A Feature Selection Approach Hybrid Grey Wolf and Heap-Based Optimizer Applied in Bearing Fault Diagnosis

CHUN-YAO LEE¹, (Member, IEEE), TRUONG-AN LE², AND YU-TING LIN¹

¹Department of Electrical Engineering, Chung Yuan Christian University, Taoyuan 320314, Taiwan
²Department of Electrical and Electronic Engineering, Thu Dau Mot University, Thu Dau Mot, Binh Duong 75000, Vietnam

Corresponding author: Chun-Yao Lee (cyl@cycu.edu.tw)

ABSTRACT An effective bearing fault diagnosis model based on machine learning is proposed in this study. The model can separate into three stages: feature extraction, feature selection, and classification. In the stage of feature extraction, multiresolution analysis (MRA) and fast Fourier transform (FFT) are applied to extract the features from the raw signal measured from the rotating machine. In the second stage, a powerful feature selection method is proposed and utilized in the stage of feature selection. The new feature selection method is based on grey wolf optimization (GWO) and heap-based optimizer (HBO) with new strategies combined. Finally, support vector machine (SVM) and linear discriminant analysis (LDA) are used as the classifier independently. To verify the capability of the proposed model, four different datasets are applied to test the model in this study, respectively University of California Irvine (UCI) benchmark dataset, bearing dataset, Case Western Reserve University (CWRU) benchmark dataset, and Machinery Failure Prevention Technology (MFPT) benchmark dataset. The proposed method is compared with the existing methods and can certify the robustness with the experiment results.

INDEX TERMS Bearing fault diagnosis, feature selection, grey wolf optimizer, heap-based optimizer.

I. INTRODUCTION

In modern industry, the rotating machine is widely used in the manufacturing factory and the most important part of a rotating machine is the rolling bearing [1]. The rotating machine failure occurs on the rolling bearing is the most common seen [2]–[4]. Rotating machine failure can lead to motor breakdown or even human injuries [5]. To detect the fault of rotating machine, or even to avoid the damage which may occur in the future is the most important task that needs to be solved. Therefore, the main purpose of this study is to investigate an effective bearing fault diagnosis approach based on machine learning.

The process of machine learning can generally separate into three stages: feature extraction, feature selection, and classification. The raw signal measured from the rotating machine is often mixed with various information and noise. The main purpose of feature extraction is to find the most important information and extract them as features, where the quality of features can significantly affect the classification accuracy [6], [7]. In recent years, several feature extraction methods have been proposed such as Hilbert-Huang transform (HHT) [8], empirical mode decomposition (EMD) [9], wavelet transform (WT) [10], discrete wavelet transform (DWT) [11], multiresolution analysis (MRA) [12], short time Fourier transform (STFT) [13], and fast Fourier transform (FFT) [14]. The MRA was proposed by Mallat in 1989. In the processing of the high-frequency signals, MRA has it good performance in decomposing the signal at different time windows and frequency bands [15], [16]. On the other hand, FFT is a well-known and useful algorithm for fault diagnosis, which is based on decomposing the discrete Fourier transform (DFT) to analyze the frequency domain [17], [18]. Thence, MRA and FFT were applied in this study.

After the stage of feature extraction, the feature set has been generated. However, with the amount of the collected data having considerable growth, the feature set extracted from the raw signal is usually accompanied by a great number of redundant features, especially high dimensional feature sets [19]. The redundant features might lead to lower classification accuracy [20], [21]. Therefore, feature selection plays a significant role in machine learning. Feature selection is a
The pre-processing stage of the data before it is fed into the classifier. With the revolution in feature selection, metaheuristic algorithms have got attention to solve optimization problems [22]. The classical algorithms are: particle swarm optimization (PSO) [23], genetic algorithm (GA) [24], ant colony optimization (ACO) [25], artificial bee colony (ABC) [26], whale optimization algorithm (WOA) [27], Harris hawks optimizer (HHO) [28], grey wolf optimizer (GWO) [29], and heap-based optimizer (HBO) [30]. GWO was proposed by S. Mirjalili et al. in 2014, which imitates the characteristics of grey wolves with their social hierarchy [29]. HBO has success in its exploitation phase but still suffers from a local trap and premature convergence problem in the exploration phase [22]. HBO was proposed by Q. Askari et al. in 2020 [30]. Inspired by the corporate rank hierarchy (CRH), HBO simulates CRH based on the heap data structure, where the heap is a nonlinear tree structure [30]. In this study, a new feature selection method was proposed based on GWO and HBO to balance the exploration phase and exploitation phase, which improved the capability of space searching.

The last stage of fault diagnosis based on machine learning is classification. In this stage, the classifiers distinguish the features in the optimal feature subset which represents the fault condition. Several classification techniques were applied in machine learning cases, such as artificial neural network (ANN) [31], support vector machine (SVM) [32], decision trees (DT) [33], random forest (RF) [34], k-nearest neighbors (k-NN) [35], principal component analysis (PCA) [36] and linear discriminant analysis (LDA) [37]. With the techniques mentioned, SVM has the capability in making decision for small number datasets and good generalization [38]. LDA can utilize in the case of classification and dimensionally reduction. LDA has good performance with complete data sets in the case of classification [39]. Thus, in this study, SVM and LDA were independently applied as the classifier. Combine with the feature extraction strategies and the proposed feature selection method, a bearing fault diagnosis model has been accomplished.

In this study, the main purpose is to investigate an effective bearing fault diagnosis approach. For this purpose, the main contributions were present as follows:

1) In the stage of feature extraction, a new strategy was proposed based on MRA and FFT, which can extract the features from the raw signal more effectively.

2) In the stage of feature selection, a new method was proposed based on the GWO algorithm and the HBO algorithm with two new strategies combined.

3) The proposed feature selection method was compared with other classical methods using the University of California Irvine (UCI) benchmark datasets and the bearing datasets to verify the capability of the method [40]. On the other hand, to verify the robustness, the classification accuracy of the proposed bearing fault diagnosis model was compared with the existing models using the Case Western Reserve University (CWRU) benchmark dataset and the Machinery Failure Prevention Technology (MFPT) benchmark dataset [41], [42].

The other sections in this study are as follows: section II discusses the process of feature extraction. Section III reviews the GWO algorithm and HBO algorithm, also the details of the proposed feature selection method are given. Section IV illustrates the procedure of the bearing fault diagnosis model. The experiment results are present in section V. In this section, UCI benchmark datasets, bearing dataset, CWRU benchmark dataset, and MFPT benchmark dataset were applied to verify the capability of the model. Finally, section 6 presents the conclusion.

II. THE FEATURE EXTRACTION METHOD

In this study, two feature extraction methods MRA and FFT were applied. The procedure of feature extraction is shown in Fig. 1, where Fig. 1a shows the raw current signal measured from the induction motor. Fig. 1b illustrates that 25 features were extracted from the raw signal by utilizing MRA and Fig. 1c illustrates that other 25 features by FFT. Table 1 describes the mathematical expressions of each feature are respectively: mean, maximum (max), kurtosis (kur), skewness (ske) and root mean square (mse). Finally, 50 features were extracted in this stage.

A. MULTiresolution ANALYSIS

The main idea of MRA is finding the discrete wavelet transform (DWT) coefficient vector to analyze the raw signal, as depicted in Fig. 2 [43]. The signal $x(t)$ decomposed into approximation coefficients and detail coefficients by using scale function and wavelet function [44]. One-dimensional filters were applied in this algorithm, respectively high-pass filter $H$ and low-pass filter $G$, and is down sampling which means it ignores the odd-indexed elements of the signal. The calculation of decomposing the raw signal $x(t)$ is shown in (1), and the calculation of scale function and wavelet function are shown in (2) and (3) respectively.

$$x(t) = \sum_{k} A_{j_0,k} \varphi_{j_0,k}(t) + \sum_{j} \sum_{k} D_{j,k} \psi_{j,k}(t)$$

$$\varphi_{j,k}(t) = \sqrt{2} \varphi(2^j t - k)$$

$$\psi_{j,k}(t) = \sqrt{2} \psi(2^j t - k)$$

where $A_j$ is approximation coefficients, and $D_j$ is detail coefficients. It is an essential issue to choose the appropriate mother wavelet, which can significantly influence the features extracted from the signal useful or not [45], [46]. In this study, the detail coefficients $d_1$ to $d_5$ were applied.

