IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 15, 2020, accepted October 27, 2020, date of publication November 2, 2020, date of current version November 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3035081

Intelligence Bearing Fault Diagnosis Model Using Multiple Feature Extraction and Binary Particle Swarm Optimization With Extended Memory

CHUN-YAO LEE, (Member, IEEE), AND TRUONG-AN LE
Department of Electrical Engineering, Chung Yuan Christian University, Taoyuan 320314, Taiwan
Corresponding author: Chun-Yao Lee (cyl@cycu.edu.tw)

ABSTRACT This article presents an effective bearing fault diagnosis model based on multiple extraction and selection techniques. In multiple feature extraction, the discrete wavelet transform, envelope analysis, and fast Fourier transform are considered. While the combined binary particle swarm optimization with extended memory is focusing on feature selection. The current signals are analyzed by discrete wavelet transform. From there, the statistical features in the time and frequency domain are extracted by two techniques: envelope analysis, fast Fourier transform. Subsequently, the binary particle swarm optimization is combined with extended memory and two proposed position update mechanisms to eliminate redundant or irrelevant features to achieve the optimal feature subset. Besides, three classifiers including naïve Bayes, decision tree, and linear discriminant analysis are applied and compared to select the best model to detect the bearing fault.

INDEX TERMS Fault diagnosis, feature extraction, feature selection, particle swarm optimization.

I. INTRODUCTION

Rotary machines are one of the most important equipment in the manufacturing industry and other fields nowadays. Many rotating machinery failures cause not only lead to huge economic losses but also potentially serious casualties. One of the most important components in rotating machinery is bearings. According to Electric Power Research Institute research, the failure rate of bearings accounts for 41% of the faults in rotary machines [1]. Therefore, in recent years, bearings fault diagnosis models have been getting more and more attention from researchers. In this study, a high-performance bearings fault diagnosis model is investigated based on a combination of techniques to extract attributes from the current signals of the induction motors. From there, an optimal feature subset is determined by the feature selection algorithm. An appropriate classifier is used to identify bearing faults via the optimal feature subset.

The technique of extracting signal attributes is also known as feature extraction which applies signal analysis methods to calculate characteristic features of signals statistics in time domain, frequency domain and time-frequency domain. The characteristic features are very important factors that contribute to the effectiveness of the classification model. Some traditional signal analysis techniques are widely used as envelope analysis (EA), fast Fourier transform (FFT). There are also some techniques developed for non-linear and non-stationary signals such as short-time Fourier transform (STFT), wavelet transform (WT), local mean decomposition (LMD), Hilbert-Huang transform (HHT). However, the above techniques have their own limitations. EA is a simple method and low computational cost but the signal envelopes tend to overshoot and undershoot [2]. This will likely affect the quality of the extracted features. FFT has high resolution in the frequency domain but is not effective in time domain. While STFT has a fixed resolution in both the frequency and time domain [3]. WT overcame the problems of STFT with high resolution in both the frequency and time domain. This is done by working with different window sizes. However, WT is not enough effective for nonlinear signals [4]. Compared with WT, LMD is essentially self-adaptive feature, and the endpoint effect of LMD involves a certain degree of inhibition, simultaneously addressing the problem of under-envelope and over-envelope [5]. Recently HHT has attracted more attention from researchers. This technique combines the empirical mode decomposition (EMD) and Hilbert transforms (HT) for analysis to analyze nonlinear and non-stationary time series. However, one problem

The associate editor coordinating the review of this manuscript and approving it for publication was Her-Terng Yau.

that still exists for HHT is that it is particularly sensitive to noise [6].

For these reasons, the combination of signal analysis techniques is a better approach to feature extraction. Discrete wavelet transform (DWT) is one alternative of WT. Therefore, DWT has full advantages of WT. In addition, DWT also has good noise reduction ability. To remove noise and extract the potential features from the current signal, the combination of DWT, FFT and EA has been investigated in this study. First, DWT decomposed the original signal into components with different frequencies. From there, the statistical characteristics in the frequency and time domain are extracted from these components by two techniques: FFT and EA.

However, among those features that are extracted by the signal processing method, it is possible that redundant or irrelevant features are included in the feature set [7]. That affects the predicted accuracy. In order to solve this problem, the feature selection technique is used to eliminate redundant and irrelevant features, finding the most optimal feature set. In recent years, evolutionary computational techniques (ECs) have been implemented for feature selection problems such as particle swarm optimization (PSO), genetic algorithm (GA), grey wolf optimizer (GWO), differential evolution (DE). However, each technique has its own shortcomings. The main drawback of PSO is the premature convergence to find the best solution and prone fall into a local optimum trap [8]. While GA owns crossover and mutation operations which are relatively directionless operations in the search space, that makes GA a high computational cost [8]. BWO also encountered the same problems as PSO, it has tended to fall into a local optimum [9]. DE also has the same shortcomings as other evolutionary algorithms. In addition, DE is also difficult to adjust control parameters for various issues [10].

The solutions presented in ECs for feature selection are number sequences, whereby the dimensionality is equal to the total number of features in the dataset [11]. The elements in number sequence can be binary numbers in binary search space or real numbers in continuous search space. In this study, the ECs for feature selection in the binary search space is considered. Therefore, the above techniques are also known as binary-PSO (BPSO) [12], binary-GA (BGA) [13], binary-GWO (BGWO) [9], [14], binary-DE (BDE) [15]. In which the optimal solution is presented as a binary string, where '1' shows the corresponding feature is selected and '0' means not selected. Almost no optimal techniques can guarantee the perfect optimal feature set except when an exhaustive search is performed. However, an exhaustive search often takes a long time because a feature set with n features includes the total number of 2^n solutions [16]. Despite this, PSO and its variants are still the optimal methods that researchers invest the most [11], [17], [18].

Unlike PSO, the velocity in BPSO does not directly determine the new position of a particle. It is converted into the binary value for the corresponding position by the transfer function. In the binary search space, new solutions are created by flipping their position entries [19]. Therefore, there is no guarantee that the new positions after going through the update process do not coincide with the previous ones. This can waste resources and increase computational costs. To solve the above problem, the BPSO-based feature selection technique with extended memory is proposed, so-called BPSO-EM. Extended memory is proposed to be used to store all historical solutions. New solutions are compared with in-memory solutions to ensure that there is no repetition and maintain high diversity among the particles. In addition, two new position update mechanisms are proposed to enhance the ability to find promising areas around the best solution and global exploration capabilities.

Besides the hard work of finding an optimal feature set, feature classification is another important part of the fault diagnosis model. The optimal feature subset from the feature selection technique is the input of the classifier, which is evaluated based on classification accuracy. Currently, neural networks are widely applied in the field of intelligent diagnosis of machinery faults [20], [21]. However, it is a complex system. Neural networks are also more computationally expensive than traditional algorithms [22]. Meanwhile, machine learning algorithms with simpler advantages but high classification efficiency, suitable for small and medium data sets. Therefore, in this study, three machine learning algorithms naïve Bayes (NB), decision tree (DT), linear discriminant analysis (LDA) were applied and their classification performance was compared to select the most effective fault diagnosis model. Based on the strengths and weaknesses of the methods analyzed above, they motivated us to propose an intelligence bearing fault diagnosis model using multiple feature extraction and binary particle swarm optimization with extended memory for feature selection was created.

The key points of this study are summarized as follows:

1) Propose a feature extraction technique from a combination of three signal analysis methods (DWT, FFT, and EA). The main advantage of this approach is simplicity and low computational complexity. This approach helps to reduce the size of raw data, extract potential features to increase diagnostic efficiency.
2) Propose a feature selection technique using BPSO with extended memory and incorporate two new position update mechanisms to enhance exploration and exploitation abilities, maintain a high diversity population, and prevent premature convergence.
3) Propose an intelligent and efficient bearing fault diagnosis model from the combination of the feature extraction, feature selection, and feature classification techniques.

