

Fault diagnosis method based on EWT and improved ConvNeXt networks

LI Jun*, MA Yi

School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

***Corresponding author:** LI Jun (lijun691201@mail.lzjtu.cn)

Received: August 6, 2025

Revised: September 20, 2025

Accepted: October 14, 2025

Abstract: Due to the interference of strong noise, feature extraction faces the challenge of limited information, which is not conducive to motor equipment fault diagnosis. This paper proposes a fault diagnosis method based on the empirical wavelet transform (EWT) and an improved ConvNeXt network. The modal components were extracted from the signals of different sensors using empirical wavelet transform, noise was removed, and then the signals were reconstructed. Secondly, the short-time Fourier transform (STFT) was used to convert the one-dimensional signal after noise reduction and reconstruction into a two-dimensional time-frequency spectrum image that enhanced signal features. Single-channel images generated by a single sensor were fused to form multi-channel images, thereby boosting the feature extraction capability of the ConvNeXt network. Additionally, the Ghost convolution module and the efficient local attention mechanism (ELA) were introduced into the ConvNeXt-T (ConvNeXt-Tiny) network, further enhancing the network's performance. Experimental validation was conducted on application examples of various fault diagnostic devices, and comparisons were made with existing mainstream deep learning methods such as SE-InceptionV3, CBAM-ResNet, and CNN-LSTM etc. Experimental results confirmed under different noise environments and variable operating conditions, the proposed method achieved better diagnostic accuracy and enhanced generalization performance.

Key words: fault diagnosis; ConvNeXt; empirical wavelet transform; short-time fourier-transformation; efficient local attention; data level fusion

0 Introduction

Amid the rapid development of industrial automation and intelligent manufacturing, fault diagnosis of motor equipment has become a critical task to ensure the normal operation of production systems and cutting down on maintenance costs. Motor faults are generally categorized into mechanical and electrical faults^[1]. Among common mechanical failures, bearing faults account for the highest proportion—approximately 41%—while electrical faults primarily include stator and rotor winding faults^[2-4]. Therefore, the timely identification of early fault characteristics in motor equipment is of significant importance for enhancing operational safety and reliability. Motor fault diagnosis approaches are primarily categorized into four types: experience-based methods, model-based methods, machine learning methods, and deep learning methods^[5]. Experience-based methods rely mainly on the domain expertise and practical knowledge accumulated by professionals over time. In contrast, model-based methods analyze and study faults by constructing physical or mathematical models^[6]. For example, López et al.^[7] introduced a fault diagnosis approach that integrated hidden

Markov models with stochastic resonance. However, due to the lack of sufficient prior knowledge, the two types of methods often exhibit low diagnostic accuracy and poor generalization under complex operating conditions. In recent years, machine learning methods are gaining increasing traction in motor fault diagnosis. Furthermore, signal processing techniques serve as the initial step, deployed to denoise the signals acquired by sensors. Subsequently, features from both the time and frequency domains are extracted, which is then leveraged in conjunction with machine learning algorithms for fault classification^[8]. Zhang et al.^[9] employed variational mode decomposition (VMD) technique to decompose bearing current signals, followed by the random forest classifier with the whale optimization algorithm, achieving excellent diagnostic accuracy for bearing faults. Yang et al.^[10] introduced a new rolling bearing status feature extraction method incorporating VMD along with improved envelope spectrum entropy (IESE). The original feature set is fused by the joint approximate diagonalization eigen (JADE), followed by support vector machine (SVM) classification to identify bearing conditions. Liu et al.^[11] proposed a novel fault diagnosis method based on adaptive weighted

symplectic geometry mode decomposition (AWSGMD-CP) with cosine difference factor (CDF) and Pearson correlation coefficient (PCC) and extreme learning machine (ELM) optimized by ant colony optimization (ACO). Although these traditional machine learning methods have shown effectiveness in diagnosing faults, they still exhibit notable limitations, including dependence on manual feature extraction and inadequate generalization capability etc.

In the other hand, the fusion of deep learning and signal processing has emerged as a prominent direction in recent fault diagnosis research. Zhao et al.^[12] applied a signal processing method to denoise and decompose signals. Subsequently, principal component analysis (PCA) and fractal box dimension analysis were employed, and the selected optimal components were input into a convolutional neural network (CNN) for bearing fault identification. Wang et al.^[13] proposed a novel method for diagnosing bearing faults in motor stator currents across different operating conditions. It leverages time-shift denoising (TSD) and discrete wavelet transform (DWT) enhanced conditional domain adaptation (CDA) techniques, and the feature learning space of the convolutional neural network (CNN) is expanded into the wavelet domain. Wang et al.^[14] proposed a new visual intelligence-based fault diagnosis method named multi-image feature extraction and attention fusion (MIFEAF). Bai et al.^[15] introduced a novel fractional Fourier transform (FRFT) -encoded feature representation scheme, integrating recursive graph (RP) coding and CNN for fault diagnosis. Zhao et al.^[16] investigated a novel small-sample fault diagnosis method for rolling bearings based on continuous wavelet transform and Siamese neural network (CWT-SNN). Dai et al.^[17] proposed a method for diagnosing bearing faults, which integrated the pelican optimization algorithm (POA), Gramian angular difference fields (GADF), and the Swin transformer network etc. Guo et al.^[18] proposed an attention-based ConvNeXt with a parallel multiscale dilated convolution residual module for the rotor fault diagnosis. However, the aforementioned approaches depend exclusively on data from a single sensor for fault feature extraction, which restricts the richness of obtainable fault information and results in low model adaptability under complex operating conditions.

To overcome the limitations of single-sensor data in providing complete information, multi-sensor data fusion has gradually emerged as an effective solution^[19-21]. Li et al.^[22] introduced a data layer fusion algorithm that leveraged correlated variance contribution (CVC) for dynamic

integration of multiple sensor signals in fault diagnosis. Guo et al.^[23] converted data from multiple sensors into RGB images and combined multi-scale learning and attention mechanisms to achieve fault diagnosis. Xiao et al.^[24] utilised graph convolutional networks to integrate multi-sensor information, thereby markedly enhancing fault identification accuracy. He et al.^[25] proposed a new method called ConvNeXt-FPN based on two-dimensional recursive fusion of multi-channel signals and combined with the convolutional block attention module (CBAM), which was used for detecting complex eccentric faults of permanent magnet synchronous generators. The experimental results confirmed its effectiveness.

We proposed a novel fault diagnosis method that integrated signal processing with multi-sensor data fusion and an improved ConvNeXt network. To demonstrate its effectiveness, the proposed method was compared with the existing deep learning methods such as SE-InceptionV3, CBAM-ResNet, and CNN-LSTM etc.

1 Data pre-processing

1.1 Empirical wavelet transform

EWT is an adaptive approach for signal processing, enabling the adaptive construction of wavelet bases driven by the spectral characteristics of the signal, thus achieving adaptive signal decomposition^[26]. The key steps involved in EWT are as follows.

Step 1 The original signal undergoes Fourier transformation, resulting in its Fourier spectrum, which is then normalized to the interval $[0, \pi]$. Within the range $[0, \pi]$, divide n frequency bands $\Lambda_n = [\omega_{n-1}, \omega_n]$ of unequal bandwidths. ω_n represents the center coordinate of each frequency band. The boundary width of each band is denoted by τ_n .

Step 2 Based on the construction method of Meyer wavelets, empirical wavelets and their corresponding empirical scaling functions are designed.

Step 3 Using the constructed experience wavelet and scale function, the original signal is convolved to obtain feature components at different frequencies. Finally, the signal is reconstructed as the sum of several modal functions at different frequencies.

1.2 EWT-STFT method

In actual operation, motor signals are often affected by strong background noise, which obscures critical fault information and hinders comprehensive feature acquisition, thereby reducing diagnostic accuracy. To address this issue, the signal was decomposed using empirical wavelet

transform, and noise was removed and the signal was reconstructed based on the comprehensive evaluation index (which combined cosine similarity, kurtosis, and envelope entropy). Additionally, the STFT was employed to fuse data from multiple sensors, consequently enhancing fault diagnosis precision. The proposed data preprocessing framework consists of the following steps.

