Journal of Engineering, Project, and Production Management
2026, 16(3), 2026-269

Rolling Bearing Fault Detection Based on Deep Residual Capsule Network

Yuzhou Zhang
Undergraduate Student, College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin, 150040, China, E-mail: Yuzhouzhangzyz@outlook.com

Production Management
Received February 27, 2026; revised April 16, 2026; accepted April 18, 2026
Available online May 4, 2026

Abstract: Bearing failure can cause equipment to stop running, reduce production efficiency, and in severe cases, even lead to safety accidents. Traditional bearing fault detection methods often lack sufficient feature extraction, limited generalization, and poor anti-noise performance. In this sense, this paper proposes a bearing failure detection method based on a Deep Residual Capsule Network (DRCN). DRCN combines the advantages of the Deep Residual Network (ResNet) and the Capsule Network (CapsNet). ResNet, by leveraging its remaining structure, can automatically learn deep fault features and achieve high robustness, thereby avoiding the limitations of manual feature extraction and alleviating the problem of network degradation. CapsNet captures the spatial hierarchical relationship between features through a dynamic routing mechanism, effectively improving the deficiency of traditional convolutional neural networks in expressing feature direction and position information. By optimizing the network structure and parameters, DRCN enhances its adaptability to small data samples and improves the generalization performance of the model. To verify this method, the DRCN model was tested in a set of bearing failure data from Case Western Reserve University (CWRU) and Jiangnan University (JNU). The results show that the proposed model achieves an accuracy of 99.13% on the CWRU dataset and 98.78% on the JNU dataset. Compared with the traditional ResNet and CapsNet methods, the DRCN proposed in this paper has higher diagnostic accuracy.

Keywords: Deep residual capsule network**,** equipment, fault diagnosis, production, rolling bearing.

DOI 10.32738/JEPPM-2026-269

1. Introduction

Bearing damage can cause equipment to die and production to be impaired, so maintenance costs go up. Malfunction can also cause secondary component damage or safety accidents. Moreover, it will reduce the accuracy and efficiency of the device and affect the quality of the product, adversely affecting the company's economy and credibility. Studies show that relevant bearing problems account for about 30 to 40 percent of all rotation machine failures (Raj and Kumar, 2024). Proper and timely identification, therefore, will be essential to ensure that equipment functions stably, increases production efficiency, and controls maintenance costs.

Traditional diagnostic pad error techniques depend on expert experience with classic signal handling techniques. This technology includes an analysis of the landscape (Wu et al., 2022), such as feng, peaks, and square roots. Frequency analysis (Jain and Bhosle, 2021) includes methods such as the Fourier transform and inclusive mediation. Frequency analysis (Du et al., 2020) is often used in small, fluid waves and in the form of EMD. Nevertheless, such approaches exhibit significant drawbacks within practical industrial applications. Traditional methods rely on artificially designed features based on prior knowledge and are not sensitive enough to early weak faults or compound faults. For instance, time-domain indicators tend to fail in high background noise environments. The frequency-domain method assumes that the signal is stable, but under variable-speed conditions, the offset of the fault characteristic frequency will lead to diagnostic failure. Although time-frequency methods can handle non-stationary signals, the wavelet transform relies on empirical basis function selection, and empirical mode decomposition has problems such as mode aliasing and endpoint effects, all of which can lead to poor feature separability. Especially in the case of compound faults, the overlap degree of manually extracted features may reach 30% to 40%, which will significantly increase the misjudgment rate of the classifier (Zhang et al., 2021). Second, Traditional methods are usually based on the assumption of constant operating conditions. However, in actual industrial scenarios, loads and rotational speeds often fluctuate, which can cause feature drift and make it difficult for static diagnostic models to adapt to dynamic operating environments (Zhou et al., 2019). Take gas turbine bearings as an example. Variations

in speed can alter the frequency of failure, disable fixed threshold models, and significantly reduce the accuracy of diagnostics. Moreover, in complex mechanical structures, the vibration signal drops rapidly with partial transmission, making it difficult to effectively distinguish the broken signal from the background noise using conventional methods. They are calculated at high frequencies, are not sensitive to noise, are inefficient in diagnosis, and become loose with time. For example, traditional methods require average multi-rotation images to absorb noise. Under strong winds, the miscalculation of the diagnostic pad fan will be more than 30%.

By deepening my research, the field of problem diagnosis has changed. Structures such as Convolutional Neural Networks (CNN), Deep Belief Network (DBN), Deep Residual Contraction Network (DRSN), Long Short-Term Memory Network (LSTM), and Capsule Neural Network (CapsNet) provide powerful learning and performance. These models are capable of autonomously learning profound and highly distinctive fault characteristics directly from raw vibration data or its processed forms. These models significantly reduce their reliance on manual prior knowledge and outperform traditional methods on numerous public bearing datasets. In a study on rolling bearing fault diagnosis, Jiang and Wang (2025) employed Convolutional Neural Networks (CNN). This model demonstrates the effectiveness in extracting relevant features from vibration signals. The experimental results show that the accuracy rates of training, validation, and testing reach 100%, 99%, and 98.9%, respectively, indicating that this method has good applicability in actual diagnosis. However, the training loss curve of this method fluctuates, the training time is relatively long, and its efficiency on large-scale datasets still needs to be improved. Zheng et al. (2024) adopted DBN and combined it with phase space reconstruction to construct the acceleration-velocity matrix input and simultaneously applied the set Ensemble Empirical Mode Decomposition (EEMD). The fault recognition rate of this model on the training set exceeds 99%, the normal state recognition rate reaches 100%, and the fault recognition rates of the mild, moderate, and severe rolling elements of the outer ring reach 93.33%, 91.33%, and 95.33%, respectively. The training time of this method is comparable to that of other input methods, and it can achieve qualitative state assessment and damage classification. However, the diagnostic accuracy fluctuates under high-noise conditions, and its ability to identify extremely weak faults is still insufficient. Zhang et al. (2024) combined Deep Residual Contraction Network (DRSN) with vibration grayscale images and Dempster-Shafer evidence theory for the fault diagnosis of rolling bearings. The average prediction accuracy rates of faults in the inner ring, rolling elements, and outer ring of this method are 96.40%, 96.82% and 95.61%, respectively, and the overall accuracy rate is 96.28%. It shows good adaptability to different working conditions. However, the diagnostic accuracy will decline under high-noise signals, and the stability of identifying extremely weak faults still needs to be improved (Zhang, 2022). Chen et al. (2025) used LSTM for fault diagnosis of rolling bearings, achieving an accuracy rate of 93.41%, a recall rate of 93.02%, and an inference time of 18.95 seconds. Moreover, the relatively shallow network depth limits its feature extraction and classification capabilities. Despite the short training time, the accuracy rate is still relatively low. Wu (2023) proposed a one-dimensional CapsNet model for fault diagnosis of rolling bearings. This model directly acts on the time-domain vibration signal and can take the original signal as input without preprocessing, thereby avoiding feature loss. The model achieves end-to-end classification through the capsule module, and its recognition accuracy on the CWRU bearing dataset reaches over 98%, verifying its diagnostic effectiveness. However, the stability of feature extraction under strong noise is still insufficient, and there is still room for improvement in diagnostic accuracy when dealing with concurrent multiple faults.

