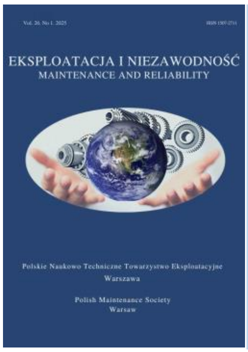


Eksploatacja i Niezawodnosc – Maintenance and Reliability
Volume 27 (2025), Issue 1
journal homepage: http://www.ein.org.pl

Article citation info:
X. Lei, N. Lu, B. Jiang, C. Wang, C. Chen, A Multi-scale Attention Mechanism Diagnosis Method with Adaptive Online Updating Based on Deep Learning under Variable Working Conditions, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2025: 27(1) http://doi.org/10.17531/ein/192975

A Multi-scale Attention Mechanism Diagnosis Method with Adaptive Online Updating Based on Deep Learning under Variable Working Conditions

Indexed by: Web of Science Group

Xue Lei[a], Ningyun Lu[a,*], Bin Jiang[a], Cunsong Wang[b], Chuang Chen[c]

[a] College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, China
[b] Institute of Intelligent Manufacturing, Nanjing Tech University, China
[c] College of Electrical Engineering and Control Science, Nanjing Tech University, China

Highlights

- Multi-scale attention method with adaptive online updating for variable conditions.
- Flexibly updates diagnostic models based on online data status.
- Adaptive weight random undersampling balances inter-class data uniformly.

Abstract

With the advance of industrial systems, the online equipment fault diagnosis has encountered many challenges such as data drift and data imbalance under varying operating conditions, thus making stable and accurate diagnosis increasingly critical. Considering the above issues, a multi-scale attention mechanism diagnosis method with adaptive model that can be updated based on deep learning has been proposed. The method is composed of four main steps: training the multi-scale offline diagnosis model, transferring the parameters of the offline model, assessing the degree of data drifting, and adaptively updating the diagnostic model. A data balance strategy with adaptive weight balances both inter-class and intra-class data. The method updates the diagnostic model flexibly according to online data status, to reduce the impact of data drifting. The method was verified on a bearing test rig, which can reproduce the common bearing faults under variable working conditions. The experimental results have shown that the proposed method can accurately and reliably identify the bearing faults.

Keywords

fault diagnosis, data imbalance, data drifting, adaptive model updating, multi-scale attention mechanism, variable working conditions

1. Introduction

In industrial systems, equipment failures have a significant impact on production efficiency and safety. Therefore, timely and accurate fault diagnosis of equipment is crucial [1-4]. Traditional fault diagnosis methods rely on mathematical and statistical signal analysis [5], [6], often supplemented by physical models and expert knowledge. However, these methods generally require stable environments and high-quality data, limiting their applicability in variable working conditions The primary challenges in such conditions include difficulty in extracting dynamic features and the scarcity of samples for fault analysis. In contrast to traditional methods, deep learning technology has emerged as a crucial tool for addressing these challenges. Recent years have seen a burgeoning adoption of data-driven methods [7], [8], particularly deep learning [9], in

(*) Corresponding author.
E-mail addresses: X. Lei (ORCID: 0009-0004-4847-8465) leixue@nuaa.edu.cn, N. Lu (ORCID: 0000-0002-9964-7677) luningyun@nuaa.edu.cn, B. Jiang (ORCID: 0000-0003-2696-5436) binjiang@nuaa.edu.cn, C. Wang (ORCID:0000-0003-0016-8578) wangcunsong@njtech.edu.cn, C. Chen (ORCID:0000-0003-2117-7860) chenchuang@njtech.edu.cn,

the fault diagnosis field [10-13]. Research in this area has introduced various methods, including Deep Belief Networks (DBN) [2] [14], Convolutional Neural Networks (CNN) [15-18], Long Short-Term Memory (LSTM) [19], Sparse Autoencoders (SAE) [20], Deep Boltzmann Machines (DBM) [21], and Bayesian deep learning (BDL) [22]. Under different operating conditions, the application of deep learning algorithms has proven to be highly effective in diagnosing faults [11]. Notably, these methods predominantly belong to offline learning, offering controlled training processes and stability in data management.

Variable working conditions necessitate comprehensive health monitoring [23]. Multiscale methods, which aggregate information across different scales, provide more exhaustive data and prove beneficial for practical applications. Currently, studies have successfully employed multiscale fault diagnosis methods to handle equipment faults under varying conditions [24-26]. While these methods can extract extensive information, they may also capture redundant data. Attention mechanisms can mitigate this by adjusting weight distributions to enhance the relevance of pertinent information and suppress irrelevant data. Combining attention mechanisms with multiscale methods optimizes the use of relevant information across different scales [24] [27]. When new data emerges, indicating changes in data distribution or patterns, models must be updated or retrained to adapt to these changes. Offline learning approaches lack the flexibility required for frequent model retraining, rendering them less suitable for scenarios with substantial online data, especially under variable operating conditions.

In practical scenarios, equipment operates in nonstationary environments, and the status of equipment changes over time or with varying working conditions. Taking aircraft engines as an example, aircraft engine bearings operate in a complex and dynamic environment, influenced by various flight phases such as takeoff, cruise, climb, and landing. These bearings experience variations in speed, load, and temperature, as well as fluctuations in vibration, impact, and flight mission requirements [28]. In such a complex and dynamic scenario, data collected from bearings may experience data drift over time or with variations in the operating conditions of the bearings [1]. In other words, the characteristics of the bearing signals may change with time or with variations in the operating conditions of the bearings. These changes may be influenced by various complex factors, making it challenging to accurately predict the current bearing state using models trained with offline data. Due to these dynamic changes, diagnostic models trained offline may struggle to adapt to data drift, affecting their effectiveness in real-time fault diagnosis. Therefore, fault diagnosis methods must ensure accuracy while accommodating practical conditions to adapt to dynamically changing data and real-time updating requirements. Adaptability is crucial for addressing system changes and uncertainties, thereby enhancing system robustness and performance [29]. Online learning methods can update models promptly, offering strong real-time capabilities and adaptability. Some researchers have begun exploring online learning for fault diagnosis under variable working conditions [30-32]. Effective utilization of online data for model updates requires a strategic approach to enhance training flexibility.

When training a fault diagnosis model, it is typically assumed that the number of data samples for each fault category is balanced. However, some researchers are investigating fault diagnosis methods that address class imbalance [27], [33], [34]. In practical scenarios, most online data collected from equipment represent normal operating states, with instances of faults being comparatively rare. The scarcity of fault state data poses a challenge for online learning methods, as these models may progressively accumulate more knowledge about normal states while receiving less information about fault conditions. This imbalance can lead to reduced sensitivity to fault characteristics, ultimately diminishing diagnostic accuracy. Therefore, designing a data balancing strategy to maintain relative equilibrium in data quantities is crucial for effective model training.

To address the challenges of data drifting and imbalance, this study introduces a multi-scale attention mechanism for fault diagnosis method with adaptive online updating grounded in deep learning theory. This method can adaptively update the diagnostic model based on the status of online data. By combining multi-scale diagnosis models with attention mechanisms, it extracts more comprehensive diagnostic information and better handles the extraction of fault features under time-varying conditions. Additionally, an adaptive weight random under-sampling strategy is proposed, based on distance measurement and data imbalance rate. This strategy balances inter-class data while ensuring a more uniform distribution of intra-class data, thereby enhancing the diagnostic model's ability to acquire comprehensive information. The proposed method effectively combines the advantages of both offline and online learning approaches to adapt to data drifting and address the imbalance in online data. Firstly, datasets are partitioned into different scales according to data distribution using distance measurement. A Deep Belief Network combined with Extreme Learning Machine (DBN-ELM) approach is then employed to train these datasets, resulting in a multi-scale fault diagnosis model. Attention mechanisms are used to assign different weights to models at each scale, yielding a robust and flexible multi-scale fault diagnosis model. Next, model transfer techniques migrate the offline fault diagnosis model into an online learning framework, where it undergoes real-time diagnosis with online data. Subsequently, the degree of data drifting is assessed by utilizing historical data. Based on the degree of data drifting and data imbalance rate, a determination is made on whether to update the online fault diagnosis model. The objective is to ensure that the model retains essential historical knowledge while exhibiting strong adaptability to changing data distributions. Once the decision to update the model is made, an adaptive weight random under-sampling strategy is employed to balance the data distribution, ensuring relative balance among data from different states. The proposed method demonstrates strengths in retaining historical knowledge, adapting to varying degrees of data drift for flexible model updates, maintaining relative balance in data quantities, and facilitating efficient transfer with fewer sample data.

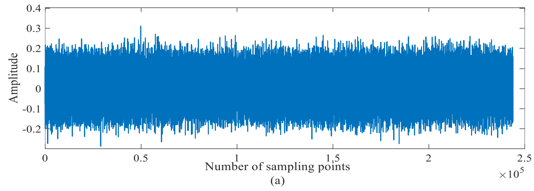
The organization of this article is as follows. Section 2 mainly introduces the preliminary work. Section 3 mainly describes the structure and general process of the proposed method. The experimental verification of the proposed method and result analysis are described in Section 4. Section 5 concludes this article.

