Multi source heterogeneous data diagnosis method of rotating machinery based on parameter collaborative optimization of multi-scale convolutional autoencoder

Xiaoli Yang
Huanghe University of Science and Technology, Zhengzhou, 450000, China
E-mail: *3397685388@qq.com*

Check for updates

Abstract. In order to fully utilize the features of multi-source heterogeneous data and effectively improve the accuracy and efficiency of fault diagnosis of rotating machinery, a multi-source heterogeneous data diagnosis method based on parameter collaborative optimization multi-scale convolutional autoencoder (MSCAE) is proposed. Firstly, multi-scale information learning is integrated into the convolutional autoencoder (CAE) to consider the temporal and spatial feature information of the diagnostic object simultaneously. To improve the training and diagnostic efficiency of MSCAE, a quantum particle swarm optimization (QPSO) module is used to perform hyperparameter optimization on it using chaos initialization and dynamic weight strategy (DWS). Besides, the sparse attention mechanism is introduced into the MSCAE model to improve the recognition rate of key fault features hidden in the original heterogeneous signals. Finally, the confusion matrix and visualization techniques are used to achieve fault classification. The experimental results demonstrate that after 100 experiments, the proposed method has an average diagnostic accuracy of 98.5 % and strong robustness to noise, providing a new method for rotating machinery fault diagnosis based on multi-source heterogeneous data.

Keywords: quantum particle swarm optimization, dynamic weight strategy, MSCAE, multi source heterogeneous data, rotating machinery.

1. Introduction

With the popularization of digital transformation in manufacturing enterprises, enterprises usually deploy multiple sensors in the industrial production process to monitor the status of equipment, and collect a large amount of multi-source heterogeneous data, such as vibration signals collected by vibration sensors and current signals collected by current sensors and other time-series data. These data exhibit characteristics such as large volume, heterogeneous sources, and low value density. Although their sources and manifestations are different, they are all descriptions of the status of the same device. Different data sources complement each other, which can avoid the state evaluation bias caused by one single sensor signal and reflect the condition of the device comprehensively from different perspectives [1].

In recent years, with the rapid development of deep learning in the fault diagnosis field [2-6], the achievements in fault diagnosis based on multi-source heterogeneous data and deep learning also have been particularly outstanding. Reference [7] proposed a bearing fault diagnosis method that combines multi-source data fusion with improved attention mechanism. By introducing attention mechanism, the information weights of different data sources can be adjusted dynamically, with good robustness and adaptability. Reference [8] proposed a selective convolutional deep residual network based on channel space attention mechanism and feature fusion. This network utilizes channel space attention mechanism to assign different weights to different channels, and integrates the features of different channels and different levels of feature information, which improves the efficiency of fault recognition. Reference [9] proposed a hybrid deep neural network designed specifically for estimating remaining useful life. This network utilizes LSTM to extract temporal features from multi-sensor data and CNN to extract spatial

features, demonstrating better generalization ability in complex prediction scenarios with increasing operating conditions and fault modes. Reference [10] utilized generative adversarial networks to extract features from multi-sensor data, and finally used variational autoencoders to achieve bearing state evaluation and remaining service life prediction, providing a new solution for equipment health management based on multi-sensor data. Reference [11] concatenated multi-sensor signals and inputted them into a multi-channel fusion feature extractor, and combined them with a bidirectional long short-term memory (BiLSTM) network to construct a bearing health index. Reference [12] proposed a convolutional autoencoder with improved loss function, which can capture the degradation process of bearings effectively, extract the health status features of bearings from multi-sensor data, and avoid the loss of local information. Reference [13] merged each sensor data by channel to achieve efficient fusion of multi-sensor data. Reference [14] proposed a multimodal convolutional neural network that can learn fault related sensitive features from vibration signals and infrared images automatically, and fuse these deep features using T-SNE (T-Distributed Stochastic Neighbor Embedding, T-SNE). Reference [15] used one-dimensional CNN to extract features from vibration signals and sound signals respectively, and fused these features. Compared with the algorithm based on single mode sensor, this method had higher diagnostic accuracy. Literature [16] proposed an improved multi head attention mechanism to extract the features of different sensor signals, and designed a bilinear model to achieve fine fusion of features, which can extract complementary fault information between different signals effectively. Reference [17] proposed a deep coupled autoencoder model for learning joint features of vibration and acoustic signals. Reference [18] alternately extracted and fused one-dimensional temporal data features and two-dimensional image data features based on stacked autoencoders, fully utilizing the correlation and complementarity of multi-source heterogeneous data. Reference [19] proposed an integrated network based on multiple CNN and BiLSTM with compression excitation mechanism in parallel. Reference [20] proposed a convolutional long short-term memory (CLSTM) network, which used CNN to capture shallow features of individual sensors, and then captures deep features through CLSTM. Reference [21] proposed a novel multi-sensor selection framework. The framework can determine adaptively which sensors were used to predict the remaining service life, and explained how different sensors affect the final result of the remaining service life prediction over time. Reference [22] proposed an end-to-end multi-sensor data fusion method based on LSTM encoder decoder structure. Reference [23] used multiple different deep belief networks (DBNs) to extract data features with different sources and structures, and proposed a weight matrix to fuse the decision results of different DBNs. Reference [24] proposed a multi-sensor data fusion method for device health indicators.

Although the above stated deep learning diagnostic models based on multi-source heterogeneous data have achieved high diagnostic accuracy, most of them have not been evaluated based on comprehensive factors such as multi-source heterogeneous feature extraction effectiveness and model diagnostic efficiency to assess the effectiveness of constructed model. Based on this, this paper proposes a multi-source heterogeneous data diagnosis method based on parameter collaborative optimization of MSCAE. Compared to most the current idea of building models based on one single factor, the hybrid model is constructed based on comprehensive factors such as multi-source heterogeneous feature extraction effectiveness, model diagnostic efficiency, and noise resistance, which is more comprehensive and applicable. The main contributions of the paper are as follows:

1) A multi-scale information learning module considering different spatial scales is integrated into CAE to construct a new model named as MSCAE, enhancing its ability to extract features from heterogeneous multi-source information while considering feature information at different spatial scales to improve its noise suppression capability.