Nevertheless, existing deep-learning-based bearing fault diagnosis methods still face several key challenges. First, the lack of spatial hierarchical representation. Mainstream CNNs can extract image-like features but loses spatial position and geometric information due to pooling operations, making it difficult to effectively model the specific spatial distribution patterns of bearing faults in time-frequency representations. Second, training and representation bottlenecks in deep networks. Although ResNet alleviates gradient vanishing and explosion issues through skip connections, its core convolutional units still have limitations in modeling spatial relationships (Cong and Zhou, 2023). Finally, insufficient robustness under small-sample and variable operating conditions. In real industrial applications, fault samples, especially those of severe faults, are often scarce. Existing models tend to overfit with small samples and perform poorly under changing operating conditions, such as load and speed, leading to significant degradation in cross-condition diagnosis scenarios.

To address the constraints of existing deep learning approaches for bearing fault detection, this work introduces a diagnostic technique utilizing a Deep Residual Capsule Network (DRCN). This framework combines the advantages of ResNet and CapsNet and aims to overcome the deficiencies of traditional and modern deep learning architectures. Residual networks help to mitigate the problems associated with gradient disappearance and model degradation by combining jump connections, and drive effects can be maintained even in deep networks. CapsNet uses vector capsules to represent entity properties and thus describes more precisely the spatial hierarchical relationships between the fault characteristics. Compared to traditional CNN methods, which rely only on the activation of scalar neurons, the DRCN proposed in this paper has a stronger representation capacity for the processing of complex structured data.

2. Datasets

2.1. Case Western Reserve University (CWRU) Bearing Dataset

The bearing fault dataset from CWRU, made publicly available by the university's Bearing Data Center, is a commonly adopted reference dataset in the field of rotating machinery fault diagnosis (Smith and Randall, 2015). The data collection is based on a precise experimental bench, whose core components include a 2-horsepower AC motor drive system, SKF/NTN series test bearings, and a high-precision accelerometer. The installation positions of the sensors cover the drive end, fan end, and base. The collected vibration signals are three-dimensional, and the sampling frequencies include 12 kHz and 48 kHz.

The dataset encompasses four operational states of bearings: normal, inner race fault, outer race fault, and rolling

element fault. Artificial flaws were created using Electrical Discharge Machining (EDM) to produce single-point defects. These defects had diameters of 0.1778mm, 0.3556mm, and 0.5334mm. For faults on the outer race, defect locations were varied at the 3, 6, and 12 o'clock positions to analyze the impact of the load zone. The experimental conditions cover a variety of combinations. The motor speeds include 1797, 1772, and 1750r/min, and the load grades are divided into 0, 1, 2, and 3 horsepower to simulate different actual working intensities.

The original data are stored in MATLAB (.mat) files, each containing drive-end vibration signals, fan-end signals, base signals, time series, and rotational speed. The filenames follow a standardized convention of "sampling frequency sensor position fault type fault size_load_index". The dataset contains over 200 independent samples, with typical subsets including 4-8 healthy samples, 23-40 inner race fault samples, 53-77 outer race fault samples, and 11-40 rolling element fault samples. Owing to its high parameter controllability, quantifiable defect severity, and comprehensive condition coverage, the CWRU dataset has become the gold standard for evaluating fault diagnosis algorithms and has been cited over ten thousand times in the past decade. Dataset access link: https://engineering.case.edu/bearingdatacenter/welcome.

2.2. Jiangnan University (JNU) Bearing Fault Dataset

This study uses the bearing fault dataset publicly released by JNU as the benchmark data. This dataset systematically records the vibration responses of deep groove ball bearings under various operating conditions (Jiang et al., 2022), covering four typical mechanical states: inner ring failure, health status, outer ring failure, and rolling element failure, encompassing the main failure modes of rotating parts machinery.

A data acquisition was conducted using a high-precision vibration acceleration sensor (model PCB 352C33, sensitivity 100mV/g, measurement range ±50g). A sampling frequency of 50 kHz was employed to capture the dynamic characteristics of the bearing, ensuring the integrity of high-frequency fault features. The sensor was mounted on the bearing housing using a magnetic base. For each operating condition, data were collected five times, with a 2-minute interval between acquisitions to mitigate the influence of random noise. The experimental setup included three rotational speed conditions, which include 600 r/min, 800 r/min, and 1000 r/min, to simulate load variations in real industrial scenarios.

The original continuous signal is divided into non-overlapping slices, each containing 1024 sampling points (corresponding to 20.48ms). This length selection meets two requirements: one is to ensure the coverage of the fault feature period under the sampling theorem, and the other is to match the input size requirements of the CNN. After segmentation, 488 independent samples were generated for each condition, and a total of 1,952 samples were obtained from the four states, forming a balanced dataset.

The labels are encoded as follows: Health status =0, inner circle failure =1, outer circle failure =2, rolling unit failure =3, thereby forming the label system required for supervised learning. The dataset partitioning follows a standard machine learning process, using an 8:2 hierarchical random partition to ensure consistent category distribution between the training set and the test set. The final dataset contains 1,562 training samples and 390 test samples. Dataset access link: https://github.com/ClarkGableWang/JNU-Bearing-Dataset.

It should be noted that no data augmentation techniques (such as rotation, scaling, or noise injection) were used in this study. The objective was to evaluate the model performance under realistic small-sample conditions, which more accurately reflects the scarcity of fault data in practical industrial environments. Data augmentation will be explored in future work.

Here, we present example samples from the JNU dataset, as shown in Fig. 1. Time-domain vibration signals of four health conditions in the JNU bearing fault dataset under the 800rpm operating condition: (a) normal condition, (b) inner race fault, (c) outer race fault, (d) rolling element fault. Each subplot shows a representative sample with a length of 1024 points.

3. Method

3.1. ResNet Development

In this study, two residual structures, ResNet-A and ResNet-B, were designed. These structures can prevent the performance degradation problem of the network as the depth increases and improve the performance of bearing fault diagnosis by fusing the depth features within the residual block (Hao et al., 2021). The specific structures of ResNet-A and ResNet-B are shown in Fig 2.

The ResNet-A architecture reduces data dimensions by employing larger strides in its convolutional layers, which decreases the total parameter count of the model. In addition, it enhances the learning advantage by using more distortion filters. In contrast, ResNet-B is combined with the specification layer and dropout layer of the specification, the latter of which provides the most powerful extraction function when multiple ResNet-B groups are piled up.

The function of the distortion layer is to extract features, which can amplify useful signal information and suppress irrelevant noise. The startup function depends on Rectified Linear Unit (ReLU), which enhances the scarcity of the network and is very useful for the training process. Leakage prevents excessive compatibility by randomly interfering with certain neurons, while the Batch Normalization (BN) layer helps alleviate the problem of gradient descent and accelerates the aggregation speed.

3.2. Construction of the Capsule Neural Network (CapsNet)

CapsNet can estimate an object's position through the capsule's vector direction and use a dynamic algorithm to determine the existence of expected entities. The advantage of this design is to classify signals of complex and delicate defects. (De

Sousa Ribeiro et al., 2024) The hierarchy of malfunction signal extraction indicates a link between the local location and the master plan, and the extrapolation of the full picture of the signals through local identification. The network structure consists of master and digital capsules (Haq et al., 2023).

