Fault Diagnosis of Rotating Machinery With Limited Expert Interaction: A Multicriteria Active Learning Approach Based on Broad Learning System

Zeyi Liu, Jingfei Zhang, Xiao He, *Senior Member, IEEE*, Qinghua Zhang, Guoxi Sun, and Donghua Zhou, *Fellow, IEEE*

***Abstract*— Recently, research on the fault diagnosis of rotating machinery, especially for the compound or unknown cases, has drawn increasing attention. Some advanced learning-based approaches have achieved good fault diagnosis performance to some degree. However, in practical applications, how to utilize prior knowledge as fully as possible for fault diagnosis with constraints of limited expert interaction remains an open issue. In this brief, a fault diagnosis methodology of rotating machinery with limited expert interaction is proposed. With related feature extraction techniques, a novel multicriteria active learning (MCAL) query strategy is designed to select the relatively valuable samples for annotation. In addition, the broad learning system (BLS) is exploited to achieve fast incrementally updating or retrain procedures with high diagnostic accuracy in different diagnosis scenarios. Several experiments are conducted on a real-world rotating machinery fault diagnosis (RMFD) experimental platform. Compared with other existing advanced approaches, the diagnosis performance of the proposal shows high stability and flexibility. The annotation cost of experts is also significantly reduced, which makes the proposal more suitable for dealing with practical problems.**

***Index Terms*— Broad learning system (BLS), fault diagnosis, multicriteria active learning (MCAL), rotating machinery.**

I. Introduction

ROTATING machinery plays a vital role in numerous practical industrial scenarios. Due to the long-term work in complex environments such as high temperature and high pressure, the occurrence of faults for such dynamic systems is inevitable in most cases [1], [2]. Hence, rotating machinery fault diagnosis (RMFD) methodologies with good effects are of great significance for improving production efficiency, reducing potential losses, and avoiding major disasters. Generally, the provided effective information of time-domain data for the rotating machinery is limited and easily affected by noise, which makes the fault characteristics not obvious. In this context, several time–frequency domain transformation methods, such as wavelet packet decomposition (WPD) [3] and empirical mode decomposition (EMD) [4], have been introduced to amplify the hidden fault features in the 1-D initial vibration signal with the signal decomposition procedure. However, there are generally many hyperparameters that needed to be defined by experience. The inappropriate selection of parameters can generally result in poor diagnostic performance.

Although the above methodologies have been successfully applied to many scenarios, there are still some difficulties to be faced for the RMFD, especially for the diagnosis of the compound or unknown faults. Typically, it is difficult and error-prone to simply utilize existing learning-based approaches to identify the relevant complex fault types since the related features are coupled or unknown. An effective way to address the above limitations is to introduce expert prior knowledge [5]. However, the cost of relying solely on expert knowledge (such as the form of annotation) for complex diagnosis tasks subjectively is high for existing approaches. Hence, the construction of the RMFD model with limited expert interaction is valuable and practical.

As a basic paradigm to solve the limited expert interaction problem mentioned above, active learning (AL) has attracted more and more attention recently for industrial applications [6], [7]. For instance, Chen *et al.* [8] proposed an uncertainty and complexity-based AL scheme to deal with the gearbox fault diagnosis with the utilization of EMD-singular value decomposition (SVD) and random forests. A cost-sensitive AL approach is also proposed with the utilization of the bidirectional gated recurrent unit (BGRU) for handling the imbalance cases [9]. In [10], an AL-based actual bearing compound fault diagnosis approach is designed based on the decoupling attentional residual network. While in [11], an AL-based fault diagnosis approach is proposed with the self-organizing cellular networks.

Nevertheless, there are still some limitations to be concerned for such advanced approaches to deal with RMFD tasks. From the perspective of the AL query strategy, the performance obtained by a single query strategy may be very bad in some fault conditions and parameter settings [12]. Namely, it is not stable to only consider a single query strategy in real-life such as RMFD. Moreover, existing multicriteria AL (MCAL)

Manuscript received 21 May 2022; accepted 12 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61733009, Grant 62033008, and Grant 61933013; and in part by the National Key Research and Development Program of China under Grant 2017YFA0700300. Recommended by Associate Editor F. You. *(Corresponding author: Xiao He.)*

Zeyi Liu and Jingfei Zhang are with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: liuzy21@mails.tsinghua.edu.cn; zhangjf17@mails.tsinghua.edu.cn).

Xiao He is with the Beijing National Research Center for Information Science and Technology (BNRist), Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: hexiao@mail.tsinghua.edu.cn).

Qinghua Zhang and Guoxi Sun are with the Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China (e-mail: fenglangren@tom.com; guoxi.sun@gdupt.edu.cn).

Donghua Zhou is with the College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China (e-mail: zdh@tsinghua.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCST.2022.3200214.

Digital Object Identifier 10.1109/TCST.2022.3200214

methods often cannot autonomously decide which single query strategies should be preferred for the current scenario, which is not flexible enough. On the other hand, classifier models with deep architecture usually require a large number of labeled samples to avoid underfitting [13], [14]. However, in practical applications, only a small number of labeled samples can be provided in the early stage of annotation. Therefore, such models have certain limitations for solving the problems considered in this brief. Moreover, their time costs for training are also relatively high, which constraints its practicality. Hence, it is crucial to design a reasonable classifier with a stable and flexible query strategy to handle the RMFD tasks with limited expert interaction. Recently, a broad learning system (BLS) inherited from the *random vector functional-link neural network* (RVFLNN) is proposed, which has been applied in the field of fault diagnosis. For instance, Fu *et al.* [15] proposed an adaptive BLS for RMFD. The BLS with principal components analysis (PCA) is also introduced to deal with the fault diagnosis task of the rotor system [16]. With the bagging technique, Wang *et al.* [17] introduced the BLS to deal with the wear fault diagnosis of aero-engines. As a result, the BLS has shown great potential for handling similar tasks.

In this context, a novel MCAL approach based on BLS is proposed in this brief for RMFD with limited expert interaction. First, various signal processing techniques such as WPD are utilized to extract vibration features. A novel MCAL query strategy is then designed to select the overall valuable samples for annotation. The BLS is introduced as a basic classifier to complete the learning process iteratively. The applicable conditions of its incrementally update and retraining procedure are also discussed and designed, which is beneficial to reduce the training costs to a certain extent with high diagnostic accuracy. Several experiments are conducted to validate the effectiveness of the proposed approach. The data used in this brief is obtained under a variety of different experimental settings, which was sourced from a real test rig of rotating machinery. The experiment results show the superiority of the proposal in handling RMFD tasks with limited expert interaction compared with other typical approaches.

The main contributions of this brief are summarized as follows.