Step 1 The vibration signals collected by the sensors are adaptively decomposed using EWT to extract a set of adaptive modal components. The comprehensive evaluation index value C of each component is calculated by

$$\begin{cases} C = \omega \left(\lambda \frac{1}{E_p} + \gamma K_v \right), \\ \omega = \left| \frac{\sum_{i=1}^N x_1(n) x_2(n)}{\sqrt{\sum_{i=1}^N x_1^2(n)} \sqrt{\sum_{i=1}^N x_2^2(n)}} \right|, \\ K_v = \frac{E[x(t) - \mu]^4}{\sigma^4}, \\ E_p = -\sum_{j=1}^N p_j \lg p_j, p_j = a(j) / \sum_{j=1}^N a(j), \end{cases} \quad (1)$$

where E_p denotes the envelope entropy of each component; K_v represents the kurtosis of each component; ω indicates the residual similarity between each component signal and the original signal; $\lambda = \gamma = 1$; μ represents the mean of the component signal; σ denotes the standard deviation of the component signal; and E denotes the mathematical expectation.

Step 2 Select the top three components with the largest comprehensive evaluation index values C for signal reconstruction.

Step 3 By performing an STFT on the reconstructed signal to generate a single-channel grey-scale time-frequency map, each single-channel grey-scale map formed by each sensor is then fused into a multi-channel image in the channel dimension.

2 Improvement of ConvNeXt model

ConvNeXt is a pure convolutional neural network model renowned for its powerful image processing capabilities. ConvNeXt adopts the ResNet architecture as its backbone and incorporates structural designs inspired by transformers for enhancement. Its core components consist of standard convolution layers, depthwise convolution layers, downsampling layers, normalization layers, and activation functions. The main body of the ConvNeXt network comprises four distinct stages. Each stage performs feature extraction by

stacking multiple ConvNeXt blocks, while spatial downsampling between stages is achieved using a 2×2 convolution with a stride of 2 to reduce resolution.

To achieve model reliability and lightweight design, we made improvements to the ConvNeXt-T model.

First, the block stacking ratio for each stage of the model is adjusted from the original 3:3:9:3 configuration to 1:1:3:1. This modification reduces the overall network depth, thereby alleviating redundant computations and improving efficiency, while still preserving sufficient representational capacity for feature extraction. Second, the standard convolution operations in both the ConvNeXt block and downsampling modules are substituted with Ghost convolutions. By generating more feature maps from a limited number of intrinsic feature maps, Ghost convolution effectively reduces the number of parameters and computational cost without significantly impairing the expressive power of the model. This substitution not only lightens the model but also enhances its suitability for lightweight deployment scenarios. Finally, an efficient layer attention module is introduced at the end of each stage. The ELA module enhances the network's ability to capture global contextual information and focus on more discriminative features, thus strengthening feature representation and improving classification accuracy. Through these structural improvements, the proposed model achieved a better balance between accuracy and efficiency.

2.1 Ghost convolution

Ghost convolution is a lightweight convolution operation^[27]. It is primarily designed to lighten neural networks by decreasing computational complexity and parameter count, while preserving effective feature extraction.

In traditional convolution, generating a large number of feature maps consumes significant computational resources, yet many of these feature maps exhibit high similarity or redundancy. Ghost convolution divides the feature map generation process into two steps. It initially applies standard convolution to obtain a limited set of intrinsic feature maps (intrinsic features), followed by applying simple linear operations (such as depth-separable convolution, pointwise convolution, or affine transformation) to generate the remaining ghost features.

2.2 Efficient local attention

ELA aims to strengthen feature extraction capabilities in deep convolutional neural networks when handling

complex visual tasks^[28]. It effectively leverages spatial information within images to strengthen the model's attention to target regions, thereby enhancing its performance in complex visual tasks.

Fig. 1 depicts the ELA module architecture. Assume that the input size of the convolutional block is $C \times H \times W$. First, average pooling is performed along each channel within two spatial extents: horizontally ($H, 1$) and vertically ($1, W$). This produces the height-dimension output $z_c^h(h)$ for channel c at position h and the width-direction output $z_c^w(w)$ for channel c at position w .

To enrich positional information along both horizontal and vertical directions, two 1D convolution operations are performed on the above pooling results. Subsequently, group normalisation and the sigmoid activation function are used to process the enhanced positional information, thereby obtaining the positional attention weights in the horizontal and vertical directions.

Finally, element-wise multiplication between the original feature map and the two-directional attention

maps produces the final output of the ELA module.

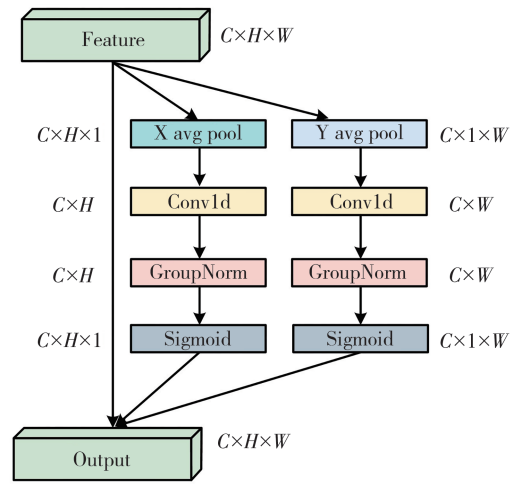


Fig. 1 Efficient local attention structure diagram

2.3 Overall framework of proposed method

The proposed fault diagnosis procedure is illustrated in Fig. 2.