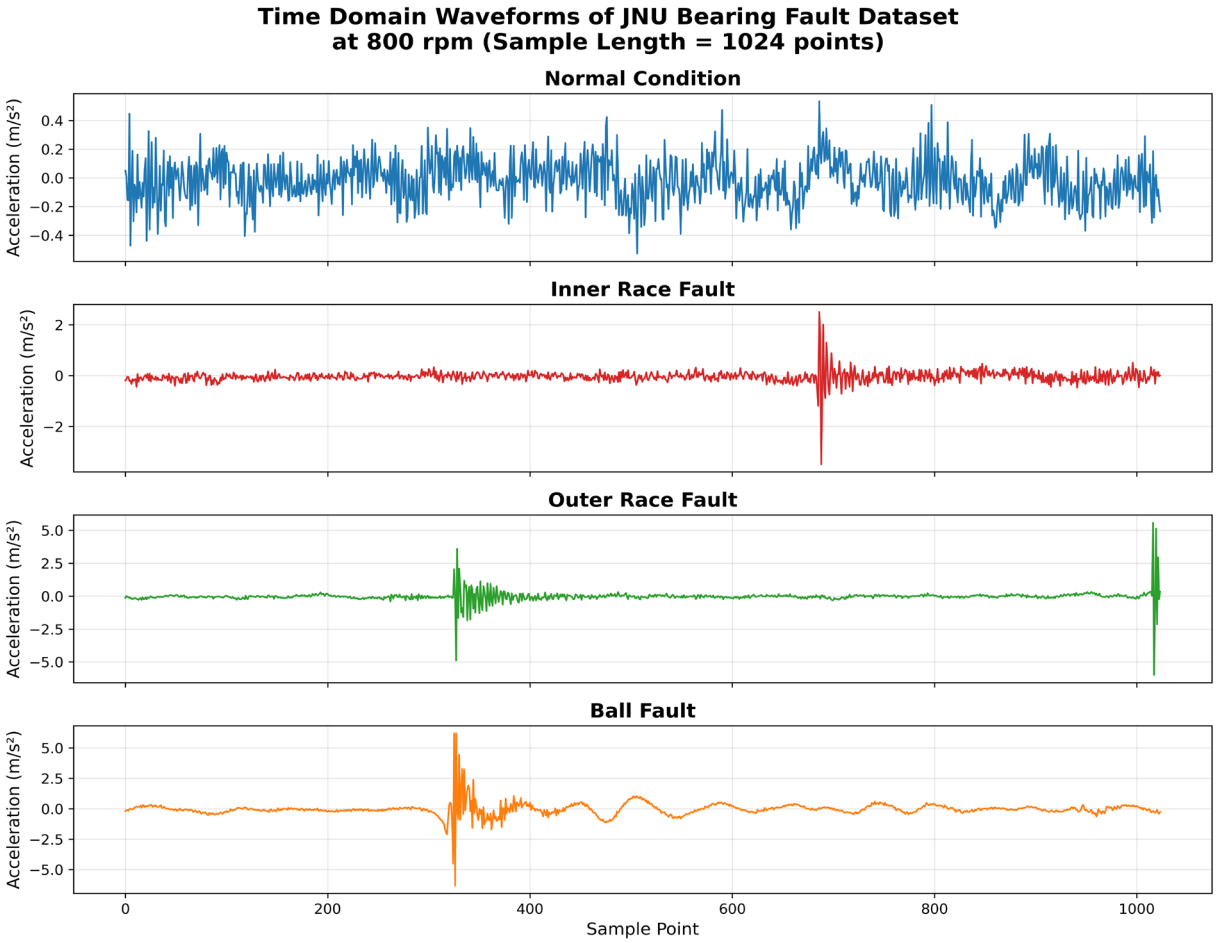

Fig. 1. Example samples from the JNU bearing fault dataset

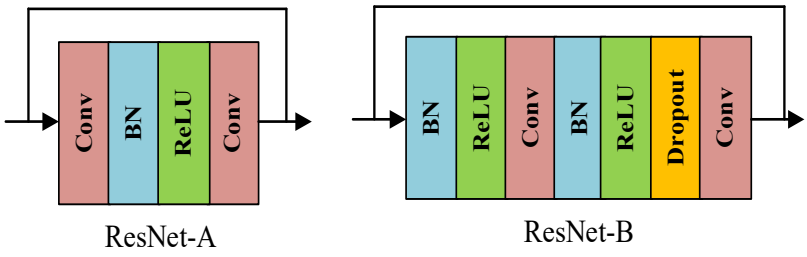

Fig. 2. Architectures of ResNet-A and ResNet-B

3.2.1. Primary capsule layer

The capsules main layer uses gathering processes to process vibration signals, extract their basic properties, and form the capsules primary vectors. The capsules retain the orientation characteristics of defect signals and effectively avoid confusion between similar waveforms (Jiao et al., 2022). Small capsules are then activated in a nonlinear way by pressing the activation function, making CapsNet sensitive to changes in direction while maintaining vector orientation characteristics.

This level is referred to as the low-dimensional information code for high-dimensional vectors as input for further abstraction and further classification of network layers. Fig. 3 illustrates the restructuring of the capsules main layer.

3.2.2. Digital capsule layer

The Digital Capsule Layer primarily functions through the dynamic routing mechanism between capsules. Its input is the reshaped sub-capsules from the primary capsule layer, and the dynamic routing algorithm activates these sub-capsules into higher-level capsules. The output consists of the lengths of class-specific capsules, and the capsule with the longest vector is selected as the recognition result (Sabour et al., 2017). The dynamic routing mechanism between capsules is illustrated in Fig. 4 and is defined as follows in Eqs. (1)-(5).

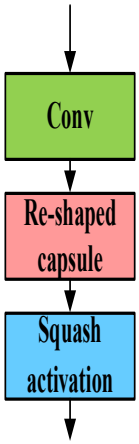

Fig. 3. Structure of the primary capsule layer

$$\hat{u}_{j|i} = W_{ij} \cdot u_i \tag{1}$$

$$c_{ij} = \frac{exp\left(b_{ij}\right)}{\sum_i exp\left(b_{ij}\right)} \tag{2}$$

$$s_j = \sum_i c_{ij} \cdot \hat{u}_{j|i} \tag{3}$$

$$v_j = squash\left(s_j\right) \tag{4}$$

$$b_{ij} = b_{ij} + v_j \cdot \hat{u}_{j|i} \tag{5}$$

Where *i* denotes the capsule of the current layer and j denotes the capsule of the next layer. We first initialize the temporary variable b_{ij} to map the weights of the high-level capsule with the low-level vector.