1) A novel MCAL approach is proposed to select the overall relatively valuable samples, which is stable and flexible. As a result, the procedure of variable parameter selection can be simplified.
2) The BLS is introduced to complete the learning process iteratively. The applicable conditions of its incrementally update and retraining procedure are also discussed and designed to reduce the training costs with high diagnostic accuracy.
3) Several experiments are conducted with the utilization of realistic RMFD data. The experimental results show that the proposed approach outperforms most of the typical approaches.

The remainder of this study is organized as follows. In Section II, the proposed methodology is presented in detail. The experimental analysis is then presented in Section III to show its effectiveness and availability. Related discussion and comparison are also mentioned. In Section IV, this brief is summarized by some concluding remarks.

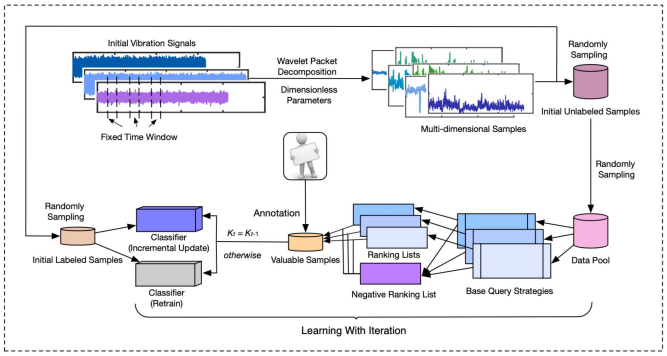


Fig. 1. Flowchart of the proposed RMFD approach with limited expert interaction.

II. Novel RMFD Approach

In this section, a novel RMFD approach is proposed with three main stages. Our task is to identify the faults effectively with limited expert iteration, even for such compounds or unknown scenarios. The flowchart is shown in Fig. 1.

A. Stage I: Data Processing and Feature Extraction

For a large RMFD experimental platform, the original data can be collected based on the sensor installed on the gearbox to sample vibration signal, which can be denoted as $S = \{s_1, s_2, \ldots, s_N\}$. In detail, s_ω represents the vibration data of the original signal at time ω. Clearly, the underlying fault information is often drowned in background noise due to the complex, nonstationary, and nonlinear nature of vibration signals. One commonly utilized approach is the WPD technique [3], which overcomes the disadvantage that the window size does not change with frequency to a certain extent.

Essentially, WPD can decompose the initial vibration signal layer by layer to obtain different frequency bands [18]. In other words, it can amplify the fault characteristics. A set of features can be obtained via reconstructing the signal at each node of the jth layer. The decomposition results of the jth layer has 2^j frequency band signals. Assume that the pth ($p \in \{0, 1, \ldots, 2^j\}$) frequency band obtained after the decomposition of the jth layer of wavelet packets of the signal at time t is defined as $S_{(j,p)}(t)$. The jth layer sub-bands $S_j^{2p-1}(t)$ and $S_j^{2p+1}(t)$ can be obtained as

$$S_j^{2p-1}(t) = 2^{1/2} \sum_{k \in Z} h(k) S_{j-1}^{p}(2\omega - k) \tag{1}$$

$$S_j^{2p+1}(t) = 2^{1/2} \sum_{k \in Z} g(k) S_{j-1}^{p}(2\omega - k) \tag{2}$$

where $h(k)$ and $g(k)$ denote the *high pass filter* (HPF) and *low pass filter* (LPF), respectively. With the reconstruction of the extracted low-frequency coefficients and high-frequency coefficients of each frequency band signal, the new reconstructed signal can be obtained. To weaken the effect of signal fluctuations due to perturbation, six dimensionless parameters

$\xi = \{\xi_1, \ldots, \xi_6\}$, including the *waveform, impulse, margin, peak, kurtosis indicator,* and *vibration intensity*, are then exploited to amplify the features furtherly. Refer to [19] for more details.

B. Stage II: Instance Query and Label Annotation

Assumed that the n-dimensional RMFD data $\mathcal{D} \in \mathbb{R}^{m\times n} = \{x_1, \ldots, x_m\}$ can be obtained as mentioned in Section II-A. For each instance x_i, we have $x_i \in \mathbb{R}^n$ where $n = 6 \times N$ and $i \in \{1, \ldots, m\}$. On this basis, an initial labeled data with $|\mathcal{L}_0|$ samples $\mathcal{L}_0 = \{(x_1, y_1), \ldots, (x_{|\mathcal{L}_0|}, y_{|\mathcal{L}_0|})\}$ and an unlabeled data with $|\mathcal{U}_0|$ samples $\mathcal{U}_0 = \{x_1, \ldots, x_{|\mathcal{U}_0|}\}$ can be obtained. Let $y_i \in \{0, 1, \ldots, K_0\}$ is the fault type of x_i and K_0 denotes the initial number of known fault types with the initial labeled data. In the framework of AL, the main task is to select the *relatively valuable* samples to be queried and annotated. The general process is to select a batch $\mathcal{Q}$ with query strategy θ in unlabeled data pool $\mathcal{P}$ where $\mathcal{P} \subset \mathcal{U}$. The experts are required to provide the actual fault types of the samples in the batch. The budget $\mathcal{B}$ is defined as the maximum number of queries for the experts under some specific constraints. Before the budget is exhausted, both $\mathcal{L}$ and $\mathcal{U}$ are updated continuously. The classifier are then re-trained to modify the model parameters based on updated $\mathcal{L}$ or $\mathcal{Q}$ incrementally. Assume that there are M different *base query strategies* (BQSs) to be considered, which refers to refer to the single-criterion query strategies integrated in the proposed MCAL in the following content. The relatively valuable samples of the ith BQS in the tth iteration can be generally formulated in the form shown in the following equation:

$$x_t^{(\theta_i)*} = \arg\max_x \mathcal{S}_t^{(\theta_i)}(\mathcal{P}_t) \tag{3}$$

where $x_t^{(\theta_i)*}$ denotes the so-called the relatively valuable samples with $\mathcal{P}_t$. $\mathcal{S}_t^{(\theta_i)}(\mathcal{P}_t)$ represents the objective score function to calculate the score under criterion θ_i for all the samples in current pool $\mathcal{P}_t$. It is worth noting that the higher the value of each sample, the greater the value of its calculated score is assumed. Moreover, the objective score function is assumed to be multiplied by -1 for achieving the effect inversion for some opposite situations. Inspired by Chen and Liu [20], the MCAL framework can be viewed as a problem of preference allocation to all the BQSs based on the ranking information. For any criteria θ_i where $\theta_i \in \{\theta_1, \theta_2, \ldots, \theta_M\}$, the corresponding ranking list in the tth iteration $R_t^{(\theta_i)}$ can be denoted as