2) The sparse attention mechanism is introduced into the MSCAE model to further improve the recognition rate of the model for key fault features.

3) To improve the training and diagnostic efficiency of MSCAE, a quantum particle swarm

optimization module is used to perform hyperparameter optimization on the MSCAE model using chaos initialization and dynamic weight strategy.

The remains of the paper are arranged as following: Section 2 is dedicated to the related theories. Section 3 presents the proposed integrated model and diagnostic process. Experiment verification and conclusion are presented in Section 4 and Section 5 respectively.

2. Related theories

2.1. Multi-scale convolutional autoencoder

The basic structure of CAE consists of an encoder and a decoder, as shown in Fig. 1. The encoder gradually reduces the spatial dimension of data through convolutional and pooling layers, while increasing the number of feature channels to extract more abstract feature representations. On the other hand, the decoder gradually restores the spatial dimension of the data through the deconvolution layer (or transposed convolution layer), while reducing the number of feature channels needed to reconstruct the original data.
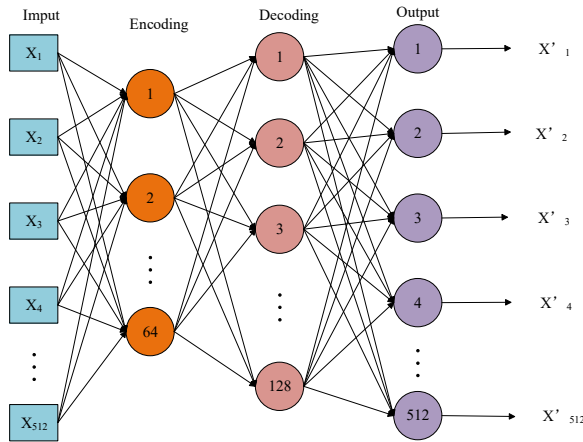

Fig. 1. The basic structure of CAE

The encoder part can be represented as follows:

$$h = f(X) = f_{encoder}(X), \tag{1}$$

where h represents the decoded feature. The encoder part can be implemented using convolution and pooling operations, which can be expressed as follows:

$$h = \sigma(Conv(X, W) + b), \tag{2}$$

where $Conv(X, W)$ performs convolution operation on the input data X. W is the convolution kernel, b represents the bias term, and σ is the activation function (which can be sigmoid or ReLU function).

The decoder part can be represented as follows:

$$X' = g(h) = g_{decoder}(h), \tag{3}$$

where X' represents the reconstructed data. Convolution and upsampling operations can be used to implement the decoder part, as follows:

$$X' = \sigma(Conv(Upsample(h), W') + b'). \tag{4}$$

In the formula, $Upsample(h)$ represents the upsampling operation on the decoded features, W' is the convolution kernel of the decoder, b' is the bias term of the decoder, and σ is the activation function.

The encoding in traditional CAE mainly uses fixed scale convolutional kernels for feature extraction. The kernel scale of each convolutional layer remains unchanged, limiting the ability to extract classification information at different resolutions and kernel scales. Especially in the case of variable operating condition, extracting deeper level features for diagnosis becomes more challenging. The structure of MS-CAE is adding a layer of kernels with different scale before the encoding layer of original CAE, as shown in Fig. 2, and corresponding multi scale information learning could be realized. Multi scale information learning can capture information more comprehensively and diversely by introducing convolution operators or pooling operations of different scales, further enhancing the model's understanding of data. Therefore, integrating multi-scale information learning modules into the encoding of CAE can help constructing new MSCAE models and discover more elastic key features. This module combines kernels of different scales and considers feature information at different spatial scales, thereby improving the ability to suppress noise and redundant feature information, and enhancing the model's representation and classification performance of data. The MSCAE model encodes information during the network training phase, and the defined loss function is as following:

$$Loss = H\big(p(x), q(x)\big) = -\sum_{x} p(x) \log q(x), \tag{5}$$

where $Loss$ refers to the loss function, $p(x)$ is the target distribution, and $q(x)$ is the estimated distribution.

The gradient of the backpropagation output layer for MSCAE can be described roughly as following:

$$\delta^L = \frac{\partial L(w,b)}{\partial z^L} = \frac{\partial L(w,b)}{\partial x^L} \otimes \sigma'(z^L), \tag{6}$$

where δ is the gradient, $\otimes$ represents the operation of the matrix, and ∂ represents solving for the partial derivative. The function of Eq. (6) is to avoid gradient explosion of MSCAE.
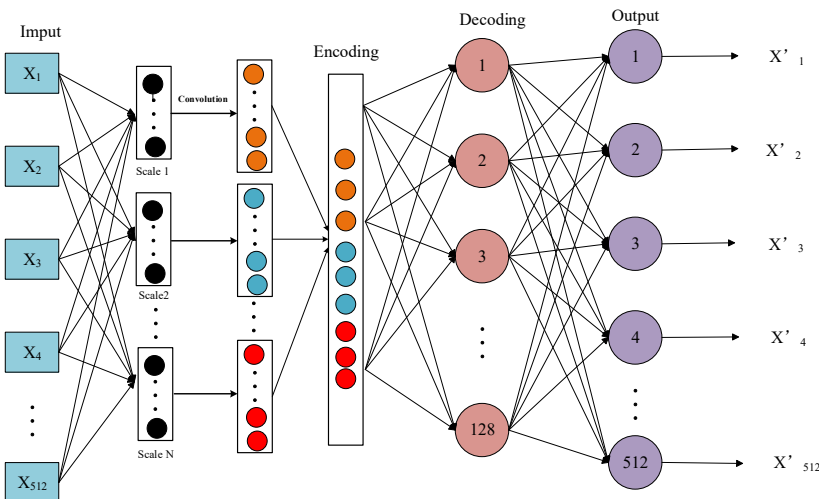

Fig. 2. The basic structure of MS-CAE

2.2. Sparse attention mechanism

Sparse attention is an optimized attention mechanism. It can map a query vector with a set of key value pairs to an output vector accurately. Sparse attention mechanisms typically include the following key steps:

Firstly, for each element in the input sequence, calculate its corresponding score to indicate the importance of the element. This score is usually calculated based on the characteristics of the input elements and some learnable parameters. Then introduce sparsity and focus attention weights to zero in some way (such as L1 regularization), so that the model only focuses on the most important elements in the input sequence. Then normalization: convert the attention score into a probability distribution, so that the sum of the attention weights of all elements is 1. Finally, it is weighted and merged by using normalized attention weights to weight the input sequence, and then a merging operation (such as summation or averaging) is performed to obtain the final output representation.