II. FEATURE EXTRACTION

The main task of this section is to extract potential features from the current signal of the induction motor in the time and frequency domain. For this purpose, the current signals are first acquired from the test motors. The feature extraction process has a combination of three signal processing techniques, where are DWT, EA, and FFT. Then, statistical

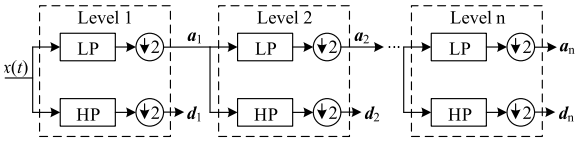

FIGURE 1. Flowchart of n-level wavelet decomposition of $x(t)$.

features are extracted as fault signatures from the time and frequency domains of each decomposed signal.

A. DISCRETE WAVELET TRANSFORM

Wavelet transform is a powerful tool for analysing transient or non-stationary signals. It has been applied in many fields [23]–[25]. In WT, the signal $x(t)$ is decomposed into components with different frequency bands which are generated by scaling and translating a mother wavelet [26].

The continuous wavelet transform (CWT) of $x(t)$ is expressed as

$$CWT(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t)\, \psi \frac{t-b}{a} dt \tag{1}$$

where $\psi(t)$ denotes the mother wavelet function, a is the scale factor and b is the translation or shifting factor.

The DWT is the discretization version of CWT, defined by

$$DWT(j,k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{+\infty} x(t)\, \psi \left(\frac{t-2^j k}{2^j} \right) \tag{2}$$

where a, b are replaced by 2^j and $2^j k$

DWT is often used to improve computational efficiency [26]. Fig. 1 shows a decomposition tree structure of DWT for n levels. The original signal $x(t)$ is passed through the low pass (LP) and high pass (HP) filter resulting in two vectors called approximate $\boldsymbol{a}$ and detailed $\boldsymbol{d}$ coefficients, respectively. The symbol $\downarrow 2$ indicating down-sampling means that it ignores the odd-indexed elements of the filtered signal [27].

B. ENVELOPE ANALYSIS AND FFT

Envelope analysis is a simple method performed as follows: First, determine all the local extrema of the input data. Then, connect all the local maxima by a cubic spline line as shown in the upper envelope. Repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelopes should cover all the data between them [4], [28].

FFT is a very efficient algorithm to calculate the discrete Fourier transform (DFT). FFT converts the signal $x(t)$ with t samples from the time domain to frequency domain on a finite number of frequency lines k, where the distance between each line is the frequency resolution [29]. The expression of FFT is shown in (3)

$$Y(k) = \sum_{n=0}^{N-1} x(t) \mathrm{e}^{-j2\pi kt/N} \tag{3}$$

C. FEATURE EXTRACTION PROCESS

When a failure occurs in the rotating machinery, the current signal is affected by a bearing fault. This leads to changes in amplitude and distribution in the time domain and frequency of the signal. In this work, the feature extraction process is performed as follows:

Step 1: DWT is first used for signal decomposition. In this article, the wavelet decomposition of the signal analyzed at level 4 is adopted. Fig. 2a illustrates the wavelet decomposition of the healthy signal at level 4. The output decomposition structure consists of the wavelet decomposition vectors: the detail coefficients $\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3$, $\boldsymbol{d}_4$, and approximation coefficient $\boldsymbol{a}_4$. Four detail coefficients are treated as input of EA and FFT techniques.

Step 2: EA is applied for each detail coefficient. The seven statistical feature parameters in time domain are calculated from the upper envelope and do the same for the lower envelope. A total of 56 features (4*2*7) are extracted from the time domain using EA technique. Fig. 2b illustrates the EA technique applied to $\boldsymbol{d}_3$ and 14 features obtained from the upper envelope and the lower envelope.

Step 3: Similar to *Step 2*, FFT is also applied to each detailed coefficient. The five statistical feature parameters in frequency domain are calculated from the frequency spectrum of the detail coefficient. A total of 20 features (4*5) are extracted from the frequency domain by the FFT technique. Fig. 2c illustrates five features obtained from the frequency spectrum of $\boldsymbol{d}_3$.

Thus, the feature dataset of each signature contains 76 features. Table 1 and Table 2 presents the mathematical expressions of these statistical feature parameters in time domain and frequency domain [30], [31]. The order and denoted of the features are arranged in Fig. 3.

TABLE 1. Definition for seven statistic parameters in time domain.

Feature	Equation
(1) Mean value	$x_{\text{mean}} = \frac{\sum_{n=1}^{N} x(n)}{N}$
(2) Standard deviation	$x_{\text{std}} = \sqrt{\frac{\sum_{n=1}^{N} (x(n)\text{-}x_{\text{mean}})^2}{N\text{-}1}}$
(3) Root mean square	$x_{\text{rms}} = \left(\frac{\sum_{n=1}^{N} \sqrt{\lvert x(n) \rvert}}{N} \right)^2$
(4) Max value	$x_{\text{max}} = \max(x(n))$
(5) Min value	$x_{\text{min}} = \min(x(n))$
(6) Kurtosis	$x_{\text{kur}} = \frac{\sum_{n=1}^{N} (x(n)\text{-}x_{\text{mean}})^4}{(N\text{-}1)x_{\text{std}}^4}$
(7) Skewness	$x_{\text{ske}} = \frac{\sum_{n=1}^{N} (x(n)\text{-}x_{\text{mean}})^3}{(N\text{-}1)x_{\text{std}}^3}$

Note: $x(n)$ is a signal series for $n = 1, 2, \ldots, N$, where N is the number of data points.

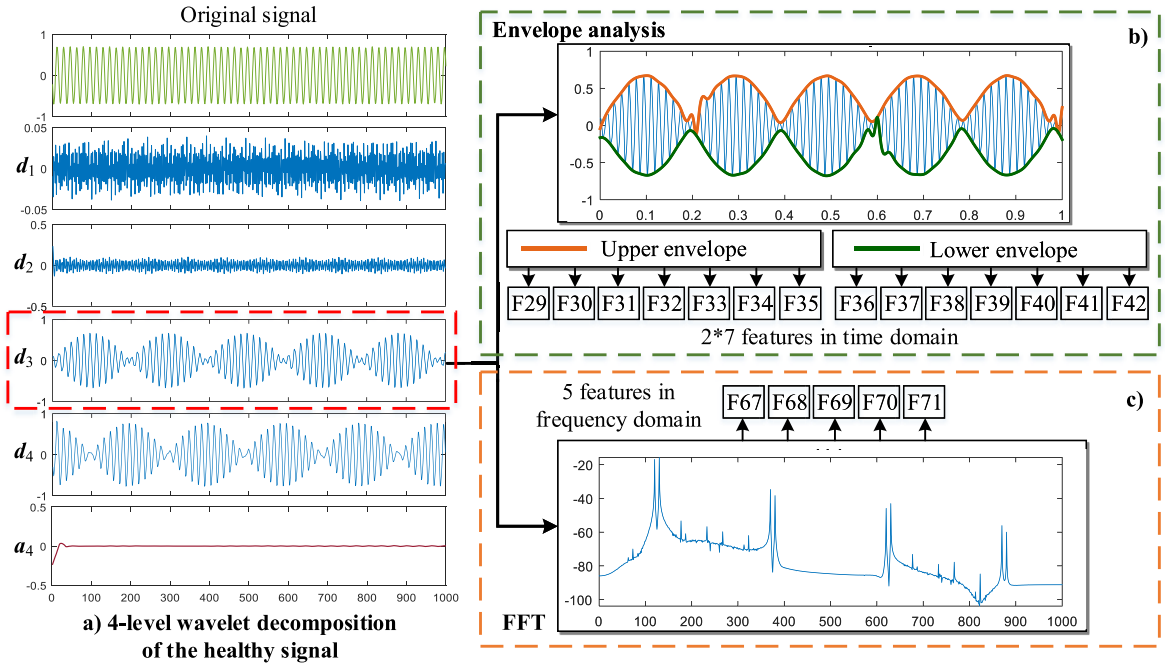

FIGURE 2. The illustration of the feature extraction process.

TABLE 2. Definition for five statistic parameters in frequency domain.