2. Preliminary work

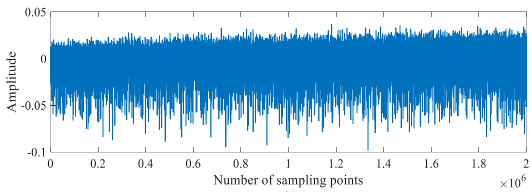
2.1. Time-varying working conditions

Fig.1 displays the vibration signals collected for bearing health, inner race, outer race, and rolling element faults under different operating conditions. Constant speed working conditions data are sourced from publicly available datasets at Case Western Reserve University (healthy state data (Fig. 1(a)), inner race fault number IR007 (Fig. 1(d)), outer race fault number OR007@3 (Fig. 1(g)), and rolling element fault number B007 (Fig. 1(j))), while time-varying speed condition data are sourced from the bearing dataset at the University of Ottawa in Canada (healthy state data numbered H-A-1 (increasing speed, Fig. 1(b)), H-B-1 (decreasing speed, Fig. 1(c)), inner race fault number I-A-1 (increasing speed, Fig. 1(e)), I-C-1 (increasing then decreasing speed, Fig. 1(f)), outer race fault number O-A-3 (increasing speed, Fig. 1(h)), O-B-2 (decreasing speed, Fig. 1(i)), rolling element fault number B-A-1 (increasing speed, Fig. 1(k)), and B-C-1 (decreasing speed, Fig. 1(l))) [35]. As shown in Fig.1, significant fluctuations in the amplitude of vibration signals are observed under different operating conditions for the same health status. This variation may reflect changes in the load, speed, and other operational conditions experienced by the bearing under different operating conditions. The differences in these amplitude fluctuations may be related to the varying vibration characteristics of the bearing under different operating conditions. For instance, during acceleration and deceleration processes, changes in inertial forces and load distribution may occur, leading to fluctuations in the vibration signal amplitude.

To identify different health status, these fluctuations in amplitude may influence the training of models and the classification of health status. Under the same health status, the diagnostic model may be disrupted by the amplitude fluctuations of vibration signals under different operating conditions, resulting in reduced accuracy in fault diagnosis. Particularly in online diagnosis of bearing health status under varying operating conditions, these fluctuations may make it challenging for the model to accurately identify the health status, as it may be affected by changes in operating conditions and lead to misjudgments.



(a)



(b)

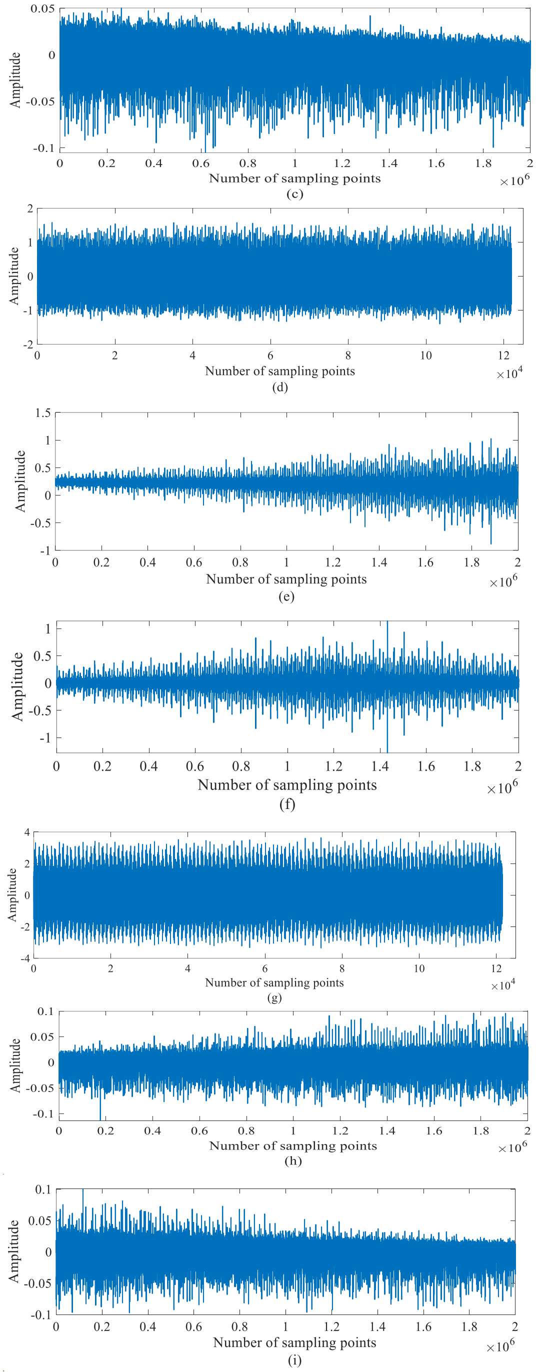


Fig. 1. Vibration signals collected under different operating conditions.

Therefore, appropriate methods need to be employed to address the influence of amplitude fluctuations in vibration signals under different operating conditions to enhance the accuracy and reliability of health state diagnosis. This may involve feature extraction and selection tailored to different operating conditions, as well as the design of health state classification models adaptable to varying operating conditions [36], [37].

2.2. Data imbalance

In engineering practice, there exists a significant imbalance in the online dataset, with the number of samples in the normal state far exceeding those in the failure state. Because there are more samples in the normal state, the model may tend to learn the features of normal states excessively.

For a binary classification problem [1], y represented the class label, health status is represented by 0, fault status is represented by 1, and the data imbalance occurs at the time t if

$$p^t(y=0) \gg p^t(y=1) \tag{1}$$

Data drifting is represented by the joint probability of equipment, the relationship between the features x and the labels y has changed over time, and the data drifting occurs at time$t_0$and time t_1 is represented by:

$$\exists x \ \ p^{t_0}(x,y) \neq p^{t_1}(x,y) \tag{2}$$

Specifically, data imbalance may cause the following issues:

(1) Bias Towards Normal State: The model is more likely to predict new samples as normal states, neglecting potential failure states.

(2) Performance Degradation: As the model fails to fully learn the features of failure states, its performance in practical applications may be affected, leading to an increase in the false-negative rate.

The degree of data imbalance can be expressed by the following formula:

$$Data\ Imbalance\ Ratio\ \ \frac{number\ of\ samples\ in\ min ority\ class}{number\ of\ samples\ in\ majority\ class} \tag{3}$$

This imbalance ratio quantifies the extent of the imbalance, the higher this ratio, the greater the degree of imbalance in the dataset.

2.3. Determination of the degree of data drifting

Some data in the online dataset exhibit varying degrees of drift compared to the offline dataset. The Mahalanobis distance can measure the dissimilarity between a data point and an entire dataset, independent of data dimensions [38]. In the context of data drift, where the distribution of data may change, rendering the existing data model inadequate, the Mahalanobis distance is employed to effectively assess the anomaly level of a data point relative to the dataset. This evaluation helps determine whether data drift has occurred. The strength of the Mahalanobis distance lies in its consideration of data correlation. By transforming the distance into a chi-square distribution, it mitigates the impact of data correlation, making it more applicable to datasets with correlated features. Therefore, when data drift involves changes in data distribution and variations in correlation, the Mahalanobis distance can be used to evaluate the extent of data drift.

There is another advantage to determining the degree of data drifting: based on the degree of data drifting, a similar degree of data be transferred to the training diagnosis model when certain statuses lack sufficient training data. In other words, another advantage of assessing the degree of data drift is that it provides a solid basis for data transfer when facing limited samples in the dataset.

To determine the degree of data drifting, this paper first utilizes the DBN-ELM offline learning model to identify the data's status. The healthy data space consists of normal data, while distinct fault data spaces encompass corresponding fault data. Subsequently, the Mahalanobis distance between the data points and the healthy data space needs to be calculated. This data originates from various fault data spaces, allowing an assessment of the degree of data drifting. The steps to calculate the Mahalanobis distance are as follows:

Step 1: Calculate the average value $\bar{x}_i$ of each datax_{ij}, n is the total number of samples, according to the following formula:

$$\bar{x}_i = \frac{\sum_{j=1}^{n} x_{ij}}{n} \tag{4}$$

Step 2: Calculate the standard deviation s_i of each data, according to the following formula:

$$s_i = \sqrt{\frac{\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2}{n-1}} \tag{5}$$

Step 3: Orthogonality the feature vector acquires thez_{ij}, and calculate its transpose matrixZ^T, according to the following formula:

$$z_{ij} = \frac{(x_{ij}-\bar{x}_i)}{s_i} \tag{6}$$

Step 4: Let the correlation matrix of the orthogonal matrix be A, each of these elements$a_{ij}$was calculated according to the following formula:

$$a_{ij} = \frac{\sum_{m=1}^{n} z_{im} z_{mj}}{n-1} \tag{7}$$

Step5: According to the following formula, calculate the Mahalanobis distance $d_{MD,j}$:

$$d_{MD,j} = Z^T A^{-1} Z \tag{8}$$

2.4. Online learning

Online learning is also known as adaptive learning. It involves integrating training data into the model as a continuous data stream and dynamically updating the model in response.

The universal mathematical description of online learning can be expressed through the lens of a sequence prediction problem.

Problem Formulation: Consider a data sequence (x_1, y_1), (x_2, y_2),…,(x_t, y_t), …, where each$x_t$corresponds to input data at time t and $y_t$is the associated label.

Model Assumption: A parameterized model, denoted byθ_t, is assumed to make predictions based on the current observed data. The model parameters change over time t .

Learning Rule: At each time step t , the model uses the current parameters θ_t to predict the label $\hat{y}_t$. The observed label y_t is then used to update the model parameters. This update rule is denoted as $\theta_{t+1} = UpdateRule(\theta_t, x_t, y_t)$, where$UpdateRule$ is the specific update rule for the given problem and model.

Objective: The goal of online learning is to incrementally improve the model by processing new data points (x_t, y_t), allowing it to adapt to dynamic changes in the data and maintain high performance in a continuously evolving environment.