The sparse attention mechanism is specifically expressed as follows based on the above stated:

Assuming the input sequence is X, the learnable weight matrix is W, where the unit is the output dimension of the attention layer. Firstly, the attention score can be calculated using the following formula:

$$score(t) = X[:, t, :]W, \tag{7}$$

where t represents the time interval of the input sequence.

The sparsity (L_1 regularization) calculation is as following:

$$loss = -\sum_{t=1}^{sequence_{length}} \log(score(t)) + \lambda ||score||_1, \tag{8}$$

where λ is the regularization coefficient. The function of Eq. (8) is to prevent overfitting.

The normalization calculation is as following:

$$a(t) = \frac{\exp(score(t))}{\sum_{t'=1}^{sequence_{length}} \exp(score(t'))}. \tag{9}$$

The final output is as following:

$$output = \sum_{t=1}^{sequence\_length} a(t)X[:, t, :]. \tag{10}$$

2.3. Parameter optimization method based on QPSO-DWS

2.3.1. Dynamic weight strategy

DWS is used to adjust the inertia weights in the QPSO algorithm to balance global and local search capabilities. The update formula for inertia weight is as following:

$$\omega = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{T} \times t, \tag{11}$$

where ω_{max} and ω_{min} are the maximum and minimum values of the inertia weight, T is the maximum number of iterations, and t is the current number of iterations.

2.3.2. Quantum particle swarm optimization

The QPSO [25] is a global optimization algorithm based on the principles of quantum mechanics. In QPSO, the formula for updating the position of particles is as following:

$$x_i^{t+1} = p + (-1)^{rand(0,1)} \cdot \alpha \cdot | M^t - x_i^t | - \ln\left(\frac{1}{u}\right), \tag{12}$$

where x_i^t is the position of the ith particle at the tth iteration:

$$p = w \cdot P_i + (1 - w) \cdot P_{gbest}, \tag{13}$$

where p_i is the individual optimal position of the particle i, P_{gbest} is the global optimal position, M^t is the average of the optimal positions of all particle individuals, α is the step factor, u is the random number in the range [0, 1], and rand(0,1) is the random generation of 0 or 1.

2.3.3. Quantum particle swarm optimization-dynamic weight strategy

The QPSO-DWS algorithm combines the QPSO algorithm with DWS, which can search the hyperparameter space and find the optimal hyperparameter combination more effectively. Specifically, in each iteration, the inertia weight is adjusted according to the DWS, and then the particle position is updated based on the QPSO position update formula to evaluate the fitness of the particles, and update the individual and global optimal positions.

3. The proposed integrated model and diagnostic process

On the basis of the proposed MSCAE, this article combines it with sparse attention mechanism to propose an ensemble model. This ensemble model combines convolution kernels of different scales to consider feature information at different spatial scales, while embedding a sparse attention mechanism specifically designed for processing multi-source signal data to capture its core features, thereby effectively improving the noise suppression and redundant feature information removal capabilities of the ensemble model.

The integrated model implants sparse attention mechanism between the convolutional layers of the autoencoder layers of MSCAE, with the specific formula as following:

$$h^{(l)} \begin{cases} = ConvBlock\left(h^{(l-1)}\right) \alpha^{(l)}, \\ = SparseAttention\left(h(l)\right) h_{att}^{(l)}, \\ = \alpha^{(l)} \odot h^{(l)} h^{(l+1)}, \\ = Pool\left(h_{att}^{(l)}\right), \end{cases} \tag{14}$$

where $\odot$ represents element multiplication operation to create attention guided feature maps before pooling.

The clustering centers and features generated by MSCAE are optimized using the following loss function:

$$L_{total} = L_{CE}(X, X') + \lambda_1 L_{cluster} + \lambda_2 R_{sparse} L_{cluster} = \sum_{i=1}^{k} \sum_{x \in Si} \|h(x) - \mu_i\|^2, \tag{15}$$

where μ_i is a learnable parameter that is updated through backpropagation.

During the training phase, the synchronization process is as follows:

Data Path: $X \; \underline{CAE + Attention} \rightarrow h \; \underline{K-means} \rightarrow \mu_i$,

Signal Path: $\mu_i \; \underline{Cluster-aware \; Decoder} \rightarrow X'$. $\qquad$ (16)

Reconstruct the decoder using cluster conditional deconvolution:

$$Deconv(h, \mu_i) = \sigma(W_i' * Upsample(h) + b_i'). \qquad (17)$$

Choosing an appropriate parameter optimization strategy is an important step in improving the efficiency of hybrid model training and diagnosis. To reduce the experimental time for parameter optimization of this model, the proposed QPSO-DWY algorithm is applied to the parameter optimization of the integrated model.

The diagnostic process based on the proposed integrated model and parameter optimization strategy is shown in Fig. 3, as follows:

1) Data preparation: Normalize the collected multi-source heterogeneous data and divide it into training and testing sets.

2) Feature extraction: Output the training set to the ensemble model for model training, and synchronously optimize the model hyperparameters based on the QPSO-DYS algorithm to complete the training of the ensemble model.

3) Classification and diagnosis: Input the test set into the trained ensemble model for fault classification.