Feature	Equation
(1) Mean frequency	$x_{\mathrm{mf}} = \frac{\sum_{k=1}^{K} y(k)}{K}$
(2) Frequency centre	$x_{\mathrm{fc}} = \frac{\sum_{k=1}^{K} f_k y(k)}{\sum_{k=1}^{K} y(k)}$
(3) Root variance	$x_{\mathrm{rv}} = \sqrt{\frac{\sum_{k=1}^{K} (f_k - x_{\mathrm{fc}})^2 y(k)}{K}}$
(4) Root mean square frequency	$x_{\mathrm{rmsf}} = \sqrt{\frac{\sum_{k=1}^{K} f_k^2 y(k)}{\sum_{k=1}^{K} y(k)}}$
(5) Root mean square ratio	$x_{\mathrm{rmsr}} = \frac{\sum_{k=1}^{K} (f_k - x_{\mathrm{fc}})^{1/2} y(k)}{K\sqrt{x_{\mathrm{rv}}}}$

Note: $y(k)$ is spectrum of the signal for $k = 1, 2, \ldots, K$, where K is the number of the spectrum lines, and f_k is the frequency value of the kth spectrum line.

III. BINARY PARTICLE SWARM OPTIMIZATION WITH EXTENDED MEMORY (BPSO-EM) FOR FEATURE SELECTION

This section starts by introducing the basic BPSO algorithm for feature selection. It then explains two mechanisms of position updating: 1) Updating the particle position based on feature weights; 2) Updating the particle position based on crossover operation. The operation of these two mechanisms combines with extended memory to create a potential solution. Finally, it presents the overall algorithm of BPSO-EM.

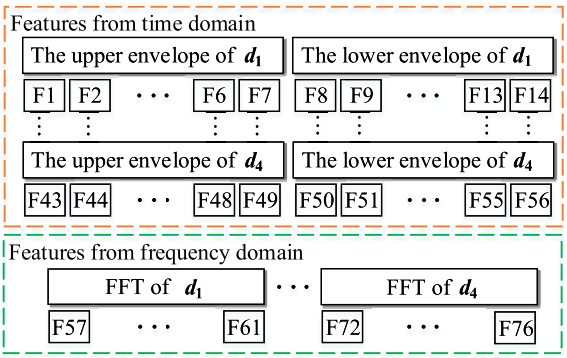

FIGURE 3. The order and denoted of features in the feature.

A. BINARY PARTICLE SWARM OPTIMIZATION

The standard BPSO was introduced by Kennedy and Eberhart in 1997 [12]. In BPSO, each particle is considered a solution in binary search space (i.e., each feature subset) is represented as a bit string, in which each position of a particle with value '1' means the corresponding feature is selected and '0' means unselected. Unlike continuous PSO, the position of a particle in BPSO is updated by flipping each position bit from the value 0 to 1 or vice versa by converting its velocity to the probability value $s(v_{\mathrm{id}})$ performed by the transfer function. The most widely used transfer function is the sigmoid function in (4). The position bit value x_{id} is updated as (5).

$$s(v_{\mathrm{id}}) = \frac{1}{1+e^{-v_{\mathrm{id}}}} \tag{4}$$

$$x_{\mathrm{id}} = \begin{cases} 1, & \text{if rand} < s(v_{\mathrm{id}}) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where *rand* is a random number uniformly distributed within the interval of [0, 1] and the dth bit of the ith particle.

For each particle (solution), the velocity v_{id} is updated as (6) [32]:

$$v_{\mathrm{id}} = wv_{\mathrm{id}} + c_1 r_1(p_{\mathrm{id}} - x_{\mathrm{id}}) + c_2 r_2(g_{\mathrm{d}} - x_{\mathrm{id}}) \tag{6}$$

where w is the inertia weight, p_{id} is dth bit of the personal best position of the ith particle, g_{d} denotes the dth bit of the global best position (i.e., the best position of the personal best in the swarm), c_1 and c_2 are the acceleration coefficients, r_1 and r_2 are random variables in the range [0, 1].

According to [19], the inertia weight value plays an important role in balancing exploration and exploration in search space. The linear time-varying inertia weight is applied in this study as shown in (7), its value decreases from 0.9 to 0.4 according to the number of iteration [33].

$$w = w_{\max} - (w_{\max} - w_{\min})(\frac{t}{T_{\max}}) \tag{7}$$

where $w_{\min}$ and $w_{\max}$ are the minimum and maximum values of inertia weight which set at 0.4 and 0.9, respectively, t is the number of iterations and $T_{\max}$ is the maximum number of iterations.

Each particle is updated its position to find the best solution of personal is called personal best (pbest) and the best solution of the swarm is called global best (gbest). In this article, the position of particles (solutions) assessed by the fitness value is the maximum classification accuracy of k-NN classifier.

In addition, a feature selection method using evolutionary algorithms is a wrapper method that uses a learning algorithm performance accuracy as a feature evaluation criterion [7]. Besides, the number of selected features is also considered to compare the effectiveness of the methods. Therefore, the update procedure pbest and gbest select classification performance as first priority in this study as proposed in [34]. First, the pbest is updated if the classification accuracy of a particle's new position is better than its pbest value. In this case, the number of selected features will not be considered as in the traditional update procedure. Second, if the fitness value of a particle's new position is equal to pbest and the number of selected features selected is smaller then the pbest will be updated by the particle's new position. This guarantees a better solution with the same fitness value but fewer numbers of selected features. The update procedure gbest is similar to pbest by comparing pbest value of particles. The update procedure is expressed as pseudocode 1 where $\mathbf{x}_i = (x_{i1}, \ldots, x_{in})$ (n is the dimension) and $\mathbf{p}_i = (p_{i1}, \ldots, p_{in})$ are the current position and personal best position of the ith particle, respectively. $\mathbf{g} = (g_1, \ldots, g_n)$ is the global best position.

B. UPDATING THE PARTICLE'S POSITION BASED ON FEATURE WEIGHTS

While the particles in the swarm have not yet found the most optimal solution, the space near the best current position of the swarm is considered a potential region [35]. Therefore, exploring solutions around this region is worth considering. This article proposed the particle's position updating mechanism based on feature weights. Each weight of a feature is calculated by ReliefF algorithm which is one of the widely used and effective methods to evaluate features [36]. The condition for implementing this mechanism is the Hamming distance S between the particle's current position and the best solution of the swarm is calculated. If half this distance does not exceed a difference threshold θ, this update mechanism is executed. The θ threshold is often set to $D/4$ (with D being the feature set length) [37]. The Hamming distance between two binary vectors of the same length is the number of positions at which the corresponding bit values are different. Calculating the Hamming distance between two binary vectors includes two steps: 1) Using binary XOR operator for two binary vectors and 2) Counting the number of 1s in the resulting vector [38].

Thus, if condition $S/2 < \theta$ (where $\theta = D/4$) is satisfied, it means that the current position of the particle is close to gbest, then different bit positions of the particle will be considered for adjustment based on their weights, so-called fine-tuning process. In addition, roulette wheel selection is applied as an effective tool to select a feature proportional to its probability. Thus, if a feature has a higher weight, its probability of getting selected is higher. However, in roulette wheel selection there may be cases where a higher feature weight may not be selected. This is to ensure the diversity of the population.

Details of the update mechanism are described as follows:

Step 1: The weighted vector of the feature set $\mathbf{w}$ is calculated by ReliefF algorithm.

Step 2: Find the difference between gbest and particle's current position $\mathbf{x}_i$ using the binary XOR operator.

Step 3: A sub-vector $\mathbf{u}$ extracted from the original feature weight vector $\mathbf{w}$ corresponding to the different positions between gbest and $\mathbf{x}_i$. Note that this mechanism only considers the position bits of $\mathbf{x}_i$ different from gbest. The rest of $\mathbf{x}_i$ remains the same.

Step 4: Calculate the probability of selection to the feature weight sub-vector. If u_{i} is the weight of ith feature in the feature weight sub-vector $\mathbf{u}$, then its probability $\mathbf{q}$ of being selected is:

$$q_{\mathrm{i}} = \frac{u_{\mathrm{i}}}{\sum_{i=1}^{m} u_{\mathrm{i}}} \tag{8}$$

where m is the number of elements in the sub-vector.