$$\mathcal{R}_t^{(\theta_i)} = \sigma\left(\mathcal{S}_t^{(\theta_i)}(\mathcal{P}_t)\right) \tag{4}$$

where $\sigma(\cdot)$ represents a sort function. Notably, samples with larger objective score values are assumed to have a lower ranking. Applying (4) for all the query criteria, M ranking lists and score lists in the tth iteration can be obtained. Furthermore, the score lists are considered to be standard due to the numerical difference of all the BQSs. In detail, we have

$$\widetilde{\mathcal{S}}_t^{(\theta_i)}(x_k) = \frac{\mathcal{S}_t^{(\theta_i)}(x_k) - \min \mathcal{S}_t^{(\theta_i)}(x)}{\max \mathcal{S}_t^{(\theta_i)}(x) - \min \mathcal{S}_t^{(\theta_i)}(x)} \tag{5}$$

where $x \in \mathcal{P}_t$.

Clearly, it is difficult to infer the implicit optimal recommendation list to be queried based on the existing BQSs. However, it is relatively easy to get a recommendation list with negative effect for query procedure (denoted as *negative ranking list* (NRL) in the following) considering the advanced effects of BQS. If any BQS has a positive correlation with the NRL, it is believed that the preference of such a BQS needs to be allocated a lower value numerically. The specific determination can be customized more flexibly by the users based on the specific RMFD tasks. For instance, a typical kind of NRL can be defined as

$$\mathcal{R}_t^{(neg)} = \sigma\left(\widetilde{\mathcal{S}}_t^{(neg)}(\mathcal{P}_t)\right) \tag{6}$$

where

$$\widetilde{\mathcal{S}}_t^{(neg)}(x_k) = 1 - \sum_{i=1}^{m} \widetilde{\mathcal{S}}_t^{(\theta_i)}(x_k)/m. \tag{7}$$

In the sequence recommendation problem, the influence of information at different positions on the similarity is different. Specifically, it is clear that: 1) if the number of samples that are at the top of the two ranking lists is larger, their correlation should be larger; 2) if the number of samples that are at the end of the two ranking lists is larger, their correlation should be relatively larger; and 3) if the number of samples with large differences in positions in the two ranking lists is large, their correlation should be small. In this context, the Jeffery divergence is introduced for solving such a task, which can evaluate the correlation effects of any pairwise ranking lists reasonably compared with other measurement. The preference allocated to BQS θ_i is expressed as

$$\gamma\left(\mathcal{R}_t^{(\theta_i)}\right) = \kappa[\mathcal{J}(\widetilde{\mathcal{R}}_t^{(\theta_i)} || \widetilde{\mathcal{R}}_t^{(neg)})]^{-\eta} \tag{8}$$

where κ and η denote the hyperparameters used to limit the numerical overall range of $\xi(\mathcal{R}_t^{(\theta_i)})$. $\mathcal{J}(\cdot)$ represents the Jeffery divergence, which is defined as

$$\begin{aligned}\mathcal{J}\left(\widetilde{\mathcal{R}}_t^{(\theta_i)} || \widetilde{\mathcal{R}}_t^{(neg)}\right) &= D_{KL}\left(\widetilde{\mathcal{R}}_t^{(\theta_i)} || \widetilde{\mathcal{R}}_t^{(neg)}\right) \\ &\quad + D_{KL}\left(\widetilde{\mathcal{R}}_t^{(neg)} || \widetilde{\mathcal{R}}_t^{(\theta_i)}\right)\end{aligned} \tag{9}$$

where

$$D_{KL}(\widetilde{\mathcal{R}}_t^{(neg)} || \widetilde{\mathcal{R}}_t^{(\theta_i)}) = \sum_{i=1}^{|\mathcal{P}_t|} \widetilde{\mathcal{R}}_t^{(neg)}(x_i) \log \frac{\widetilde{\mathcal{R}}_t^{(neg)}(x_i)}{\widetilde{\mathcal{R}}_t^{(neg)}(x_i)}. \tag{10}$$

Specifically, $\widetilde{\mathcal{R}}_t^{(\theta_i)}$ and $\widetilde{\mathcal{R}}_t^{(neg)}$ denotes the normalization of ranking lists. With (4), the $\widetilde{\mathcal{R}}_t^{(\theta_i)}(x_i)$ can be calculated as

$$\widetilde{\mathcal{R}}_t^{(\theta_i)}(x_i) = \frac{|\mathcal{R}_t^{(\theta_i)}(x_i) - \mathcal{P}_t| + 1}{|\mathcal{P}_t|}. \tag{11}$$

On this basis, the overall score value of any sample x_k in $\mathcal{P}_t$ can be defined as

$$\widetilde{\mathcal{S}}_t(x_k) = \sum_{i=1}^{m} \gamma\left(\mathcal{R}_t^{(\theta_i)}\right) \widetilde{\mathcal{S}}_t^{(\theta_i)}(x_k). \tag{12}$$

Then the $\mathcal{Q}_t$ in the proposed MCAL framework can be formulated. The samples to be queried in the tth iteration is expressed as

$$x^{*(t)} = \arg\max_x \widetilde{\mathcal{S}}_t(\mathcal{P}_t). \tag{13}$$

C. Stage III: Classifier Design

With the $\mathcal{Q}_t$, the next step is to design the classifier. For real-life RMFD tasks, an ideal classifier under the AL framework needs to have the appropriate model capacity with fast training speed to reduce the cost of annotation and training simultaneously. In this brief, the BLS is utilized to deal with such an issue since the closed-form solution can be obtained while having a faster training speed.

For any labeled data $L_0 \in \mathbb{R}^{|\mathcal{L}_0|\times n}$ and encoded output $\mathcal{Y}_0 \in \mathbb{R}^{|\mathcal{L}_0|\times K_0}$, the mapping relationship between the input nodes and ith features nodes can be obtained as

$$Z_i^0 = \phi\big(L_0 W_{e_i}^0 + \beta_{e_i}^0\big) \tag{14}$$

where $Z_i^0 \in \mathbb{R}^{N_1\times N_f}$, $i \in \{1, 2, \ldots, N_1\}$ and $Z^0 = [Z_1^0, Z_2^0, \ldots, Z_{N_1}^0]$. N_f represents the number of feature nodes for any group. $W_{e_i}^0$ and $\beta_{e_i}^0$ denotes the weights and bias, respectively, which are randomly generated. As mentioned in [21], the linear inverse problem can be applied. The initial $W_{e_i}^0$ can be fine-tuned to obtain related better features due to the advantages of sparse autoencoder characteristics [7], [22]. For the jth group of enhancement nodes, it can be obtained as

$$H_j^0 = \xi\Big(Z^0 W_{h_j}^0 + \beta_{h_j}^0\Big) \tag{15}$$

where $H_j^0 \in \mathbb{R}^{N_2\times N_e}$, $j \in \{1, 2, \ldots, N_2\}$, and $H^0 = [H_1^0, H_2^0, \ldots, H_{N_2}^0]$. N_e represents the number of enhancement nodes for any group. Similarly, $W_{h_j}^0$ and $\beta_{h_j}^0$ also denotes the weights and bias, respectively, which are randomly generated. $\xi(\cdot)$ represents a nonlinear mapping such as *tanh* function.