This is just a simple mathematical description, specific online learning algorithms and update rules may vary based on the particular problem and application scenario.

2.5. Model transfer

Model-based transfer learning is a method that leverages previously acquired knowledge to enhance learning performance on a new task. This approach involves transferring a model trained on a source domain to a target domain, accelerating the learning process, and improving model performance.

In model-based transfer learning, we consider two domains: the source domain and the target domain [11]. Suppose we have trained a model (referred to as the source model) on the source domain. The source model has learned valuable features and knowledge from the source domain. Now, the goal is to transfer these useful features and knowledge to the target domain to improve learning tasks in the target domain.

The primary idea of transfer learning is to adjust the parameters of the source model to adapt to the data distribution in the target domain. This way, the source model can provide

a good starting point for the target task, followed by slight adjustments or fine-tuning on the target domain to meet specific task requirements.

The general steps of model-based transfer learning are as follows:

Step 1: Source Model Training: Train a robust model on the source domain using a large amount of labeled data.

Step 2: Model Transfer: Transfer the model learned on the source domain to the target domain.

Step 3: Target Domain Fine-Tuning: Use limited labeled data from the target domain to fine-tune the source model, adapting it to the requirements of the target task.

Step 4: Performance Evaluation: Evaluate the performance of the transferred model on the target domain.

In this way, model-based transfer learning effectively utilizes knowledge learned from the source domain to enhance learning performance in the target domain.

3. The proposed method

Considering the presence of data drifting and imbalance, when diagnosing the real-time data status, the offline fault diagnosis model might not be adaptable to data drifting. To leverage the strengths of both offline learning and online learning methods, this paper proposes a fault diagnosis approach based on online model transfer.

The steps of the proposed method are outlined as follows:

Step 1: The offline learning stage: Train the offline multi-scale attention mechanism fault diagnosis model using the DBN-ELM algorithm, extracting valid features from historical data during the training process.

Step 2: Evaluating the degree of data drifting: After the establishment of the offline fault diagnosis model, labeled data is employed to determine the extent of drifting.

Step 3: Model transfer stage: Utilizing the model transfer method, migrate the parameters of the offline fault diagnosis model to the online diagnosis model before training it with online data.

Step 4: Model update stage: Based on the degree of data drifting, decide whether to update the online fault diagnosis model. If updating is necessary, incorporate new online data into the training set and retrain the fault diagnosis model.

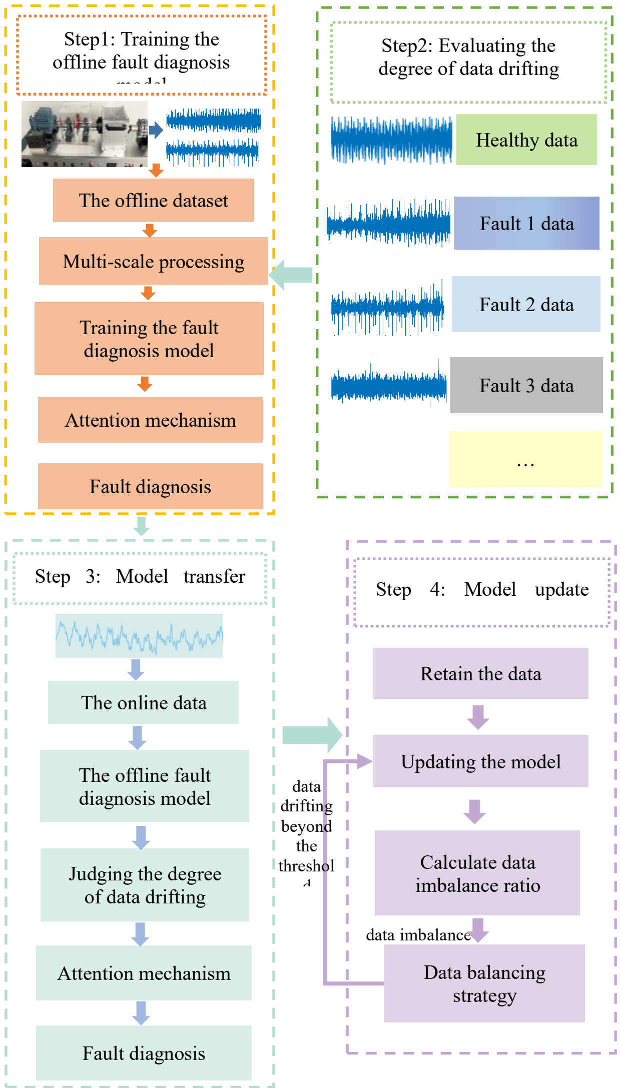


Fig. 2. The flow chart of the proposed method.

Using vibration data from various bearing faults as examples, the flow chart of the proposed method is shown in Fig. 2.

3.1. The offline learning stage

The offline fault dataset includes data from various operational states. The dataset is divided into training and testing datasets: the training dataset is employed to train the offline model, while the testing dataset is used to evaluate the model's performance. After training the offline model, the training parameters are retained. These model parameters can be directly transferred when the model needs to be transferred. The DBN is one of the classical deep learning algorithms. It is combined with the Back Propagation (BP) neural network to extract and classify features. The process of model training requires iteration.

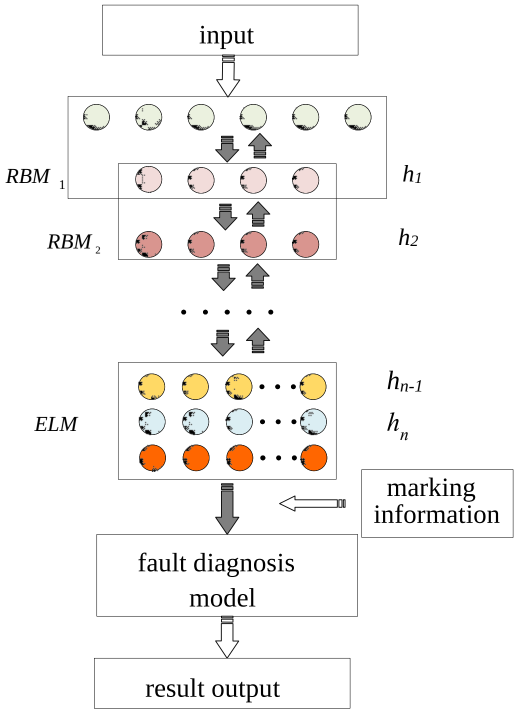
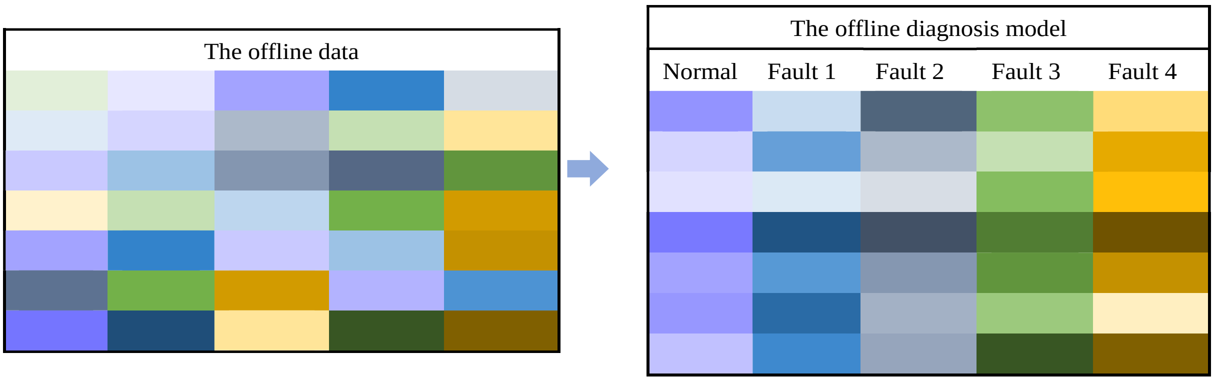


Fig. 3. The structure of DBN-ELM.

However, the DBN-ELM algorithm integrates the DBN and Extreme Learning Machine (ELM) algorithms without the need for iteration. The ELM algorithm can learn quickly and exhibits good generalization performance. As a result, the DBN-ELM algorithm is well-suited for training an offline diagnosis model to extract valid features. The structure of the DBN-ELM is illustrated in Fig. 3.

The DBN algorithm is employed to extract features, while the ELM algorithm is used for data classification. As shown in Fig. 4 (a), distinct colored blocks represent different fault types, and the color depth of the blocks represents the degree of data drifting. The normal status is represented by the purple blocks, while the other colored blocks represent fault status data. The task of offline learning is to identify different fault statuses. As Fig. 4 (b) shows, the process is represented by identifying the different colored blocks.

In this paper, the multi-scale approach involves partitioning the data into several datasets based on differences in data distribution. Subsequently, small-scale fault diagnosis models are trained separately for each dataset. In the training phase of the offline diagnosis model, the overall training dataset is divided into multiple training subsets based on data distribution.



(a) (b)

Fig. 4. The stage of offline learning.

As shown in Fig. 5, the original dataset and each subset are trained separately using DBN-ELM, obtaining diagnosis models at multiple different scale distributions. The model for the $i-th$ scale is denoted as Mi. Based on the proximity of the diagnostic data to each training set and using the attention mechanism, different weights are assigned to each scale model. The closer the distance, the greater the weight. The model trained on the overall data is denoted as M, and the outputs of the multi-scale subsets are fused with the output of the model trained on the overall dataset to make a final diagnostic decision.