Step 5: Apply the roulette wheel selection process for selecting potentially useful features based on their weights. The features are selected as the wheel is spun many times and each time selecting a value of the weighted probability vector. The number of spins is the number

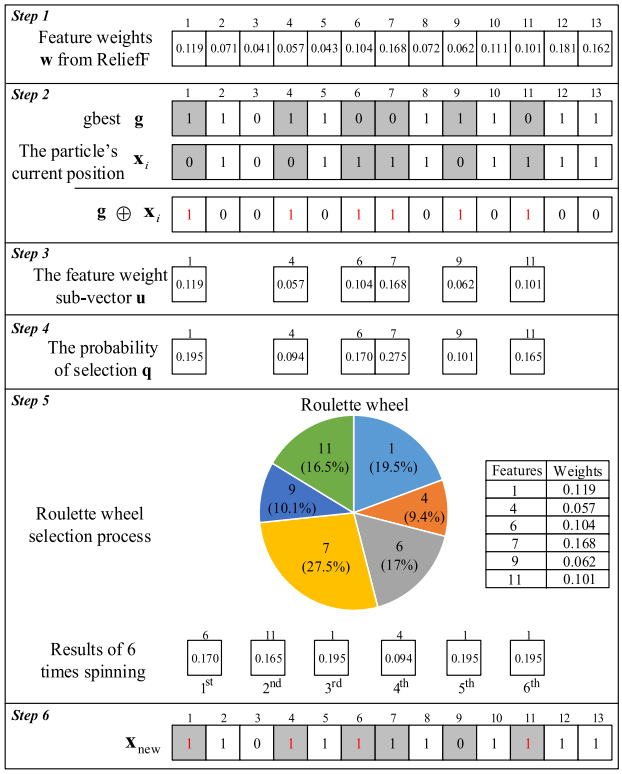

FIGURE 4. The example of the particl's position updating mechanism based on feature weights.

of different bits between two particles gbest and $\mathbf{x}_i$ to the difference between their positions.

Step 6: The positions selected from the roulette wheel selection process are re-mapped to the corresponding positions of $\mathbf{x}_i$ creating a particle's new position $\mathbf{x}_{\text{new}}$.

An example of this update mechanism is illustrated in the Fig. 4. Assume that a feature weight vector of a feature set $\mathbf{w}$ calculated using a ReliefF method consists of 13 elements. gbest and the particle's current position $\mathbf{x}_i$ are a 13-bits binary bit string the same to the length of the feature set. In Step 2, values '1' of the result vector gbest $\oplus$ $\mathbf{x}_i$ indicates that there are differences between g_{best} and X. In Step 3, a feature weight sub-vector feature $\mathbf{u}$ was extracted from $\mathbf{w}$ corresponding to 6 different positions (1, 4, 6, 7, 9, 11). Then, the probability of each element in $\mathbf{u}$ is then calculated by (8). Step 5 is to make a roulette wheel and spin 6 times. As a result, features in positions 6, 11, 1, 4, 1 and 1 are selected for 6 times spinning where 3rd, 5th, 6th select the same feature. Finally, $\mathbf{x}_{\text{new}}$ is restructured based on the selected feature indexes (1, 4, 6, 11). Other positions are retained their original values.

C. UPDATING THE PARTICLE'S POSITION BASED ON CROSSOVER OPERATION

Inspired by genetic algorithms: The good parents were much more likely than ordinary parents to produce offspring of ability. Indeed, the combination of two good solutions highly possible to create more effective solutions. Crossover operation is a powerful tool for doing that. This operation is capable

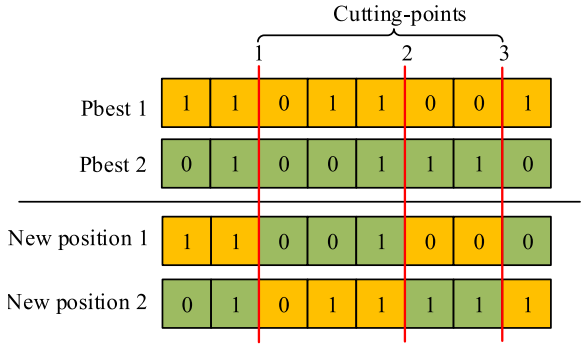

FIGURE 5. The illustration of the three-points crossover.

of exploring new areas in the search space, enhancing global search capabilities, ensuring population diversity, and preventing premature convergence. The three-points crossover operation is applied in this article, wherein a pair of personal best solutions selected randomly from the set of personal best positions of the particles by the roulette wheel selection technique based on their fitness values. Three-point crossover proceeds by cutting the pair of personal best solutions at three random points and swapping the segments to create two new solutions (positions). Then one of two is chosen to replace the particle's current position. Fig. 5 illustrates the three-points crossover.

D. THE PROPOSED BPSO-EM METHOD

The proposed extended memory used in BPSO-EM, which has a role to store all positions (solutions) that have been evaluated by the fitness function. Position update mechanisms are executed which means the particle moved to a new position. The new position of the particle will be compared with each position stored in the extended memory. If the new position of the particle matches the solutions in memory, position update mechanisms will work to update the new position that reaches more promising areas in the search space. New solutions are always checked with extended memory to ensure new positions are not the same as those previously evaluated. This process avoids adding a noticeable amount of computation because evaluating solutions is a time-consuming operation to calculate [39]. On the other hand, combining BPSO with EM helps solve the local trap problem and further improves its performance. Therefore, BPSO-EM is a new approach for feature selection proposed in this article. However, the size of the extended memory EM will depend on the size of the data set and the maximum number of iterations of the algorithm. That may increase the model's computation cost. The procedure of the proposed method to identify the best solution (the optimal feature subset) is shown in Fig. 6. Pseudocode 2 shows the process of checking and updating positions according to the two mechanisms described in the previous subsection.

A ρ threshold is set to optimize the algorithm, preventing it from falling into an infinite loop. Especially low-dimensional datasets, because if a feature dataset has only a few features,

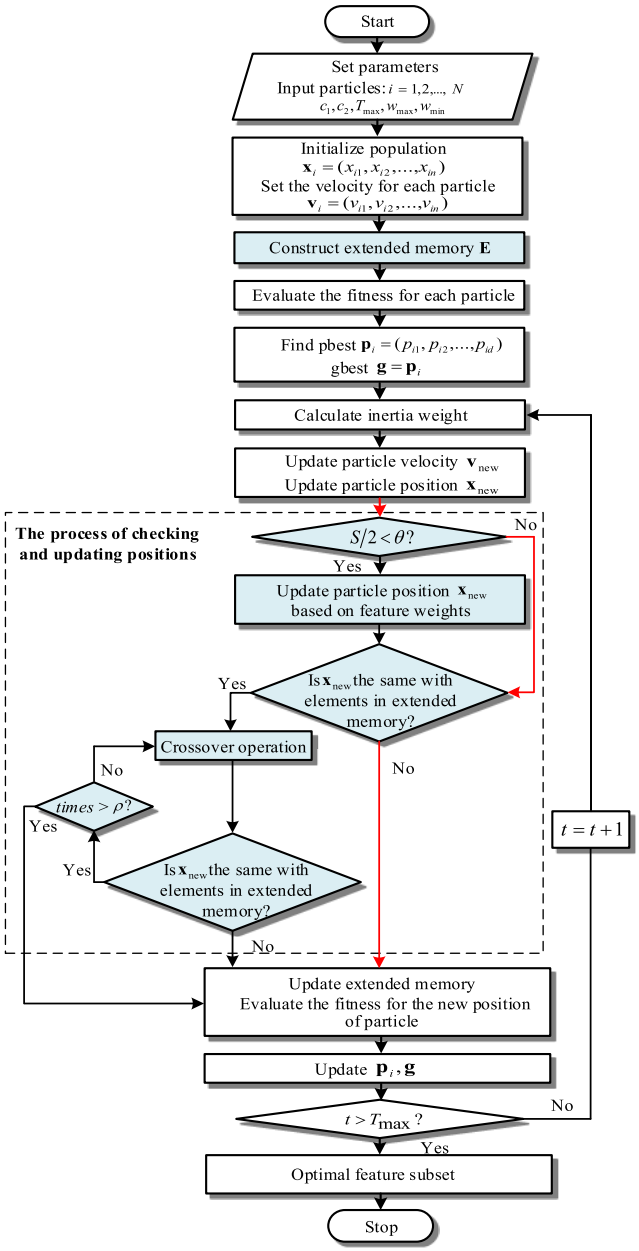

FIGURE 6. The flowchart of the proposed BPSO-EM method.

the total number of solutions is very small. Therefore, it will take a long time to check and find the satisfying position provided that it is not the same as the positions in the extended memory. The selection of ρ threshold is user-defined. Note that this threshold value will affect the computation time of the algorithm. In this study, $\rho = 100$ is used for all data sets used to evaluate the effectiveness of this algorithm.