B. FAST FOURIER TRANSFORM

To obtain the information from the frequency domain, the signal $x(t)$ transforms from the time domain into the frequency domain on a finite number of frequency lines $f$ by using FFT [47]. The calculation of FFT is shown in (4). In this
In the stage of encircling, the wolves surround the prey and the prey will be at the center of wolves. Each wolf kept a certain distance from the prey. To simulate the situation, the distance $\vec{D}$ between the wolf and the prey is shown in (5). While the position of the wolves changes continuously according to the prey, shown in (6).

$$\vec{D} = \left| \vec{C} \times \vec{X}_p(t) - \vec{X}(t) \right|$$  \hspace{1cm} (5)

$$\vec{X}(t + 1) = \vec{X}_p(t) - \vec{A} \times \vec{D}$$  \hspace{1cm} (6)

where $t$ is the current number of iterations, $\vec{A}$ and $\vec{C}$ are coefficient vectors, $\vec{X}_p$ is the position of prey, and $\vec{X}(t)$ is the position of wolves. The calculation of $\vec{A}$ and $\vec{C}$ are shown in (7) and (8).

$$\vec{A} = 2\vec{\alpha} \times \vec{r}_1 - \vec{a}$$  \hspace{1cm} (7)

$$\vec{C} = 2\vec{\beta} \times \vec{r}_2$$  \hspace{1cm} (8)

where $\vec{r}_1$ and $\vec{r}_2$ are stochastic vectors in $[0, 1]$, and the calculation of coefficient $\vec{a}$ is shown in (9). Coefficient $\vec{a}$ decreased from 2 to 0 gradually while $t$ increases, and $\vec{A}$ is decided by $\vec{a}$. When $\left| \frac{\vec{A}}{\vec{r}_1} \right| < 1$, represents the search agent alternate from exploration phase into the exploitation phase.

$$\vec{a} = 2 - \frac{2 \times t}{T}$$  \hspace{1cm} (9)

where $T$ is the max number of iterations.
In the stage of hunting, the position of wolves will update according to (10).

\[
\vec{X}(t + 1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}
\]

(10)

where \(\vec{X}_1, \vec{X}_2\) and \(\vec{X}_3\) are decided by the distance between \(\alpha, \beta, \delta\) and the prey respectively, shown in (11), (12), and (13).

\[
\vec{X}_1 = \vec{X}_a - \vec{A}_1 \times \vec{D}_\alpha
\]

(11)

\[
\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \times \vec{D}_\beta
\]

(12)

\[
\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \times \vec{D}_\delta
\]

(13)

where \(\vec{X}_a, \vec{X}_\beta, \vec{X}_\delta\) are the position of \(\alpha, \beta, \delta\) respectively, \(\vec{D}_\alpha, \vec{D}_\beta\) and \(\vec{D}_\delta\) are the distance between \(\alpha, \beta, \delta\) and the prey respectively, shown in (14), (15), and (16).

\[
\vec{D}_\alpha = \left| \vec{C}_1 \times \vec{X}_a - \vec{X} \right|
\]

(14)

\[
\vec{D}_\beta = \left| \vec{C}_2 \times \vec{X}_\beta - \vec{X} \right|
\]

(15)

\[
\vec{D}_\delta = \left| \vec{C}_3 \times \vec{X}_\delta - \vec{X} \right|
\]

(16)

B. HEAP-BASED OPTIMIZER

In the heap structure, each heap node represents an individual in a corporation. The highest-level node is the chief executive officer (CEO) of the company, which represents the most potential solution. The nodes at lower levels are subordinates and at higher levels are senior executives. HBO has three types of interaction between the heap nodes, respectively: 1) Interact with immediate boss. 2) Interact with colleagues. 3) Self-contribution [30]. The heap updates continuously with the interaction between heap nodes to complete the searching task. To hybrid with GWO correctly, the solution generated based on HBO will be represented as the \(\bar{\text{sol}}\) in this study.

1) INTERACT WITH IMMEDIATE BOSS

In the heap, the parent node can be regarded as the immediate boss. The interaction with the immediate boss is shown in (17).

\[
\bar{\text{sol}}_i^k(t) = \frac{\vec{B}^k + \gamma \vec{k} \cdot \vec{B}^k - \vec{X}_i(t)}{\left| \vec{B}^k - \vec{X}_i(t) \right|}
\]

(17)

where \(\bar{\text{sol}}\) is the solution generated, \(\vec{B}\) is the parent node, \(\vec{k}\) is the \(k\) th bit in the stochastic vector \(\vec{\lambda}\) and \(\gamma\) is the parameter. \(\vec{\lambda}\) is calculate in (18) and \(\gamma\) is calculate in (19).

\[
\vec{\lambda}^k = 2r - 1
\]

(18)

\[
\gamma = \left\lfloor \frac{t \mod \frac{T}{C}}{\frac{T}{C}} \right\rfloor
\]

(19)

where \(r\) is stochastic number which evenly distributed in \([0, 1]\), \(T\) is the max number of iterations, \(C\) is a parameter defined by users. In this case, the value of \(C\) is defined as \(T/25\).

2) INTERACT WITH COLLEAGUES

The heap nodes with the same level can be regarded as colleagues. The interaction with colleagues is shown in (20).

\[
\bar{\text{sol}}_i^k(t) = \begin{cases}
\vec{X}_i^k(t), & p \leq p_1 \\
\vec{X}_i^k(t) + \gamma \vec{k} \cdot \vec{X}_i^k(t), & p > p_1 \text{ and } p \leq p_2 \\
\vec{X}_i^k(t), & p > p_2 \text{ and } p \leq p_3 \text{ and } f(\vec{S}_i) < f(\vec{X}_i(t)) \\
\vec{X}_i^k(t), & p > p_2 \text{ and } p \leq p_3 \text{ and } f(\vec{S}_i) \geq f(\vec{X}_i(t))
\end{cases}
\]

(20)

where \(\vec{S}\) is the vector of colleague node, \(f\) is objective function which calculate the fitness value of search agent.

3) SELF CONTRIBUTION

The self-contribution is shown in (21).

\[
\bar{\text{sol}}_i^k(t) = \vec{X}_i(t)
\]

(21)

To combine three types of interaction, roulette wheel is applied to balance the exploration phase and exploitation phase. The probability of roulette wheel is divided into \(p_1, p_2\) and \(p_3\) based on the number of iterations \(t\). The combination of interactions is shown in (22).

\[
\bar{\text{sol}}_i^k(t) = \begin{cases}
\vec{X}_i^k(t), & p \leq p_1 \\
\vec{X}_i^k(t) + \gamma \vec{k} \cdot \vec{X}_i^k(t), & p > p_1 \text{ and } p \leq p_2 \\
\vec{X}_i^k(t), & p > p_2 \text{ and } p \leq p_3 \text{ and } f(\vec{S}_i) < f(\vec{X}_i(t)) \\
\vec{X}_i^k(t), & p > p_2 \text{ and } p \leq p_3 \text{ and } f(\vec{S}_i) \geq f(\vec{X}_i(t))
\end{cases}
\]

(22)

where \(p\) is evenly distributed stochastic number in \([0, 1]\), the calculation of probabilities \(p_1, p_2\) and \(p_3\) are shown in (23), (24) and (25).

\[
p_1 = 1 - \frac{t}{T}
\]

(23)

\[
p_2 = p_1 - \frac{1 - p_1}{2}
\]

(24)

\[
p_3 = p_2 + \frac{1 - p_1}{2} = 1
\]

(25)