4) Evaluation and visualization: using t-SNE for dimensionality reduction, model performance evaluation, and parameter optimization.
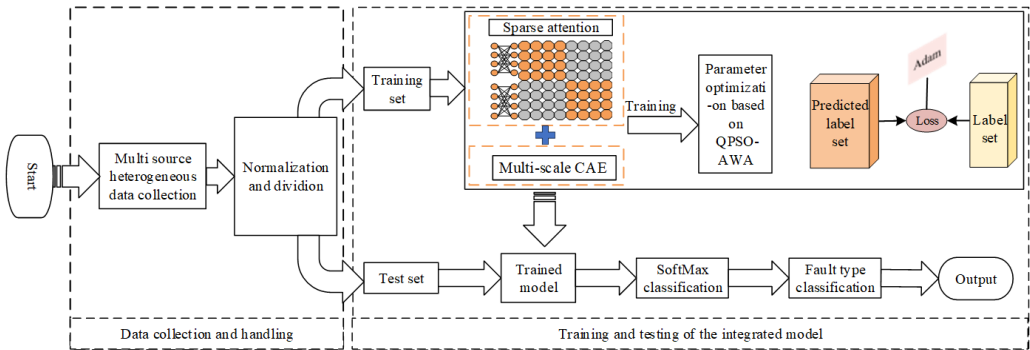

Fig. 3. Diagnosis flow chart based on the proposed hybrid model

4. Experiment

The method was validated using the NSK 6205 DDU bearing dataset from the Korea Mechanical Research Institute [26]. The experimental testing system consists of a three-phase induction motor, torque sensor, gearbox, two bearing seats, one rotor, and one hysteresis brake, as shown in Fig. 4. The bearing seats are designed in an easy disassembly mode to facilitate the installation of bearings of different running states, and the corresponding heterogeneous data of each running states are collected separately. Five operating states of rolling bearings including healthy bearing, Inner race fault (0.3 mm), Inner race fault (1.0 mm), Outer race fault (0.3 mm), Inner race fault (0.3 mm) are conducted, and the five states are represented by 0, 1, 2, 3 and 4 respectively. Fig. 5 shows the points of failure from the faulty bearings.

The test runs at speed of 3010 RPM, with torque applied to braking and load of 0, 2, and 4 Nm, respectively. Acoustic data is collected without load. Variable load test simulates bearing damage occurring on inner or outer race. The heterogeneous data of five types of running states include vibration, acoustic, and current data of healthy bearings under 0 Nm load, temperature data under

2 Nm load, inner and outer race fault data of 0.3 mm and 1.0 mm under 0 Nm load, and temperature data under 2 Nm load. The time-domain waveforms of the collected signals are shown in Fig. 6.
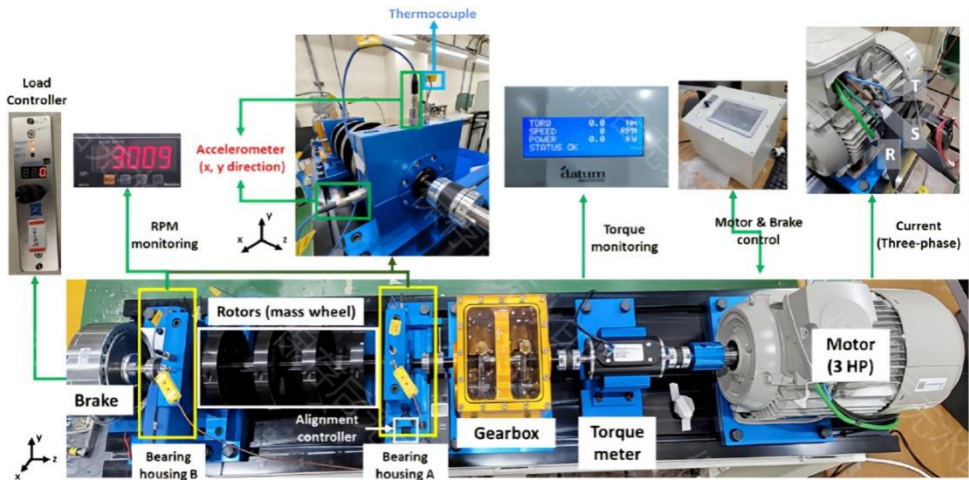

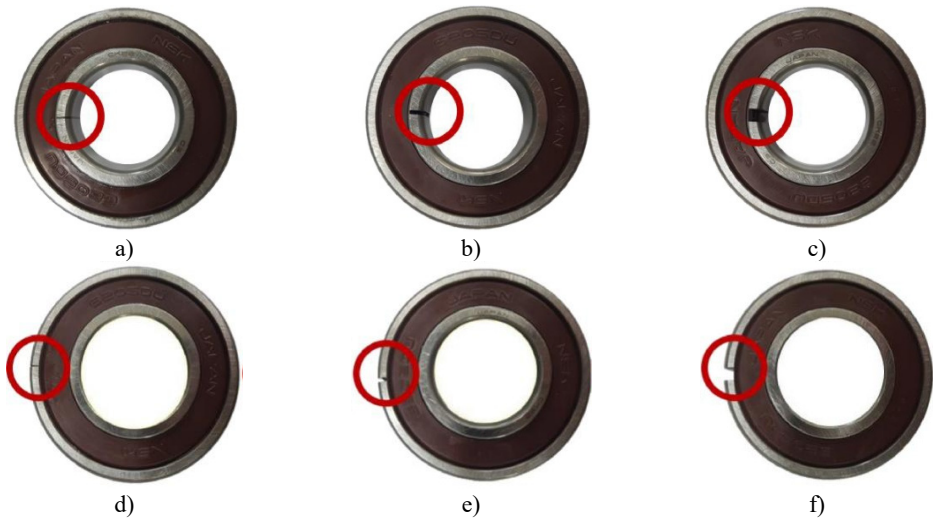**Fig. 4.** The test rig

a)   b)   c)

d)   e)   f)