In BPSO-EM, the traditional updating mechanism of BPSO is maintained if the new position has a Hamming distance with gbest that does not satisfy condition $S/2 < \theta$ and is not the same as in-memory solutions (see subsection III.B). The direction of the red arrow in Fig. 6 shows the process of checking and updating positions is ignored.

Pseudocode 1 The Update Procedure pbest and gbest With Classification Performance as First Priority

```
if f(x_i) > f(p_i)
    p_i = x_i; //update pbest
    else if f(x_i) ==f(p_i)and |x_i| < |p_i|
            p_i = x_i; //update pbest
    end
end
if f(p_i) >f(g)
    g =p_i; //update gbest
     else if f(p_i) ==f(g)and |p_i| < |g|
            g =p_i; //update gbest
     end
end
```

A simple classifier, k-NN, is used with the number of nearest neighbor $k = 1$ and 10-folds cross-validation to evaluate each particle. Euclidean distance is used to compute k closest neighbors. The fitness function is based on the classification accuracy of k-NN classifier, which is defined as follows [40]:

$$Fitness = \frac{N_{\text{True}}}{N_{\text{True}} + N_{\text{False}}} \times 100\% \tag{9}$$

where N_{True} is the number of true predicted instances, N_{False} is the number of false predicted instances.

IV. INTELLIGENCE MODEL FOR BEARING FAULT DIAGNOSIS

This section proposes an intelligence model for bearing fault diagnosis based on the current signals of the induction motor. It is proposed based on a combination of three processing signal techniques DWT, EA, and FFT for feature extraction, binary particle swarm optimization with extended memory for feature selection, and a suitable classier. This model is depicted in Fig. 7 and consists of three important stages as follows:

Stage 1: The current signals are acquired from test motors, which are decomposed by DWT. The wavelet decomposition of the current signals at level 4 using the order 4 Symlets wavelet [41], [42]. Four detail coefficients from the decomposition are considered as inputs for EA and FFT techniques. Each detail coefficient obtains 14 features in time domain from the upper envelope and the lower envelope and 5 features in frequency domain from its spectrum. A total of 76 features were extracted during this approach.

Stage 2: The process of optimizing the feature set obtained from *Stage 1* is done through an efficient stochastic optimization algorithm based on a combination of BPSO and extended memory. As a result, redundant and irrelevant features are eliminated. The optimal feature subset will improve the accuracy of the diagnostic model.

Stage 3: The optimal feature subset is the input data, which is provided separately for three well-known classifiers: NB, DT, and LDA. The accuracy classifications are used

Pseudocode 2 The Process of Checking and Updating Positions

input: Current position of particle **x**
Global best position **g**
Extended memory **E**
output: New position of particle $\mathbf{x}_{new}$
$S \leftarrow$ calculate Hamming distance between **x** and **g**;
if $S/2 < \theta$ *//where* $\theta = D/4$ (*D being the feature set length).*
  x $\leftarrow$ position update mechanism based on feature weights;
end
$L \leftarrow$ the size of extended memory **E**;
while $i < L$
  $i = i + 1$;
  if x is the same with $\mathbf{e}_i$ of extended memory *//compare* ***x*** *to each value in the extended memory.*
    x $\leftarrow$ position update mechanism based on crossover operation;
    $times = times + 1$;
    $i = 0$; *//reset i to check the changed position with the first value in extended memory.*
  end
  if $times > \rho$
    $times = 0$; *break*; *//to avoid an infinite loop.*
  end
end
E $\leftarrow$ **x**; *//update* ***x*** *to extended memory.*
$\mathbf{x}_{new} \leftarrow \mathbf{x}$; *//update* $\mathbf{x}_{new}$.

TABLE 3. Description of 6 benchmark datasets.

Datasets	Features	Instances	Classes
Wine	13	178	3
Vehicle	18	94	4
Segmentation	19	2310	7
BreastEW	30	569	2
Ionosphere	34	351	2
Sonar	60	208	2

to evaluate the efficiency and compatibility of classifiers with bearings data to select an optimal model for detecting bearings faults.

V. EXPERIMENT RESULTS

In this section, the experiments are divided into two case studies.

Case study 1: To examine the effectiveness of the proposed feature selection algorithm and evaluate its applicability not only in the bearing fault diagnosis model but also in the pattern recognition field. It presents the results of evaluating the BPSO-EM algorithm for feature selection based on benchmark datasets. Its results are compared with well-known evolutionary computation paradigms for feature selection and also compared with state-of-art feature selection algorithms.

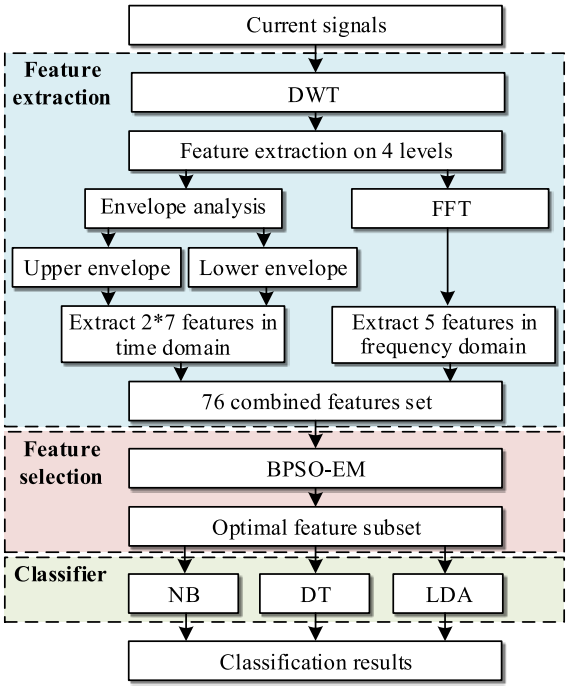

FIGURE 7. Intelligence model for bearing fault diagnosis.

Case study 2: The effectiveness of the proposed bearing fault diagnosis model is evaluated and analyzed based on the dataset acquired from test motors including healthy motor and bearing failure motors. This real dataset has named the bearings dataset.

A. CASE STUDY 1: BENCHMARK DATASETS

1) DATA DESCRIPTION

Usually, the datasets used in fault motor diagnostic models are low-dimensional data sets [40], [43]. Therefore, six benchmark datasets with small sizes were selected from the UCI Machine Learning Repository [44] to perform in this study including Wine Sonar, Vehicle, Segmentation Ionosphere, BreastEW, Ionosphere, and Sonar. The description of the 6 benchmark datasets is shown in Table 3.

2) PARAMETER SETTING AND EVALUATION CRITERION

In this subsection, parameters are assigned values for all our experiments. All parameters used in the proposed BPSO-EM algorithm are presented as follow: number of particles: 10, maximum number of iterations: 100, $c_1 = c_2 = 2.05$, $w_{min} = 0.4$, $w_{max} = 0.9$, $\rho = 100$ (see subsection III.D). In addition, the parameters of four evolutionary computation algorithms used for comparison including binary particle swarm optimization (BPSO), genetic algorithm (GA), binary grey wolf optimizer (BGWO), binary differential evolution (BDE), which are presented in Table 4.