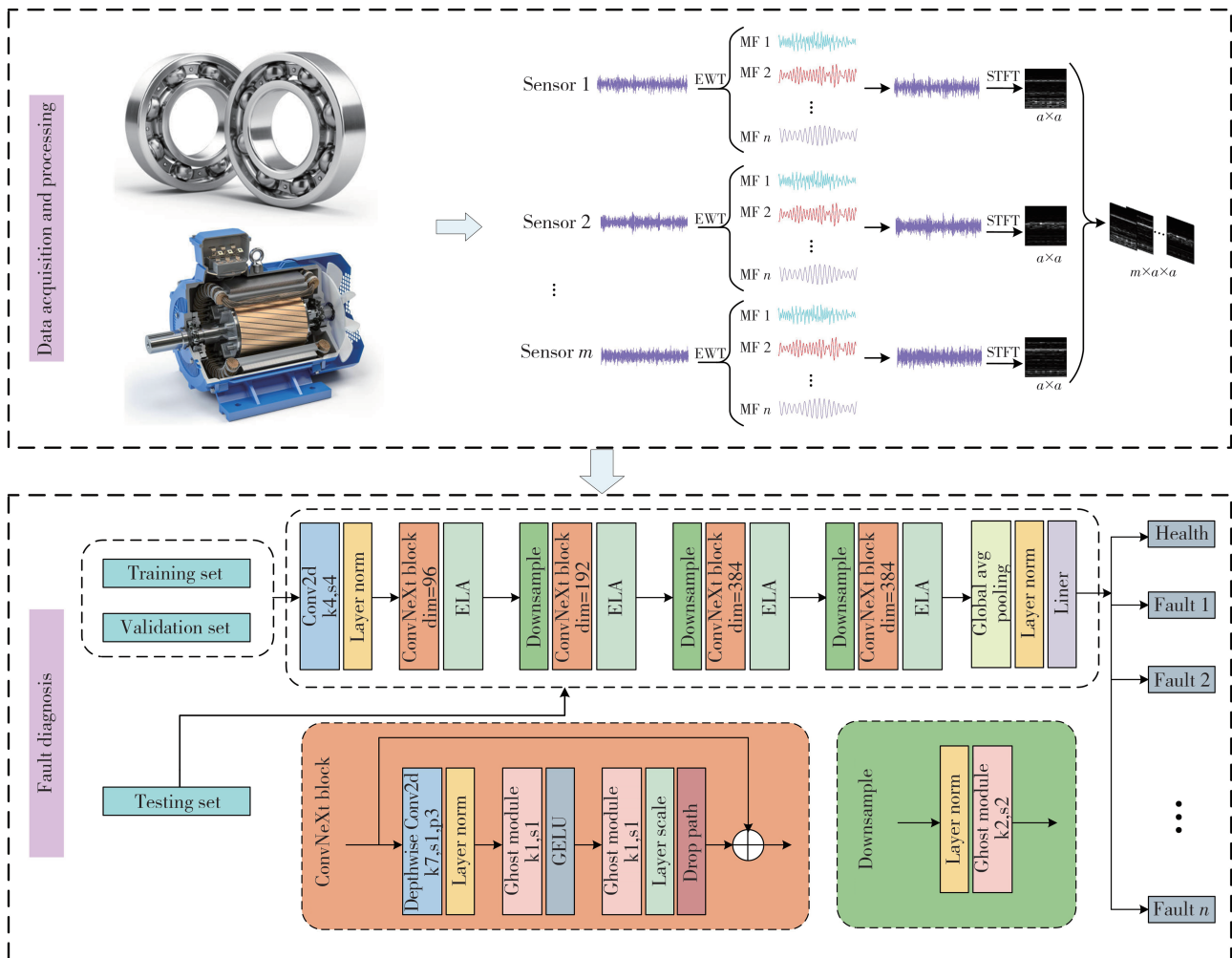


Fig. 2 Proposed method's overall architecture

Firstly, for signals collected from multiple sensor sources, empirical wavelet transform is used for signal

denoising and reconstruction processing to effectively extract key fault feature components. Secondly, the

reconstructed signals are converted into 2D time-frequency maps using STFT, and these are fused along the channel axis to extract comprehensive signal characteristics. Thirdly, the network architecture of the ConvNeXt-T model is modified to enhance its performance under diverse operating conditions. Finally, training and testing are conducted to validate the model's performance, thereby validating its effectiveness and superiority.

3 Results and discussion

The reliability and generalisation ability of the proposed method were demonstrated through two experimental case studies conducted on the CWRU bearing dataset^[29] and the synchronous motor fault dataset^[30].

During the training process, the ADAM optimization algorithm was employed, with cross-entropy as the loss function. 30 iterations were conducted, the learning rate was set at 0.000 1, and the batch size was 32. Each experiment was conducted with five Monte Carlo simulations on the test dataset, and the mean of the results was adopted as the final result.

Signal-to-noise ratio (SNR), root mean square error (RMSE), and energy percentage (EP) are selected to evaluate the reconstructed signal.

$$\text{SNR} = 10 \lg \left(\frac{\sum_{i=1}^N x(t)^2}{\sum_{i=1}^N (x(t) - x'(t))^2} \right), \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x(t) - x'(t))^2}, \quad (3)$$

$$\text{EP} = \frac{\sum_{i=1}^N |x'(t)|^2}{\sum_{i=1}^N |x(t)|^2}, \quad (4)$$

where N is the number of data points.

Accuracy, precision, recall, and $F1$ -score are selected as classification performance evaluation indicators for fault diagnosis.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (5)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (7)$$

$$F1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (8)$$

where TP represents the number of actual faults detected

by the model and accurately identified; TN is the number of actual non-faults that are correctly classified as normal; FP is the number of actual non-faults that are incorrectly classified as faults; FN is the number of actual faults that are incorrectly classified as normal.

3.1 CWRU bearing dataset experiments and analysis

The bearing dataset created by CWRU conducts fault simulation on bearings using the electrical discharge machining technology. Vibration signals are acquired under four distinct load conditions corresponding to motor powers of 0 kW, 0.736 kW, 1.471 kW, and 2.207 kW. The fault types specifically include three categories: rolling element faults, outer ring faults, and inner ring faults. The fault diameters are 0.177 8 mm, 0.355 6 mm, and 0.533 4 mm. In this experiment, vibration signals from the drive end (DE) and fan end (FE) are selected at a sampling frequency of 12 kHz to construct bearing datasets for the DE and FE sensors. Each dataset includes nine different fault types and one normal state. Each fault and normal data category includes 1 000 labeled samples, divided into training set, validation set, and test set, in a ratio of 70%, 20%, and 10%, respectively. Furthermore, this experiment considers four operating conditions corresponding to motor speeds of 187.96 rad/s (0 kW), 185.34 rad/s (0.736 kW), 183.26 rad/s (1.471 kW), and 181.16 rad/s (2.207 kW), which are consistent with the experimental setup of the CWRU dataset.

3.1.1 Using EWT for signal denoising and reconstruction

Using the vibration signals from the DE and FE as an example, when the inner ring failed under operating condition 0 (failure diameter 0.177 8 mm), a noisy signal characterized by an SNR of -4 dB was constructed by introducing Gaussian white noise. To assess the denoising capability of the EWT, it was compared with variational modal decomposition (VMD), empirical modal decomposition (EMD), discrete wavelet transform (DWT), and others. In VMD, the mode number K and penalty factor are set to 7 and 2 000 via grid search, respectively. In DWT, the wavelet basis and decomposition level are empirically set to sym8 and 5, respectively. EWT adaptively divides the frequency band boundaries based on spectral local maximum detection.

Table 1 presents the experimental results. The EWT denoising results exhibit a higher SNR, a lower RMSE, and a larger EP than the EMD, VMD, and DWT methods, indicating that the EWT denoising effect is

significantly superior to other denoising methods.

Table 1 Noise reduction evaluation index results

Method	SNR/dB	RMSE	EP/%
EMD-DE	10.57	0.279 8	92.32
EMD-FE	11.13	0.123 9	96.80
VMD-DE	13.17	0.061 9	94.42
VMD-FE	13.54	0.061 2	94.85
DWT-DE	13.20	0.061 6	95.42
DWT-FE	13.71	0.060 4	95.61
EWT-DE	14.31	0.060 2	95.59
EWT-FE	14.75	0.054 9	96.72

3.1.2 Verification of effectiveness of signal denoising and reconstruction using STFT

The EWT demonstrates exceptional noise reduction and reconstruction capabilities in high-noise environments. We used the high-quality signals obtained through EWT-based noise reduction and reconstruction to conduct further performance comparison studies on time-frequency image conversion methods. To systematically demonstrate the advantages of STFT in constructing time-frequency images, several widely adopted image transformation techniques, such as continuous wavelet transform (CWT), recursive plots (RP), and Markov transform fields (MTF), were compared with short-time Fourier transform (STFT). All methods were classified using the original ConvNeXt-T model, with the dataset consisting of DE and FE data under operating condition 0, and the image resolution size was 64×64 .

According to the experimental data in Table 2, STFT achieves superior performance across all four evaluation metrics on both DE and FE datasets, indicating that the two-dimensional time-frequency plots generated by STFT better represent signal features. Additionally, the time-frequency plots generated by the four methods on DE and FE were concatenated across channels to create dual-channel images, which were then classified. These results are depicted in Fig. 3, illustrating that STFT consistently outperforms the other methods in classification accuracy under all four operating conditions.