Fig. 5. Bearing by crack size: a) inner race 0.3 mm, b) inner race 1.0 mm, c) inner race 3.0 mm, d) outer race 0.3 mm, e) outer race 1.0 mm, f) outer race 3.0 mm
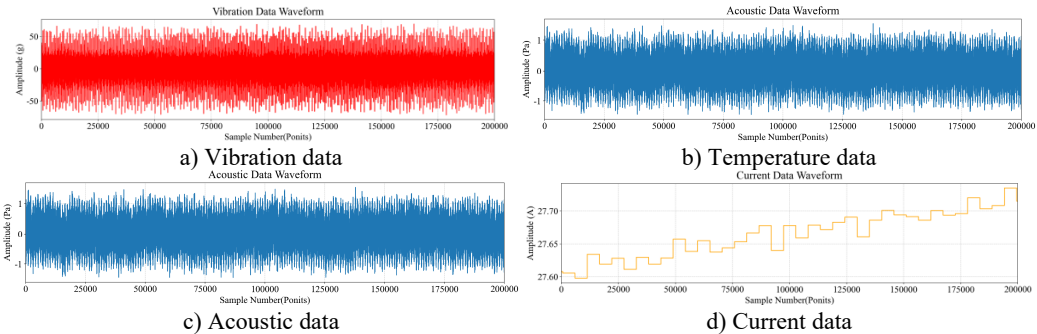

a) Vibration data   b) Temperature data

c) Acoustic data   d) Current data

Fig. 6. Time-domain waveforms of five kinds of heterogeneous data

In this study, standardization and sliding window methods are used to preprocess the network input data. Due to the large amount of single fault data, a window length of 4096 data points and a step size of 64 are used to create training data, which is randomly divided into datasets for training and testing. The details of multi-source sample division are shown in Table 1. Among them, the sliding window method is mainly used to extract multiple overlapping samples from the input sequence (as shown in Fig. 7). Then 38×4 sets of heterogeneous sample sets are chosen for each state, that is to say that there are all 760×4 sets of heterogeneous sample sets obtained for the five running states in the four loading states. Select 2900 sets of sample sets from 760×4 sets randomly and divide them into training set sand testing sets in 7:3 ratio, and the corresponding number of training sets and test sets 2030 (406×5) and 870 (174×5) respectively.

Table 1. Division table of training and test data

Data type	Running states	Loads	Total samples	Partitioning ratio of training set	Partitioning ratio of test set
Temperature	Healthy bearing Inner race fault (0.3 mm) Inner race fault (1.0 mm) Outer race fault (0.3 mm) Outer race fault (1.0 mm)	2 Nm	200000	70 %	30 %
Acoustic	Healthy bearing Inner race fault (0.3 mm) Inner race fault (1.0 mm) Outer race fault (0.3 mm) Outer race fault (1.0 mm)	0 Nm	200000	70 %	30 %
Vibration	Healthy bearing Inner race fault (0.3 mm) Inner race fault (1.0 mm) Outer race fault (0.3 mm) Outer race fault (1.0 mm)	4 Nm	200000	70 %	30 %
Current	Healthy bearing Inner race fault (0.3 mm) Inner race fault (1.0 mm) Outer race fault (0.3 mm) Outer race fault (1.0 mm)	4 Nm	200000	70 %	30 %

Fig. 8 shows the diagnostic results. The confusion matrix and t-SNE classification showed good results without misclassification. Based on Fig. 8(a) and 8(b), the following phenomenon could be observed: the optimization step size is dynamically adjusted through the cyclic learning rate (CLR) strategy. The model converges rapidly in the early stage of training and avoids

overfitting in the later stage, and the verification loss tends to be stable. Based on Fig. 8(c) and 8(d): the confusion matrix shows that the classification accuracy rate of the model for all categories reaches 100 %, and the clear clustering effect of feature embedding further verifies the feature extraction ability of the model. The self-attention mechanism improves the feature fusion effect significantly, the QPSO-DYS optimization accelerates the convergence speed, and the dynamic learning rate strategy reduces the risk of overfitting effectively.
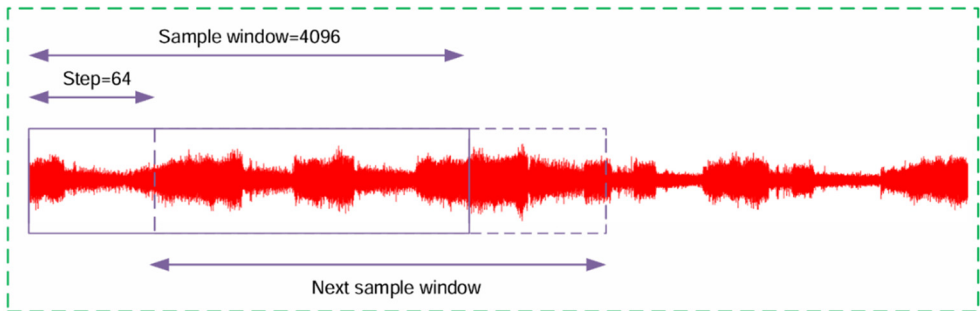

Fig. 7. Sliding window data augmentation

a) The curves of precision and loss rate

b) The variable learning rate curve

c) The confusion matrix diagram
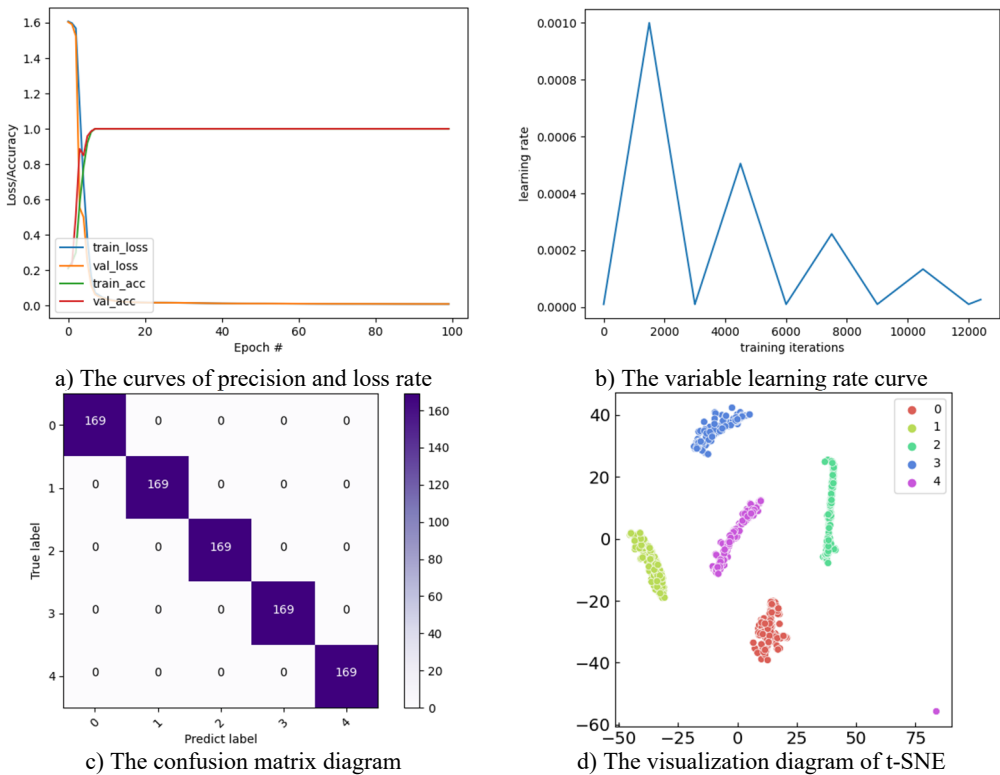
d) The visualization diagram of t-SNE