Currently, feature selection methods are categorized into two groups: 1) Wrappers; 2) Filter approach [11]. The wrapper approaches use a classifier to evaluate the selected feature. In contrast, a filter feature selection approach is independent of any classification algorithm. In this study, BPSO-EM is

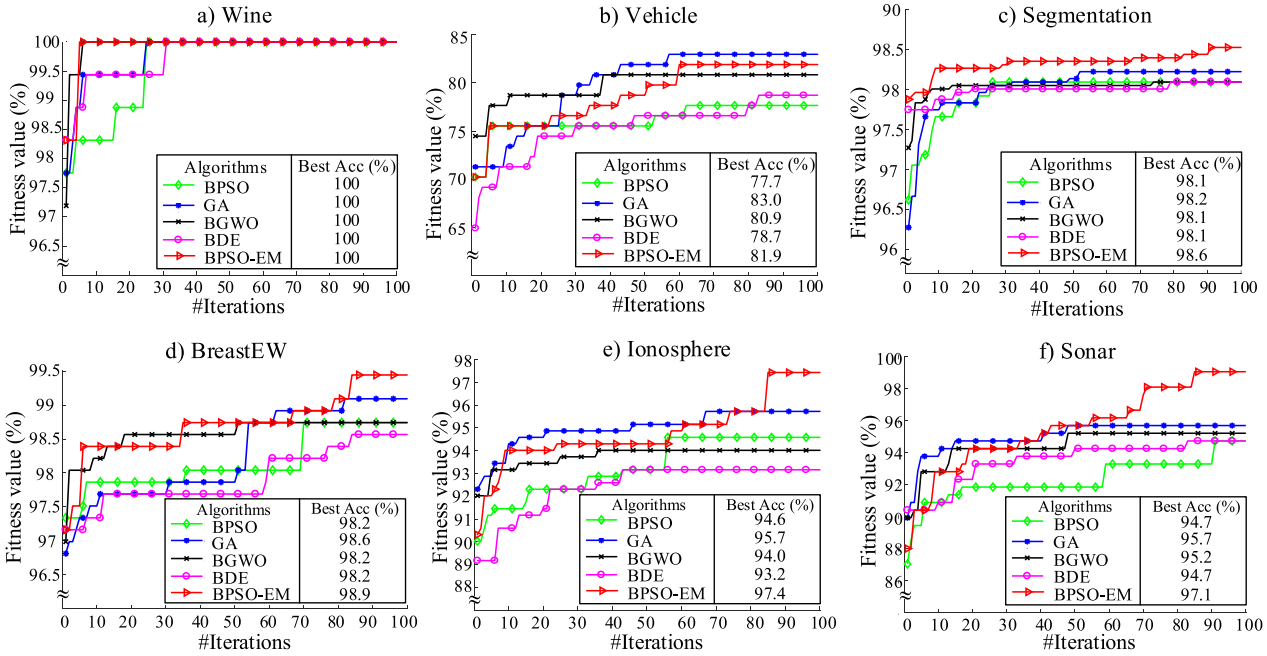

FIGURE 8. Convergence curves for BPSO-EM and well-known evolutionary computation paradigms.

TABLE 4. Parameter setting of four evolutionary computation algorithms.

BPSO	GA	BGWO	BDE
Number of particles:10 Maximum number of iterations: 100 $c_1 = c_2 = 2.05$ $w_{min} = 0.4$ $w_{max} = 0.9$	Number of chromosomes: 10 Maximum number of iterations: 100 Crossover rate: 0.8 Mutation rate: 0.01	Number of wolves: 10 Maximum number of iterations: 100	Number of vectors:10 Maximum number of iterations: 100 Crossover rate: 0.9

TABLE 5. Comparison between BPSO-EM and the well-known evolutionary computation paradigms.

Datasets	BPSO		GA		BGWO		BDE		BPSO-EM	
	Avg Acc	Avg No.F	Avg Acc	Avg No.F	Avg Acc	Avg No.F	Avg Acc	Avg No.F	Avg Acc	Avg No.F
Wine	99.4	**6.8**	99.4	8.7	99.1	9.4	99.4	9.4	**99.8**	8.4
Vehicle	75.0	8.7	76.8	8.8	74.6	12.4	74.7	12.8	**79.5**	**8.0**
Segmentation	98.0	**8.5**	98.1	10.4	97.6	13.7	97.7	13.8	**98.4**	9.3
BreastEW	97.6	15.9	98.0	**15.1**	97.5	21.9	97.6	22.2	**98.3**	16.2
Ionosphere	93.3	**12.7**	94.2	14.5	92.6	20.8	92.7	21.2	**95.5**	13.9
Sonar	93.1	**29.6**	**95.4**	30.4	93.5	41.9	93.4	45.6	94.9	30.4

Avg Acc: The average classification accuracy (%).
Avg No.F: the average number of selected features.

considered as a wrapper feature selection method, which uses k-NN classifier in the feature subset evaluation step via the best classification performance. There are usually two criteria used to evaluate the effectiveness of a feature selection method, such as 1) Maximize classification accuracy; 2) Minimize the number of selected features [11]. In this study, the classification accuracy criterion is our first priority. This criterion also conforms to the requirements of the high-accuracy fault diagnosis model. Equation (9) presents the fitness value of BPSO-EM, which is also the classification accuracy criterion.

3) COMPARISON WITH THE WELL-KNOWN EVOLUTIONARY COMPUTATION PARADIGMS

The experimental datasets are simulated by a personal computer (PC) with Intel Core i3-7100 3.9Ghz (4CPUs), 16GB RAM, and Matlab 2017a. In this subsection, four the algorithms for feature selection are adopted to compare and evaluate the effectiveness of the proposed BPSO-EM algorithm. They represent categories of ECs including BPSO, BGWO as popular algorithms in swarm intelligence, GA as a typical algorithm in evolutionary algorithms, and another algorithm as BDE.

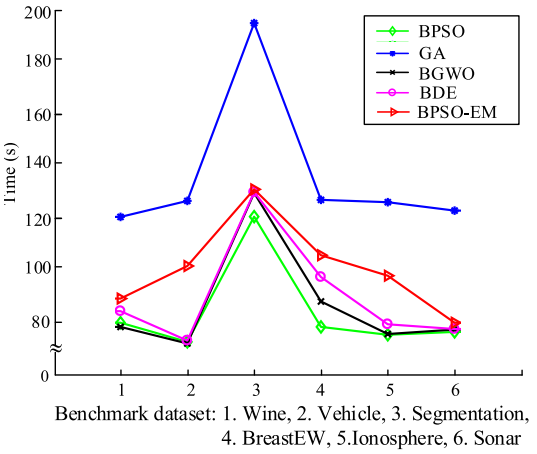

FIGURE 9. The average computational time of five algorithms for six datasets.

The results of algorithms are quantitatively compared using the following metrics:

- Convergence curves of the best-so-far solutions.
- The average classification accuracy and the average number of selected features are calculated for the 30 independence runs.
- The average computational time is calculated for the 30 independence runs, where the maximum number of iterations is 100 for 1 run.

Observe the convergence curves of the best solutions so far in Fig. 8, the BPSO-EM has a higher classification accuracy than 4 other algorithms on 4 of 6 benchmark datasets including Segmentation, BreastEW, Ionosphere, and Sonar. In particular, the proposed algorithm is much more efficient than BDE on Ionosphere and Sonar with the differences of 4.2% and 2.4%, respectively. While GA has reached the highest classification accuracy on Vehicle dataset. All five algorithms achieve 100% classification accuracy on Wine dataset.

The average classification accuracy (Avg Acc) and the average number of selected features (Avg No.F) are presented in Table 5. The results of the better algorithm for each data are made bold. The results show that BPSO-EM achieves better average classification accuracy than the four compared algorithms on 5 of 6 benchmark datasets including Wine, Vehicle, Segmentation, BreastEW, and Ionosphere. Especially on Vehicles and Ionosphere, BPSO-EM is significantly better than BGWO with the differences of 4.9% and 2.6%, respectively. Considering the average number of selected features, BPSO achieved the smallest average number of selected features on 4 of 6 data sets. While BPSO-EM is most effective on Vehicle dataset. The above results show that the two criteria for average classification accuracy and the average number of selected features are often contradictory. However, the ability to select the optimal feature subset of the proposed method is better than the comparison algorithms based on the average classification accuracy.