HBO applied a 3-ary heap in the structure of CRH, which means up to three subordinates that each heap node can managed. In order to make the heap work correctly, the heap represents in the form of an array. The index of the parent of node \(i\) is in a d-ary heap is shown in (26).

\[
\text{parent}(i) = \text{floor} \left[ \frac{i + 1}{d} \right]
\]

(26)

where \(\text{floor}\) is the floor function. In d-ary heap, each node can have up to \(d\) child node. The calculation of the index that represents the \(j\)th child of node \(i\) is shown in (27).

\[
\text{child}(i,j) = d \times i - d + j + 1
\]

(27)
In order to verify the position of each node, the nodes belonging to different levels are divided into different depths. For example, a child node has one more depth than its parent node, and nodes located at the same depth as their colleague nodes. The depth of the CEO node is 0. The depth of node $i$ is shown in (28).

$$\text{depth}(i) = \lfloor \log_d(d \times i - i + 1) \rfloor - 1$$  \hspace{1cm} (28)

For a node $i$, the index of its colleague node is randomly generated in the interval between $\left[ \frac{d \cdot \text{depth}(i) - 1}{d - 1} + 1, \frac{d \cdot \text{depth}(i) - 1}{d - 1} \right]$. In order to keep the heap correct, HBO search upward to insert the node $i$ as shown in pseudo-code 1. Finally, complete the establishment of heap architecture shown in pseudo-code 2.

---

**pseudo-code 1** Heapify_up(i)

**Input:** $i$ (index of node that is going to heapify)

Assumed that the other nodes were heap property

1. while $i \neq \text{root}$ and heap[i].key<heap[parent(i)].key
2. swap(heap[i], heap[parent(i)])
3. $i = \text{parent}(i)$
4. end

**pseudo-code 2** Build_heap

**Input:** $X_i(i = 1, 2, 3, \ldots, N), N$ ($N$ is the length of array)

1. for $i$=1:N
2. heap[i].value=i
3. heap[i].key=f($X_i$)
4. Heapify_up(i)
5. end

---

C. THE PROPOSED FEATURE SELECTION METHOD

Feature selection is a binary problem, in other words, which is accomplished by transforming the bits between 1 and 0. For example, converting 1 to 0 means the feature is deleted, and converting the bit value from 0 to 1 means the feature is selected. Therefore, all the bits must be represented in binary. A binary version with sigmoid function is used to deal with feature selection problem. Each bit converts into 0 or 1 is based on (30) [48].

$$X_{id}^{i+1} = \begin{cases} 1, & \text{if} \ \text{sigmoid} \left( \frac{X_1 + X_2 + X_3}{3} \right) \geq \text{rand}_2 \\ 0, & \text{otherwise} \end{cases}$$  \hspace{1cm} (30)

where $\text{rand}_2$ is stochastic number which evenly distributed in $[0, 1]$, and $\text{sigmoid}$ is the sigmoid function shown in (31).

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-10(x-0.5)}}$$  \hspace{1cm} (31)

GWO has its superior performance of its exploitation phase, but still prone to fall into local traps [22]. Therefore, in this study, a new approach HBBGWO which hybrid GWO and HBO was proposed. The proposed method improved the capability of both global and local searching. In addition, to increase the opportunity of escaping from the local traps, two strategies were proposed in this study. The two strategies operated while the condition is corresponded. The detail of two strategies and the proposed feature selection method have illustrated in the subsection: 1) Updating solution based on crossover operation. 2) Updating solution based on mutation. 3) A new feature selection approach HBBGWO is proposed.

---

1) UPDATING SOLUTION BASED ON CROSSOVER OPERATION

Crossover operation is one of the operators of the GA algorithm which can enhance the capability of global searching. The main process of crossover operation is to choose two parent genes and stochastic cut into several parts, then combine them into a new solution, where the length of the solution is the same as its parents [49]. Therefore, good parents are important in this process. Choosing good parents can prone to generating a good solution.

The new $\text{sol}$ was stored as the record after the interaction. In the first iteration, $X_a$ was saved as the record. Otherwise, the record update after the interaction every time. To implement the crossover operation at the appropriate occasion, a threshold was established before the crossover operation implemented. First, compare the $\text{sol}$ with the record. Crossover operation implement while the $\text{sol}$ is more than 90% similar with the record. In this study, three-point crossover operation was applied. The process of crossover operation is shown in Fig. 3. Roulette wheel was also applied to choose one of the parents from the heap, and operate crossover with the $\text{sol}$. The weight of the roulette wheel is set as the fitness value of each heap node so the good parent has more chance to be choose. In this study, the $k$-NN classifier is adopted to calculate the fitness value. Only consider to reduce the error rate, the calculation of fitness value is shown in (32) [50].

$$\text{fitness} = \frac{N_T}{N_T + N_F} \times 100\%$$  \hspace{1cm} (32)

where $N_T$ is the number of instances that is true predicted, $N_F$ is the number of instances that is false predicted.
2) UPDATING SOLUTION BASED ON MUTATION

Mutation is one of the other operators based on genetic algorithm [49]. The process of mutation is shown in Fig. 4. The bit which has been choice converting from 0 to 1 or from 1 to 0. Different from crossover operation, the purpose of mutation is to increase the diversity of solutions, which can increase the ability of local searching.

After implementing the crossover operation, check again if the sol is still the same as the record. The mutation implemented while the similarity of the sol and the record achieve 100%, this threshold can assist the searching in the space more detail. Another threshold was established to balance exploration and exploitation as shown in (33). If the random number rand3 conforms with (33), choose 10% of bits randomly to implement mutation. The random number rand3 is evenly distributed in [0, 1]. Otherwise, 5% of bits will be chosen to implement mutation. Based on (33), there has no chance to choose 10% of bits for mutation after t exceeds over 76% of T, which can speed up the searching progress during the early stage and slow down in the later stage.

\[ \text{rand}_3 \geq 1.3 \times \frac{t}{T} \]  

(33)

3) THE PROPOSED HBBGWO ALGORITHM

After the interaction, the sol will be utilized to update the wolves position. If the fitness value of the sol greater than or equal to the fitness value of \( \alpha \), then the sol replace \( \alpha \). If the fitness value of the sol less than \( \alpha \) and greater than or equal to \( \beta \), then the sol replace \( \beta \), and so on. If the fitness value of the sol less than \( \delta \), then the position of the wolves will not update. The details of updating the position of the wolves are shown in pseudo-code 3.

Finally, a new feature selection method HBBGWO was accomplished. Fig. 5 shows the flow chart of HBBGWO and the details are explained as follows:

Step 1: Set the parameters up (population \( N \), user defined parameter \( C \), maximum iteration \( T \)).

Step 2: Initialize the position of the wolves \( \vec{X}_i \) by using (29).

Step 3: Calculate the fitness value of \( \vec{X}_i \) by using (32).
Step 4: Determine the position of wolves \( \tilde{X}_\alpha, \tilde{X}_\beta, \tilde{X}_\delta \) respectively. In this study, the wolf with the highest fitness value will determine as \( \alpha \), where the second one is \( \beta \), and the third one is \( \delta \).

Step 5: Store \( \tilde{X}_\alpha \) as the record.

Step 6: Update the position of wolves \( \tilde{X}_i \) by using (6) and (10), and convert into binary by using (30).

Step 7: Calculate the fitness value of \( \tilde{X}_i \) by using (32).

Step 8: Build the heap based on pseudo-code 1 and pseudo-code 2.

Step 9: Calculate the parameter \( \gamma \) by using (19).

Step 10: Interact between the heap nodes by using (22) and generate a new solution \( \text{sol} \), then convert the \( \text{sol} \) into binary by using (30).

Step 11: Check the similarity of the \( \text{sol} \) and the record. If the similarity over 90\%, go to Step 12 to implement crossover operation. Otherwise, go to Step 17 to update the record as the \( \text{sol} \).

Step 12: Implement crossover operation and update the \( \text{sol} \).

Step 13: Check the similarity of \( \text{sol} \) with the record again. If the \( \text{sol} \) as same as the record, go to Step 14 to check if the condition is meet the (33). Otherwise, go to Step 17 to update the record as the \( \text{sol} \).