If the distance of the data does not exceed the existing dataset, only the subset model at the scale of the data is used. The results obtained from the subset model are then combined with those from the entire training set model for decision-making. If the distance of the data exceeds the original dataset, then based on the distance of the data to each scale dataset, reference is made to an attention mechanism to allocate corresponding weights to the models of each scale. The smaller the distance, the greater the weight assigned. The total weight of all scales equals 1. Combine the results obtained from each scale with the entire training set for decision fusion.

The weight allocation strategy are as follows:

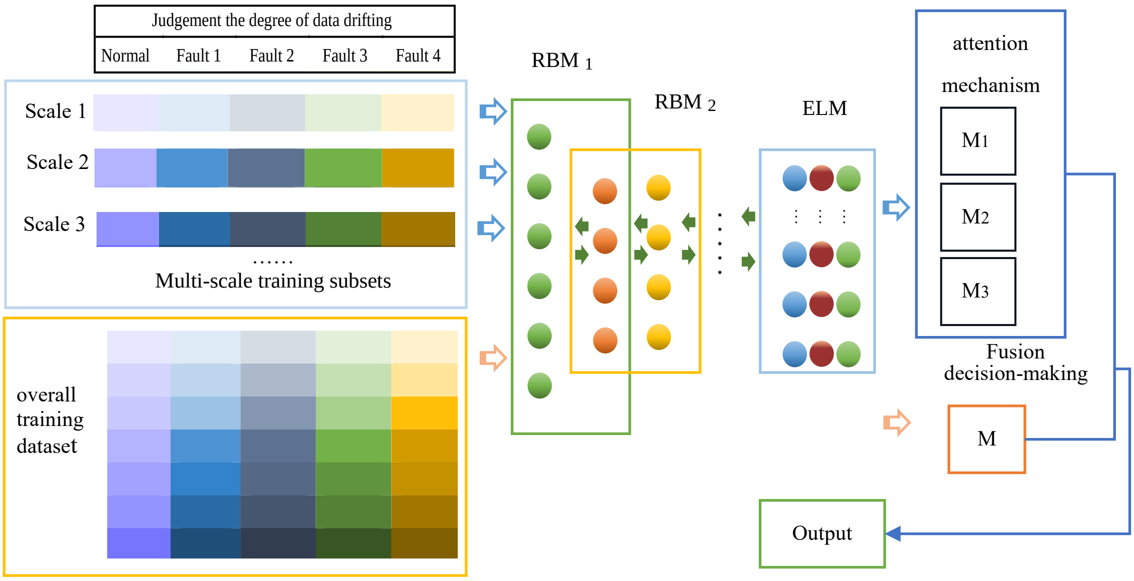


Fig. 5. Training of the multi-scale fault diagnosis model.

Scenario 1: Data belongs to a subset

If the data x belongs to the i-th subset D_i, it also belongs to the entire dataset D. The final diagnostic result y_f is

a combination of the output of the subset model M_i and the output of the overall model M, with equal weights.

$$y_f = 0.5y_i + 0.5y \tag{9}$$

where y_i is the output of model M_i and y is the output of model M.

Scenario 2: Data does not belong to any subset

If the data x does not belong to any subset D_i, it is not part of the entire dataset D. There is a total of n scales. The final diagnostic result y_f is a weighted combination of the outputs of all subset models and the output of the overall model M, with equal total weights for the subsets and the overall model.

The final diagnostic result y_f is then:

$$y_f = 0.5(\sum_{i=1}^{n} w_i\, y_i) + 0.5y \tag{10}$$

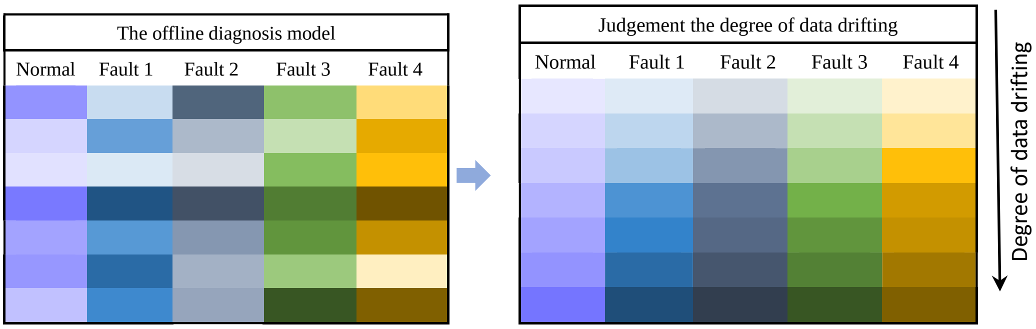
Combining the two scenarios, the comprehensive decision formula for the final diagnostic result y_f can be summarized as:

$$y_f = \begin{cases} 0.5y_i + 0.5y, & if\ x \in D_i \\ 0.5(\sum_{i=1}^{n} w_i\, y_i) + 0.5y, & if\ x \notin D_i \end{cases} \tag{11}$$

where i ranges from 1 to n.

3.2. Evaluating the degree of data drifting

The trained offline fault diagnosis model is capable of identifying the fault status of historical data. Normal status data is categorized into the health data space, and the degree of data drifting is determined by calculating the Mahalanobis distance between fault data and the health data space. A larger value of the Mahalanobis distance indicates a greater degree of data drifting. The data can be divided based on this distance, allowing for categorization into different degrees of drifting.



(a) (b)

Fig. 6. Judging the degree of data drifting.

Fig. 6 (a) depicts the data status after training the offline fault diagnosis model, with blocks of the same color belonging to the same class. However, the degree of data drifting has not yet been determined for the same-class data. Fig. 6 (b) illustrates the result of evaluating the degree of data drifting, where the color depth of blocks represents the extent of data drifting.

3.3. Model transfer

In this paper, model transfer refers to the use of the DBN-ELM algorithm to train balanced historical offline data, obtain a high-performance offline fault diagnosis model, and determine the relevant parameter settings of this diagnostic model. Because this model is trained using offline data, it can effectively extract fault knowledge, which represents valid knowledge left after training on historical data. Based on this model, online data is utilized to update the new knowledge. The model transfer enables the retention of effective knowledge from historical data and facilitates the transfer of offline diagnosis models. For new data, model transfer occurs during each dataset update, requiring distance-based multi-scale attention mechanism for diagnostic decision-making.

3.4. The stage of online learning

Based on the model transfer method, the offline fault diagnosis model is transferred to online learning, using the online data to train the online fault diagnosis model. This preserves the valid knowledge while allowing the model to be updated with new information. Online learning converts the online data into training data within the model and performs model updates. Whether to update the online fault diagnosis model is determined by the extent of data drift and the imbalance ratio. Online learning is efficient and flexible, making it suitable for scenarios with a large amount of data and data drifting.

It should be noted that before online model updating, it is necessary to balance the training data quantity. If the drift degree of new data falls within the range of the old data, the dataset is not updated. However, when the drift degree of new data exceeds a certain threshold, model updating is required.

By training the model, a drift threshold θ is obtained. When the drift level θ_t at time t does not exceed this threshold, the model is not updated. However, when the new data's drift level θ_t surpasses the threshold, the model is updated. This model update rule is denoted as：

$$UpdateRule = \begin{cases} 0, \theta_t \leq \theta \\ 1, \theta_t > \theta \end{cases} \tag{12}$$

Where $UpdateRule$ is the specific update rule for the fault diagnosis mode, 0 is defined as not updating the model, and 1 is defined as updating the model.

In this case, to maintain data balance, it is necessary to update the ratio of various classes in the dataset. This involves performing random undersampling on the larger class to balance the dataset. The ratio of undersampling is usually adjusted based on the imbalance level of the data and the specific problem. The goal of undersampling is to reduce the number of majority class samples to balance the quantity of samples between the majority and minority classes, thereby improving the model's performance.

By reducing the number of majority class (normal state) samples, the balance among sample quantities is improved, enhancing the model's sensitivity to fault states. By reducing sample numbers, computational and time costs during model training is lowered. Undersampling can reduce redundant information in training data, allowing the model to focus more on learning critical features. Random undersampling does pose a risk of information loss, especially when the boundary between normal and fault states is ambiguous. The introduction of a multi-scale attention mechanism further enhances the model's ability to capture fault information by assigning different weights to highlight important features, thereby mitigating the impact of information loss.

There is no fixed standard for the specific undersampling ratio, and it needs to be adjusted according to the actual situation. In experiments, it is common to use the current data with various undersampling ratios and evaluate the model's performance through methods like cross-validation to identify the optimal undersampling strategy that suits the problem.

Before this, the corresponding Mahalanobis distance for each data point had been calculated. Each class dataset was divided equally based on the number of samples in the training set. The target of undersampling was set to the number of samples in each class of the original dataset. Since only one class of data may exceed the balance range, even if there are new data points in other classes, they may not exceed the range. For a particular class, let $Lmin$be the minimum distance and $Lmax$ be the maximum distance. Then the distance difference is $L = Lmax - Lmin$. Taking normal data as an example, after dividing the normal data equally, the distance difference of each portion of data is calculated. The distance difference of the i-th portion of normal data is defined as Li. Then, weights are assigned based on the distance difference of each equally divided portion of data. The weight obtained for the i-th portion of normal data is Li/L, where a larger weight for a portion of data implies a smaller likelihood of being deleted in the weighted random undersampling process. This is because

a larger weight indicates a higher probability of retaining the sample.