Hence, the mapping between initial labeled data and output with such architecture can be obtained as

$$\begin{aligned}
&\Phi^0\big(\mathcal{L}_0, W_e^0, \beta_e^0, W_h^0, \beta_h^0\big)\\
&\quad= \big[Z_1^0, \ldots, Z_{N_1}^0 | \xi\big(Z^0 W_{h_1}^0 + \beta_{h_1}^0\big) + \cdots\\
&\qquad + \xi\Big(Z^0 W_{h_{N_2}}^0 + \beta_{h_{N_2}}^0\Big)\Big] W_{N_2}^0\\
&\quad= \big[Z_1^0, \ldots, Z_{N_1}^0 | H_1^0, \ldots, H_{N_2}^0\big] W_{N_2}^0\\
&\quad= \big[Z^0 | H^0\big] W_{N_2}^0 = \mathcal{Y}_0
\end{aligned} \tag{16}$$

where $\mathcal{Y}_0$ represents the encoded label of $\mathcal{L}_0$.

Hence, we have

$$W_{N_2}^0 = \big[Z^0 | H^0\big]^{\dagger} \mathcal{Y}_0 \tag{17}$$

where $[Z^0|H^0]^{\dagger}$ represents the pseudo-inverse operation of $[Z^0|H^0]$. In detail, $W_{N_2}^0$ can be solved by using ridge regression approximation of $[Z^0|H^0]^{\dagger}$ with the following equation, which is defined as:

$$W_{N_2}^0 = \Big(\lambda I + \big[Z^0|H^0\big]\big[Z^0|H^0\big]^T\Big)^{-1}\big[Z^0|H^0\big]^T \mathcal{Y}_0 \tag{18}$$

where λ denotes the further constraints on the sum of the squared weights $W_{N_2}^0$ and I denotes the identity matrix. If $\lambda \to 0$, the inverse problem degenerates into the least square problem and leads the solution to original pesuedoinverse. For simplicity, let A^0 denote $[Z^0|H^0]$ in the following. We have

$$W_{N_2}^0 = \lim_{\lambda\to 0}\Big(\lambda I + A^0\big(A^0\big)^T\Big)^{-1}\big(A^0\big)^T \mathcal{Y}_0. \tag{19}$$

So far, the pretrained classifier Φ_0 with $\mathcal{L}_0$ is obtained. It is worth noting that $K_t \neq K_{t-1}$ represents that the expert believes that $\mathcal{Q}_t$ contains a new unknown fault type. In this case, current pool $\mathcal{P}_t$ is suggested to be used for a complete retraining process before the annotation budget is exhausted. While $K_t = K_{t-1}$ represents that all the fault types have already been observed before the tth iteration. In this case, the incremental update process should be considered to save the time for retraining the model. With the pretrained learner Φ_{t-1} and corresponding W_e^{t-1}, β_e^{t-1}, W_h^{t-1} and β_h^{t-1}, we have

$$\begin{aligned}
A_t = \Big[&\phi\big(\mathcal{Q}_t W_{e_1}^{t-1} + \beta_{e_1}^{t-1}\big), \ldots, \phi\Big(\mathcal{Q}_t W_{e_{N_1}}^{t-1} + \beta_{e_{N_1}}^{t-1}\Big)\\
&|\xi\big(Z^{t-1} W_{h_1}^{t-1} + \beta_{h_1}^{t-1}\big), \ldots, \xi\Big(Z^{t-1} W_{h_{N_1}}^{t-1} + \beta_{h_{N_2}}^{t-1}\Big)\Big]
\end{aligned} \tag{20}$$

where

$$Z^{t-1} = \Big[\phi\big(\mathcal{Q}_t W_{e_1}^{t-1} + \beta_{e_1}^{t-1}\big), \ldots, \phi\Big(\mathcal{Q}_t W_{e_{N_1}}^{t-1} + \beta_{e_{N_1}}^{t-1}\Big)\Big]. \tag{21}$$

Hence, the updating mapping matrix can be formulated as

$$A_t' = \Big[A_{t-1}' \;\Big|\; A_t^T\Big]^T \tag{22}$$

where A_{t-1}' denotes the updating results for $\mathcal{P}_{t-1}$. As mentioned in [21], the associated pseudo-inverse updating algorithm can be defined as

$$\big(A_t'\big)^{\dagger} = \Big[\big(A_{t-1}'\big)^{\dagger} - B_t D_t^T \Big| B_t\Big] \tag{23}$$

where

$$D_t^T = A_t^T\big(A_{t-1}'\big)^{\dagger} \tag{24}$$

$$C_t = A_t^T - D_t^T A_{t-1}' \tag{25}$$

and

$$B_t^T = \begin{cases} C_t^{\dagger}, & \text{if} C_t \neq 0;\\ (1 + D_t^T D_t)^{-1}\big(A_{t-1}'\big)^{\dagger} D_t, & \text{if} C_t = 0. \end{cases} \tag{26}$$

As a result, the updated weight is defined as

$$W_{N_2}^t = W_{N_2}^{t-1} + \big(\mathcal{Y}_t^T - A_t^T W_{N_2}^{t-1}\big) B_t \tag{27}$$