Step 14: Check if the condition is meet the (33). If “Yes”, go to Step 15, 10\% of the \( \text{sol} \) implement mutation. Otherwise, go to Step 16, 5\% of the \( \text{sol} \) implement mutation.

Step 15: Choose 10\% of the \( \text{sol} \) stochastic to implement mutation.

Step 16: Choose 5\% of the \( \text{sol} \) stochastic to implement mutation.

Step 17: Update the record as the \( \text{sol} \). The record only store the latest \( \text{sol} \) that generated after the interaction.

Step 18: Calculate the fitness value of the \( \text{sol} \) by using (32).

Step 19: Update the position \( \tilde{X}_\alpha, \tilde{X}_\beta, \) and \( \tilde{X}_\delta \) based on pseudo-code 3.

Step 20: Check if \( t \) exceeds the maximum number of iterations. If “Yes”, stop the process. Otherwise, go to Step 6 to start the next iteration.

pseudo-code 3 Updating Position of the Wolves

1: \textbf{if} fitness\(_{\text{sol}} \geq \) fitness\(_{\alpha}\) \textbf{then}
2: fitness\(_{\alpha} = \) fitness\(_{\text{sol}}; /\text{update fitness}_{\alpha}\)
3: \( X_\alpha = \text{sol}; /\text{update } \alpha \) position
4: \textbf{else if} fitness\(_{\text{sol}} < \) fitness\(_{\alpha} \& \) fitness\(_{\text{sol}} \geq \) fitness\(_{\beta}\) \textbf{then}
5: fitness\(_{\beta} = \) fitness\(_{\text{sol}}; /\text{update fitness}_{\beta}\)
6: \( X_\beta = \text{sol}; /\text{update } \beta \) position
7: \textbf{else if} fitness\(_{\text{sol}} < \) fitness\(_{\beta} \& \) fitness\(_{\text{sol}} \geq \) fitness\(_{\delta}\) \textbf{then}
8: fitness\(_{\delta} = \) fitness\(_{\text{sol}}; /\text{update fitness}_{\delta}\)
9: \( X_\delta = \text{sol}; /\text{update } \delta \) position
10: \textbf{end if}

IV. THE BEARING FAULT DIAGNOSIS MODEL

The bearing fault diagnosis model can generally separate into three stages: feature extraction, feature selection, and classification. Fig. 6 depicts the flow of the model. In the stage of feature extraction, MRA and FFT were applied to extract the features from the raw signal. 25 features were extracted by using the MRA, another 25 features were extracted by using FFT. Finally, total of 50 features were extracted in this stage.

In the stage of feature selection, a powerful feature selection method HBBGWO was proposed and applied. The proposed method can eliminate the redundant features more efficiently and robustly, which can increase the classification accuracy. In this stage, an optimal feature subset was generated.

In the stage of classification, SVM and LDA were applied to distinguish the features of normal motor bearing and different severities of fault motor bearing. SVM is a widely used classifier because of its advantages in resolving small sample and nonlinear problems. The basic idea of SVM is to map the input samples to a high dimensional space and find hyperplanes that can best separate all the classes while training [51]. After that, the hyperplane is used to separate different classes between testing samples. The parameters can deeply influence the classification performance, 10-fold cross validation is applied in this study [52], [53]. The main idea of LDA is to find a projection vector to enlarge the distance between the samples from the same class [54]. The type of a sample can be separate based on the line while predicting. In this stage, the optimal feature set was provided for the classifier to accomplish the task.

V. EXPERIMENTAL VERIFICATION AND RESULTS

In this study, the hardware equipment includes a personal computer (4 CPUs Intel Core i5-6500, 16GB of RAM), and the software of the experiment is MATLAB 2017-a version. In this study, four different datasets were applied to certify...
the capability and the robustness of the proposed motor fault diagnosis approach, which is explained as follows.

Case Study 1: The UCI benchmark datasets were adopted to verify the searching capability of the proposed feature selection method. In this case study, the proposed feature selection method was compared with three other methods: BGWO, BPSO and GA.

Case Study 2: The bearing dataset was applied in this case study to verify the capability and the robustness of the model. The raw signal is the current signal measured from the experimental induction motors. In this case study, the proposed bearing fault diagnosis model with the proposed feature selection method was compared with three other methods mentioned in case study 1.

Case Study 3: The benchmark dataset provided by the CWRU machine learning repository was applied to certify the capability and the robustness of the model. In this case study, the proposed bearing fault diagnosis model was compared with other state-of-art methods.

Case Study 4: Another benchmark dataset provided by the MFPT society was applied to verify the robustness of the model. In this case study, the proposed bearing fault diagnosis model was compared with other state-of-art methods.

A. PARAMETERS SETTINGS

The proposed HBBGWO is a wrapper approach of feature selection, which need a classification method as a predictor. The performance of the predictor can be used as the objective function to evaluate the fitness value of the feature subset [55]. In this study, the $k$-NN algorithm is applied to evaluate the fitness value which was described in subsection III.C.1. The $k$-NN algorithm is one of the simplest machine learning algorithm, which can evaluate the fitness value rapidly. The nearest neighbors $k$ of $k$-NN algorithm is set to 5, and the number of cross-validation $k$-fold is defined as 10. Table 2 shows the parameters setting of the proposed feature selection method and three other methods (BGWO, BPSO and GA). All the parameters are used throughout four case studies in this study.

### TABLE 2. Parameters settings of the proposed method and the three other feature selection methods.

|                      | The proposed method | BGWO       | BPSO       | GA          |
|----------------------|---------------------|------------|------------|-------------|
| Number of wolves:    | 10                  | 10         | 10         | 10          |
| Maximum number       | 100                 | 100        | 100        | 100         |
| of iterations $T$:   |                     |            |            |             |
| $d = 3$              |                     |            |            |             |
| $C = T / 25$         |                     |            |            |             |

### TABLE 3. The details of the UCI benchmark datasets.

| Datasets     | Features | Instances | Classes |
|--------------|----------|-----------|---------|
| BreastEW     | 30       | 569       | 2       |
| Ionosphere   | 34       | 351       | 2       |
| PenglungEW   | 325      | 73        | 7       |
| Segmentation | 19       | 2310      | 2       |
| Sonar        | 60       | 208       | 2       |
| Vehicle      | 18       | 946       | 4       |

and Vehicle. The details of each dataset are shown in Table 3. BreastEW is breast cancer Wisconsin (diagnostic) dataset that predicts whether the cancer is benign or malignant. Ionosphere is radar data used to classify radar signals returned from the ionosphere as good or bad. PenglungEW has a larger dimension with seven classes which has higher sensitivity. Segmentation is the dataset in which the instances were drawn randomly from a database of seven outdoor images. The images were hand segmented to create a classification for every pixel. Sonar is a dataset obtained by bouncing off a sonar signal from a metal cylinder or a cylindrical rock at various angles and under different conditions. Vehicle is the dataset of vehicle silhouettes, the images were acquired by a camera looking downwards at the model vehicle from a fixed angle of elevation.

2) COMPARISON WITH OTHER FEATURE SELECTION METHODS

To verify the proposed method comprehensive, three criterions were established: 1) The average fitness value. 2) The average number of selected features. 3) The average operating times. The average fitness value can reflect the space searching capability of the proposed method, and the number of selected features can significantly influence the efficacy of the classifier. Finally, the operating times can immediately reflect the computation cost. The results are average number after 30 times independent implementation. Table 4 shows the results of each method. The best results of each criterion are bolded. In terms of the fitness value, the proposed method got the best on Ionosphere, PenglungEW, Segmentation, Sonar, and Vehicle. The average fitness values of the proposed method are only lower than BGWO 0.03% on BreastEW but higher than BGWO 0.2%,
2.19%, 0.46%, 0.35%, and 0.22% on each other datasets respectively. For all the datasets, the proposed method got the higher average fitness values than BPSO and GA, especially better than GA. Therefore, the proposed method has better space searching capability than the three other methods.