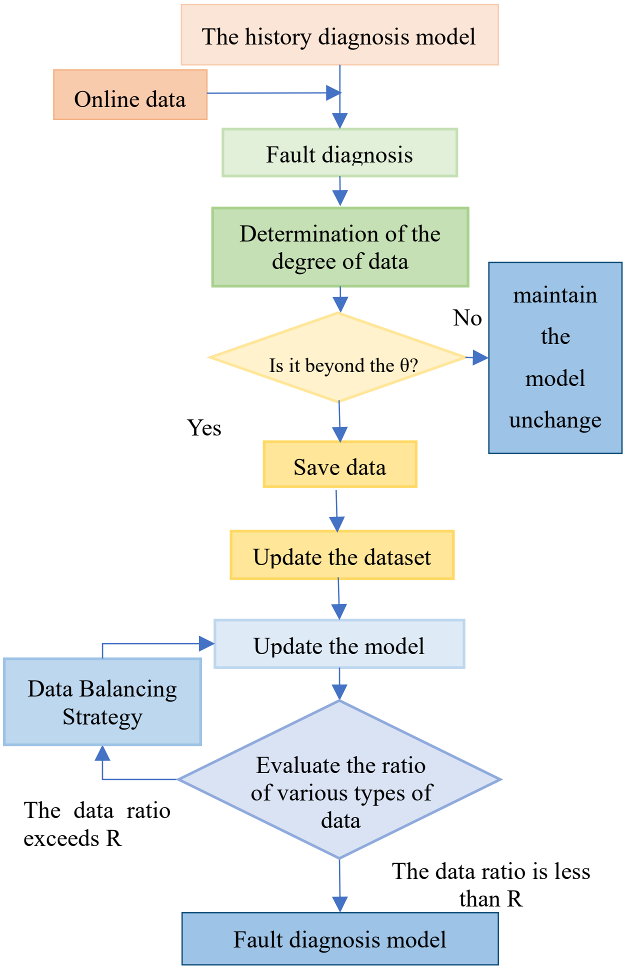


Fig. 7. The adaptive online updating process of the model.

Through variable-weight data balancing, it is possible to maintain balance between inter-class data while also achieving a more uniform distribution of intra-class data, thereby aiding in the establishment of stable and reliable diagnostic models.

In the case of the diagnostic process for online data D, assuming a predefined imbalance ratio R that ensures model performance, the adaptive online updating process for the fault diagnosis model is as follows:

Step 1: Utilize the existing model to diagnose the health status of D.

Step 2: Determine if the drift degree of online data D exceeds a predefined threshold.

Step 3: If the drift degree of D exceeds the threshold, retrain the diagnostic model.

Step 4: If the drift degree of D does not exceed the threshold, store the data.

Step 5: Calculate the proportions of each data class. If the ratio between normal data and fault data exceeds R, implement a data balancing strategy to balance the data, restoring it to a 1:1 ratio.

Step 6: Retrain the diagnostic model with the restored balanced dataset to obtain the updated diagnostic model.

The adaptive online updating process of the model is shown in Fig. 7.

4. Experimental verification

4.1. Experiment setups and Datasets descriptions

4.1.1. The bearing test table setups

In industrial production equipment, rotating machinery typically accounts for a significant proportion, as they are indispensable components in many industrial processes. Bearings, in particular, are critical components of rotating machinery. Approximately 40% to 50% of failures in rotating equipment occur in bearings. This makes timely diagnosis of bearing faults extremely important. And in practical scenarios, bearings operate in nonstationary environments, and the status of bearings changes over time or with varying equipment operating conditions. This makes the operational data of bearings highly suitable for validating the effectiveness of the method proposed in this paper.

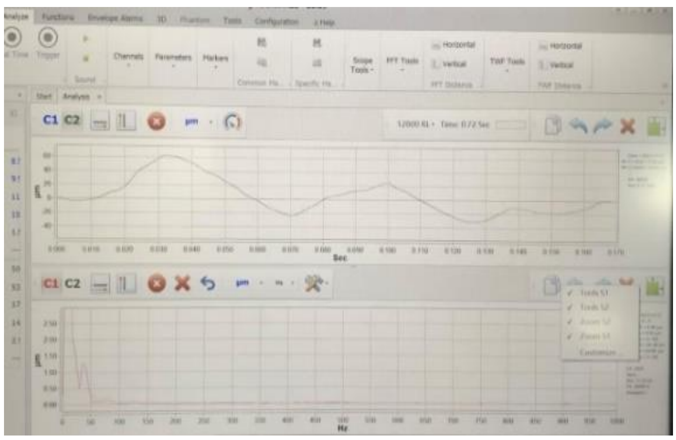


Fig. 8. The bearing test table.

This paper validates the proposed method using a bearing fault dataset collected under varying working conditions. The data used in the experimental section were collected from our laboratory test table. The bearing test table can reproduce the common bearings faults under variable working conditions. The bearing test table is shown in Fig. 8. The bearing test table is composed of a three-phase asynchronous motor, frequency converter, coupling, friction support kit, parallel gearbox, electromagnetic powder brake, and other components. This test bench can simulate most real machine bearing faults. It includes a speed measurement module, laser signal transmitter, and speed display module. The speed and load can be adjusted in real time. Real-time vibration signals can be collected through a vibration sensor (accelerometer) and directly transmitted to the computer-end software. The speed and load vary continuously over a period, simulating the changing operating conditions of bearings in both healthy and faulty states. The purpose of this experiment is to collect data under different variable speed and load conditions for bearing health and fault states. The dataset includes normal status, outer raceway fault, inner raceway fault, roller fault, blade deformation fault, gear missing teeth fault. It includes variable speed status, such as speed increasing (abbreviated as SI), speed increasing and then decreasing (abbreviated as SID), speed decreasing (abbreviated as SD), speed decreasing and then increasing (abbreviated as SDI), variable load status, such as load increasing (abbreviated as LI), load increasing and then decreasing (abbreviated as LID), load decreasing (abbreviated as LD), load decreasing and then increasing (abbreviated as LDI). Three sets of data are taken for each status.

A total of two datasets were used in this paper to simulate historical data and online data. The first dataset is an offline fault dataset, which includes a portion of data with different statuses and provides sufficient data for training the diagnosis model. The second dataset is the online dataset. Both the offline and online datasets include data not only under the same working conditions but also under different working conditions. As shown in Fig. 9. Offline dataset and online dataset not only include some data under the same working conditions but also include some data under different working conditions.

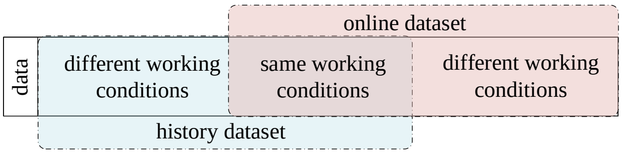


Fig. 9. Offline dataset and online dataset.

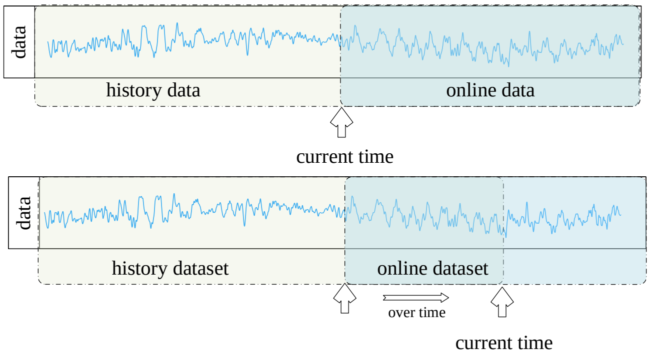


Fig. 10. The offline dataset and online dataset.

Fig. 10 displays the offline dataset and online dataset. To simulate the imbalance in online data, the amount of normal status data is greater than that of fault data. Additionally, the amount of data with different degrees of drifting varies.

4.1.2. Determination of update threshold

To confirm the effect of different degrees of data drift on fault diagnosis, the preliminary dataset was used as the training dataset, and datasets with varying degrees of drift were used as the testing dataset. A total of 7 data drift datasets were used, numbered from small to large according to the degree of drift. Table 1 displays the detailed information of these datasets. Fig. 11 shows the average accuracy of the 20 test iterations for each dataset. The results indicate that as the data difference increases, the test accuracy decreases. The diagnostic accuracy has notably decreased from Dataset 4 to Dataset 5. Therefore, determining whether to update the model is based on the degree of drift in Dataset 4. Thus, it is necessary to determine a suitable update threshold for model updating based on the actual conditions of different datasets.

Table 1. Introduction of the bearing dataset.

Dataset	Bearing Health status	Label	Number of data
preliminary data	normal	1	5 000
	outer raceway fault	2	5 000
	inner raceway fault	3	5 000
	roller fault	4	5 000
	blade deformation fault	5	5 000
	gear missing teeth fault	6	5 000
data drifting dataset [2]-data drifting dataset [8]	normal	1	1 000
	outer raceway fault	2	1 000
	inner raceway fault	3	1 000
	roller fault	4	1 000
	blade deformation fault	5	1 000
	gear missing teeth fault	6	1 000

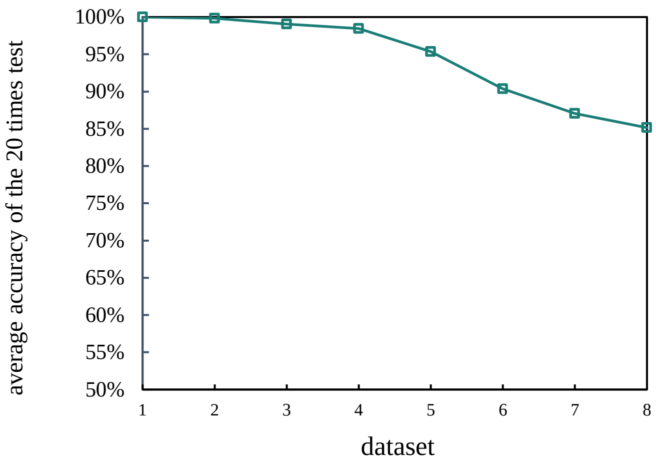


Fig. 11. The average accuracy of the 20 times test for each dataset.