Table 2 Diagnostic results of different image transformation methods

Method	Accuracy/%	Precision/%	Recall/%	F1-score/%
CWT-DE	94.43	95.24	94.43	94.39
CWT-FE	94.90	95.42	94.90	94.90
RP-DE	93.13	93.55	93.13	93.10
RP-FE	93.61	94.16	93.61	93.52
MTF-DE	92.03	93.04	92.03	92.04
MTF-FE	92.31	93.07	92.31	92.06
STFT-DE	95.41	95.67	95.41	95.39
STFT-FE	96.47	96.75	96.47	96.48

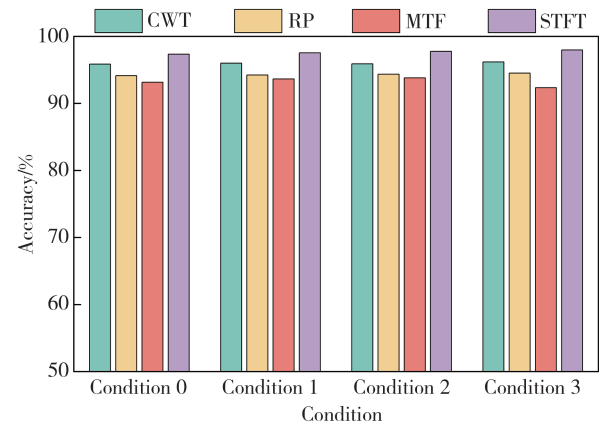


Fig. 3 Comparison of image conversion methods under four working conditions

3.1.3 Ablation experiments

To further investigate the contributions of the ghost convolution module and the ELA module to the improved ConvNeXt-T network, ablation experiments were conducted on the first dataset. Specifically, four model variants were designed: 1) the baseline ConvNeXt-T without any additional modules, 2) ConvNeXt-T with only the ghost module, 3) ConvNeXt-T with only the ELA module, and 4) ConvNeXt-T with both ghost and ELA.

Table 3 presents the results of ablation experiments for the ghost convolution and ELA modules. The results show that the ghost convolution module effectively reduces model parameters and training time while maintaining comparable accuracy. The ELA module improves classification accuracy by enhancing feature representation, albeit with a slight increase in computational cost. When the ghost convolution and ELA are used together, they can achieve the best overall performance, obtaining the highest accuracy with lower model complexity, thus demonstrating their complementary advantages.

Table 3 Comparison of ablation experiment results on CWRU dataset

Method	Accuracy/%	Parameters/M	Training time/s
ConvNeXt-T	95.31	39.33	394
ConvNeXt-T+Ghost	95.11	36.39	371
ConvNeXt-T+ELA	96.25	39.47	412
ConvNeXt-T+Ghost+ELA	96.42	36.53	385

3.1.4 Performance comparison between fusion and single-source methods

To further validate the effectiveness of multi-sensor data fusion in enhancing fault diagnosis performance, we conducted comparative experiments by inputting DE, FE, and the fused data (DE-FE) into the improved ConvNeXt model. The model's performance curves (loss and accuracy) for different data inputs under operating condition 0 are presented in Fig. 4. The model

using fused dual-sensor data exhibits a faster decline in loss values, more stable convergence, and higher accuracy, indicating that multi-source data fusion significantly strengthens the model's learning efficacy and generalisation capabilities. In contrast, the loss curve of the single-sensor model shows significant fluctuations, and its accuracy is lower than that of the model using the fused data from both sensors. This reflects the limitations of single-sensor data due to its insufficient information content.

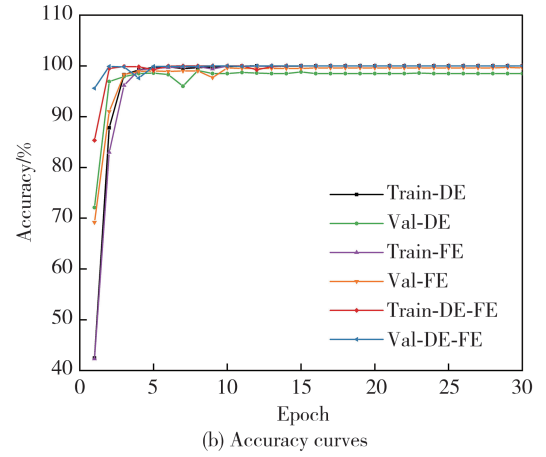
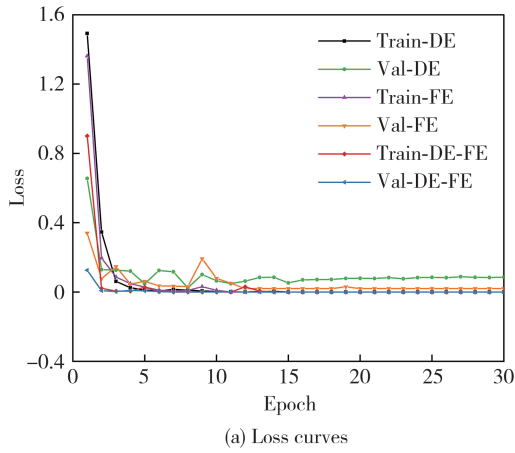


Fig. 4 Loss curves and accuracy curves of different methods under working condition 0

Table 4 shows the diagnostic accuracy and average values under four operating conditions for different data input methods. The model that fuses data from two sensors has significantly higher accuracy than the single-sensor model under all operating conditions, demonstrating that multi-sensor data fusion significantly strengthens the model's robustness and diagnostic accuracy in complex conditions.

Table 4 Diagnostic results for different data entry methods

Method	Accuracy/%				Average/%
	Condition 0	Condition 1	Condition 2	Condition 3	
ICovNeXt-DE	97.08	98.64	98.10	97.90	97.93
ICovNeXt-FE	97.89	98.60	98.02	97.70	98.05
ICovNeXt-DE-FE	99.92	99.86	100	99.86	99.91

3.1.5 Comparison of different model performance

We further conducted comparative experiments to systematically analyse the comprehensive fault diagnosis performance of the proposed method under complex operating conditions and evaluate its superiority over mainstream advanced models. To this end, a comparative analysis was conducted between the proposed method and several models, such as EfficientNetV2, ShuffleNetV2, CNN-LSTM^[31], CBAM-ResNet^[32], and SE-InceptionV3^[33] under four different operating conditions.

Table 5 demonstrate that the proposed method attains accuracy rates of 99.92%, 99.86%, 100%, and 99.86% across the four operating conditions, respectively, with an average accuracy rate of 99.91%, significantly outperforming the other comparison models. This demonstrates that the proposed method is more robust and generalizable when applied to complex environments. Additionally, the proposed method completes training in only 374 s, indicating that it combines high accuracy with computational efficiency.

Table 5 Comparison of performance of different models

Method	Accuracy/%				Average/%	Training time/s
	Condition 0	Condition 1	Condition 2	Condition 3		
EfficientNetV2	99.48	99.42	99.52	99.44	99.47	604
ShuffleNetV2	98.50	98.60	99.16	98.48	98.69	391
CNN-LSTM ^[31]	99.22	98.82	99.24	99.04	99.08	498
CBAM-ResNet ^[32]	99.42	99.38	99.50	99.44	99.44	549
SE-InceptionV3 ^[33]	99.00	98.86	99.32	98.70	98.97	449
ICovNeXt	99.92	99.86	100	99.86	99.91	374

Fig. 5 illustrates the confusion matrices of different models under operating condition 0. In the confusion

matrix, the predicted class labels are shown along the columns, while the true class labels are shown along the

rows. Correct predictions appear in the diagonal cells. Fig. 5 shows that the comparison model exhibits some misclassification in a few fault categories, while the proposed method has only 1 sample misclassified. Fig. 6 shows t-SNE visualizations of different models under

operating condition 0. It shows that the feature clustering of the comparison model exhibits local overlap, while the proposed method effectively distinguishes fault samples of different categories with clearer clustering between fault types.