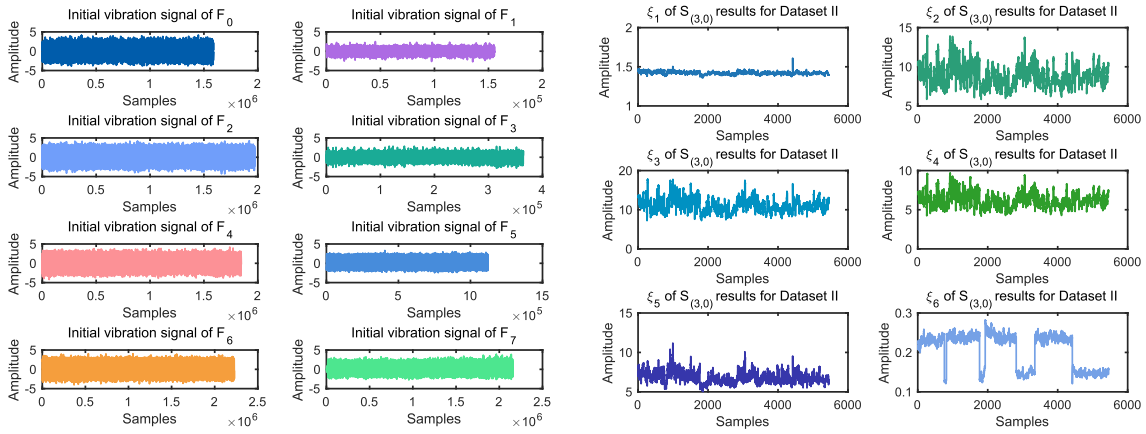
where $\mathcal{Y}_t$ represents the encoded label of $\mathcal{Q}_t$. Clearly, for any test dataset $\mathcal{D}_{test}$, the encoded of predicted fault types $\hat{\mathcal{Y}}_{\mathcal{D}_{test}}$ in the tth iteration can be determined as

$$\hat{\mathcal{Y}}_{\mathcal{D}_{test}} = \Phi_t(\mathcal{D}_{test}, W_e^t, \beta_e^t, W_h^t, \beta_h^t) = A_{(t)\mathcal{D}_{test}}^T W_{N_2}^t \tag{28}$$

where $A_{(t)\mathcal{D}_{test}}$ denotes the mapping results of $\mathcal{D}_{test}$ with the pretrained model Φ_t.

III. Experiment

In this section, several experiments of RMFD are conducted in detail to verify the effectiveness and availability of the proposal.



(a) (b)

Fig. 2. Schematic of related data in experiments. (a) Initial vibration signals for the normal type and seven fault types. (b) Examples of selected reconstructed signals with six dimensionless parameters of Dataset II.

TABLE I
DESCRIPTIONS OF RELATED FAULT TYPES OF THE RMFD

Type	*Description*	*Type*	*Description*
F_0	Bearing and gearbox are in the normal state.	F_4	Outer-ring of the bearing is deficient.
F_1	Bearing is normal and pinion is deficient.	F_5	Outer-ring of bearing with. the deficiency of pinion.
F_2	Inner-ring of bearing is deficient.	F_6	Some balls on the bearing are missing and gearbox is normal.
F_3	Inner-ring of bearing with the deficiency of pinion.	F_7	Absence of balls on the bearing with the deficiency of pinion.

TABLE II
DETAILS OF USED DATASETS IN EXPERIMENTS I AND II

Parameter	*Dataset I*	*Dataset II*	*Dataset III*
TWL	L	L	L
WPD Layer	1	3	3
Dimension	6	24	24
Step	256	256	512
Scale	5456	5456	2728

Notes: TWL denotes the time-window length where $L = 4096$.

A. Data Collection

Data measured in different operating modes are collected from a large-scale RMFD experimental platform. A realistic test rig is utilized to generate vibration signals which are conducted in *Guangdong Province Key Laboratory of petrochemical equipment fault diagnosis* [19], which is a centrifugal multistage impeller blower with key components such as a three-phase induction motor, gearbox, bearing, load, and shaft. Gear wheels and pinions are engaged and abrade mutually inside the gearbox. As shown in Fig. 2(a), the 1-D initial vibration signals for the normal type and seven fault types can be collected. Six dimensionless parameters are used as mentioned in Section II-A to complete the feature extraction task. Several examples are shown in Fig. 2(b). Furthermore, the descriptions of related fault types of the RMFD are illustrated in Table I. Specifically, the WPD technique is utilized to extract the features contained in the collected initial vibration signal S. In addition, the details of used datasets in Experiments I and II are summarized in Table II. Notably, the data for dimensionality reduction is obtained by downsampling the original vibration data by 1/8, which is beneficial to reduce the time complexity of data processing to a certain extent.

B. Experimental Setting

To illustrate the effectiveness of the RMFD method proposed in this brief, three different kinds of relevant experimental designs are conducted. As demonstrated in Section II-B, any single-criterion AL query strategies can be integrated into the proposed MCAL framework. In this brief, two advanced informativeness-measure approaches such as *uncertain sampling* (US) [23], *query-by-committee* (QBC) [24] and one advanced approach that representativeness-measure approach such as *K-center* [25] are utilized.

Remark 1: In detail, the original vibration signal is divided into multiple samples by a fixed time window. The related vibration signal of collected sample data can be recorded with a fault database. If experts believe that a specific sample needs to be annotated, it is feasible to perform various related techniques (such as spectrum analysis) by tracing back its corresponded original vibration signal. Experts can then evaluate the fault type of the corresponding fault sample. The results of these evaluations will be provided as labels to the classifier for training.

1) ***Experiment I****: Effects of Dimension and Scale for Query Strategies:* In this experiment, *Datasets I–III* are utilized. Two kinds of classes (i.e., four and six) are set for the sub-experiments, respectively. Notably, the number of data samples for various faults is different whereas the samples for compound faults are usually relatively small.

Five query strategies are considered to be compared with the proposed MCAL query strategy, which are as follows.

1) *US* [23]: Informativeness-measure AL approach (single criterion).
2) *QBC* [24]: Informativeness-measure AL approach (single criterion).