Fig. 8. Diagnosis results

To verify the advantages of the proposed method, the following models are used for comparison: WDCNN, ResNet, WDCNN-VLR-AE, WDCNN-VLR-SE, WDCNN-SE-AE (the hyperparameter settings and other details of these comparison methods are set in the same way as in their respective literature). Fig. 9 shows the average diagnostic accuracy of each model after multiple experiments. The model presented in this article exhibits the highest accuracy of 98.5 % and the lowest loss rate of 1.7 %. Specifically, the accuracy of our model is significantly higher

than other compared models, indicating that it has stronger discriminative ability and higher accuracy in fault classification tasks. Meanwhile, the loss rate of our model is significantly lower than other models, indicating that it can converge more effectively during training and has better stability.

ResNet and WDCNN perform relatively poorly. Their accuracy is 25.7 % and 27.6 % lower than this model, indicating poor feature extraction ability and difficulty in fully utilizing key information buried in the data. In addition, the loss rates of ResNet and WDCNN are 11.1 % and 12.6 % higher than the model in this paper, which may be related to their model structure and optimization algorithm, leading them to easily fall into local optima, thereby affecting overall performance.

In contrast, WDCNN-VLR-AE, WDCNN-VLR-SE, and WDCNN-SE-AE showed a 10.5 %, 14.6 %, and 7.5 % decrease in accuracy compared to our model on the NSK 6205 DDU dataset, while the loss rate increased by 5 %, 8.2 %, and 4.1 %, respectively, demonstrating an improvement in feature extraction, classification, and outlier resolution capabilities for the dataset.
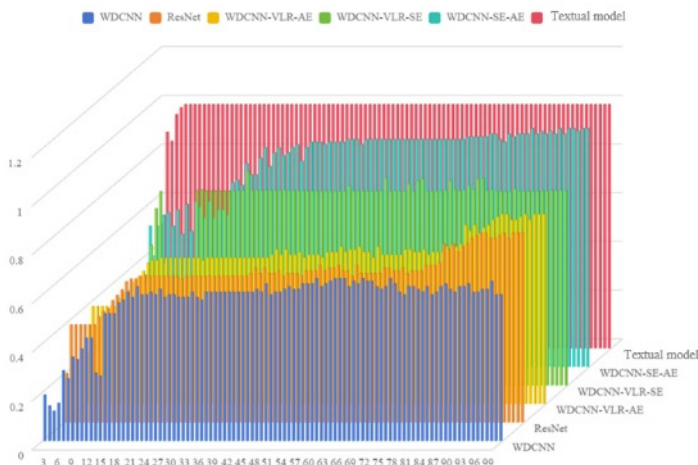
a) The accuracy rate of kinds of models after multiple experiments
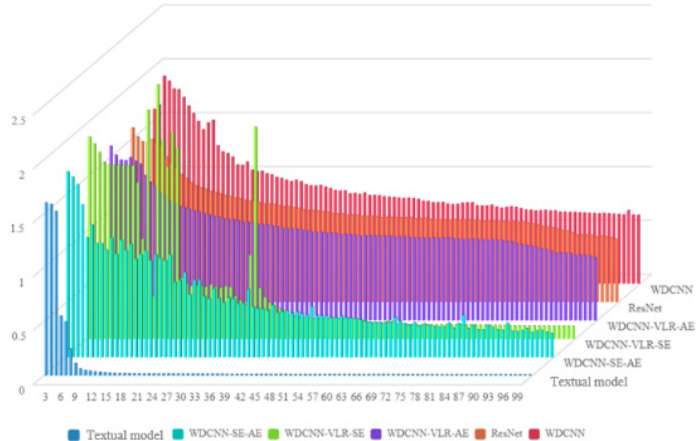
b) The loss rate of kinds of models after multiple experiments

Fig. 9. Comparison results

The radar chart shown in Fig. 10 further demonstrates the accuracy and loss rate of the dataset based on the proposed model after ten times of experiments. The comparison results show that the model proposed in this paper exhibits excellent performance under the same operating conditions.

Specifically, the model presented in this article was able to maintain the lowest loss rate and highest accuracy in ten experiments, with significantly higher accuracy than other models and significantly lower loss rate than others. This indicates that it has better stability and convergence when dealing with multi-source heterogeneous data.
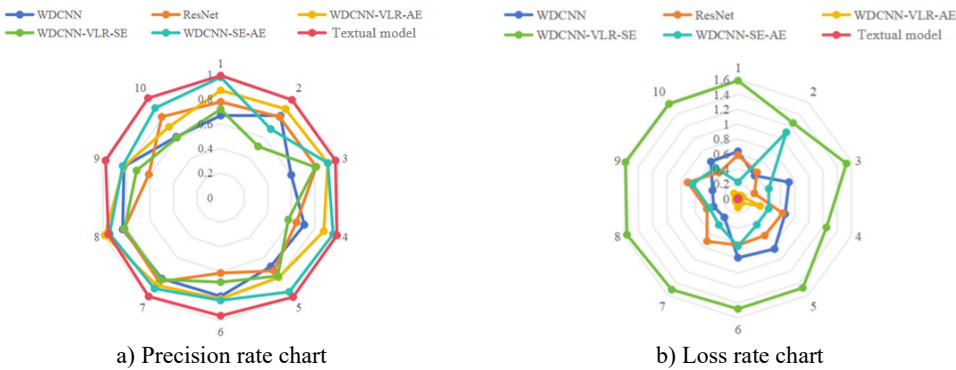