On the other hand, in terms of the average number of selected features. The results of the proposed method have better than the three other compared methods on all the datasets. The average number of selected features of the proposed method have 0.4, 0.5, 59.4, 1.3, 3.4, and 0.2 less than BGWO on six datasets respectively. Besides that, the results of the proposed method have substantially better than BPSO and GA. The significant improvement in the number of the proposed method helps the classifier accomplish its detecting task more efficiently.

Finally, the results of average operating times are shown in Fig 7. The average operating times of the proposed method are the shortest on all the six datasets, less than BGWO 7.8 sec. (9.8%), 5.4 sec. (6.7%), 7.5 sec. (9.0%), 13.0 sec. (15.5%), 2.8 sec. (3.7%) and 5.2 sec. (6.9%) on six datasets respectively. The average operating times of GA are the greatest on every dataset. The interactions between the heap nodes are very simple, and the only record of the $sol$ simplifies the updating of the wolves position. The lesser operating times can reduce the computing cost directly.

C. CASE STUDY 2: BEARING DATASET OF CURRENT SIGNAL

1) DESCRIPTION OF THE DATASETS

The raw current signals are measured from the experimental induction motor with normal and three different types of fault. Fig. 8 shows the experimental hardware, respectively a 2hp induction motor (4 poles, 380V, 60Hz), a dynamometer (11kW, 2 rpm, 69 Hz), a data acquisition, and a personal computer. Fig. 9 shows the bearing with three different severities of artificial defect. Fig. 9a shows the square hole located on the inner race. Three different sizes of hole respectively 0.53mm $\times$ 0.53mm, 0.53mm $\times$ 1.24mm and 0.53mm $\times$ 1.96mm, which respectively shows in Fig. 9b, Fig 9c, and Fig. 9d. Normal condition and each type of fault collect 100 samples, totally 400 samples are collected, which has the length of 2000 data points. The rate of sampling is defined as 1k Hz. The ratio of training dataset and testing dataset is 70:30. Both the training dataset and testing dataset were random split from the optimal feature subset. The same process of splitting training dataset and testing dataset is also applied in Case study 3 and Case study 4. The raw signal measured from induction motor is usually full with noises in reality. In order to simulate the real situation more realistically, Gaussian white noise was added to the raw signal at five different levels ($\infty$ dB, 30dB, 20dB, 10dB, 0dB). The signal to noise ratio (SNR) is the indication of the noise level, which means that $\infty$ dB indicates no noise and 0dB indicates the largest noise. The robustness of the bearing fault diagnosis model was verified with different level of noise.

2) COMPARISON WITH OTHER FEATURE SELECTION METHODS

The optimal feature subset was generated after feature selection, then fed to SVM and LDA respectively. Fig. 10 shows the optimal convergence curves of each noise level. Fig. 10a shows the curves while under $\infty$dB. Although every feature selection method got 100% of fitness value, the proposed HBBGWO reached 100% with the least number of iterations. Fig. 10b shows the curves while under 30dB of noise. BGWO, GA, and the proposed method reached 99.50% of fitness value and BGWO reached the top slightly faster than the proposed method. Fig. 10c shows the curves while under 20dB

| Datasets            | BGWO       | BPSO       | GA         | The proposed |
|---------------------|------------|------------|------------|--------------|
| Avg. Fit. (%)       | Avg. No. F.|
| BreastEW            | 98.12      | 13.1       | 97.85      | 17.1         | 95.27       | 14.5        | 98.09       | 12.7         |
| Ionosphere           | 93.45      | 6.4        | 91.22      | 9.9          | 91.18       | 11.5        | 93.65       | 5.9          |
| PenglungEW           | 92.10      | 138.4      | 91.74      | 164.6        | 92.92       | 162.4       | 94.29       | 79.0         |
| Segmentation         | 92.57      | 7.5        | 91.94      | 9.1          | 92.87       | 9.4         | 93.03       | 6.2          |
| Sonar               | 92.50      | 20.8       | 89.74      | 28.5         | 89.31       | 29.2        | 92.85       | 17.4         |
| Vehicle             | 74.57      | 7.3        | 73.05      | 9.2          | 68.37       | 9.5         | 74.79       | 7.1          |
of noise. Both GA and the proposed method reached 96.50% of fitness value, but GA converge faster than the proposed method. Fig. 10d shows the curves while under 10dB of noise, where the proposed method got the highest 87.00% of fitness value and can reach the top first. Finally, Fig. 10e shows the curves while under 0dB of noise. The proposed method still got the highest 74.50% of fitness value, higher than BGWO 1.75%, BPSO 5.75%, and GA 2.50%. This figure shows the robustness of the proposed method while under greater noise, also a better optimal feature subset will provide to the classifier. Table 5 shows the optimal features of each method under different levels of noise. The proposed method contains the least number of features while under each level of noise, which points out that the good capability of selecting features in this case, where the fewer number of features can reduce the computation cost of the classifier. The indication of each feature is described in Table 1.

In this case study, the major purpose is to verify the capability and robustness of the bearing fault diagnosis model.
Two criterions were established in this case study: 1) The average classification accuracy. 2) The average operating times. Table 6 shows the average accuracy using the SVM classifier. Every method reached 100% while under dB of noise except the accuracy of without feature selection (without FS), and the accuracy decreased when the noise got greater. The accuracy of feature selection with any method has significantly higher than the accuracy without feature selection while under each level of noise, which points out the importance of the step of feature selection. While under 30dB of noise, GA and the proposed method got the same accuracy of 99.20%, higher than BGWO 0.87%. While under 20dB of noise, the proposed method got the highest 95.83% of accuracy, higher than BGWO 4.16%. While under 10dB of noise, the proposed method reached the highest 83.33% of accuracy, higher than BGWO 0.83%. Finally, the proposed method got the highest 71.67% of accuracy while under 0dB of noise, higher than BGWO 5%. The results show the robustness of the proposed method while the noise gets larger. Table 7 shows the average accuracy of using the LDA classifier. The accuracy with feature selection is still higher than the accuracy without feature selection under every level of noise. Both BGWO and the proposed method reached 100% of accuracy while under dB of noise. The performance of the proposed method is worse than BPSO while under 30dB of noise, and worse than BPSO while under 20dB of noise. The proposed method got the highest accuracy under 10dB and 0dB, which shows robustness under greater noises again. Compare Table 6 and Table 7, the highest accuracy of using the SVM classifier under each level of noise is higher than using the LDA classifier, except under dB of noise. Therefore, the SVM classifier is more suitable to apply in this case study. The average operating times of each method under different levels of noise are depicted in Fig. 11. The proposed method has the best performance under each level of noise. The average operating times are lesser than BGWO of 12.6 sec. (14.6%), 7.0 sec. (8.6%), 7.1 sec. (8.6%), 12.4 sec. (14.6%) and 7.4 sec. (9.3%) while under each level of noise respectively. GA has the longest average operating time and is significantly longer than the proposed method.

In this case study, the capability and the robustness of the proposed bearing fault diagnosis model were verified. On the other hand, the SVM classifier is more suitable than LDA classifier in this case study.

D. CASE STUDY 3: CWRU BENCHMARK DATASET
1) DESCRIPTION OF THE DATASET
CWRU benchmark dataset is one of the most favorable datasets to verify the performance of the proposed model. The
 experimental hardware includes the 2hp induction motor, the load motor, and the torque encoder. The raw signals are the vibration signal measured from the induction motor, which includes four different levels of load (0hp, 1hp, 2hp, 3hp), three different fault locations (inner race, outer race, ball), three different defect diameters (0.007 inches, 0.014 inches, 0.021 inches), and the sampling rate is 12kHz. The main purpose is to identify the severities and the location of the bearing fault. Thus, the different motor load is not considered in this study. Table 8 shows the details of the CWRU benchmark dataset. While under normal condition, the signal is cut into 100 samples under each level of load, a total of 400 samples are collected. Except for the normal condition, the signals are cut into 50 samples under each level of loads with three different defect diameters, a total of 200 samples are collected. The length of the samples is all 2000 data points. Finally, the data is separated into 10 classes.