TABLE III
COMPARISON RESULTS OF ACC ± SD VALUES IN EXPERIMENT I

Dataset	*Strategies*	*4-Class*		*Strategies*	*6-Class*		*Overall Average Rank*
		Test Rate = 0.2	*Test Rate* = 0.3		*Test Rate* = 0.2	*Test Rate* = 0.3	
Dataset I	US	0.492 ± 0.06	0.528 ± 0.04	US	0.279 ± 0.02	0.267 ± 0.03	5.00
	QBC	0.546 ± 0.00	0.535 ± 0.01	QBC	0.348 ± 0.00	0.337 ± 0.01	2.00
	KCenter	0.522 ± 0.03	0.537 ± 0.03	KCenter	0.315 ± 0.01	0.304 ± 0.02	3.50
	DWUS	0.477 ± 0.03	0.482 ± 0.02	DWUS	0.274 ± 0.01	0.280 ± 0.01	5.75
	RS	0.523 ± 0.01	0.511 ± 0.01	RS	0.328 ± 0.01	0.341 ± 0.03	3.00
	Proposal*	0.539 ± 0.02	0.553 ± 0.01	Proposal*	0.339 ± 0.00	0.341 ± 0.01	**1.75**
Dataset II	US	0.683 ± 0.01	0.691 ± 0.01	US	0.563 ± 0.01	0.640 ± 0.01	5.00
	QBC	0.720 ± 0.002	0.704 ± 0.01	QBC	0.596 ± 0.02	0.661 ± 0.01	2.75
	KCenter	0.704 ± 0.01	0.704 ± 0.03	KCenter	0.608 ± 0.01	0.666 ± 0.01	2.25
	DWUS	0.666 ± 0.04	0.628 ± 0.02	DWUS	0.565 ± 0.01	0.625 ± 0.01	5.75
	RS	0.704 ± 0.02	0.690 ± 0.03	RS	0.597 ± 0.01	0.654 ± 0.01	3.75
	Proposal*	0.709 ± 0.02	0.715 ± 0.01	Proposal*	0.621 ± 0.01	0.663 ± 0.01	**1.50**
Dataset III	US	0.689 ± 0.02	0.684 ± 0.03	US	0.558 ± 0.01	0.527 ± 0.03	5.00
	QBC	0.714 ± 0.01	0.719 ± 0.01	QBC	0.591 ± 0.01	0.574 ± 0.01	2.25
	KCenter	0.701 ± 0.02	0.690 ± 0.01	KCenter	0.598 ± 0.01	0.596 ± 0.00	2.25
	DWUS	0.643 ± 0.03	0.644 ± 0.03	DWUS	0.545 ± 0.04	0.550 ± 0.01	5.75
	RS	0.695 ± 0.03	0.679 ± 0.02	RS	0.590 ± 0.01	0.589 ± 0.01	4.00
	Proposal*	0.712 ± 0.01	0.703 ± 0.02	Proposal*	0.601 ± 0.02	0.591 ± 0.02	**1.75**

3) *K-Center Approach* (KCenter) [25]: Representativeness-measure AL approach (single criterion).
4) *Density-Weighted US* (DWUS) [26]: Informativeness and representativeness-measure AL approach (multiple criteria).
5) *Random Sampling* (RS): Select the samples randomly.

Furthermore, the annotation of the initial small number of labeled samples ($|\mathcal{L}_0| = 0.01 * |\mathcal{U}_0|$, $\mathcal{P}_0 = \mathcal{U}_0$) is incomplete. In other words, several classes of fault samples are likely to be unknown by the classifier at the initial stage of manual annotation. For the *four-class* scenario, we set $|\mathcal{Q}_t| = 3$. While for the *six-class* scenario, $|\mathcal{Q}_t| = 6$. For all the experiments, we set $\mathcal{B} = 50$. Moreover, three random tests are employed for such tasks. The average accuracy (ACC) and standard deviation (SD) are utilized as the evaluation of the results.

2) ***Experiment II****: Effects of Classifier With the Proposed Query Strategy:* In this experiment, *Dataset III* is mainly considered. The aim is to illustrate the effectiveness of the BLS compared to different classifiers under the proposed multicriteria query strategy. In detail, the design of incrementally updating scheme is followed by the contexts mentioned in Section II-C where $N_f = N_e = 100$ and $N_1 = N_2 = 10$. In this case, logistic regression (LR) and support vector machine (SVM) with three different kernel functions, including *Linear* (SVM-I), *Polynomial* (SVM-II), and *radial basis function (RBF)* (SVM-III), are considered for comparison. The relevant other parameters are consistent with the default parameters in *scikit-learn* application programming interface (API). More setting details of the utilized AL implementation can be referred to [27].

Four diagnosis tasks (two-class, four-class, six-class, and eight-class) are conducted. $|\mathcal{Q}_t|$ is set as 2, 6, 10 and 15, respectively. The annotation of the initial small number of labeled samples ($|\mathcal{L}_0| = 0.01 * |\mathcal{U}_0|$, $\mathcal{P}_0 = \mathcal{U}_0$) is still incomplete.

C. Results and Comparison

The results of Experiment I with the overall average rankings of such six query strategies are reported in Table III. As stated in Section III-B, the US and QBC are two single-criterion AL informativeness-measure query strategies while the KCenter is a single-criterion AL representativeness-measure query strategy. DWUS is a MCAL query strategy preferred to the relatively informative and representative samples, which is identical to the proposed methodology essentially.

For different datasets, the overall average rankings of the proposal are 1.75, 1.50, and 1.75. As a comparison, DWUS has an overall average rank of 5.75 on all three datasets while the US strategy is also less effective. One important reason is that the effects of sample feature extraction in these datasets are different. On the whole, the effect of US relies too much on the effect of time-frequency domain feature extraction, which makes its effect is very poor in some datasets with insufficient features (even less effective than RS). Although DWUS is a multicriteria query strategy, it inherits the shortcomings of the US method to a certain extent due to its inflexibility. In contrast, the effect of KCenter and QBC is better overall. Such methods are less affected by the effect of feature extraction. The former relies on the selection of the core set at each iteration [25].

Moreover, the learning curves of different settings in Experiment I are shown in Figs. 3 and 4. It is worth noting that $|\mathcal{P}_t|$ is small in the early stage of the active query process (i.e., when t is small). In this context, there is less effective information in the data pool that can be exploited in this case, which causes a certain degree of fluctuation regarding the learning curves. Namely, the performance of the learning curve in the early stages of the active query process is less important than in the later stages of the active query process. As discussed in [12] and [20], the representativeness-measure strategies can generally receive relatively better effect in the early stage, while the performance of informativeness-measure strategies

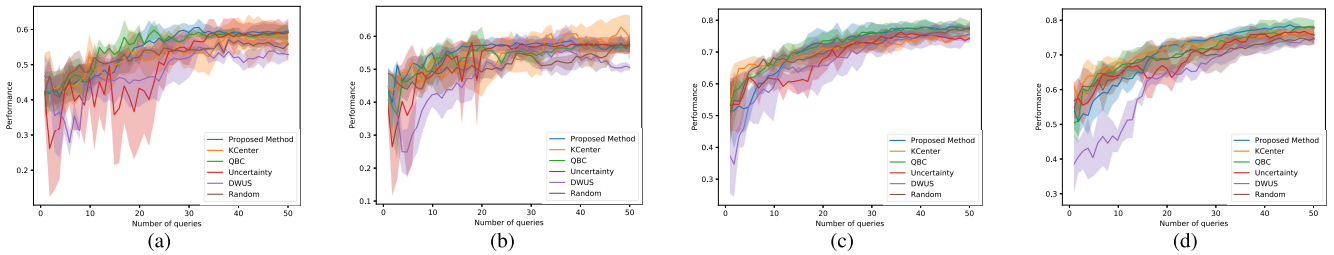


Fig. 3. Learning curves of different settings in *Experiment I* for four-class RMFD. (a) *Dataset I* where test ratio is 0.2. (b) *Dataset I* where test ratio is 0.3. (c) *Dataset II* where test ratio is 0.2. (d) *Dataset II* where test ratio is 0.3.