Eq. (1) represents an affine transformation, mapping the feature vector u_i from a lower layer to the higher-level feature vector $\hat{u}_{j|i}$ through the affine matrix W_{ij}. Eq. (2) calculates the coupling coefficients c_{ij}, which represent the weights assigned to all higher-level capsules, the sum of all c_{ij} for each i equals 1. The dynamic routing algorithm determines these coupling coefficients to extract global features, effectively performing a global pooling operation. Eq. (3) computes the weighted sum of input vectors based on the coupling coefficients, yielding the output vector s_j. The squash activation function in Eq. (4) normalizes the vector to obtain v_j. Steps (2)-(5) are iteratively repeated to update the routing coefficients. The inner product between $\hat{u}_{j|i}$ and v_j is then calculated to measure vector similarity. When these two vectors are highly aligned, their inner product becomes large, thereby increasing b_{ij} and strengthening the coupling relationship. Consequently, the length of the corresponding feature capsule grows proportionally. This process can be summarized by Algorithm 1, as shown in Table 1.

3.3. Construction of the DRCN

By combining ResNet with CapsNet and leveraging ResNet's feature extraction capability along with CapsNet's dynamic routing mechanism, a new DRCN is developed. The ECG signal first passes through a pre-activation module consisting of a convolutional layer, a batch normalization layer, and a ReLU activation layer. It is then fed into the ResNet-C architecture for feature learning. The extracted features are reshaped into master vector capsules via the Primary Capsule Layer of CapsNet. These primary capsules are activated into higher-level capsules by the Digital Capsule Layer, enabling fault classification through the dynamic routing mechanism CapsNet.

Fig. 5. illustrates the complete DRCN framework. The network integrates three identical ResNet-C modules. Each module contains one ResNet-A block, followed by two ResNet-B blocks, a batch normalization layer, and a ReLU activation layer. The CapsNet part includes a primary capsule layer and a digit capsule layer. The final fault detection is based on the magnitude of the output capsule vectors.

Table 1. Dynamic routing algorithm

Algorithm 1: Dynamic routing process
Input: Low-level capsules u_i, number of iterations T Output: High-level capsules v_j
1: Initialize $b_{ij} = 0$
2: Perform affine transformation using Eq. (1) to obtain prediction vectors $\hat{u}_{j\|i}$
3: for i in T do:
4: Compute coupling coefficient c_{ij} via Eq. (2)
5: Compute output vector s_j using Eq. (3)
6: Apply activation via Eq. (4) to obtain capsule output v_j
7: Update routing variable using Eq. (5)
8: return v_j

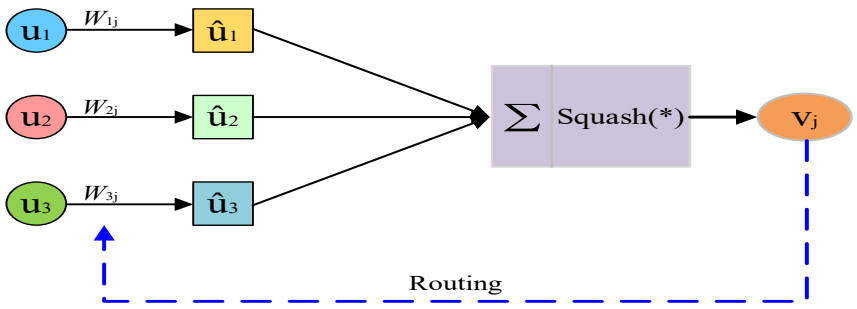

Fig. 4. Dynamic routing mechanism between capsules

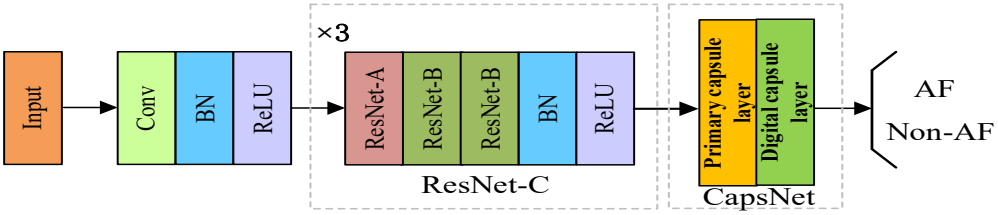

Fig. 5. Architecture of the DRCN

From the perspective of mathematical mapping, the overall computational process of DRCN can be expressed as:

$$y = H_{caps}(H_{ResNet-C}(H_{stem}(X)))$$

where X denotes the input signal, and the final output of the network is a binary classification result:

$$y = \{y_{AF} | y_{Non-AF}\}$$

First, the input signal passes through a pre-activation module (stem):

$$h_1 = \sigma(BN(Conv(X)))$$

where $Conv$ denotes the convolution operation, BN denotes Batch Normalization, and σ represents the ReLU activation function.

Subsequently, the features are fed into the ResNet-C structure, which consists of three cascaded residual blocks:

$$RC = R_3 \circ R_2 \circ R_1$$

For the i-th residual block, the output is defined as:

$$h^{(i)} = h^{(i-1)} + F_i\left(h^{(i-1)}\right)$$

where the residual mapping is given by:

$$F_i(h) = \sigma\left(\mathrm{BN}\left(\mathrm{Conv}_2\left(\sigma\left(\mathrm{BN}\left(\mathrm{Conv}_1(h)\right)\right)\right)\right)\right)$$

The three stages are specifically expressed as:

Stage A:

$$h_2 = h_1 + F_A(h_1)$$

Stage B:

$$h_3 = h_2 + F_B(h_2)$$

Stage C:

$$h_4 = h_3 + F_C(h_3)$$

The output of ResNet-C is:

$$h_5 = \sigma\left(\mathrm{BN}(h_4)\right)$$

Next, the features are fed into the capsule network. First, through the primary capsule layer:

$$u = \mathrm{PrimaryCapsule}(h_5)$$

Then into the digital capsule layer:

$$v = \mathrm{DigitalCapsule}(u)$$

The output of the k-th capsule is defined as:

$$v_k = \mathrm{squash}\left(\sum_j c_{jk} W_{jk} u_j\right)$$

Finally, classification is performed through two separate branches. The AF branch is:

$$y_{AF} = f_{AF}(v) = \sigma(W_{AF} v + b_{AF})$$

The Non-AF branch is:

$$y_{Non-AF} = f_{Non-AF}(v) = \sigma(W_{Non-AF} v + b_{Non-AF})$$

The mathematical innovation of the proposed fusion strategy can be summarized as follows.

First, a spatially preserving mapping from features to capsules is introduced. In conventional CNN-based methods, global pooling is typically applied after feature extraction to compress feature maps into scalar representations, which inevitably leads to the loss of spatial positional information. In contrast, the proposed fusion strategy directly reshapes the feature maps output by ResNet-C into primary capsule vectors, where each spatial location corresponds to an independent capsule. In this way, the spatial distribution of fault-related features in the time frequency domain is effectively preserved.

From a mathematical perspective, this mapping can be formulated as $R^{H' \times W' \times C'} \rightarrow R^{K \times C'}$, where $K = H' \times W'$ denotes the number of spatial locations and C'represents the capsule dimension, thereby achieving a lossless spatial transformation from feature maps to capsule vectors.