2) COMPARISON WITH THE STATE-OF-ART STUDY
In recent years, several bearing fault diagnosis methods were proposed whether based on machine learning or deep learning. Chen et al. proposed multi-scale convolution neural network and long short-term memory (MCNN-LSTM) model, using raw vibration signals as input for an automatic feature learning neural network which extracted different frequency signal characteristics utilizing two convolutional neural networks, then long short-term memory was utilized to identify the fault type according to the learned features [56]. Li et al. proposed a rolling bearing fault diagnosis method which combined wavelet threshold denoising, genetic algorithm optimization variational mode decomposition (GA-VMD) and the whale optimization algorithm based on the von Neumann topology optimization least squares SVM (VNWOA-LSSVM) [57]. Yuan et al. proposed a rolling bearing fault diagnosis method based on convolutional neural network and support vector machine (CNN-SVM), using the continuous wavelet transform to convert one-dimensional original vibration signals into two-dimensional time-frequency images, then input for the constructed model [58]. Zuo et al. proposed a bearing fault diagnosis method spiking neural network (SNN), as the third generation neural network, is tailored for fault diagnosis in rotating machinery [59]. Zhang et al. proposed a self-supervised joint learning (SSJL) fault diagnosis method to reduce the dependent on labeled data, which based on three-channel vibration images [60].

In this case study, the proposed bearing fault diagnosis model was compared with the state-of-art model. Table 9 shows the classification accuracy of the proposed method and other state-of-art methods also using the CWRU benchmark dataset. The proposed model got the 99.55% of accuracy with the SVM classifier, and 97.09% accuracy with the LDA classifier. The proposed model with the SVM classifier got the highest accuracy in this case study, higher than the proposed model with the LDA classifier. In this case study, the operating time of the proposed method is 107.5 seconds, higher than the operating times of case study 2 because the complexity of CWRU benchmark dataset is higher than the bearing dataset. The proposed method not only reached a

![FIGURE 11.](image)

**TABLE 8.** The process of the CWRU benchmark dataset.

| Fault location | Loads (hp) | Defect diameters (inches) | Samples | Classes |
|----------------|------------|---------------------------|---------|---------|
| Normal         | 0/1/2/3    | 0.007                     | 200     | 1       |
| Inner race     | 0/1/2/3    | 0.014                     | 200     | 3       |
| Ball           | 0/1/2/3    | 0.021                     | 200     | 4       |
| Outer race     | 0/1/2/3    | 0.007                     | 200     | 5       |

**TABLE 6.** The average accuracy using the SVM classifier.

| Methods       | Average accuracy (%) | dB   | 30dB | 20dB | 10dB | 0dB |
|---------------|-----------------------|------|------|------|------|-----|
| Without FS    | 99.17                 | 97.50| 93.33| 75.83| 63.33|
| BGWO          | 100.00                | 98.33| 91.67| 82.50| 66.67|
| BPSO          | 100.00                | 97.50| 95.00| 81.67| 70.00|
| GA            | 100.00                | 99.20| 95.00| 82.50| 65.83|
| Proposed      | 100.00                | 99.20| 95.83| 83.33| 71.67|

**TABLE 7.** The average accuracy using the LDA classifier.

| Methods       | Average accuracy (%) | dB   | 30dB | 20dB | 10dB | 0dB |
|---------------|----------------------|------|------|------|------|-----|
| Without FS    | 99.58                | 97.44| 92.33| 78.94| 63.33|
| BGWO          | 100.00               | 98.17| 94.89| 82.64| 67.75|
| BPSO          | 99.97                | 98.94| 94.72| 80.53| 65.56|
| GA            | 99.83                | 98.36| 95.28| 82.58| 67.72|
| Proposed      | 100.00               | 98.47| 95.14| 83.00| 69.33|
higher accuracy than the SNN model but also can detect more classes. Therefore, the proposed bearing fault diagnosis model has the better capability that is suitable for appliance in practical tasks of fault diagnosis.

E. CASE STUDY 4: MFPT BENCHMARK DATASET

1) DESCRIPTION OF THE DATASET

MFPT is another benchmark dataset widely used in fault diagnosis. The details of the MFPT benchmark dataset after the process are shown in Table 10. The raw signals are vibration signals measured from the experimental motor, which includes baseline, outer race fault, and inner race fault three different health conditions. The sampling frequency of the baseline condition is 97656Hz with 270 pounds of loads, and the signal is cut into 600 samples. The sampling frequency of outer races failure and inner race failure are both 48828Hz with respectively 0, 50, 100, 150, 200, 250, and 300 pounds of loads. The signals were cut into 50 samples with each level of load, a total of 350 samples under seven types of load. The length of samples is all 2000 data points. Finally, the data is separated into 3 classes.

2) COMPARISON WITH OTHER FEATURE SELECTION METHODS

In this case study, the main purpose is also to compared the proposed bearing fault diagnosis model with the state-of-art method. Yuan et al. proposed the model CNN-SVM, which is illustrated in the last subsection [58]. Sun et al. proposed the second-order time-reassigned multi synchrosqueezing transform (STMSST) [61], which based on Gaussian modulated linear group delay (GLGD) model to deal with the vibration signals of fault features from those obtained time-frequency images. Wang et al. proposed a lightweight convolutional neural network (LCNN) method, which can largely satisfy the need of lesser parameter amount and storage space with high accuracy [62]. Zuo et al. proposed a bearing fault diagnosis method SNN, which illustrated in the last subsection [59]. Jiang et al. test the method based on Chen-Lee Chaotic and chaotic dynamic error maps were used in analysis and feature status identification [63].

In this case study, the proposed bearing fault diagnosis model was compared with the state-of-art methods using the MFPT dataset. Table 11 shows the results of five existing fault diagnosis methods and the proposed method. The proposed model got 99.49% of accuracy while using the SVM classifier, and 98.95% of accuracy while using the LDA classifier. Although the accuracy of the proposed model with the SVM classifier is lower than the accuracy of the LCNN method but still higher than four other methods. From the results of case study 2, case study 3, and case study 4, the proposed model with the SVM classifier is more suitable than LDA, and the capability and robustness of the proposed model are certified. On the other hand, the operating time of the proposed method using MFPT benchmark dataset is 87 seconds, which points out that the complexity of MFPT benchmark dataset is lower than CWRU benchmark dataset.

VI. CONCLUSION

In this study, a new bearing fault diagnosis model was proposed. In the stage of feature extraction, the model use MRA and FFT to extract the features from the raw signal. In the stage of feature selection, a new method HBOGWO was proposed. HBOGWO is an enhanced method, that hybrid the GWO algorithm, the HBO algorithm, and two strategies of updating position. Finally, SVM and LDA were adopted respectively as the classifier to distinguish the optimal feature subset provided by the feature selection methods.

Four different datasets were applied to verify the capability and the robustness of the proposed method. In Case study 1, the UCI benchmark dataset was applied to certify the performance of the proposed feature selection methods. The results show that the space searching capability of the proposed method is better than the BGO, BPSO, and GA. On the other hand, the number of selected features and the operating times are both lesser than three other methods. In Case study 2, the bearing dataset measured from induction motors was adopted to verify the bearing fault diagnosis model. Both
the performance of the selected features and the operating times of the proposed method are better than the three other methods, the proposed model also reached the highest classification accuracy with the SVM classifier. In this study, the proposed method can still detect the bearing fault with noises and different sources of signals more effectively and with robustness.

REFERENCES

[1] D. T. Hoang and H. J. Kang, “A motor current signal-based bearing fault diagnosis using deep learning and information fusion,” IEEE Trans. Instrum. Meas., vol. 69, no. 6, pp. 3325–3333, Jun. 2020.

