, J., Mo, Z., Zhang, H., and Miao, Q.: A deep learning method for bearing fault diagnosis based on time-frequency image, *IEEE Access*, 7, 42373–42383, <https://doi.org/10.1109/access.2019.2907131>, 2019.
- Wang, Z. and Ji, S.: Smoothed dilated convolutions for improved dense prediction, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, United Kingdom, 19–23 August 2018, ACM, 2486–2495, <https://doi.org/10.1145/3219819.3219944>, 2018.
- Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S.: Cbam: Convolutional block attention module, *Proceedings of the European conference*

- on computer vision (ECCV), Munich, Germany, 8–14 September 2018, Springer, 3–19, https://doi.org/10.1007/978-3-030-01234-2_1, 2018.
- Yang, Y., Yu, D., and Cheng, J.: A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM, *Measurement*, 40, 943–950, <https://doi.org/10.1016/j.measurement.2006.10.010>, 2007.
- Ye, Z. and Yu, J.: AKSNet: A novel convolutional neural network with adaptive kernel width and sparse regularization for machinery fault diagnosis, *J. Manuf. Syst.*, 59, 467–480, <https://doi.org/10.1016/j.jmsy.2021.03.022>, 2021.
- Yen, G. G. and Lin, K.-C.: Wavelet packet feature extraction for vibration monitoring, *IEEE T. Ind. Electron.*, 47, 650–667, <https://doi.org/10.1109/ijcnn.1999.836202>, 2000.
- Yu, Y. and Junsheng, C.: A roller bearing fault diagnosis method based on EMD energy entropy and ANN, *J. Sound Vib.*, 294, 269–277, <https://doi.org/10.1016/j.jsv.2005.11.002>, 2006.
- Zhang, W., Peng, G., Li, C., Chen, Y., and Zhang, Z.: A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, *Sensors*, 17, 425, <https://doi.org/10.3390/s17020425>, 2017.
- Zhang, W., Li, C., Peng, G., Chen, Y., and Zhang, Z.: A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, *Mech. Syst. Signal Pr.*, 100, 439–453, <https://doi.org/10.1016/j.ymsp.2017.06.022>, 2018.
- Zhao, M., Zhong, S., Fu, X., Tang, B., and Pecht, M.: Deep residual shrinkage networks for fault diagnosis, *IEEE T. Ind. Inform.*, 16, 4681–4690, <https://doi.org/10.1109/tii.2019.2943898>, 2019.
- Zhu, Z., Peng, G., Chen, Y., and Gao, H.: A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis, *Neurocomputing*, 323, 62–75, <https://doi.org/10.1016/j.neucom.2018.09.050>, 2019.