Second, the synergistic coupling between residual semantics and dynamic routing. In the proposed fusion framework, the dynamic routing mechanism is not applied to raw signals or shallow features. Instead, it operates on residual semantic features that have been deeply extracted by ResNet-C. Specifically, the update of the coupling coefficients in the dynamic routing algorithm,

$$c_{ij} \leftarrow c_{ij} + \hat{\mathrm{u}}_{j|i} \cdot v_j$$

is performed within a semantically rich residual feature space. This enables the routing procedure to conduct iterative agreement based on higher-level feature representations, rather than on low-level raw signals.

Third, the structured modeling capability of part–whole hierarchies. Through the coupling of the aforementioned two-level network architecture, DRCN achieves hierarchical modeling from low-level local features (e.g., individual impulse waveforms) to high-level global fault patterns (e.g., complete vibration characteristics of inner-race faults). Specifically, the lower-level capsules u_i represent local features, while the higher-level capsules $\mathrm{v_j}$ represent holistic fault entities, and the routing mechanism learns the spatial compositional relationships between them. This "part-to-whole" hierarchical representation capability constitutes a mathematical property that is not possessed by traditional CNNs or by standalone ResNet or CapsNet models.

3.4. Experimental Results and Analysis

3.4.1. Evaluation metrics

To assess the model's effectiveness, this research uses accuracy, precision, recall, and the F1-score as performance indicators. Their formal definitions are provided below (Shi, 2023).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \tag{9}$$

Accuracy refers to the ratio of the accurately sorted samples to the total number of bearing samples. Accuracy refers to the truly defective part of an item that has been identified as defective. The recall rate represents the model's ability to correctly detect fault samples in all actual fault scenarios. The F1 score is a harmonious average of accuracy and recall and provides a balanced measure of the diagnostic performance of the model. True Positive (TP) is the number of incorrect samples that have been correctly detected. False Negative (FN) is the number of errors classified as normal. A True Negative (TN) is a normal sample that has been correctly identified. FP (False Positive) means that a normal sample is wrongly marked as defective.

3.4.2. Parameter settings

The model employs the cross-entropy loss function to quantify the discrepancy between its predictions, and the ground truth labels, as defined in reference (Zhou et al., 2019).

$$Loss = -\sum_{i=1}^{n} y_i \, log \, \hat{y}_i \tag{10}$$

In Eq. (10), n denotes the total number of samples, L is the classification loss, and y_i and $\hat{y}_i$ represent the true and predicted labels of the i-th sample, respectively.

This training uses Adam optimization and learning speed 0.0001. Take early suspension measures to avoid overtraining and stop training if the verification loss does not improve for 8 consecutive weeks. The dropout rate is set to 0.3. The structure of the network and the associated hyperparameters are determined by a detailed combination of grid search and subsequent manual work. All convolutions start with 16 filters and a kernel size of 9.

The ResNet-A component reduces the model's complexity by setting the initial convolution stride to 2. The second and third modules of the ResNet-A have doubled the number of filters to include a wider range of features. The main capsule layer of the capsule network reconfigures the characteristics of up to eight up to 306 capsules. The next digital capsule layer contains the same 16-size capsules as the target category (2). Perform dynamic routing processing for a single iteration.

The detailed parameter configuration of the DRCN is shown in Table 2. All experiments were conducted on the NVIDIA Tesla T4 GPU server, and the model was developed based on Python 3.6.5 and PyTorch 1.10.1 frameworks.

4. Results

4.1. Case Western Reserve University Bearing Fault Dataset

This study uses the experimental data collected from the drive end of the bearing test bench. To verify the proposed method, four datasets (A, B, C, and D) were constructed, each corresponding to different load conditions, and each sample contains 2048 data points. Detailed dataset information is shown in Table 3.

Here we present example samples from the JNU dataset, as shown in Fig.6. Time-domain waveforms of drive-end vibration signals from the CWRU bearing dataset (12 kHz, 0 HP load): (a) normal condition; (b) inner race fault (0.007 inches); (c) outer race fault (0.007 inches at the 6 o'clock position); (d) rolling element fault (0.007 inches).

To verify the effectiveness of the method, five representative intelligent fault diagnosis models were selected for comparative experiments, including 1D Convolutional Neural Network (1D-CNN), LSTM, DBN, DRSN, and CapsNet (Zhou et al., 2019). All models were evaluated on the four datasets listed in Table 3. The detailed comparison of diagnostic accuracy rates across different models is shown in Table 4 (Jing and Wu, 2024).

As shown in Table 4, the diagnostic accuracy rates of the proposed DRCN on datasets A through D is 99.13%, 98.52%, 98.87%, and 99.32%, respectively, all higher than those of other comparison models. The outcomes confirm that DRCN reliably outperforms both conventional approaches and alternative deep learning architectures in diagnosing rolling bearing faults, affirming its effectiveness and robustness.

Here we present example images of different types of bearing faults in Fig. 7.

Table 2. Parameter configuration of the DRCN

Module	Network layer	Kernel size	Number of kernels	Stride	Capsule dimension	Capsule dimension
Pre-activation Module	Pre-activation Layer	9	16	1	/	/
ResNet-C1	ResNet-A	9	16	2	/	/
	ResNet-B	9	16	1		
	ResNet-B	9	16	1		
ResNet-C2	ResNet-A	9	32	2	/	/
	ResNet-B	9	32	1		
	ResNet-B	9	32	1		
ResNet-C3	ResNet-A	9	64	2	/	/
	ResNet-B	9	64	1		
	ResNet-B	9	64	1		
CapsNet	Primary Capsule Layer	9	16	/	8	306
	Digital Capsule Layer	/	/	/	16	2

4.2. Jiangnan University Bearing Fault Dataset

Given the characteristics of the bearing fault dataset and the research objectives, the following key preprocessing steps should be performed before constructing a deep learning classification model. Signal segmentation divides the continuous vibration signal under 600 r/min conditions into non-overlapping segments of 1024 sampling points, each corresponding to 20.48ms, generating standardized time slices that ensure each sample fully reflects the fault characteristic period. To eliminate high-frequency environmental disturbances while retaining the character of impulse failures, noise counter measurements were filtered using medium-moving filters and low-wave thresholds. (for example, using z-score) to eliminate differences in sensor scale. Frequency characteristics were constructed using the Short-Time Fourier Transform (STFT) or the Continuous Wavelet Transform (CWT) to convert the time band transmission signals into frequency to be used in the CNN input structure. The reconstruction data has been separated into (1,562 samples) for the reconstruction data and (390) for the test sample, to maintain a reasonably balanced category distribution and avoid data leaks.

As illustrated in Fig. 8, the proposed model attains an accuracy of 98.78%, with precision, recall, and F1-score reaching 98.32%, 97.41%, and 97.86%, respectively. These metrics collectively indicate robust classification performance, verifying the model's efficacy given the employed dataset and preprocessing methods.

4.3. Model Comparison

To verify the recognition performance of the proposed DRCN model against other fault diagnosis models, this study compares the fault detection accuracy reported by competing approaches. To ensure an equitable comparison, all experimental outcomes presented here are derived from the Case Western Reserve University bearing fault dataset, with a detailed comparison provided in Table 5.