The average computational time of each algorithm is shown in Fig. 9. It's easy to realize that GA is the most time-consuming algorithm of all the datasets. When BPSO-EM took more time than the other three algorithms (BPSO, BGWO, BDE) on the three datasets as Vehicle, BreastEW, and Ionosphere. On the Wine, Segmentation, and Sonar datasets, the computational times of the algorithms are similar, except for GA.

In general, with significant classification performance, the proposed approach highly possible to compete with other state-of-art algorithms for feature selection. This is satisfied with our priority criteria (classification performance) set in this study.

4) COMPARISON WITH THE STATE-OF-ART FEATURE SELECTION ALGORITHMS

In order to an objective assessment of the ability of the proposed approach in the feature selection problem, a comparison between BPSO-EM and the results cited from the state-of-art feature selection algorithms which represent recent research directions on this problem has been carried out. Brief description of state-of-art feature subset selection algorithms is as follows: 1) A hybrid feature selection method is proposed seagull optimization algorithm (SOA) and thermal exchange optimization (TEO) is named SOA-TEO3 [45]; 2) An approach to improve the PSO algorithm on the update mechanism and in conjunction with the spiral mechanism to enhance local search around the optimal solution called HPSO-SSM [46]; 3) BGOA-M proposed in [47] is based on a combination of grasshopper optimization algorithm and mutation operator to enhance the exploration ability of the algorithm; 4) A binary variant of the butterfly optimization algorithm (BOA) proposed in [48] to improve the exploration ability in the feature search space, which is called S-bBOA; 5) A hybrid algorithm based on a combination of the whale optimization algorithm (WOA) and the simulated annealing algorithm (SA) called WOASAT-2, which is proposed to enhance exploitation ability by searching the optimal solution in a promising area [49]. All five optimization algorithms are categorized as swarm intelligent algorithms. The two indicators are used in this comparison including the average classification accuracy and the average number of the selected features. Table 6 presents the comparison result of the proposed approach with SOA-TEO3, HPSO-SSM, BGOA-M, S-bBOA, and WOASAT-2. The results of the better algorithm for each data are made bold. Compared with other feature selection algorithms, it can be easily deduced from the results in Table 6 that the proposed approach achieved better classification accuracy than the other comparison algorithms on the Wine, Segmentation, and BreastEW datasets. On the Sonar and Ionosphere datasets, the proposed approach reached better classification accuracy than the other algorithms but slightly worse than WOASAT-2. However, the proposed approach did not really excel in the lowest number of selected feature criteria. In summary, based on the classification performance of the proposed algorithm which is worth considering for the feature selection problem.

TABLE 6. Comparison between BPSO-EM and the state-of-art feature selection algorithms.

Datasets	Algorithms	Avg Acc (%)	Avg No.F
Wine	SOA-TEO3[45]	90.94	**3.81**
	HPSO-SSM [46]	99.38	4.43
	BGOA-M [47]	98.88	4.40
	S-bBOA [48]	98.43	6.20
	WOASAT-2 [49]	99.00	6.40
	BPSO-EM	**99.80**	8.40
Vehicle	SOA-TEO3[45]	**91.71**	**6.07**
	BGOA-M [47]	77.04	9.60
	BPSO-EM	79.50	8.00
Segmentation	HPSO-SSM [46]	97.56	**8.07**
	BPSO-EM	**98.40**	9.30
BreastEW	SOA-TEO3[45]	94.55	9.11
	HPSO-SSM [46]	94.89	**6.76**
	BGOA-M [47]	96.97	12.50
	S-bBOA [48]	97.09	16.80
	WOASAT-2 [49]	98.00	11.60
	BPSO-EM	**98.30**	16.20
Ionosphere	SOA-TEO3[45]	90.53	7.58
	HPSO-SSM [46]	92.57	**7.10**
	BGOA-M [47]	94.58	11.46
	S-bBOA [48]	90.70	16.20
	WOASAT-2 [49]	**96.00**	12.80
	BPSO-EM	95.50	13.90
Sonar	SOA-TEO3[45]	93.97	27.17
	BGOA-M [47]	91.47	26.80
	S-bBOA [48]	93.62	32.80
	WOASAT-2 [49]	**97.00**	**26.40**
	BPSO-EM	94.90	30.40

Avg Acc: The average classification accuracy (%).
Avg No.F: the average number of selected features.

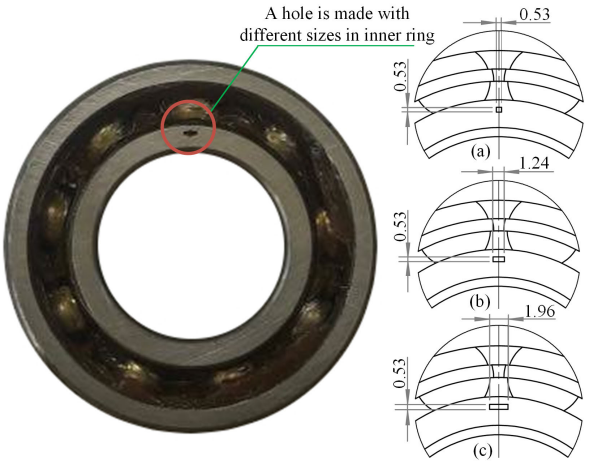

FIGURE 10. Bearing artificial fault with different dimensions. a) square hole with a height × width (0.53mm × 0.53mm) b) square hole with a height × width (0.53mm × 1.24mm) c) square hole with a height × width (0.53mm × 1.96mm).

B. CASE STUDY 2: BEARING DATASET

1) EXPERIMENTAL SETUP AND DATA ACQUISITION

This work focuses on detecting the bearing fault of the induction motor. Bearings of the test motors are artificially made a square hole on the inner ring with different dimensions as shown in Fig. 10. The stator current signals are measured from a healthy motor and three types of bearings

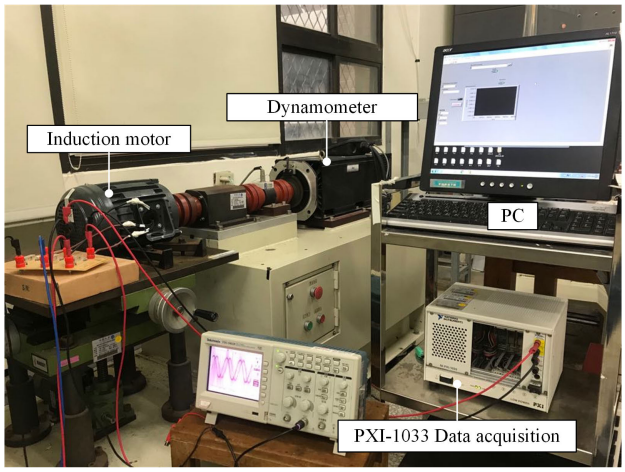

FIGURE 11. The experimental hardware architecture model.

TABLE 7. Results of the feature selection algorithms on bearings dataset.

Algorithms	Avg Acc (%)	Avg No.F
BPSO	96.5	30.3
GA	98.0	31.5
BGWO	94.4	48.8
BDE	94.4	48.9
BPSO-EM	**99.7**	**27.8**

Avg Acc: The average classification accuracy (%).
Avg No.F: the average number of selected features.

fault motors. The experimental hardware architecture model is shown in Fig. 11 including AC motors (3 phase induction motor, 4 poles, 2 HP, 380VAC, 60Hz), data acquisition equipment (NI PXI-1033), dynamometer (69Hz, 11kW, 2000rpm) and PC. Besides, there are also power supply systems and dynamometer control cabinet. There are 200 instances collected corresponding to four cases (a healthy motor and three faulty motors). Each instance contained 2000 data points that are sampled within 2 seconds.

2) TECHNIQUES IMPLEMENTATION

This subsection is presented based on the 3 stages of the intelligent bearing fault diagnosis model proposed in section IV. In stage 1, the 200 instances from test motors are analyzed by DWT, EA, and FFT techniques. There are 76 potential features that are extracted by the feature extraction process as presented in section II. A bearing fault dataset from 4 types of testing motors is established with 76 features, 200 instances, and 4 classes. The size of this dataset is similar to the benchmark datasets used above.