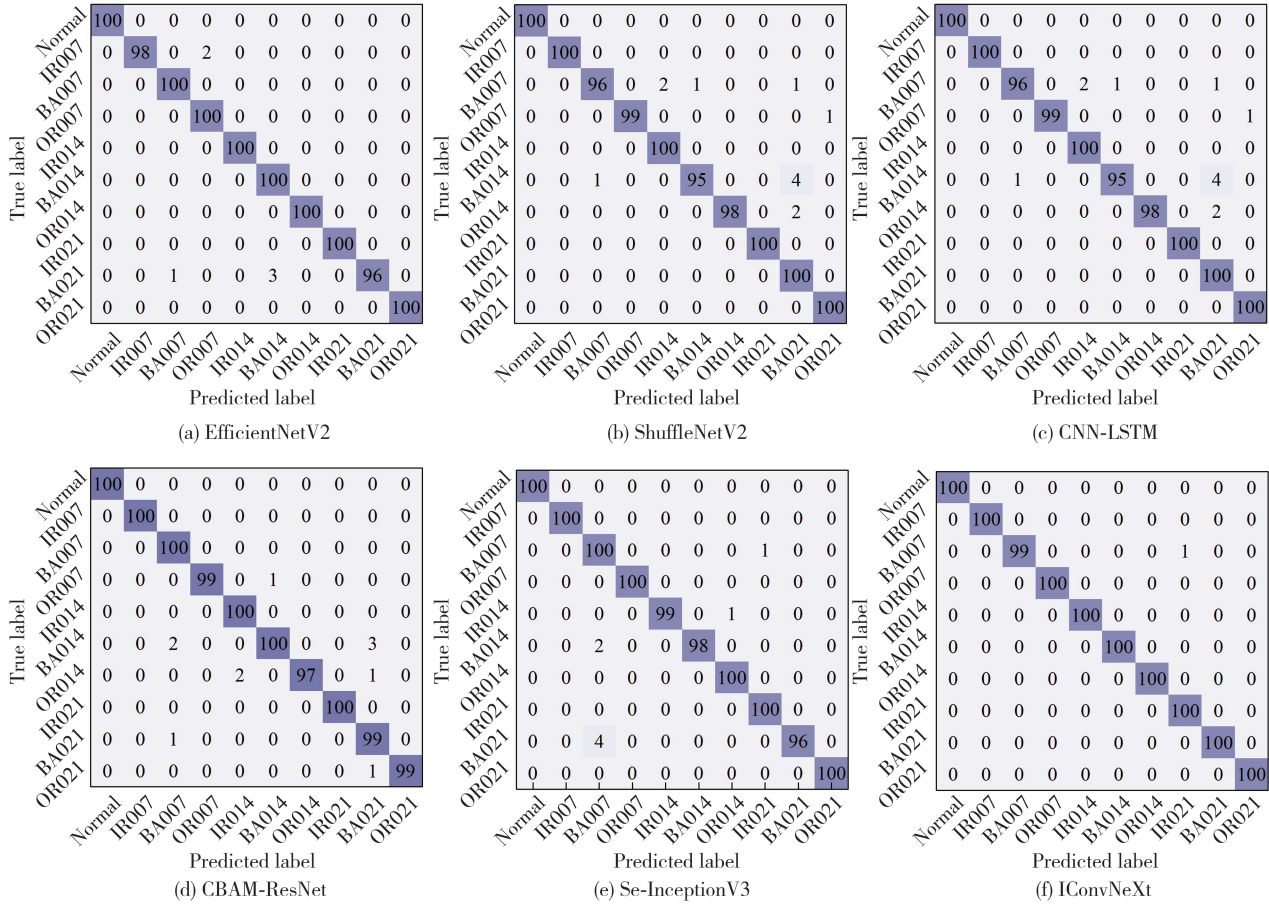


Fig. 5 Confusion matrix diagram of different methods under working condition 0

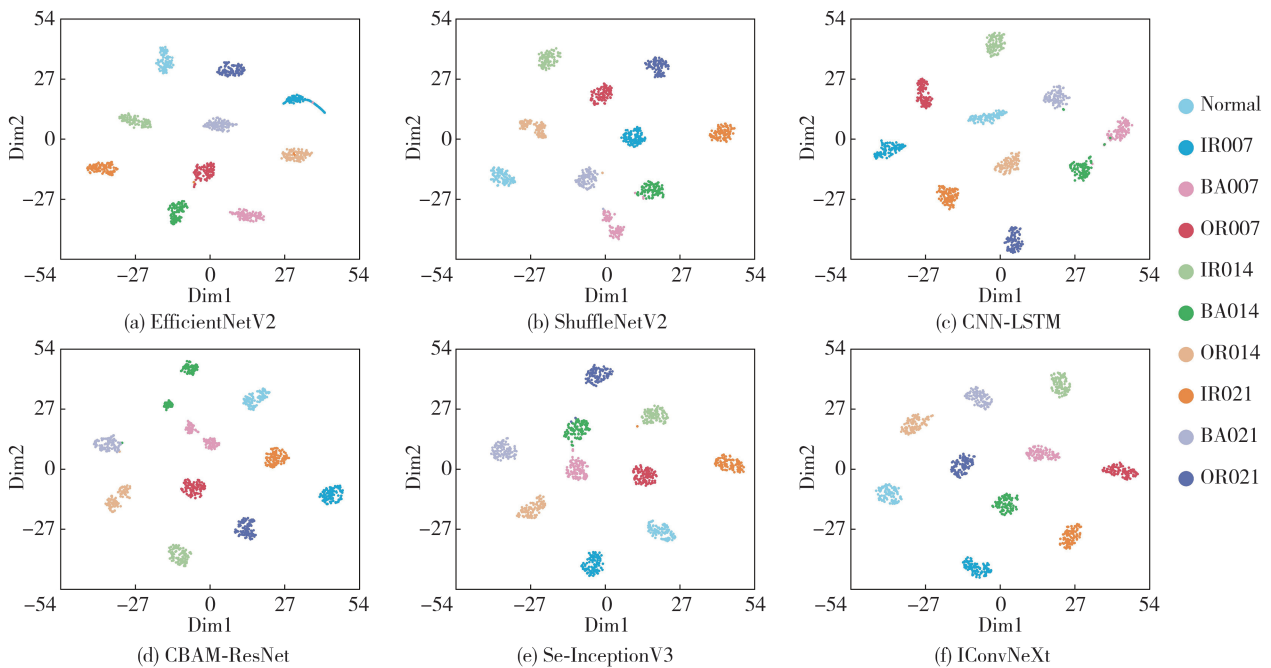


Fig. 6 T-SNE visualization results of different methods under working condition 0

3.1.6 Comparison of model performance under different noise environments

To assess the performance of the proposed method under noise interference, we selected EfficientNetV2, ShuffleNetV2, CNN-LSTM^[31], CBAM-ResNet^[32], and SE-InceptionV3^[33] as comparison models and tested their fault diagnosis performance under different signal-to-noise ratios (−4 dB, −2 dB, 2 dB, 4 dB) in operating condition 0. Table 6 shows the comparison results of accuracy and training time for each method.

At low signal-to-noise ratios (e. g., −4 dB), the proposed method exhibits a marked advantage, attaining 98.70% accuracy. The accuracy of this model (98.70%) is 0.92 percentage points higher than that of the second-best EfficientNetV2 (97.78%); while the accuracy of ShuffleNetV2 is only 93.44%. At a higher signal-to-noise ratio of 4 dB, the proposed method achieves an accuracy of 99.74%, exceeding that of the comparison models (98.82% to 98.98%) as its performance continues to enhance with increasing SNR. With an impressive average accuracy of 99.29%, the proposed method distinctly outperforms the other models, validating its stability and generalisation ability in noisy environments. In terms of training efficiency, the proposed method completes training in just 386 s, significantly faster than EfficientNetV2 (667 s) and CNN-LSTM (595 s), and close to the lightweight model ShuffleNetV2 (346 s). These results confirm that structural optimization enables the proposed method to maintain both high accuracy and computational efficiency.