[2] S. Singh and N. Kumar, “Detection of bearing faults in mechanical systems using stator current monitoring,” IEEE Trans. Ind. Informat., vol. 13, no. 3, pp. 1341–1349, Jun. 2017.

[3] C. Li, J. Valente de Oliveira, M. Cerrada, D. Cabrera, R. V. Sanchez, and G. Zurita, “A systematic review of fuzzy formalisms for bearing fault diagnosis,” IEEE Trans. Fuzzy Syst., vol. 27, no. 7, pp. 1362–1382, Jul. 2019.

[4] B. Dolenc, P. Boškoski, and D. Jaričić, “Distributed bearing fault diagnosis based on vibration analysis,” Mech. Syst. Signal Process., pp. 66–67, pp. 521–532, Jan. 2016.

[5] H. Wang, “A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN,” IEEE Trans. Meas. Instrum., vol. 69, no. 5, pp. 2377–2389, Aug. 2020.

[6] C. Zhang, Y. He, L. Yuan, and S. Xiang, “Analog circuit incipient bearing fault diagnosis method using DBN based features extraction,” IEEE Access, vol. 6, pp. 23053–23064, 2018.

[7] Y. Q. Chen, O. Fink, and G. Sansavini, “Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction,” IEEE Trans. Ind. Electron., vol. 65, no. 1, pp. 561–569, Jun. 2018.

[8] E. Elbouchikhi, V. Choqueuse, Y. Amirat, M. E. H. Benbourid, and S. Turri, “An efficient Hilbert–Huang transform-based bearing faults detection in induction machines,” IEEE Trans. Energy Convers., vol. 32, no. 2, pp. 401–413, Jun. 2017.

[9] X. Ye, Y. Fu, J. Shen, R. Feng, and G. Zhai, “An improved empirical mode decomposition based on adaptive weighted rational quartic spline for rolling bearing fault diagnosis,” IEEE Access, vol. 8, pp. 123813–123827, 2020.

[10] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zou, J. Yuan, B. Chen, and Z. He, “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.

[11] R. X. Chen, X. Huang, L. X. Yang, X. Y. Xu, X. Zhang, and Y. Zhang, “Intelligent fault diagnosis method of planetary gearboxes based on convolution network and discrete wavelet transform,” Comput. Ind., vol. 106, pp. 48–59, Apr. 2019.

[12] Z. Huang, Z. Wang, and H. Zhang, “Multilevel feature moving average ratio method for fault diagnosis of the microgrid inverter switch,” IEEE/CAA J. Autom. Sinica, vol. 4, no. 2, pp. 177–185, Apr. 2017.

[13] H.-Y. Chen and C.-H. Lee, “Vibration signals analysis by explainable artificial intelligence (XAI) approach: Application on bearing faults diagnosis,” IEEE Access, vol. 8, pp. 134246–134256, 2020.

[14] M. Jalayer, C. Orsenigo, and C. Vercellis, “Fault detection and diagnosis for rotating machinery: A model based on convolutional LSTM, fast Fourier and continuous wavelet transforms,” Comput. Ind., vol. 125, Feb. 2021, Art. no. 103378.

[15] J.-D. Wu and C.-H. Liu, “Investigation of engine fault diagnosis using discrete wavelet transform and neural network,” Expert Syst. Appl., vol. 35, no. 3, pp. 1200–1213, Oct. 2008.

[16] C. Pang and M. Kezunovic, “Fast distance relay scheme for detecting symmetrical fault during power swing,” IEEE Trans. Power Del., vol. 25, no. 4, pp. 2206–2211, Oct. 2010.

[17] G. Devadasu and M. Sushama, “A novel multiple fault identification with fast Fourier transform analysis,” in Proc. Int. Conf. Emerg. Trends Eng., Technol. Sci. (ICETETS), Vadukkotta, India, Feb. 2016, pp. 1–5.

[18] B. Cai, Y. Zhao, H. Liu, and M. Xie, “A data-driven fault diagnosis methodology in three-phase inverters for PMSM drive systems,” IEEE Trans. Power Electron., vol. 32, no. 7, pp. 5590–5600, Jul. 2017.

[19] X. Yu, F. Dong, E. Ding, S. Wu, and C. Fan, “Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection,” IEEE Access, vol. 6, pp. 3715–3730, 2017.

[20] M. Kang, M. R. Islam, J. Kim, J. M. Kim, and M. Pecht, “A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics,” IEEE Trans. Ind. Electron., vol. 63, no. 5, pp. 3299–3310, May 2016.

[21] J. R. Vergara and P. A. Estévez, “A review of feature selection methods based on mutual information,” Neural Comput. Appl., vol. 24, no. 1, pp. 175–186, Jan. 2014.

[22] R. Al-Wajih, S. J. Abdulkadir, N. Aziz, Q. Al-Tashi, and N. Talpur, “Hybrid binary grey wolf with Harris hawks optimizer for feature selection,” IEEE Access, vol. 9, pp. 31662–31677, 2021.

[23] Y. Cheng, Z. Wang, W. Zhang, and G. Huang, “Pulse flame optimization algorithm to solve the deconvolution problem for rolling element bearing fault diagnosis,” ISA Trans., vol. 90, pp. 244–267, Jul. 2019.

[24] Z. Liao, L. Song, P. Chen, and S. Zuo, “An automatic filtering method based on an improved genetic algorithm with application to rolling bearing fault signal extraction,” IEEE Sensors J., vol. 17, no. 19, pp. 6340–6349, Oct. 2017.

[25] M. Ma, C. Sun, and X. Chen, “Deductive deep belief networks with ant colony optimization for health status assessment of machine,” IEEE Trans. Instrum. Meas., vol. 66, no. 12, pp. 3115–3125, Dec. 2017.

[26] J.-M. Huang, R.-J. Wai, and G.-J. Yang, “Design of hybrid artificial bee colony algorithm and semi-supervised extreme learning machine for PV fault diagnoses by considering dust impact,” IEEE Trans. Power Electron., vol. 35, no. 7, pp. 7086–7099, Jul. 2020.

[27] X. Zhang, Z. Liu, Q. Miao, and L. Wang, “Rolling bearing fault diagnosis using a whale optimization algorithm-optimized orthogonal matching pursuit with a combined time–frequency atom dictionary,” Mech. Syst. Signal Process., vol. 107, pp. 29–42, Jul. 2018.

[28] W. Fu, K. Shao, J. Tan, and K. Wang, “Fault diagnosis for rolling bearings based on composite multiscale fine-sorted dispersion entropy and SVM with hybrid mutation SCA-HHO algorithm optimization,” IEEE Access, vol. 8, pp. 13086–13104, 2020.

[29] S. Mirjalili, S. Mirjalili, and A. Lewis, “Grey wolf optimizer,” Adv. Eng. Softw., vol. 69, pp. 46–61, Mar. 2014.

[30] Q. Askari, M. Saeed, and I. Younas, “Heap-based optimizer inspired by corporate rank hierarchy for global optimization,” Expert Syst. Appl., vol. 161, Dec. 2020, Art. no. 113702.

[31] W. Chine, A. Mellet, V. Ligth, A. Malek, G. Sulligoi, and A. M. Pavan, “A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks,” Renew. Energy, vol. 90, pp. 501–512, May 2016.

[32] M. Cui, Y. Wang, X. Lin, and M. Zhong, “Fault diagnosis of rolling bearings based on improved autoencoder and support vector machine,” IEEE Sensors J., vol. 21, no. 4, pp. 4927–4937, Feb. 2021.

[33] L. Song, P. Chen, and H. Wang, “Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery,” IEEE Trans. Instrum. Meas., vol. 67, no. 8, pp. 1887–1899, Aug. 2018.

[34] S. S. Roy, S. Dey, and S. Chatterjee, “Auto correlation aided random forest classifier based bearing fault detection framework,” IEEE Sensors J., vol. 20, no. 18, pp. 10792–10800, Sep. 2020.

[35] Q. Wang, S. Wang, B. Wei, W. Chen, and Y. Zhang, “Weighted K-NN classification method of bearings fault diagnosis with multi-dimensional sensitive features,” IEEE Access, vol. 9, pp. 45428–45440, 2021.