4.1.3. The setting of the experiment and dataset description

A total of five experiments were conducted in this study. Because experiments were conducted in advance to assess the impact of data imbalance on the fault diagnosis model, it was observed that the accuracy of fault diagnosis started to significantly decrease when the imbalance ratio among different classes exceeded 1:3. Therefore, the initiation condition for the data balancing strategy is when the data imbalance ratio exceeds 1:3. Taking into account the imbalance in the online data, the experiment set a data imbalance ratio of 1:10 between normal data and various types of fault data. The first experiment involves training the multi-scale offline fault diagnosis model. The second experiment aimed to verify the proposed method. The third and fourth experiments aimed to verify the advantage of updating the model according to the degree of data drifting. Firstly, the offline fault dataset was used to train the offline fault diagnosis model without considering the degree of data drifting. Subsequently, the online dataset was divided into two datasets, with half of the data having a smaller degree of data drifting and the other half having a larger degree of data drifting. Finally, the model was utilized to diagnose the online dataset. The fifth experiment aimed to verify the advantages of offline learning and involved training the diagnosis model using the online dataset. Table 2 shows the composition of the dataset.

Table 2. the composition of the dataset.

Bearing health status	Speed status	Load status	Label
N	SI	-	1
	SD	-	
	SID	-	
	SDI	-	
	-	LI	
	-	LD	
	-	LID	
	-	LDI	
OFR	SI	-	2
	SD	-	
	SID	-	
	SDI	-	
	-	LI	
	-	LD	
	-	LID	
	-	LDI	
IF	SI	-	3
	SD	-	
	SID	-	
	SDI	-	
	-	LI	
	-	LD	
	-	LID	
	-	LDI	
RF	SI	-	4
	SD	-	
	SID	-	
	SDI	-	
	-	LI	
	-	LD	
	-	LID	
	-	LDI	

Bearing health status	Speed status	Load status	Label
BDF	SI	-	5
	SD	-	
	SID	-	
	SDI	-	
	-	LI	
	-	LD	
	-	LID	
	-	LDI	
GMTF	SI	-	6
	SD	-	
	SID	-	
	SDI	-	
	-	LI	
	-	LD	
	-	LID	
	-	LDI	

4.2. Experimental results

4.2.1. Results of the proposed method

The results of the five experiments are presented in Table 3. The training outcome of the offline fault diagnosis model (the first experiment) is also shown in Table 3 and Fig. 12. The second experiment aimed to verify the proposed method. Firstly, the offline fault dataset was used to train the offline fault diagnosis model. Because the offline dataset is balanced and possesses an adequate number of training data, the offline fault diagnosis model performs well. The decision to update the model relies on the extent of data drifting and the imbalance ratio. Due to the continuous updating of the online learning model, diagnostic performance remains satisfactory. The results of the first experiment are presented in Table 3 and Fig. 13. The third and fourth experiments aimed to verify the advantage of updating the model according to the degree of data drifting. Firstly, the offline fault dataset was used to train the offline fault diagnosis model without considering the degree of data drifting. Subsequently, the online dataset was divided into two datasets, with half of the data exhibiting a smaller degree of data drifting, and the other half displaying a larger degree of data drifting. The offline fault diagnosis model was employed to diagnose the online dataset. In Table 3, Fig. 14, and Fig. 15, the outcomes of both experiments are not highly satisfactory. This is attributed to the differing data drifting degrees in the two datasets, leading to varying diagnostic results. The diagnostic effectiveness of the dataset with a significant data drifting degree is inferior to that of datasets with a minor data drifting degree. The absence of model updates can impact accuracy and stability. The fifth experiment aimed to verify the advantages of offline learning and involved training the diagnosis model using the online dataset. The results of the fourth experiment are presented in Table 3 and Fig. 16. The outcomes are less than satisfactory, pointing out that neglecting to combine the offline diagnosis model affects the fault diagnosis results. Fig. 18 displays box plots depicting the results of 20 experiments conducted for each different experiment setting. From Fig. 17, it can be observed that the proposed method exhibits higher stability and fewer outliers. The box plot for the proposed method is relatively narrow, indicating lower performance fluctuations. Therefore, it can be concluded that this method demonstrates superior accuracy and stability compared to other test experiments.

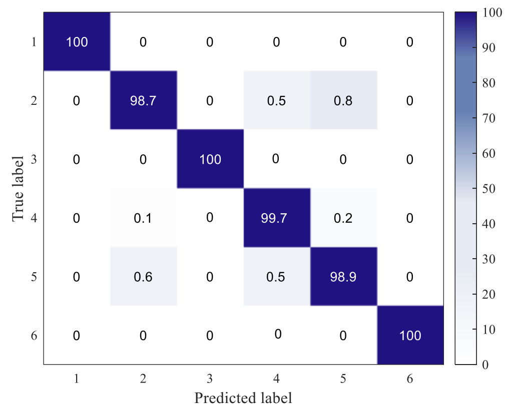


Fig. 12. Fault diagnosis confusion matrix of experiment 1.

Table 3. Results of fault diagnosis.

experiment	Offline dataset: Average accuracy (%)	Offline dataset: Standard deviation	Offline dataset: Time(s)	experiment	Online dataset: Average accuracy (%)	Online dataset: Standard deviation	Online dataset: Time(s)
1	99.55	0.45	237.65	2	98.93	0.29	132.92
				3	93.17	0.73	45.95
				4	90.41	1.72	45.87
	-	-	-	5	88.59	3.67	77.66

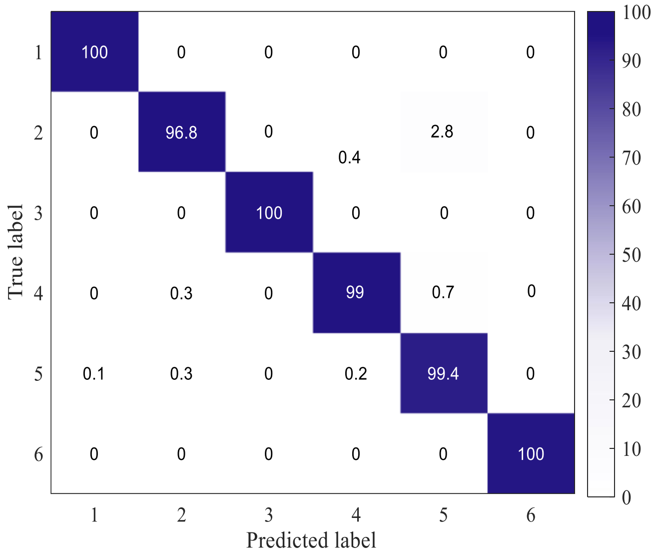


Fig. 13. Fault diagnosis confusion matrix of experiment 2.

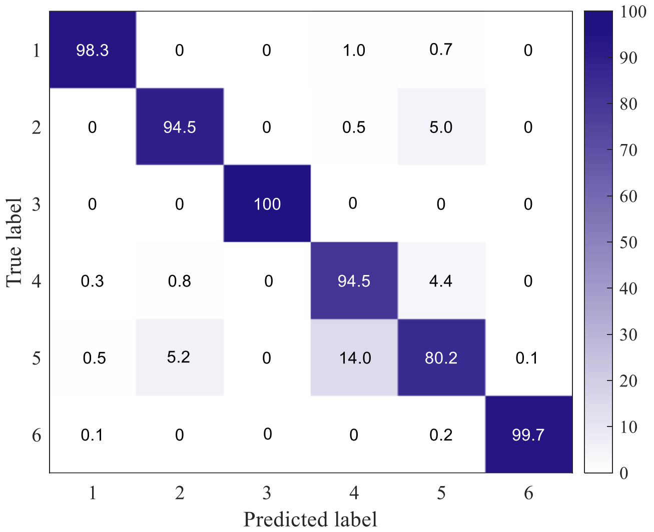


Fig. 14. Fault diagnosis confusion matrix of experiment 3.

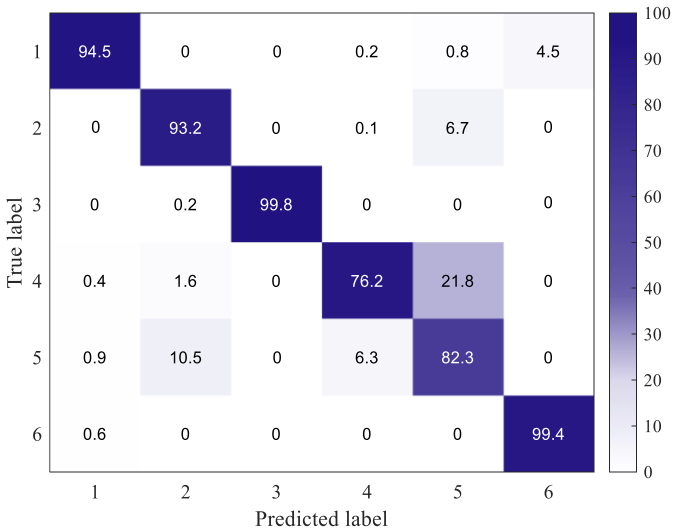


Fig. 15. Fault diagnosis confusion matrix of experiment 4.