Table 3. Experimental datasets

Dataset	Training samples	Validation samples	Test samples	Load/HP
A	1200	400	400	3
B	1200	400	400	2
C	1200	400	400	1
D	1200	400	400	0

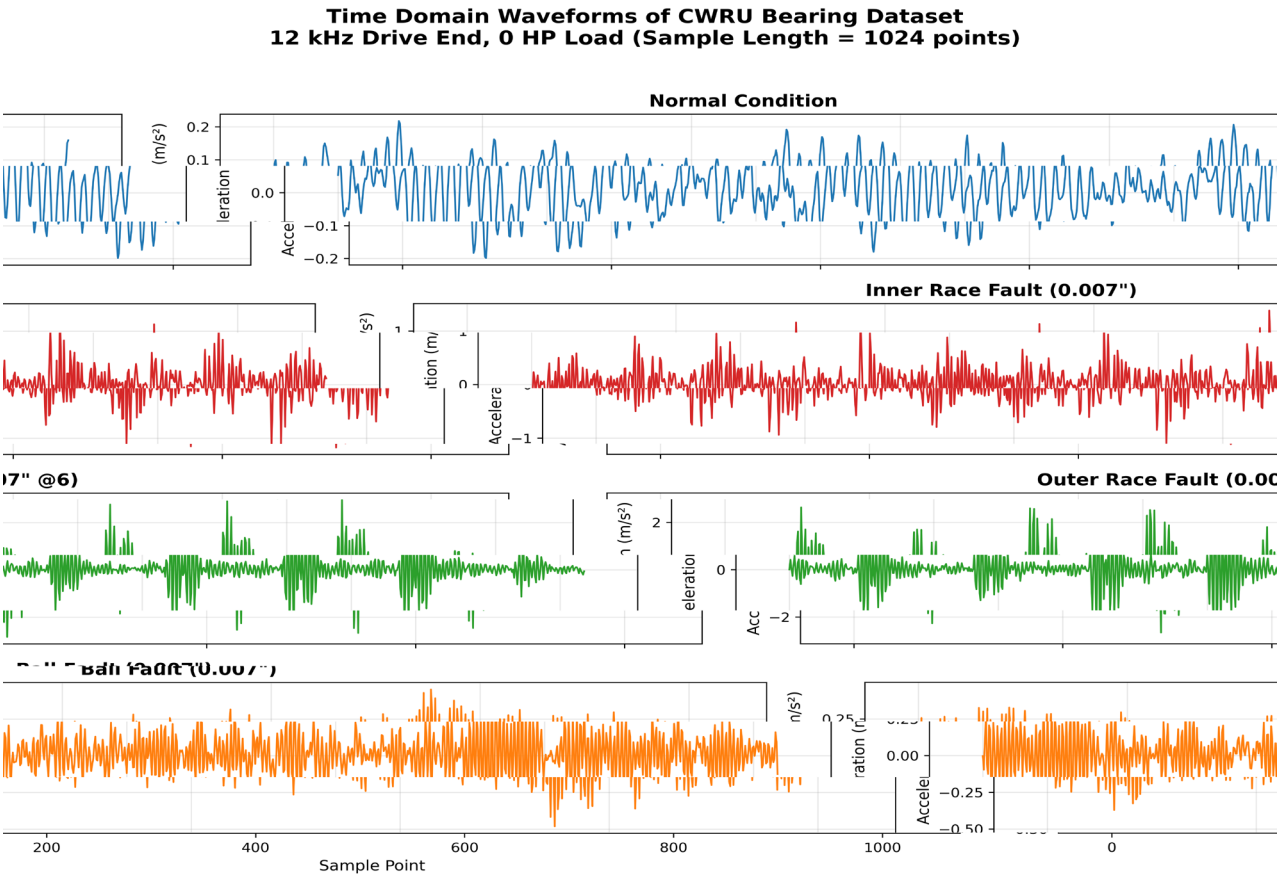

Fig. 6. Example samples from the CWRU bearing fault dataset

Table 4. Comparison of diagnostic accuracy for different fault diagnosis models

Method	Dataset A	Dataset B	Dataset C	Dataset D
1DCNN	92.47%	93.83%	95.76%	94.88%
LSTM	94.89%	97.32%	98.54%	96.63%
DBN	96.78%	97.67%	96.89%	97.56%
DRSN	97.63%	98.34%	97.45%	98.30%
CapsNet	97.95%	97.82%	98.37%	99.03%
DRCN (This study)	99.13%	98.52%	98.87%	99.32%

Fig. 7. Images of different types of bearing faults

To verify the recognition performance of the DRCN model in bearing fault diagnosis, this study compares its fault detection accuracy on the CWRU dataset with that of other diagnostic models. Table 5 lists the results of different methods: 1D-CNN is 91.50%, LSTM is 93.56%, DBN is 98.78%, DRSN is 97.20%, CapsNet is 98.90%, while the DRCN proposed in this paper reaches 99.13%, showing the best performance.

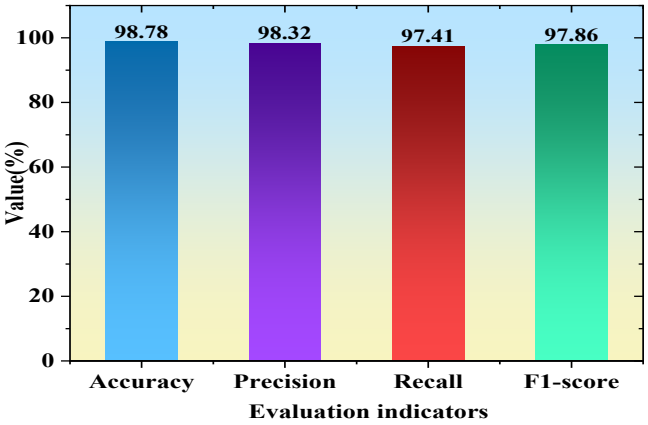

Fig. 8. Classification accuracy on the Jiangnan University bearing fault dataset

The reason why DRCN achieves higher accuracy than 1D-CNN, LSTM, DBN, DRSN and CapsNet is that it collaboratively combines the residual connection of ResNet with the capsule unit of the capsule network, precisely corresponding to the core requirement of bearing diagnosis, extracting deep and weak fault features from vibration sequence signals. Model the structured relationships among features and resist environmental noise interference (Chai et al., 2024). Residual joins alleviate the problems of vanishing gradients and feature degradation in deep models through identity mapping, enabling deeper networks to gradually strip away noise and capture nested multi-scale failure modes, thereby overcoming the limitations of 1D-CNN and basic CapsNet in depth and feature extraction capabilities (Zaeemzadeh et al., 2020). LSTM models are prone to gradient explosion in deeper stacks and have weaker local feature extraction (Al-Selwi et al., 2023). DBNs suffer from complex deep training and a tendency to overfit (Jiang et al., 2020). Capsule units output vector representations and aggregate features via dynamic routing, capturing hierarchical and spatiotemporal relations among fault attributes (e.g., frequency, amplitude, phase). This avoids the broken inter-feature dependencies in scalar-output CNNs, the difficulty of LSTMs in capturing multi-frequency relations at the same time step, the confusion caused by DRSN's lack of structured coding, and the limited discriminability of distribution-focused DBNs (Wang and Chen, 2024). In noisy environments, the two components act jointly: residual modules extract weak features, while capsule vector lengths filter valid signals, thereby enhancing noise robustness, overcoming the noise sensitivity of 1D-CNN/LSTM/DBN and the weak-feature limitations of vanilla CapsNet, and ultimately achieving higher diagnostic accuracy.