[36] H. Zhao, J. Zheng, X. Yu, and W. Deng, “Fault diagnosis method based on principal component analysis and broad learning system,” IEEE Access, vol. 7, pp. 99263–99272, 2019.

[37] R. Z. Haddad and E. G. Strangas, “On the accuracy of fault detection and separation in permanent magnet synchronous machines using MCSA/MVSA and LDA,” IEEE Trans. Energy Convers., vol. 31, no. 3, pp. 924–934, Sep. 2016.
[38] Y. Li, M. Xu, Y. Wei, and W. Huang, “A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree,” Measurement, vol. 77, pp. 80–94, Jan. 2016.

[39] M. R. Feldesman, “Classification trees as an alternative to linear discriminant analysis,” Amer. J. Phys. Anthropol., vol. 119, no. 3, pp. 257–275, Oct. 2002.

[40] UCI Machine Learning Repository. Accessed: Sep. 5, 2019. [Online]. Available: http://archive.ics.uci.edu/ml

[41] W. A. Smith and R. B. Randall, “Rolling element bearing diagnostics using the case Western Reserve University data: A benchmark study,” Mech. Syst. Signal Process., vol. 64–65, pp. 100–131, Dec. 2015.

[42] E. Bechhoefer (2016). A Quick Introduction to Bearing Envelope Analysis. MFPT Data. [Online]. Available: http://www.mftp.org/FaultData/FaultData.htm.Set

[43] C.-S. Felix, F.-G. Walter, G.-C. Raul, and C.-F. Dora, “Multiresolution analysis based on Mallat’s pyramidal algorithm applied to GPR data,” in Proc. 15th Int. Conf. Ground Penetrating Radar, Jun. 2014, pp. 647–650.

[44] C.-Y. Lee and M.-S. Wen, “Establish induction motor fault diagnosis system based on feature selection approaches with MRA,” Processes, vol. 8, no. 9, p. 1055, Aug. 2020.

[45] V. Vakharia, V. Gupta, and P. Kankar, “A multiscale permutation entropy based approach to select wavelet for fault diagnosis of ball bearings,” J. Vib. Control, vol. 21, no. 16, pp. 3123–3131, Feb. 2014.

[46] H. He, Y. Tan, and Y. Wang, “Optimal wavelet selection for ECG noise reduction using a comprehensive entropy criterion,” Entropy, vol. 17, no. 12, pp. 6093–6109, Sep. 2015.

[47] D. Strömbergsson, P. Marklund, K. Berglund, and P. Larsson, “Bearing monitoring in the wind turbine drivetrain: A comparative study of the FFT and wavelet transforms,” Wind Energy, vol. 23, no. 6, pp. 1381–1393, Jun. 2020.

[48] E. Emary, H. M. Zawbaa, and A. E. Hassanien, “Binary grey wolf optimization approaches for feature selection,” Neurocomputing, vol. 172, pp. 371–381, Jan. 2016.

[49] C. Kim, R. Batra, L. Chen, H. Tran, and R. Ramprasad, “Polymer design using genetic algorithm and machine learning,” Comput. Mater. Sci., vol. 186, Jan. 2021, Art. no. 110067.

[50] M. Van and H.-J. Kang, “Wavelet kernel local Fisher discriminant analysis with particle swarm optimization algorithm for bearing defect classification,” IEEE Trans. Instrum. Meas., vol. 64, no. 12, pp. 3588–3600, Dec. 2015.

[51] H. Zhao, Y. Gao, H. Liu, and L. Li, “Fault diagnosis of wind turbine bearing based on stochastic subspace identification and multi-kernel support vector machine,” J. Mod. Power Syst. Clean Energy, vol. 7, no. 2, pp. 350–356, Mar. 2019.

[52] V. Dave, S. Singh, and V. Vakharia, “Diagnosis of bearing faults using multi fusion signal processing techniques and mutual information,” Indian J. Eng. Mater. Sci., vol. 27, no. 4, pp. 878–888, 2020.

[53] H. Zheng and L. Zhou, “Rolling element bearing fault diagnosis based on support vector machine,” in Proc. 2nd Int. Conf. Consum. Electron., Commun. Nerv. (CECCNet), Yichang, China, Apr. 2012, pp. 544–547.

[54] J. Wen, X. Fang, J. Cui, L. Fei, and K. Yan, “Robust sparse linear discriminant analysis,” IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 2, pp. 390–403, Feb. 2018.

[55] C. Khammassi and S. Krichen, “A GA-LR wrapper approach for feature selection in network intrusion detection,” Comput. Secur., vol. 70, pp. 255–277, Sep. 2017.

[56] X. Chen, B. Zhang, and D. Gao, “Bearing fault diagnosis base on multi-scale CNN and LSTM model,” J. Intell. Manuf., vol. 5, pp. 1–17, Jun. 2020.

[57] J. Li, W. Chen, K. Han, and Q. Wang, “Fault diagnosis of rolling bearing based on GA-VMD and improved WOA-LSSVM,” IEEE Access, vol. 8, pp. 166753–166767, 2020.

[58] L. Yuan, D. Lian, X. Kang, Y. Chen, and K. Zhai, “Rolling bearing fault diagnosis based on convolutional neural network and support vector machine,” IEEE Access, vol. 8, pp. 137395–137406, 2020.

[59] L. Zhao, L. Zhang, Z.-H. Zhang, X.-L. Luo, and Y. Liu, “A spiking neural network-based approach to bearing fault diagnosis,” J. Manuf. Syst., vol. 61, pp. 714–724, Oct. 2021.

[60] W. Zhang, D. Chen, and Y. Kong, “Self-supervised joint learning fault diagnosis method based on three-channel vibration images,” Sensors, vol. 21, no. 14, p. 4774, Jul. 2021.

[61] G. Sun, Y. Gao, Y. Xu, and W. Feng, “Data-driven fault diagnosis method based on second-order time-reassigned multisynesqueezing transform and evenly mini-batch training,” IEEE Access, vol. 8, pp. 120859–120869, 2020.

[62] Y. Wang, J. Yan, Q. Sun, Q. Jiang, and Y. Zhou, “Bearing intelligent fault diagnosis in the industrial Internet of Things context: A lightweight convolutional neural network,” IEEE Access, vol. 8, pp. 87329–87340, 2020.

[63] B.-L. Jian, X.-Y. Su, and H.-T. Yau, “Bearing fault diagnosis based on chaotic dynamic errors in key components,” IEEE Access, vol. 9, pp. 53509–53517, 2021.

CHUN-YAO LEE (Member, IEEE) received the Ph.D. degree from the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, in 2007. From 1998 to 2008, he was a Distribution System Designer with the Engineering Division, Taipei City Government, and CECl Engineering Consultants, Inc., Taiwan. From 2004 to 2006, he was a Visiting Scholar with the Department of Electrical Engineering, University of Washington, Seattle, WA, USA, sponsored by the Ministry of Science and Technology, Taiwan. Since 2008, he has been a Faculty Member of electrical engineering with Chung Yuan Christian University, Taoyuan, Taiwan, where he is currently a Full Professor. His research interests include power distribution, optimization algorithms, and damage diagnosis. He received the National Excellent Teachers Award of Taiwan, in 2020.

TRUONG-AN LE received the M.S. degree in electronic engineering from the University of Transport and Communications, Vietnam, in 2012, and the Ph.D. degree from the Department of Electrical Engineering, Chung Yuan Christian University, Taiwan, in 2022. He is currently a Faculty Member at the Engineering and Technology Institute with Thu Dau Mot University, Binh Duong, Vietnam. His research interests include fault detection, optimization algorithms, machine learning, and deep learning.

YU-TING LIN received the B.S. degree in electronic engineering from Chung Yuan Christian University, Taoyuan, Taiwan, in 2020, where he is currently pursuing the M.S. degree in electrical engineering. His research interests include fault detection and optimization algorithms.