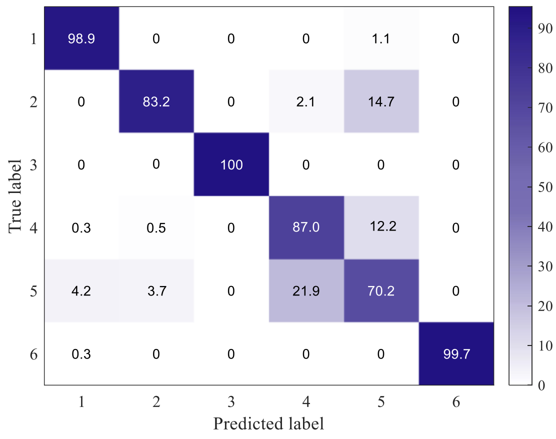


Fig. 16. Fault diagnosis Confusion matrix of experiment 5.

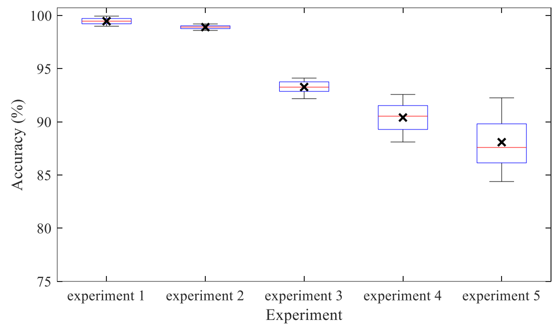


Fig. 17. Results of 20 times for each of the 5 experiments.

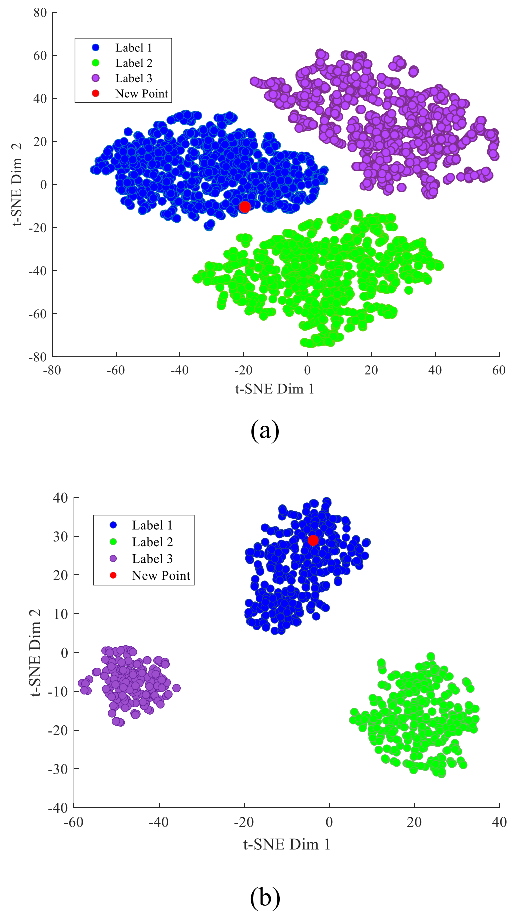


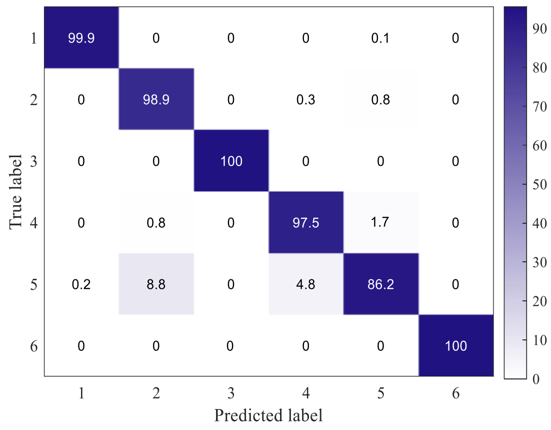
Fig. 18. Comparison before and after data rebalancing.

This section uses the t-SNE visualization method to display the data before and after rebalancing. To present the results clearly, only three categories of data are visualized. A new data point is marked in the healthy data, and its changes before and after rebalancing are recorded. Fig. 18 visually shows the changes in the data distribution of the three bearing health states before and after rebalancing. Here, 1 represents normal state, 2 represents inner ring fault, and 3 represents outer ring fault. The data changes can be seen intuitively from the Fig. 18.

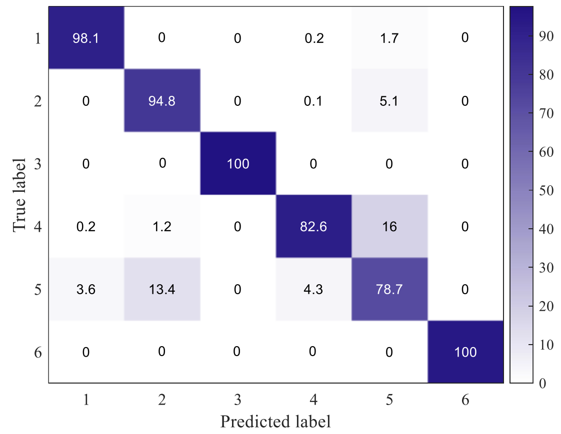
4.2.2. The advantage of evaluating data drift for fault diagnosis under limited sample conditions

To demonstrate the advantages of evaluating data drift for fault diagnosis under limited sample conditions, a transfer learning experiment was conducted in this section. The target diagnostic dataset comprised a few-sample data with consistent drift levels. A portion of this data was allocated for the target domain dataset, while the remaining portion was reserved for the testing dataset. Other fault data that did not undergo the evaluation of the degree of data drift was employed as the source domain, labeled as Source Domain Dataset 1. In this section, the classical Transfer Component Analysis (TCA) method was employed to perform transfer learning between Source Domain Dataset 1 and the target domain dataset. The transferred data was then used to train a fault diagnosis model, followed by testing its performance on the testing dataset. This experiment was denoted as Experiment 1.

Subsequently, a dataset with the same number as Source Domain Dataset 1 but with minor differences in the degree of data drift from the data with few samples was employed as Source Domain Dataset 2. Similarly, the TCA method was applied for data transfer, resulting in a fault diagnosis model trained on the transferred data. The model's performance was assessed using the testing dataset, marking this experiment as Experiment 2. The confusion matrices of these two experiments are shown in Figs. 19 (a) and (b). The results of the fault diagnosis experiments indicated that training the model with data closely resembling the few-sample data resulted in higher accuracy. This underscores the importance of evaluating data drift, as it assists in identifying more similar data under few-sample conditions, thus enabling the utilization of methods like transfer learning for enhanced diagnostic outcomes.



(a)



(b)

Fig. 19. Fault diagnosis confusion matrix of experiments A and B.

5. Conclusion

In this study, a multi-scale attention mechanism diagnosis method with adaptive online updating based on deep learning theory is proposed to effectively address the impact of data drift and online data imbalance on the performance of online fault diagnosis models in industrial systems.

The proposed method demonstrates significant advantages in experiments, particularly in:

(1) Multi-scale attention mechanism for comprehensive feature extraction: The paper proposed a diagnostic method that leveraged a multi-scale attention mechanism based on deep learning. This mechanism enhanced the model's ability to capture important features across varying working conditions.

(2) Adaptive online updating: The method incorporated adaptive online updating, allowing the diagnostic model to flexibly update according to the status of online data. This ensured the model remained effective even under data imbalance and varying working conditions.

(3) Effective data balancing strategy: The paper introduced an adaptive weight random undersampling strategy that effectively balanced both inter-class and intra-class data. This resulted in a more uniform intra-class data distribution, enhancing the model's diagnostic performance.

(4) Comprehensive experimental validation: The effectiveness of the proposed method was validated through

a series of five experiments, demonstrating its robustness and applicability across different scenarios, including datasets with small and large degrees of data drift.

Experimental results clearly indicate that the proposed method accurately and reliably identifies bearing faults despite data drift and online data imbalance, offering a flexible and effective solution for fault diagnosis in industrial systems. Future research directions may include algorithm optimization, domain expansion, and further extensive practical validation to further demonstrate its robustness and reliability under different operating conditions.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62020106003, Grant 61873122, Grant 62203213, and Grant 62303217, in part by the Aero Engine Corporation of China Industry-university-research cooperation project under Grant HFZL2020CXY011, in part by the the Research Fund of State Key Laboratory of Mechanics and Control of Mechanical Structures under Grant MCMS-I-0121G03, in part by the National Key Research and Development Program of China under Grant 2021YFB3301300, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20220332, in part by the China Scholarship Council under Grant 202206830124, and in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 23KJB510006.