a) Precision rate chart   b) Loss rate chart

Fig. 10. Radar chart results

In terms of model diagnosis and performance evaluation, we have recorded the following key indicators to comprehensively reflect the efficiency and performance of the model. The computer configuration used in the experiment is: Intel Core AMD RyzenTM9-5900HX processor, with a clock speed of 3.3 GHz, equipped with NVIDIA GeForce RTX 3080 graphics card (16 GB of video memory), the system is equipped with 32GB of memory, runs Windows 11 Home Edition, and is equipped with solid-state memory, with a total capacity of about 2TB. In order to demonstrate the advantages of the method compared to the comparative method from multiple perspectives, Table 2 presents score statistics from multiple dimensions such as time efficiency and F1 score.

Table 2. Comparison table of performance evaluation parameters

Method	Training times	Inference speed	F1-score	Recall
Propose method	7 mins	1.251 ms	98.28 %	93 %
WDCNN	29 mins	2.552 ms	96.4 %	92 %
ResNet	38 mins	5.89 ms	95.7 %	90 %
WDCNN-VLR-AE	57 mins	8.35 ms	94.2 %	88 %
WDCNN-VLR-SE	52 mins	4.56 ms	92.9 %	87 %
WDCNN-SE-AE	63	3.630	92.2 %	92 %

In order to compare the advantages of optimizing parameters in this article, several parameter optimization models were compared, as shown in Table 3. From the table, it can be seen that the parameter optimization model in this article saves a lot of training time relatively, and the final training accuracy and loss rate are only reduced by 0.1 % and 0.2 %, and the loss rate is only increased by 0.2 % and 0.2 % compared to manual parameter tuning one by one.

Table 3. Comparison of optimizing parameters

Optimization scheme	Test time	Precision	Loss ratio
Manual input test	–	98.6 %	0.92 %
PSO	62 min	92.4 %	1.52 %
QPSO	45 min	95.5 %	1.21 %
QPSO-AWA	30 min	98.5 %	0.94 %

5. Conclusions

This paper proposes a hybrid diagnostic model that integrates MSCAE and sparse attention

mechanism, and applies it to fault diagnosis based on multi-source heterogeneous features. To overcome the shortcomings of fixed scale convolution kernels in traditional convolutional autoencoder for feature extraction of multi-source information, the proposed hybrid model integrates a multi-scale information learning module into traditional convolutional autoencoder to construct MSCAE, enhancing its ability to extract features from heterogeneous multi-source information while considering feature information at different spatial scales to improve its noise suppression capability. Introduce sparse attention mechanism into MSCAE to improve the model's recognition rate of key fault features. At the same time, a quantum particle swarm optimization module is used to perform hyperparameter optimization on the hybrid model using chaos initialization and dynamic inertia weight strategy, improving the training and diagnostic efficiency of the model, and the problem of the randomness of diagnostic results caused by human selection of hyperparameters is also solved. Compared to most the current idea of building models based on one single factor, the hybrid model is constructed based on comprehensive factors such as multi-source heterogeneous feature extraction effectiveness, model diagnostic efficiency, and noise resistance, which is more comprehensive and applicable. The effectiveness and superiority of the model are verified through experiment and comparative studies. The superiority of the hyperparameter optimization algorithm was verified through ablation experiments.

In future research, we will attempt to apply the method on the diagnosis of gear and bearing faults under variable operating conditions such as variable speed, or the scenario where both variable load and variable speed occur simultaneously.

Acknowledgements

Data availability

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

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] L. Ren et al., "Data-driven industrial intelligence: current status and future directions," *Computer integrated manufacturing systems*, Vol. 28, No. 7, pp. 1914–1939, Jul. 2022, https://doi.org/10.13196/j.cims.2022.07.001

[2] H. Zhao, X. Yang, B. Chen, H. Chen, and W. Deng, "Bearing fault diagnosis using transfer learning and optimized deep belief network," *Measurement Science and Technology*, Vol. 33, No. 6, p. 065009, Jun. 2022, https://doi.org/10.1088/1361-6501/ac543a

[3] Y. Kaya, F. Kuncan, and H. M. Ertunç, "A new automatic bearing fault size diagnosis using time-frequency images of CWT and deep transfer learning methods," *Turkish Journal of Electrical Engineering and Computer Sciences*, Vol. 30, No. 5, pp. 1851–1867, Jul. 2022, https://doi.org/10.55730/1300-0632.3909

[4] M. Kuncan, "An Intelligent approach for bearing fault diagnosis: combination of 1D-LBP and GRA," *IEEE Access*, Vol. 8, pp. 137517–137529, Jan. 2020, https://doi.org/10.1109/access.2020.3011980

[5] H. Yang, X. Li, and W. Zhang, "Interpretability of deep convolutional neural networks on rolling bearing fault diagnosis," *Measurement Science and Technology*, Vol. 33, No. 5, p. 055005, May 2022, https://doi.org/10.1088/1361-6501/ac41a5

[6] E. Akcan, M. Kuncan, K. Kaplan, and Y. Kaya, "Diagnosing bearing fault location, size, and rotational speed with entropy variables using extreme learning machine," *Journal of the Brazilian Society of*

Mechanical Sciences and Engineering, Vol. 46, No. 1, Dec. 2023, https://doi.org/10.1007/s40430-023-04567-2

[7] Z. Xing, Y. Liu, Q. Wang, and J. Li, "Multi-sensor signals with parallel attention convolutional neural network for bearing fault diagnosis," *AIP Advances*, Vol. 12, No. 7, p. 07502, Jul. 2022, https://doi.org/10.1063/5.0095530