Table 5. Accuracy comparison of different models on the CWRU bearing dataset

Dataset	Reference	Method	Accuracy (%)
CWRU bearing dataset	(Jing and Wu, 2024)	1D-CNN	91.50
	(Jing and Wu, 2024)	LSTM	93.56
	(Zhang et al., 2022)	DBN	98.78
	(Tang et al., 2023)	DRSN	97.20
	(Zhao and Chai, 2024)	CapsNet	98.90
	This study	DRCN	99.13

4.4. Ablation Study

To further validate the effectiveness of each DRCN module, an ablation study was conducted. The results show that combining ResNet-C with CapsNet outperforms using either ResNet-C or CapsNet alone, confirming the effectiveness of the proposed approach. The ablation results are reported in Table 6.

As shown in Table 6, DRCN outperforms the other models across all metrics, achieving 98.78% accuracy, 98.32% precision, 97.41% recall, and 97.86% F1-score. By comparison, ResNet exhibits lower performance (e.g., 98.12% accuracy, 97.56% precision), while CapsNet (with 95.68% accuracy and 96.49% precision, among other figures) also falls behind the proposed method. These results prove the superior performance of DRCN and verify the effectiveness of the integration scheme of ResNet-C and CapsNet.

Table 6. Results of the ablation study

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ResNet	98.12	97.56	96.81	97.18
CapsNet	95.68	96.49	95.93	96.21
DRCN	98.78	98.32	97.41	97.86

4.5. Managerial Implications

The DRCN model proposed in this study is not only a technical contribution but, more importantly, provides a practical decision-making tool for predictive maintenance in engineering management. In real-world production, unplanned downtime caused by bearing failures is one of the major sources of maintenance costs. Traditional maintenance strategies face a dilemma: scheduled maintenance often leads to unnecessary part replacements and labor waste, while reactive maintenance entails production interruptions and costly emergency procurement. The value of DRCN lies in its ability to identify weak fault features from vibration signals at an early stage and to classify fault types with an accuracy exceeding 99%, enabling managers to move away from the course "periodic replacement" approach toward precise, condition-based maintenance.

In terms of integration into existing maintenance workflows, DRCN can be deployed within online monitoring systems for critical equipment. Enterprises only need to install vibration sensors on bearing housings to continuously collect vibration signals and feed them into the DRCN model in real time. The model can then automatically output the health status of the bearing and the corresponding fault probabilities. When the predicted fault probability exceeds a predefined threshold, the system can automatically issue alerts to the maintenance platform, enabling maintenance personnel to schedule inspections during planned downtime and avoid unexpected failures and production interruptions.

At present, this study has been validated only on public datasets and has not yet been deployed in real industrial environments. However, the design of DRCN fully considers industrial applicability: the model requires only single-channel vibration signals as input, supports sampling frequencies ranging from 12 kHz to 50 kHz, and maintains acceptable computational complexity (with an inference time of approximately 0.1 seconds per sample). Future work will involve collaboration with industry partners to conduct on-site validation on rotating equipment such as fans and pumps, gradually accumulating experience in industrial deployment.

If validated at an industrial scale, DRCN could bring quantifiable economic benefits to enterprises. For example, in a production line with an annual output of one million tons, reducing one hour of unplanned downtime can recover approximately 100,000 RMB in production losses. Assuming DRCN enables a 48-hour early warning for a major fault, it could directly prevent losses of around 4.8 million RMB. In terms of false alarm rates, compared with traditional 1D-CNN methods, DRCN can reduce the false alarm rate by approximately 50%, which corresponds to about 10 fewer unnecessary inspections per 100 alerts. Given an inspection cost of 2,000 RMB per instance, this translates into annual savings of tens of thousands of RMB in maintenance costs.

From a broader perspective, deep learning based diagnostic methods represented by DRCN are reshaping the decision-making paradigm in equipment management from "experience-based periodic maintenance" to "data-driven precision maintenance." Managers no longer rely on fixed maintenance schedules but dynamically adjust maintenance plans based on model-predicted fault probabilities. This not only reduces maintenance costs but also enhances equipment availability and operational safety. This transformation represents the key link between the theoretical contributions and managerial value of this study: the deep feature extraction capability of residual networks and the spatial relationship modeling capability of capsule networks jointly support this shift in decision-making logic.

5. Conclusion

The results show that the intelligent DRCN-based fault diagnosis method introduced by this operation has succeeded in solving the important limitations of existing methods in roller bearing diagnosis (learning of inappropriate characteristics, low overall performance, sensitivity to noise, etc.). The experimental evidence strongly demonstrates its contribution to the renovation and its superiority. The main innovation of this study is the integration of the residual architecture of ResNet and CapsNet design based on the capsule. The residual connections in ResNet help mitigate issues related to gradient vanishing and network degradation during deep network training, thereby enabling stable and effective extraction of subtle fault-related patterns from vibration signals. The vector capsule and dynamic routing mechanism of CapsNet enhance the modeling ability of the spatial structure and hierarchical relationship of fault features, making up for the shortcomings of traditional CNN scalar output, which is difficult to associate with features, and a single model that cannot balance feature depth and structural representation accuracy. On the CWRU and JNU public bearing datasets, the fault identification accuracy of DRCN reached 99.13% and 98.78%, respectively, which was significantly better than that of mainstream models such as 1D-CNN, LSTM, DBN, DRSN, and a single CapsNet. The ablation experiment further confirmed that the synergistic effect of the residual structure and the capsule module enabled it to maintain excellent performance in multiple dimensions, such as accuracy, precision, recall rate, and F1 score, fully demonstrating the rationality and superiority of the fusion architecture. In summary, DRCN demonstrates excellent robustness and generalization ability under complex working conditions with multiple loads and fault types, providing a new idea for the structural optimization of deep learning in the field of bearing fault diagnosis. Moreover, it has significant engineering applications and promotion potential due to its stable performance.

5.1. Limitations and Future Work

This study analyzes vibration signals collected by a single sensor and does not fully leverage the value of multi-sensor fusion, such as acoustic or temperature signals, in real industrial scenarios. The next phase of research aims to incorporate multi-source sensor data, including sound, current, and temperature, to create a multi-source input module. Using cross-modal attention or feature alignment strategies, collaborative diagnosis of multiple physical quantities can be achieved, thereby improving the detection of compound faults and early weak faults. Although DRCN has demonstrated excellent performance on public datasets, its computational complexity remains relatively high. Future research will focus on optimizing dynamic routing algorithms, such as using approximate or single routing to reduce iterations, and applying network pruning, quantization, or knowledge distillation techniques to decrease the number of parameters and computational costs, which will enhance deployment efficiency, particularly in edge application devices. The experimental data are limited to two public datasets, and systematic cross-dataset generalization validation has not been conducted. Therefore, the applicability of the model in broader industrial scenarios remains to be further verified.