Table 6 Comparison of accuracy of each method at different signal-to-noise ratios for condition 0

Method	Accuracy/%				Average/%	Training time/s
	Signal-to-noise/dB					
	−4	−2	2	4		
EfficientNetV2	97.78	98.18	98.92	98.98	98.47	667
ShuffleNetV2	93.44	97.10	98.62	98.82	97.00	346
CNN-LSTM ^[31]	94.80	97.66	98.62	98.86	97.49	595
CBAM-ResNet ^[32]	97.32	98.42	98.84	98.96	98.39	559
SE-InceptionV3 ^[33]	97.34	98.20	98.80	98.88	98.31	426
ICovNeXt	98.70	99.24	99.48	99.74	99.29	386

3.2 Experiments and analysis of synchronous motor electrical fault datasets

The synchronous motor electrical fault dataset is constructed based on a three-phase synchronous motor system driven by an inverter. The dataset comprises five types of motor electrical faults: open-phase faults located between the inverter and motor; inter-phase short-circuit

or leakage current faults; faults affecting the neutral point; rotor excitation voltage disconnection faults; and variations in the rotor excitation current. Experimental data were collected in a time-series format for four types of signal features, including stator three-phase voltage, stator three-phase current, rotor current, and motor speed, forming a four-dimensional multi-channel signal dataset. According to the original dataset study, the combination of stator three-phase voltage and stator three-phase current achieves 97% classification accuracy, while the combination involving rotor current with stator voltage or stator current reaches 100%. In this experiment, we selected stator three-phase voltage and stator three-phase current as the multi-sensor input to the model, aiming to explore whether this combination can further enhance classification performance. Each dataset contained five fault types and one normal state, divided into training, validation, and test subsets at 70%, 20%, and 10% proportions.

3.2.1 Ablation experiments

To validate the module's efficacy, supplementary ablation experiments were conducted. The experimental setup remains consistent with the preceding study, with results presented in Table 7. The baseline ConvNeXt-T model is served as the reference standard. Incorporating the ghost convolution module reduces model parameters and training time, though diagnostic accuracy consequently declines. Whilst the ELA module slightly increases computational overhead, it enhances feature representation capabilities and improves diagnostic accuracy. When the two modules are used together, the model reaches the best balance, achieving both higher accuracy and lower complexity. This shows that ghost convolution and ELA work well with each other, and it also proves that the overall model design is reliable.

Table 7 Comparison of ablation experiment results on synchronous motor electrical fault dataset

Method	Accuracy/%	Parameters /M	Training time/s
ConvNeXt-T	94.62	39.33	564
ConvNeXt-T+Ghost	94.38	36.39	516
ConvNeXt-T+ELA	95.76	39.47	572
ConvNeXt-T+Ghost+ELA	95.89	36.53	549

3.2.2 Performance comparison between fusion and single-source methods

To demonstrate the advantage of fusing data from multiple sensors for fault diagnosis enhancement, this section compares and analyzes three data input methods based on the improved ConvNeXt model: (a) stator three-phase voltage, (b) stator three-phase current,

and (a-b) the fusion of the two. The loss and accuracy curves of the model under various data input methods are illustrated in Fig.7.

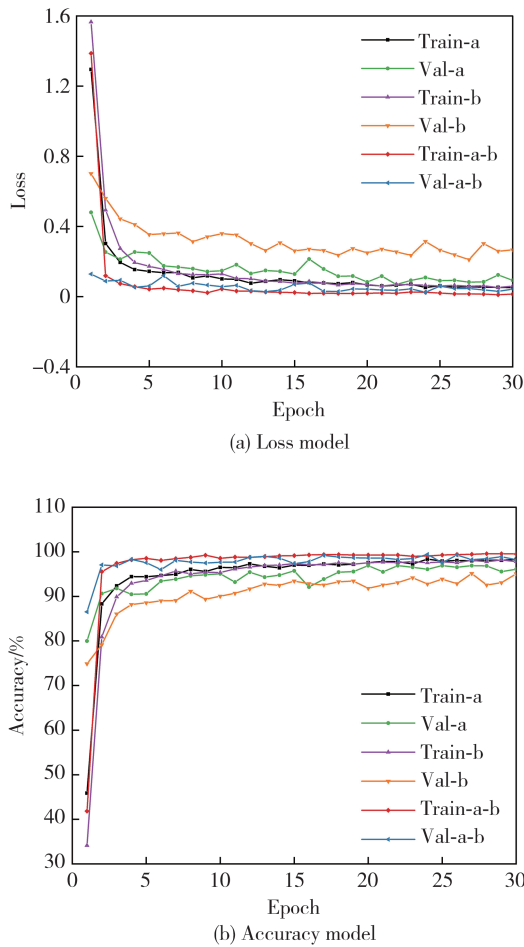


Fig. 7 Graphs of model loss and accuracy for different data input methods

The fused data (a-b) exhibits faster convergence and higher accuracy during both training and validation. Table 8 presents the model's diagnostic accuracy across various data inputs. The accuracy rates of single voltage data (a) and current data (b) are 97.22% and 96.49%, respectively, while the accuracy rate of the combined

data (a-b) is 99.33%, validating that multi-sensor data fusion effectively enhances the model's ability to characterise complex fault modes through complementary feature extraction.

Table 8 Diagnostic results for different data entry methods

Method	Accuracy/%
ICovNeXt-a	97.22
ICovNeXt-b	96.49
ICovNeXt-a-b	99.33

3.2.3 Comparison of different model performance

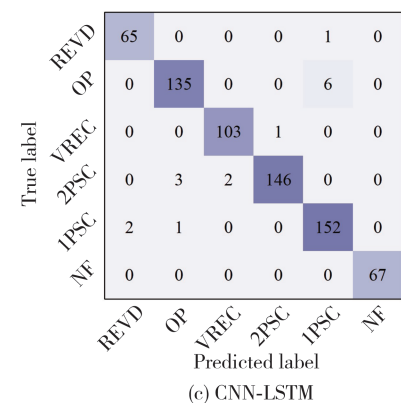
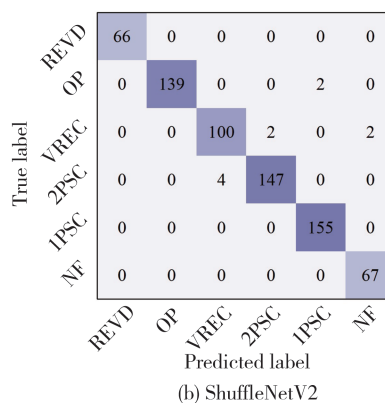
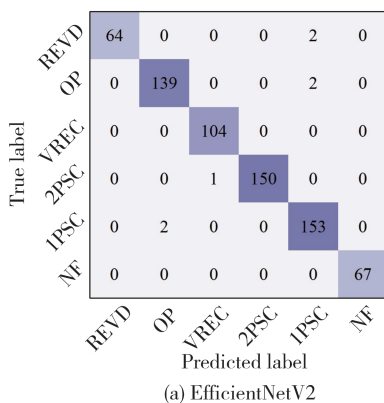
We evaluate the effectiveness of the proposed method by comparing it with popular models such as EfficientNetV2 and ShuffleNetV2. The results are detailed in Table 9.

Table 9 Comparison of performance of different models

Method	Accuracy/%	Training time/s
EfficientNetV2	98.40	750
ShuffleNetV2	98.19	576
CNN-LSTM ^[31]	98.33	609
CBAM-ResNet ^[32]	98.26	830
SE-InceptionV3 ^[33]	98.23	560
ICovNeXt	99.33	580

The proposed method achieves a significantly higher average accuracy of 99.33% compared to other models, while the training time (580 s) remains at a relatively low level. While EfficientNetV2 and CBAM-ResNet achieves relatively high accuracies of 98.40% and 98.26%, respectively, their training times are as high as 750 s and 830 s, resulting in significantly increased computational costs. ShuffleNetV2 and SE-InceptionV3 demonstrate better training efficiency (576 s and 660 s, respectively), but their accuracies are both below 99.00%.

Fig. 8 shows the confusion matrices of different models. It indicates that the proposed method achieved 100.00% classification accuracy in almost all categories, with only two samples misclassified in the OP and 2PSC categories. In contrast, other methods exhibited varying degrees of confusion.