[8] S. Zhang, Z. Liu, Y. Chen, Y. Jin, and G. Bai, "Selective kernel convolution deep residual network based on channel-spatial attention mechanism and feature fusion for mechanical fault diagnosis," *ISA Transactions*, Vol. 133, pp. 369–383, Feb. 2023, https://doi.org/10.1016/j.isatra.2022.06.035

[9] A. Al-Dulaimi, S. Zabihi, A. Asif, and A. Mohammadi, "A multimodal and hybrid deep neural network model for Remaining Useful Life estimation," *Computers in Industry*, Vol. 108, pp. 186–196, Jun. 2019, https://doi.org/10.1016/j.compind.2019.02.004

[10] D. Verstraete, E. Droguett, and M. Modarres, "A deep adversarial approach based on multi-sensor fusion for semi-supervised remaining useful life prognostics," *Sensors*, Vol. 20, No. 1, p. 176, Dec. 2019, https://doi.org/10.3390/s20010176

[11] Y. Yue, C. Liu, and T. Liu, "Deep fusion neural network for health indicator construction of bearings," *Journal of electronic measurement and instrumentation*, Vol. 35, No. 7, pp. 44–52, 2021, https://doi.org/10.13382/j.jemi.b2003550

[12] Y. B. Shen et al., "Bi-LSTM neural network for remaining useful life prediction of bearings," *Journal of vibration engineering*, Vol. 34, No. 2, pp. 411–420, Apr. 2021, https://doi.org/10.16385/j.cnki.issn.1004-4523.2021.02.022

[13] H. P. Liang, J. Cao, and X. Q. Zhao, "Remaining useful life prediction method for bearing based on parallel bidirectional temporal convolutional network and bidirectional long and short-term memory network," *Control and decision*, Vol. 39, No. 4, pp. 1288–1296, Apr. 2024, https://doi.org/10.13195/j.kzyjc.2023.0152

[14] Z. Yuan, L. Zhang, and L. Duan, "A novel fusion diagnosis method for rotor system fault based on deep learning and multi-sourced heterogeneous monitoring data," *Measurement Science and Technology*, Vol. 29, No. 11, p. 115005, Nov. 2018, https://doi.org/10.1088/1361-6501/aadfb3

[15] X. Wang, D. Mao, and X. Li, "Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network," *Measurement*, Vol. 173, p. 108518, Mar. 2021, https://doi.org/10.1016/j.measurement.2020.108518

[16] D. Wang, Y. Li, Y. Song, L. Jia, and T. Wen, "Bearing fault diagnosis method based on complementary feature extraction and fusion of multisensor data," *IEEE Transactions on Instrumentation and Measurement*, Vol. 71, pp. 1–10, Jan. 2022, https://doi.org/10.1109/tim.2022.3212542

[17] M. Ma, C. Sun, and X. Chen, "Deep coupling autoencoder for fault diagnosis with multimodal sensory data," *IEEE Transactions on Industrial Informatics*, Vol. 14, No. 3, pp. 1137–1145, Mar. 2018, https://doi.org/10.1109/tii.2018.2793246

[18] J. X. Xu, H. Ma, and X. K. Feng, "Bearing fault diagnosis method based on feature level fusion of multi-source heterogeneous spatial data," *Machinery design and manufacture*, Vol. 9, pp. 150–159, 2021, https://doi.org/10.19356/j.cnki.1001-3997.2021.09.034

[19] Z. L. Cao and C. M. Ye, "Prediction of bearing remaining useful life based on parallel CNN-SE-Bi-LSTM," *Application research of computers*, Vol. 38, No. 7, pp. 2103–2107, Jul. 2021, https://doi.org/10.19734/j.issn.1001-3695.2020.08.0224

[20] S. Wan, X. Li, Y. Zhang, S. Liu, J. Hong, and D. Wang, "Bearing remaining useful life prediction with convolutional long short-term memory fusion networks," *Reliability Engineering and System Safety*, Vol. 224, p. 108528, Aug. 2022, https://doi.org/10.1016/j.ress.2022.108528

[21] M. Kim, J.-R. C. Cheng, and K. Liu, "An adaptive sensor selection framework for multisensor prognostics," *Journal of Quality Technology*, Vol. 53, No. 5, pp. 566–585, Oct. 2021, https://doi.org/10.1080/00224065.2021.1960934

[22] Y. Zhao and Y. Wang, "Remaining useful life prediction for multi-sensor systems using a novel end-to-end deep-learning method," *Measurement*, Vol. 182, pp. 109685–109685, Sep. 2021, https://doi.org/1016/j.measurement.2021.109685

[23] J. Yan, Y. Hu, and C. Guo, "Rotor unbalance fault diagnosis using DBN based on multi-source heterogeneous information fusion," in *Procedia Manufacturing*, Vol. 35, pp. 1184–1189, Jan. 2019, https://doi.org/10.1016/j.promfg.2019.06.075

[24] Y. Ge, J. Wu, J. Qin, L. Ma, and J. Ding, "Remaining useful life prediction based on multi-source sensor data fusion under multi working conditions," in *Lecture Notes in Electrical Engineering*, Vol. 1, pp. 710–718, Mar. 2022, https://doi.org/10.1007/978-981-19-0572-8_92

[25] M. Ou, H. Wei, Y. Zhang, and J. Tan, "A dynamic Adam based deep neural network for fault diagnosis of oil-immersed power transformers," *Energies*, Vol. 12, No. 6, p. 995, Mar. 2019, https://doi.org/10.3390/en12060995

[26] W. Jung, S.-H. Kim, S.-H. Yun, J. Bae, and Y.-H. Park, "Vibration, acoustic, temperature, and motor current dataset of rotating machine under varying operating conditions for fault diagnosis," *Data in Brief*, Vol. 48, p. 109049, Jun. 2023, https://doi.org/10.1016/j.dib.2023.109049

Xaoli Yang received bachelor's degree in Huanghe University of Science and Technology, Zhengzhou, China, in 2008. Now she works at Huanghe University of Science and Technology. Her current research interests include computer software, fault diagnosis based deep learning and so on.