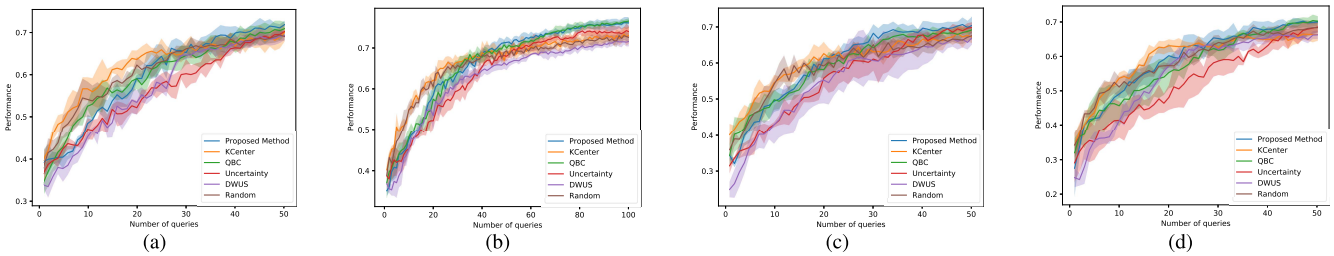


Fig. 4. Learning curves of different settings in *Experiment I* for six-class RMFD. (a) *Dataset II* where test ratio is 0.2. (b) *Dataset II* where test ratio is 0.3. (c) *Dataset III* where test ratio is 0.2. (d) *Dataset III* where test ratio is 0.3.

will surpass them in the later stage with the increase of $\mathcal{L}_t$. Considering that the average ACC $\pm$ SD of full process is utilized as the metric approach, the advantages of the proposed strategy do not seem to be particularly obvious as reported in Table III. However, it can usually reach a higher level earlier. As a result, less annotation budget is required in most cases, which further proves that our method can be more flexibly adapted to various data scenarios.

In Table IV, the results of Experiment II are reported. Although deep neural network (DNN)-based models may have higher diagnostic accuracy, the large time cost of the training procedure is unavoidable. Furthermore, although these shallow classifiers such as LR and SVM with various kernel functions can be trained quickly to some extent, more iterations are required generally to achieve the expected performance, which actually increases the cost of labeling. Considering that the cost of annotation in the actual RMFD task is also high, the fewer number of iterations to achieve the expected performance is also meaningful. Hence, it is essential to trade off the time cost of the training procedure and annotation cost in the AL framework simultaneously.

As mentioned in Section II-C, (27) is used for incremental model update if no new fault types appear in $\mathcal{Q}_t$. In this context, the *Greville theorem* is simply needed to calculate the *M-P* pseudo-inverse once in each iteration, which greatly reduces the required time complexity. Moreover, the diagnostic accuracy of BLS can also be guaranteed even with the limited labeled data with the utilization of the proposed query strategy. The consumed time for different classifiers in such four scenarios is represented by the time consumption corresponding to the SVM-I (denoted as $\mathcal{T}$). The test ratio is set as 0.3. MCAL* represents the proposed MCAL approach. As shown in Table IV, the more complex the categories of the dataset, the smaller the difference between BLS and shallow classifiers in terms of time cost. In detail, the time cost of BLS is about 1.49 times that of SVM-I for the two-class RMFD task. For the eight-class RMFD case, the corresponding ratio is reduced to 1.06. As shown in Fig. 5, the learning curves of different settings in Experiment II are summarized. The performance of BLS dominated almost the entire stage of different situations under the limited annotation budget based on the proposed query strategy, which demonstrates

TABLE IV
DETAILED AVERAGE ACC $\pm$ SD RESULTS OF EXPERIMENT II

Datasets	*Method*	*Average ACC $\pm$ SD*	*Time*
Dataset II 2-class	MCAL* + BLS	0.996 $\pm$ 0.00	1.49 $*$ $\mathcal{T}$
	MCAL* + LR	0.960 $\pm$ 0.02	1.02 $*$ $\mathcal{T}$
	MCAL* + SVM-I	0.973 $\pm$ 0.01	1.00 $*$ $\mathcal{T}$
	MCAL* + SVM-II	0.945 $\pm$ 0.01	0.99 $*$ $\mathcal{T}$
	MCAL* + SVM-III	0.925 $\pm$ 0.00	1.01 $*$ $\mathcal{T}$
Dataset II 4-class	MCAL* + BLS	0.882 $\pm$ 0.02	1.25 $*$ $\mathcal{T}$
	MCAL* + LR	0.750 $\pm$ 0.03	1.01 $*$ $\mathcal{T}$
	MCAL* + SVM-I	0.758 $\pm$ 0.02	1.00 $*$ $\mathcal{T}$
	MCAL* + SVM-II	0.714 $\pm$ 0.00	1.00 $*$ $\mathcal{T}$
	MCAL* + SVM-III	0.563 $\pm$ 0.02	1.01 $*$ $\mathcal{T}$
Dataset II 6-class	MCAL* + BLS	0.817 $\pm$ 0.01	1.14 $*$ $\mathcal{T}$
	MCAL* + LR	0.620 $\pm$ 0.02	1.01 $*$ $\mathcal{T}$
	MCAL* + SVM-I	0.651 $\pm$ 0.01	1.00 $*$ $\mathcal{T}$
	MCAL* + SVM-II	0.595 $\pm$ 0.01	1.02 $*$ $\mathcal{T}$
	MCAL* + SVM-III	0.420 $\pm$ 0.02	1.01 $*$ $\mathcal{T}$
Dataset II 8-class	MCAL* + BLS	0.747 $\pm$ 0.02	1.06 $*$ $\mathcal{T}$
	MCAL* + LR	0.566 $\pm$ 0.00	0.99 $*$ $\mathcal{T}$
	MCAL* + SVM-I	0.584 $\pm$ 0.01	1.00 $*$ $\mathcal{T}$
	MCAL* + SVM-II	0.432 $\pm$ 0.01	1.01 $*$ $\mathcal{T}$
	MCAL* + SVM-III	0.303 $\pm$ 0.01	1.01 $*$ $\mathcal{T}$