This study not only demonstrates the high diagnostic accuracy of the DRCN for bearing faults but also highlights the untapped potential of data-driven approaches in industrial digitalization. First, deep models can effectively capture critical features, suggesting that enterprises can leverage existing sensor data for efficient monitoring and early warning. Integrating deep learning methods with engineering management can optimize maintenance decisions and resource allocation, guiding managers to shift from experience-driven to data-driven, precise operations. Finally, the scalability of this approach offers new possibilities for integrating multi-source sensor data, cross-plant knowledge transfer, and real-time edge deployment, opening broad prospects for industrial equipment intelligence and predictive maintenance.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The authors used ChatGPT for language editing and formatting. All content has been reviewed by the authors, who take full responsibility for the manuscript.

Reference

Al-Selwi, S., M., Hassan, M., F., Abdulkadir, S., J., and Muneer, A. (2023). LSTM inefficiency in long-term dependencies regression problems. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 30(3), 16-31.

Chai, J., Zhao, X., and Cao, J. (2024). Small-sample fault diagnosis study of rolling bearings based on a residual parameterised convolutional capsule network. *Insight—Non-Destructive Testing and Condition Monitoring*, 66(4), 215-225.

Yong-Zhan, C., H., E., N., Jian-Ling, Q., U., Xiao-Fei, W., A., N., G., and Yuan-Xin, W., A., N., G. (2025). Rolling Bearing Fault Diagnosis Based on CNN-LSTM with Time Sequential Memory Enhancement. *Noise and Vibration Control,* 45(1), 105.

Cong, S., and Zhou, Y. (2023). A review of convolutional neural network architectures and their optimizations. Artificial Intelligence Review, 56(3), 1905-1969.

De Sousa Ribeiro, F., Duarte, K., Everett, M., Leontidis, G., and Shah, M. (2024). Object-centric learning with capsule networks: A survey. *ACM Computing Surveys*, 56(11), 1-291.

Du, W., T., Zeng, Q., Shao, Y., M., Wang, L., M., and Ding, X., X. (2020). Multi-scale demodulation for fault diagnosis based on a weighted-EMD de-noising technique and time–frequency envelope analysis. Applied Sciences, 10(21), 7796.

Hao, X., Zheng, Y., Lu, L., and Pan, H. (2021). Research on intelligent fault diagnosis of rolling bearing based on improved deep residual network. *Applied Sciences*, 11(22), 10889.

Haq, M., U., Sethi, M., A., J., and Rehman, A., U. (2023). Capsule network with its limitation, modification, and applications—A survey. *Machine Learning and Knowledge Extraction,* 5(3), 891-921.

Jain, P., H., and Bhosle, S., P. (2021). Study of effects of radial load on vibration of bearing using time-domain statistical parameters. In IOP *Conference Series: Materials Science and Engineering,* 1070(1), 012130. IOP Publishing.

Jiang, W., and Wang, Y., X. (2025). Research on rolling bearing fault diagnosis method based on convolutional neural network. *Marine Electric and Electronic Technology*, 45(7), 7-10.

Jiang, X., W., Yan, T., H., Zhu, J., J., He, B., Li, W., H., Du, H., P., and Sun, S., S. (2020). Densely connected deep extreme learning machine algorithm. *Cognitive Computation,* 12(5), 979-990.

Jiang, Y., Xie, J., Meng, L., and Jia, H. (2022). Multiple working condition bearing fault diagnosis method based on channel segmentation improved residual network. *Electronics,* 12(1), 145.

Jiao, Y., Qi, H., and Wu, J. (2022). Capsule network assisted electrocardiogram classification model for smart healthcare. Biocybernetics and Biomedical Engineering, 42(2), 543–555.

Jing, S., T., and Wu, D., S. (2024). Fault diagnosis of dual-path rolling bearing based on LSTM-CNN. *Journal of Shenyang Ligong University,* 43(1), 44-49.

Raj, K., K., Kumar, S., and Kumar, R., R. (2024). Systematic review of bearing component failure: Strategies for diagnosis and prognosis in rotating machinery. *Arabian Journal for Science and Engineering,* 1-23.

Sabour, S., Frosst, N., and Hinton, G., E. (2017). Dynamic routing between capsules. *Advances in Neural Information Processing Systems,* 30.
Shi, D., Y. (2023). Research on atrial fibrillation recognition technology based on deep learning (Master's thesis, Shandong University of Technology).
Smith, W., A., and Randall, R., B. (2015). Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mechanical Systems and Signal Processing,* 64, 100-131.
Tang, S., Y., Tong, J., Y., Zheng, J., D., Pan, H., Y., and Wu, Y. (2023). Improved deep residual shrinkage network used for bearing fault diagnosis. *Journal of Vibration and Shock,* 42(18), 217-224.
Wang, Y., and Chen, L. (2024). A multi-scale spatial–temporal capsule network based on sequence encoding for bearing fault diagnosis. *Complex and Intelligent Systems,* 10(5), 6189-6212.
Wu, G., Yan, T., Yang, G., Chai, H., and Cao, C. (2022). A review on rolling bearing fault signal detection methods based on different sensors. *Sensors*, 22(21), 8330.
Wu, K. (2023). Research on the fault diagnosis method of rolling bearing based on capsule network (Master's thesis, Hunan University of Science and Technology).
Zaeemzadeh, A., Rahnavard, N., and Shah, M. (2020). Norm-preservation: Why residual networks can become extremely deep? *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 43(11), 3980-3990.
Zhang, K., Ma, C., Xu, Y., Chen, P., and Du, J. (2021). Feature extraction method based on adaptive and concise empirical wavelet transform and its applications in bearing fault diagnosis. *Measurement,* 172, 108976.
Zhang, X., D. (2022). Research on rolling bearing fault identification based on deep residual shrinkage network (Master's thesis, Shenyang University of Technology).
Zhang, X., G., Ding, H., Wang, X., B., and Yang, L., L. (2022). Study on fault diagnosis of rolling bearing by deep residual network. *Machinery Design and Manufacture,* (1), 77-80.
Zhao, X., Q., and Chai, J., X. (2024). Improved convolutional capsule network method for rolling bearing fault diagnosis. *Journal of Vibration Engineering,* 37(5), 885-895.
Zhao, Y., Fan, Y., Li, H., and Gao, X. (2022). Rolling bearing composite fault diagnosis method based on EEMD fusion feature. *Journal of Mechanical Science and Technology,* 36(9), 4563-4570.
Zhou, D., Zhao, Y., Wang, Z., He, X., and Gao, M. (2019). Review on diagnosis techniques for intermittent faults in dynamic systems. *IEEE Transactions on Industrial Electronics*, 67(3), 2337-2347.
Zhou, Y., Wang, X., Zhang, M., Zhu, J., Zheng, R., and Wu, Q. (2019). MPCE: a maximum probability based cross entropy loss function for neural network classification. *IEEE Access,* 7, 146331-146341.

Zhang Yuzhou is currently a 2023 undergraduate student majoring in Mechatronic Engineering at Northeast Forestry University. He has a solid academic foundation in mechanical electronics, automation and related fields. He has participated in academic lectures, discipline competitions and practical training in mechanical design, intelligent control, robotics and mechatronics. He has a strong interest in scientific research and engineering practice. His areas of interest include mechatronics, intelligent control, robotics, and mechanical design.