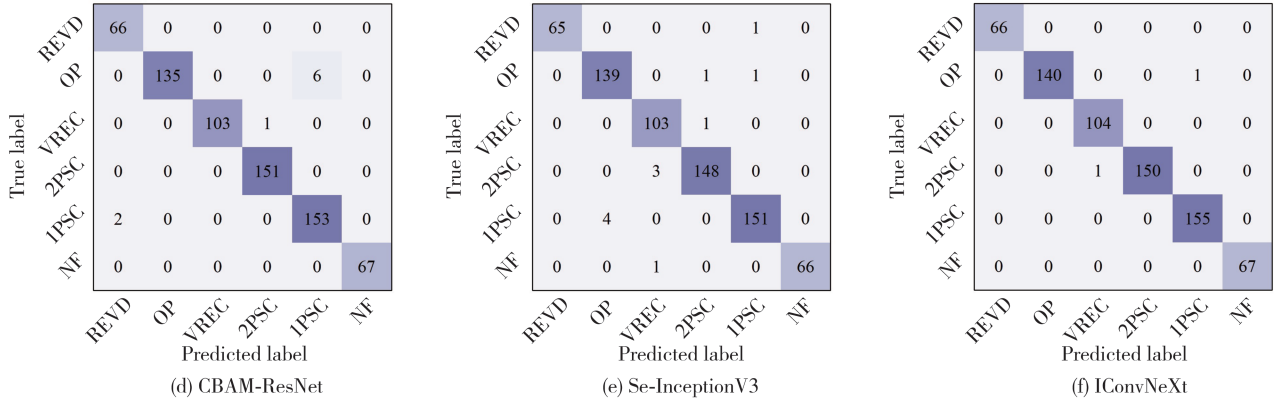


Fig. 8 Confusion matrix of different methods

The t-SNE visualisation results of different models are shown in Fig. 9. The features extracted by the proposed method exhibit distinct intra-class clustering and inter-class separation characteristics in the low-dimensional space, with clear boundaries between

different fault categories, indicating the model's strong representational capability. In contrast, the feature clusters of other comparison models exhibit local overlaps, confirming their deficiency in classification accuracy.

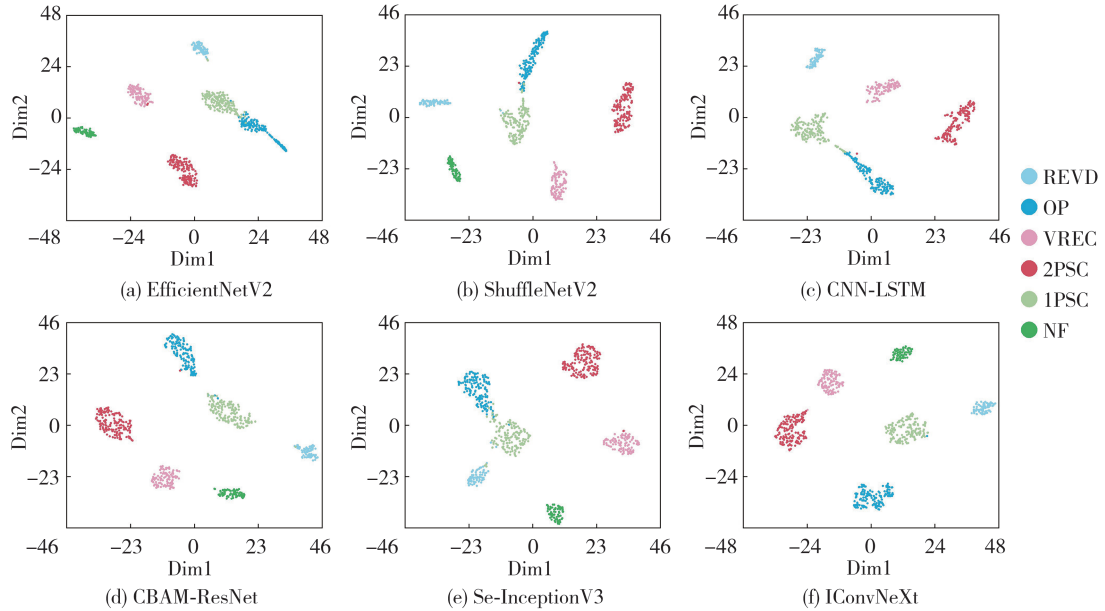


Fig. 9 Visualisation results of t-SNE with different methods

3.2.4 Comparison of model performance under different noise environments

To demonstrate the robustness of the proposed method against noise, a performance comparison among various models was performed under different noise levels. Table 10 shows the experimental data. Significant improvements in noise robustness and computational efficiency are achieved by the proposed method. Under noise conditions ranging from -4 dB to 4 dB, the proposed method achieves accuracy rates of 98.33% , 98.88% , 99.28% , and 98.92% , respectively, with an average accuracy rate of 98.85% , exceeding all comparison models. Additionally, the proposed method requires only 559 s for training, representing a 31.5% improvement in efficiency

compared to the larger-parameter CBAM-ResNet (816 s), while maintaining a training speed comparable to that of lightweight models (ShuffleNetV2: 590 s).

Table 10 Comparison of performance of different methods under different signal-to-noise ratios

Method	Accuracy/%				Average/%	Training time/s
	Signal-to-noise/dB					
	−4	−2	2	4		
EfficientNetV2	95.88	95.06	95.26	96.81	95.75	670
ShuffleNetV2	96.96	97.58	96.58	97.25	97.09	590
CNN-LSTM ^[31]	97.55	97.56	98.06	98.30	97.86	598
CBAM-ResNet ^[32]	98.17	98.47	98.36	98.44	98.36	816
SE-InceptionV3 ^[33]	97.34	98.20	98.80	98.00	98.09	550
IConvNeXt	98.33	98.88	99.28	98.92	98.85	559

4 Conclusions

This paper addresses the issues of low diagnostic accuracy and poor generalisation capability in motor equipment fault diagnosis under noisy interference and complex operating conditions. We proposed a diagnostic method based on EWT and an improved ConvNeXt network. By combining EWT with a comprehensive evaluation metric of cosine similarity-kurtosis-envelope entropy, we effectively addresses the challenge of extracting features from multi-source signals in noisy environments. Additionally, by leveraging a fusion mechanism of multi-sensor time-frequency images across channel dimensions, the method enhances the global representation of fault features. In terms of model improvement, by incorporating the ghost convolution module and ELA, the model achieves reduced computational complexity without compromising the feature extraction strengths of the ConvNeXt-T network. Through various practical fault diagnosis scenarios, the method's effectiveness has been thoroughly demonstrated, showing excellent adaptability to changing operating states and noise environments.

Nevertheless, diagnostic methods utilizing deep learning generally depend on a substantial amount of labelled samples for model training, while fault data in industrial environments are typically scarce and imbalanced. Under small-sample conditions, models are prone to overfitting, resulting in a marked drop in diagnostic accuracy. Future research will focus on small-sample learning methods to mitigate the overfitting issues caused by the limited availability of motor fault data in industrial scenarios and enhance the generalization capability of diagnostic models under limited sample conditions.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No.12172157), the Key Project of Natural Science Foundation of Gansu Province (No.25JRRA150), and Lanzhou Science and Technology Plan Project (No.2023-1-16).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

- [1] KUMAR P, HATI A S. Review on machine learning algorithm based fault detection in induction motors. *Archives of Computational Methods in Engineering*, 2021, 28(3): 1929-1940.