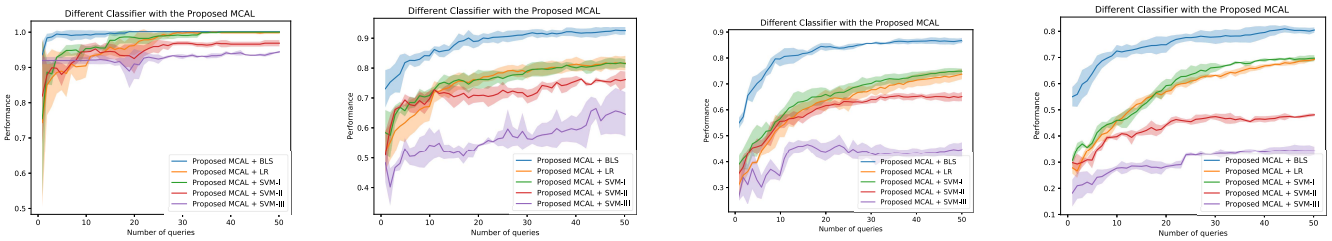


Fig. 5. Learning curves of different classifiers in *Experiment II* with the Proposed MCAL. (a) *Dataset III* with two-class RMFD where $|\mathcal{Q}_t| = 2$. (b) *Dataset III* with four-class RMFD where $|\mathcal{Q}_t| = 6$. (c) *Dataset III* with six-class RMFD where $|\mathcal{Q}_t| = 10$. (d) *Dataset III* with eight-class RMFD where $|\mathcal{Q}_t| = 15$.

the effectiveness and availability of the proposed diagnosis scheme.

IV. Conclusion

In this brief, a novel RMFD methodology of rotating machinery with limited expert interaction has been proposed. In detail, a stable and flexible MCAL scheme has been designed. The BLS has been introduced to complete the learning process iteratively with the design of applicable conditions. Several related real-life RMFD experiments have been conducted to show the effectiveness and availability compared with the existing studies.

Additionally, the representation of features regarding compound faults is of great significance for improving the diagnostic accuracy, which is one of our ongoing works.

References

[1] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.

[2] M. L. Corradini and G. Orlando, "A robust observer-based fault tolerant control scheme for underwater vehicles," *J. Dyn. Syst., Meas., Control*, vol. 136, no. 3, May 2014, Art. no. 034504.

[3] J. Chen *et al.*, "Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 70, pp. 1–35, Mar. 2016.

[4] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, nos. 1–2, pp. 108–126, Feb. 2013.

[5] L. Chang, Z. Zhou, Y. You, L. Yang, and Z. Zhou, "Belief rule based expert system for classification problems with new rule activation and weight calculation procedures," *Inf. Sci.*, vol. 336, pp. 75–91, Apr. 2016.

[6] Y. Huang, D. Xu, M. Tan, and H. Su, "Adding active learning to LWR for ping-pong playing robot," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 4, pp. 1489–1494, Jul. 2013.

[7] T. Zhao, Y. Zheng, J. Gong, and Z. Wu, "Machine learning-based reduced-order modeling and predictive control of nonlinear processes," *Chem. Eng. Res. Des.*, vol. 179, pp. 435–451, Mar. 2022.

[8] J. Chen, D. Zhou, Z. Guo, J. Lin, C. Lyu, and C. Lu, "An active learning method based on uncertainty and complexity for gearbox fault diagnosis," *IEEE Access*, vol. 7, pp. 9022–9031, 2019.

[9] P. Peng, W. Zhang, Y. Zhang, Y. Xu, H. Wang, and H. Zhang, "Cost sensitive active learning using bidirectional gated recurrent neural networks for imbalanced fault diagnosis," *Neurocomputing*, vol. 407, pp. 232–245, Sep. 2020.

[10] Y. Jin, C. Qin, Y. Huang, and C. Liu, "Actual bearing compound fault diagnosis based on active learning and decoupling attentional residual network," *Measurement*, vol. 173, Mar. 2021, Art. no. 108500.

[11] M. Chen, K. Zhu, R. Wang, and D. Niyato, "Active learning-based fault diagnosis in self-organizing cellular networks," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1734–1737, Aug. 2020.

[12] X. Zhan, H. Liu, Q. Li, and A. B. Chan, "A comparative survey: Benchmarking for pool-based active learning," in *Proc. 30th Int. Joint Conf. Artif. Intell.*, Aug. 2021, pp. 4679–4686.

[13] Y. Wang, Z. Pan, X. Yuan, C. Yang, and W. Gui, "A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network," *ISA Trans.*, vol. 96, pp. 457–467, Jan. 2020.

[14] X. Qing, J. Jin, Y. Niu, and S. Zhao, "Time–space coupled learning method for model reduction of distributed parameter systems with encoder–decoder and RNN," *AIChE J.*, vol. 66, no. 8, 2020, Art. no. e16251.

[15] Y. Fu, H. Cao, and X. Chen, "Adaptive broad learning system for high-efficiency fault diagnosis of rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[16] H. Zhao, J. Zheng, J. Xu, and W. Deng, "Fault diagnosis method based on principal component analysis and broad learning system," *IEEE Access*, vol. 7, pp. 99263–99272, 2019.

[17] M. Wang, Q. Ge, H. Jiang, and G. Yao, "Wear fault diagnosis of aeroengines based on broad learning system and ensemble learning," *Energies*, vol. 12, no. 24, p. 4750, 2019.

[18] Q. Hu, A. Qin, Q. Zhang, J. He, and G. Sun, "Fault diagnosis based on weighted extreme learning machine with wavelet packet decomposition and KPCA," *IEEE Sensors J.*, vol. 18, no. 20, pp. 8472–8483, Oct. 2018.

[19] J. Zhang, Q. Zhang, X. He, G. Sun, and D. Zhou, "Compound-fault diagnosis of rotating machinery: A fused imbalance learning method," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 4, pp. 1462–1474, Jul. 2021.

[20] Y. Zhao, Z. Shi, J. Zhang, D. Chen, and L. Gu, "A novel active learning framework for classification: Using weighted rank aggregation to achieve multiple query criteria," *Pattern Recognit.*, vol. 93, pp. 581–602, Sep. 2019.

[21] C. L. P. Chen and Z. L. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.

[22] D. Masti and A. Bemporad, "Learning nonlinear state–space models using autoencoders," *Automatica*, vol. 129, Jul. 2021, Art. no. 109666.

[23] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in *SIGIR*. London, U.K.: Springer, 1994, pp. 3–12.

[24] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proc. 5th Annu. Workshop Comput. Learn. Theory (COLT)*, 1992, pp. 287–294.

[25] O. Sener and S. Savarese, "Active learning for convolutional neural networks: A core-set approach," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–13.

[26] B. Settles, "Curious machines: Active learning with structured instances," Ph.D. thesis, Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, 2008.

[27] Y.-P. Tang, G.-X. Li, and S.-J. Huang, "ALiPy: Active learning in Python," 2019, *arXiv:1901.03802*.