References

1. Malialis K, Panayiotou CG, M. M. Polycarpou. Online Learning With Adaptive Rebalancing in Nonstationary Environments. IEEE Transactions on Neural Networks and Learning Systems 2021; 32(10): 4445-4459, https://doi.org/10.1109/TNNLS.2020.3017863.
2. Niu G, Wang X, Golda M, Mastro S, Zhang B. An optimized adaptive PReLU-DBN for rolling element bearing fault diagnosis. Neurocomputing 2021; 445: 26-34, https://doi.org/10.1016/j.neucom.2021.02.078.
3. Jiang B, Zhang K, Shi P. Integrated Fault Estimation and Accommodation Design for Discrete-Time Takagi–Sugeno Fuzzy Systems With Actuator Faults. IEEE Transactions on Fuzzy Systems 2011; 19(2): 291-304, https://doi.org/10.1109/TFUZZ.2010.2095861.
4. Lu N, Jiang B, Meng X, et al. Diagnosis, Diagnosticability Analysis, and Test Point Design for Multiple Faults Based on Multisignal Modeling and Blind Source Separation. IEEE Transactions on Systems, Man, and Cybernetics: Systems 2017; 1-12, https://doi.org/10.1109/TSMC.2017.2758802.
5. Ciabattoni L, Ferracuti F, Freddi A, Monteriù A. Statistical Spectral Analysis for Fault Diagnosis of Rotating Machines. IEEE Transactions on Industrial Electronics 2018; 65(5): 4301-4310, doi: 10.1109/TIE.2017.2762623.
6. Rai V K, Mohanty A R. Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform. Mechanical Systems and Signal Processing 2007; 21(6): 2607-2615, DOI: 10.1016/j.ymssp.2006.12.004.
7. Chu F, Liao S, Hao L, Wang P, Liu Y, Wang F. Operating Performance Assessment Method for Industrial Process with Slowness Principle-Based LSTM Network. Engineering Applications of Artificial Intelligence 2023; 123(part: A). DOI: 10.1016/j.engappai.2023.106208.
8. Chu F, Mo S, Hao L, Lu N, Wang F. Operating Performance Assessment of Complex Nonlinear Industrial Process Based on Kernel Locally Linear Embedding PLS. IEEE Transactions on Automation Science and Engineering 2024; 21(1):593-605. DOI: 10.1109/tase.2022.3230687.
9. Chen C, Tao G, Shi J, Shen M, Zhu Z. A Lithium-ion battery degradation prediction model with uncertainty quantification for its predictive maintenance. IEEE Transactions on Industrial Electronics 2024; 71(4), 3650-3659. DOI: 10.1109/TIE.2023.3274874.
10. Liu R, Yang B, Zio E, Chen X. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing 2018; 108: 33-47, https://doi.org/10.1016/j.ymssp.2018.02.016.
11. Zhao B, Zhang X, Hai L, et al. Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions. Knowledge-Based Systems 2020; 199: 105971, DOI: 10.1016/j.knosys.2020.105971.
12. Zhang Y, Ding J, Li Y, Ren Z, Feng K. Multi-modal data cross-domain fusion network for gearbox fault diagnosis under variable operating

conditions. Engineering Applications of Artificial Intelligence 2024; 133: 108236, https://doi.org/10.1016/j.engappai.2024.108236.
13. Wan L, Ning J, Li Y, et al. Intelligent fault diagnosis via ring-based decentralized federated transfer learning. Knowledge-Based Systems 2024; 284: 111288, https://doi.org/10.1016/j.knosys.2023.111288Zhao C, Zio E, Shen W. Domain generalization for cross-domain fault diagnosis: Anapplication-oriented perspective and a benchmark study. Reliability Engineering & System Safety 2024; 245: 109964, https://doi.org/10.1016/j.ress.2024.109964.
14. Zhang C, He Y, Jiang S, Wang T, Yuan L, Li B. Transformer Fault Diagnosis Method Based on Self-Powered RFID Sensor Tag, DBN, and MKSVM. IEEE Sensors Journal 2019; 19(18): 8202-8214, DOI: 10.1109/JSEN.2019.2919868.
15. Yang B, Lei Y, Jia F, et al. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. Mechanical Systems and Signal Processing 2019; 122: 692-706, DOI: 10.1016/j.ymssp.2018.12.051.
16. Li X, Jiang H, Xie M, Wang T, Wang R, Wu Z. A reinforcement ensemble deep transfer learning network for rolling bearing fault diagnosis with Multi-source domains. Advanced Engineering Informatics 2022; 51: 101480, ISSN 1474-0346, https://doi.org/10.1016/j.aei.2021.101480
17. Huo C, Jiang Q, Shen Y, Zhu Q, Zhang Q. Enhanced transfer learning method for rolling bearing fault diagnosis based on linear superposition network. Engineering Applications of Artificial Intelligence 2023; 121: 105970, https://doi.org/10.1016/j.engappai.2023.105970.
18. Plakias S, Boutalis YS. Fault detection and identification of rolling element bearings with Attentive Dense CNN. Neurocomputing 2020; 405: 208-217, https://doi.org/10.1016/j.neucom.2020.04.143.
19. Maged A, Lui CF, Haridy S, Xie M. Variational AutoEncoders-LSTM Based Fault Detection of Time-Dependent High Dimensional Processes. International Journal of Production Research 2024; 62(4): 1092-1107. doi:10.1080/00207543.2023.2175591.
20. Weng Q, Ren S, Zhu B, Jin Y, Si F. Reconstruction-based stacked sparse auto-encoder for nonlinear industrial process fault diagnosis. Eksploatacja i Niezawodność-Maintenance and Reliability 2024; 26(1). DOI:10.17531/ein/175873.
21. Peng W, Ye ZS, Chen N. Bayesian Deep-Learning-Based Health Prognostics Toward Prognostics Uncertainty. IEEE Transactions on Industrial Electronics 2020; 67(3): 2283-2293, March, DOI: 10.1109/TIE.2019.2907440.
22. Li C, Sanchez RV, Zurita G, et al. Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. Neurocomputing 2015; 168: 119-127. https://doi.org/10.1016/j.neucom.2015.06.008.
23. Song B, Liu Y, Fang J, et al. An optimized CNN-BiLSTM network for bearing fault diagnosis under multiple working conditions with limited training samples. Neurocomputing 2024; 574: 127284, https://doi.org/10.1016/j.neucom.2024.127284.
24. Wang X, Yao Y, Gao C. Wasserstein Distance- EEMD Enhanced Multi-Head Graph Attention Network for Rolling Bearing Fault Diagnosis Under Different Working Conditions. Eksploatacja i Niezawodność-Maintenance and Reliability 2024; 26(2). DOI:10.17531/ein/184037.
25. Huang Z, Zhao X. A novel multi-scale competitive network for fault diagnosis in rotating machinery. Engineering Applications of Artificial Intelligence 2024; 128: 107441. https://doi.org/10.1016/j.engappai.2023.107441.
26. Wang J, Ren H, Shen C, Huang W, Zhu Z. Multi-scale style generative and adversarial contrastive networks for single domain generalization fault diagnosis. Reliability Engineering & System Safety 2024; 243: 109879. https://doi.org/10.1016/j.ress.2023.109879.
27. Wang R, Chen Z, Zhang S, et al. Dual-attention generative adversarial networks for fault diagnosis under the class-imbalanced conditions. IEEE Sensors Journal 2021; 22(2): 1474-1485, https://doi.org/10.1109/JSEN.2021.3131166.
28. Wang C, Lu N, Cheng Y, et al. A Data-Driven Aero-Engine Degradation Prognostic Strategy. IEEE Transactions on Cybernetics 2019; PP(99): 1-11. DOI: 10.1109/TCYB.2019.2938244.
29. Jiang B, Shen Q, Shi P. Neural-networked adaptive tracking control for switched nonlinear pure-feedback systems under arbitrary switching. Automatica 2015; 61: 119-125, https://doi.org/10.1016/j.automatica.2015.08.001.
30. Wang H, Zheng J, Xiang J. Online bearing fault diagnosis using numerical simulation models and machine learning classifications. Reliability Engineering & System Safety 2023; 234: 109142, https://doi.org/10.1016/j.ress.2023.109142.
31. Xu Q, Zhu B, Huo H, et al. Fault diagnosis of rolling bearing based on online transfer convolutional neural network. Applied Acoustics 2022; 192: 108703, https://doi.org/10.1016/j.apacoust.2022.108703.
32. Mao W, Ding L, Liu Y, et al. A new deep domain adaptation method with joint adversarial training for online detection of bearing early fault. ISA Transactions 2022; 122: 444-458, https://doi.org/10.1016/j.isatra.2021.04.026.

33. Zhao Y, Zhang X, Wang J, et al. A new data fusion driven-sparse representation learning method for bearing intelligent diagnosis in small and unbalanced samples. Engineering Applications of Artificial Intelligence 2023; 117: 105513, https://doi.org/10.1016/j.engappai.2022.105513.
34. Luo W, Yang W, He J, et al. Fault diagnosis method based on two-stage GAN for data imbalance. IEEE Sensors Journal 2022; 22(22): 21961-21973, https://doi.org/10.1109/JSEN.2022.3211021.
35. Huang H, Baddour N. Bearing vibration data collected under time-varying rotational speed conditions. Data in brief 2018; 21: 1745-1749, https://doi.org/10.1016/j.dib.2018.11.019.
36. Zuber N, Bajrić R. Gearbox faults feature selection and severity classification using machine learning. Eksploatacja i Niezawodność - Maintenance and Reliability 2020;22(4):748-756. DOI:10.17531/ein.2020.4.19.
37. Li J, Liu Z, Qiu M, Niu K. Fault diagnosis model of rolling bearing based on parameter adaptive VMD algorithm and Sparrow Search Algorithm-Based PNN. Eksploatacja i Niezawodność-Maintenance and Reliability 2023; 25(2). DOI:10.17531/ein/163547.
38. Hou Y, Wu P, Wu D. An operating condition information-guided iterative variational mode decomposition method based on Mahalanobis distance criterion for surge characteristic frequency extraction of the centrifugal compressor. Mechanical Systems and Signal Processing 2023; 186: 109836, https://doi.org/10.1016/j.ymssp.2022.109836.