- [2] ALSHORMAN O, IRFAN M, ABDELRAHMAN R B, et al. Advancements in condition monitoring and fault diagnosis of rotating machinery: a comprehensive review of image-based intelligent techniques for induction motors. *Engineering Applications of Artificial Intelligence*, 2024, 130: 107724.
- [3] LUO P E, YIN Z G, YUAN D S, et al. An intelligent method for early motor bearing fault diagnosis based on Wasserstein distance generative adversarial networks meta learning. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 3517611.
- [4] LIU P J, GUO Z C, HE L, et al. Fault diagnosis method of AC motor rolling bearing based on heterogeneous data fusion of current and infrared image. *Journal of Measurement Science and Instrumentation*, 2024, 15(4): 558-570.
- [5] LIANG P F, WANG W H, YUAN X M, et al. Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment. *Engineering Applications of Artificial Intelligence*, 2022, 115: 105269.
- [6] WANG S Q, FENG Z G. Multi-sensor fusion rolling bearing intelligent fault diagnosis based on VMD and ultra-lightweight GoogLeNet in industrial environments. *Digital Signal Processing*, 2024, 145: 104306.
- [7] LÓPEZ C, NARANJO Á, LU S L, et al. Hidden Markov model based stochastic resonance and its application to bearing fault diagnosis. *Journal of Sound and Vibration*, 2022, 528: 116890.
- [8] YAN R Q, SHANG Z G, XU H, et al. Wavelet transform for rotary machine fault diagnosis: 10 years revisited. *Mechanical Systems and Signal Processing*, 2023, 200: 110545.
- [9] ZHANG H Y, ZHENG Y Q, LU J S. A novel bearing current signal diagnosis method combining variational modal decomposition and improved random forests. *The Review of Scientific Instruments*, 2024, 95(2): 025111.
- [10] YANG Y, LIU H, HAN L J, et al. A feature extraction method using VMD and improved envelope spectrum entropy for rolling bearing fault diagnosis. *IEEE Sensors Journal*, 2023, 23(4): 3848-3858.
- [11] LIU F Z, WANG H M, LI W, et al. Fault diagnosis of rolling bearing combining improved AWSGMD-CP and ACO-ELM model. *Measurement*, 2023, 209: 112531.
- [12] ZHAO K C, XIAO J Q, LI C, et al. Fault diagnosis of rolling bearing using CNN and PCA fractal based feature extraction. *Measurement*. 2023, 223: 113754.
- [13] WANG X D, LIU Z L, DAI M H, et al. Time-shift denoising combined with DWT-enhanced condition domain adaptation for motor bearing fault diagnosis via current signals. *IEEE Sensors Journal*, 2024, 24(21): 35019-35035.
- [14] WANG J, SHAO H D, HE J, et al. A novel interpretable fault diagnosis method using multi-image feature extraction and attention fusion. *Pattern Recognition Letters*, 2025, 189: 38-47.
- [15] BAI R X, MENG Z, XU Q S, et al. Fractional Fourier and

- time domain recurrence plot fusion combining convolutional neural network for bearing fault diagnosis under variable working conditions. *Reliability Engineering & System Safety*, 2023, 232: 109076.
- [16] ZHAO X F, WANG L B, YANG M S, *et al.* A novel small-sample fault diagnosis method for rolling bearings via continuous wavelet transform and Siamese neural network. *IEEE Sensors Journal*, 2024, 24 (15): 24988-24996.
- [17] DAI X, YI K, WANG F L, *et al.* Bearing fault diagnosis based on POA-VMD with GADF-Swin Transformer transfer learning network. *Measurement*, 2024, 238: 115328.
- [18] GUO B S, QIAO Z H, ZHANG N, *et al.* Attention-based ConvNeXt with a parallel multiscale dilated convolution residual module for fault diagnosis of rotating machinery. *Expert Systems with Applications*, 2024, 249: 123764.
- [19] WANG K J, WANG W Q, ZHAO Y B, *et al.* Multisensor fault diagnosis via Markov chain and Evidence theory. *Engineering Applications of Artificial Intelligence*, 2023, 126: 106851.
- [20] LIN B, ZHU G H, ZHANG Q H, *et al.* A novel framework for bearing fault diagnosis across working conditions based on time-frequency fusion and multi-sensor data fusion. *Measurement Science and Technology*, 2024, 35(12): 126205.
- [21] WENG C Y, LU B C, GU Q, *et al.* A novel multisensor fusion transformer and its application into rotating machinery fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 3507512.
- [22] LI X M, WANG Y X, YAO J C, *et al.* Multi-sensor fusion fault diagnosis method of wind turbine bearing based on adaptive convergent viewable neural networks. *Reliability Engineering & System Safety*, 2024, 245: 109980.
- [23] GUO H Y, YU W, ZHANG X G, *et al.* Intelligent mechanical fault diagnosis using multiscale residual network and multisensor fusion. *Measurement Science and Technology*, 2024, 35(11): 116007.
- [24] XIAO X Q, LI C S, HE H X, *et al.* Rotating machinery fault diagnosis method based on multi-level fusion framework of multi-sensor information. *Information Fusion*, 2025, 113: 102621.
- [25] HE Y L, LIU Z H, ZHANG W, *et al.* Eccentricity fault diagnosis of permanent magnet synchronous generators based on 2D recursive fusion graph and CBAM-ConvNeXt-FPN. *Measurement Science and Technology*, 2025, 36(5): 056110.
- [26] GILLES J. Empirical wavelet transform. *IEEE Transactions on Signal Processing*, 2013, 61(16): 3999-4010.
- [27] TANG T, QIU C H, YANG T Y, *et al.* A novel lightweight relation network for cross-domain few-shot fault diagnosis. *Measurement*, 2023, 213: 112697.
- [28] XU W, WAN Y. ELA: efficient local attention for deep convolutional neural networks. 2024: arXiv: 2403. 01123.
- [29] SMITH W A, RANDALL R B. Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study. *Mechanical Systems and Signal Processing*, 2015, 64: 100-131.
- [30] SUN Z L, MACHLEV R, WANG Q C, *et al.* A public dataset for synchronous motor electrical faults diagnosis with CNN and LSTM reference classifiers. *Energy and AI*, 2023, 14: 100274.
- [31] ERTARGIN M, YILDIRIM O, ORHAN A. Mechanical and electrical faults detection in induction motor across multiple sensors with CNN-LSTM deep learning model. *Electrical Engineering*, 2024, 106(6): 6941-6951.
- [32] CHANG M, YAO D C, YANG J W. Intelligent fault diagnosis of rolling bearings using efficient and lightweight ResNet networks based on an attention mechanism (September 2022). *IEEE Sensors Journal*, 2023, 23(9): 9136-9145.
- [33] XU L F, TEOH S S, IBRAHIM H. A deep learning approach for electric motor fault diagnosis based on modified InceptionV3. *Scientific Reports*, 2024, 14: 12344.

基于 EWT 和改进 ConvNeXt 网络的故障诊断方法

李 军*, 马 轶

兰州交通大学 自动化与电气工程学院, 甘肃 兰州 730070

摘 要: 由于强噪声的干扰, 特征提取面临信息受限的挑战, 不利于电机设备故障诊断。本文提出一种基于经验小波变换(EWT)和改进 ConvNeXt 网络的故障诊断方法。首先, 通过经验小波变换方法分别提取来自不同传感器信号的模态分量, 去除噪声后进行信号重构。其次, 使用 STFT 将降噪重构后的一维信号转换为增强信号特征的二维时频谱图像, 并将单一传感器生成的单通道图像进行融合, 形成多通道图像, 以提升 ConvNeXt 网络的特征提取能力。同时, 将 Ghost 卷积模块和高效局部注意力机制(ELA)引入到 ConvNeXt-T(ConvNeXt-Tiny)网络中, 增强了网络的性能。针对不同故障诊断设备数据集进行了实验验证, 并与现有主流深度学习方法如 SE-InceptionV3、CBAM-ResNet 和 CNN-LSTM 等进行了对比, 结果表明, 在不同噪声环境和变工况的试验中, 本文方法具有较高的诊断准确率和泛化能力。

关键词: 故障诊断; ConvNeXt; 经验小波变换; 短时傅里叶变换; 高效局部注意力机制; 数据级融合

引用格式: LI Jun, MA Yi. Fault diagnosis method based on EWT and improved ConvNeXt networks. *Journal of Measurement Science and Instrumentation*, 2026, 17(2): 307-319. DOI: 10.62756/jmsi.1674-8042.2026026