In stage 2, the proposed feature selection algorithm is applied to the bearing dataset. In addition, four other feature selection algorithms (BPSO, GA, BGWO, BDE) are also calculated and compared effectively on this dataset. The results shown in Table 7 include two criteria to evaluate the performance of feature selection algorithms where average classification accuracy as first priority. The proposed approach reached average classification accuracy higher than the other algorithms with 30 independence runs. In particular,

TABLE 8. Details of the best optimal feature subset on the bearings dataset.

Algorithms	Features	Feature indicators (F)
BPSO	28	2, 3, 5, 6, 8, 9, 10, 12, 15, 20, 22, 23, 29, 30, 33, 35, 36, 37, 38, 41, 42, 43, 44, 45, 50, 57, 64, 66.
GA	31	2, 3, 4, 8, 9, 10, 11, 15, 19, 22, 23, 24, 25, 28, 30, 31, 33, 34, 35, 43, 44, 45, 51, 52, 57, 58, 60, 62, 64, 66, 71.
BGWO	51	1, 2, 3, 5, 7, 8, 9, 10, 11, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 36, 37, 38, 40, 41, 42, 43, 44, 50, 51, 52, 53, 54, 57, 60, 61, 63, 64, 66, 67, 74, 75, 76.
BDE	41	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 15, 17, 18, 19, 22, 23, 24, 26, 28, 33, 34, 38, 40, 41, 42, 43, 44, 45, 46, 47, 50, 51, 52, 53, 56, 57, 59, 60, 62, 64, 72.
BPSO-EM	13	1, 2, 3, 8, 10, 15, 17, 23, 44, 52, 56, 69, 71.

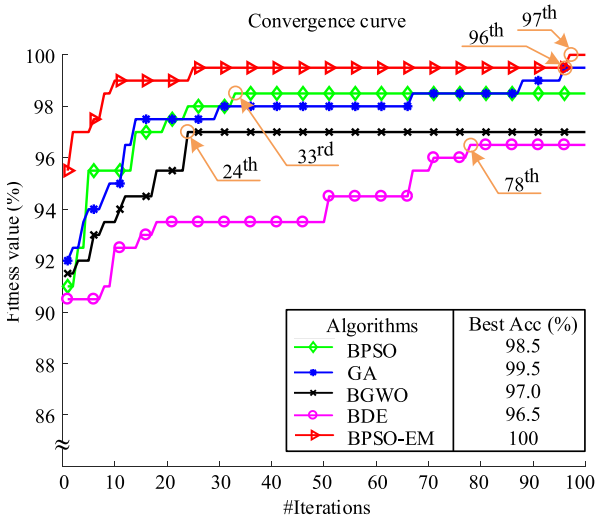

FIGURE 12. The convergence curve of the best solution.

the proposed algorithm is much more efficient than BGWO and BDE with the differences of 5.3%. The convergence curve of the best solution is shown in Fig. 12. The optimal feature subset of the best solution is detailed in Table 8. Based on classification performance criteria, BPSO-EM achieves 100% accuracy. It is obviously better than the other algorithms are compared. Based on the number of selected feature of the optimal feature subset, BPSO-EM has achieved 13 features. In other words, the proposed approach can eliminate 83% of non-essential (redundant and irrelevant) features compared to the original feature set. In addition, Fig. 12 also presents the convergence rate of the algorithms. In which BGWO, BPSO converge at the 24^{th} and 33^{rd} iteration, respectively. This shows that the weaknesses of BGWO and BPSO algorithms as presented in the literature are premature convergence and at risk of falling into local optimal traps. Meanwhile, BPSO-EM converges at the 97^{th} iteration, showing its global exploration capabilities and preventing premature convergence based on two proposed position update mechanisms.

The third stage of the bearing fault diagnosis model is to use the best feature (optimal set of features) to provide three classifiers (NB, DT, LDA). Five optimal feature subsets of the proposed feature selection approach, BPSO, GA, BGWO, BDE and the original feature set (without feature selection) that was classified and evaluated in this experiment. The robustness of the NB, DT and LDA classifiers is also

TABLE 9. Average classification accuracy using NB classifier.

FS algorithms	No.F	Avg Acc %			
		∞dB	40dB	30dB	20dB
Without FS	76	89.9	84.5	80.2	61.1
BPSO	28	92.7	88.5	81.9	58.7
GA	31	94.1	90.3	82.7	59.5
BGWO	51	95.0	90.8	83.8	60.1
BDE	41	97.2	92.1	84.6	63.4
The proposed	13	**99.7**	**98.4**	**92.3**	**66.0**

Avg Acc: The average classification accuracy (%).
No.F: The number of selected features.

TABLE 10. Average classification accuracy using DT classifier.

FS algorithms	No.F	Avg Acc %			
		∞dB	40dB	30dB	20dB
Without FS	76	98.6	97.5	85.5	67.5
BPSO	28	98.6	97.3	86.3	67.1
GA	31	98.9	97.7	85.0	67.4
BGWO	51	99.0	97.7	86.1	69.0
BDE	41	99.1	97.7	86.5	64.8
The proposed	13	**99.4**	**98.1**	**88.9**	**70.2**

Avg Acc: The average classification accuracy (%).
No.F: The number of selected features.

TABLE 11. Average classification accuracy using LDA classifier.

FS algorithms	No.F	Avg Acc %			
		∞dB	40dB	30dB	20dB
Without FS	76	98.6	98.4	91.1	73.7
BPSO	28	98.9	98.5	94.6	78.7
GA	31	98.9	98.6	94.0	77.6
BGWO	51	99.2	98.8	93.6	76.6
BDE	41	99.3	98.0	92.7	77.7
The proposed	13	**100.0**	**98.9**	**94.8**	**78.9**

Avg Acc: The average classification accuracy (%).
No.F: The number of selected features.

considered in the noise condition. Environmental interferences are simulated by adding Gaussian white noise to the input signals at different levels. Noise level is indicated by signal to noise ratio (SNR). That means the smaller the SNR value, the greater the effect of noise. The average classification accuracy is calculated after 100 training times. The classification results using NB, DT and LDA are presented in Tables 9, 10 and 11, respectively. The proposed approach reached higher classification accuracy than other algorithms in all three DT, NB and LDA classifiers. Under normal conditions (SNR $=$ ∞dB), the accuracy performance of the original feature set (Without FS) is the lowest in all three classifiers. This proves that the feature selection algorithms used are necessary. The LDA classifier achieved a better classification performance than NB, DT but the differences are

not too much. As the noise level increases, the performance of three classifiers decreases. However, under higher noise levels, the robustness of the LDA classifier has been remarkably effective compared to NB and DT. The most obvious is in the case of SNR = 20dB, the classification accuracy of all algorithms is more than 73.7% using LDA while the accuracy is lower than 66% and 70.2% using NB and DT, respectively. The same case to occur with SNR = 30dB. Therefore, BPSO-EM combined with LDA is the best option with the bearing fault diagnosis model. From the above analysis, it can be concluded that the combination of feature extraction solution, feature selection and feature classification proposed in this study created a high effective detecting bearing fault model.

VI. CONCLUSION

Among rotary machine problems, bearings failure accounts for the largest proportion [1]. Therefore, a highly effective fault diagnosis model is introduced in this study. The model adopts simple but effective feature extraction techniques including DWT, EA, FFT, and an algorithm selects important features from the BPSO algorithm combined with the extended memory to improve classification efficiency. The experiment results showed that a feature dataset with 76 features was extracted from the original dataset using proposed feature extraction techniques. The optimal feature subset achieved 13 features based on the proposed BPSO-EM feature selection algorithm. The best fault diagnosis model achieved 100% classification efficiency by using the LDA classifier. In addition, the performance of the proposed feature selection algorithm is compared and the results are better than the well-known evolutionary computation paradigms. In general, proven effectiveness allows the proposed model to be applied not only in the diagnosis of bearing faults but also in the pattern recognition field. However, the computational time is still a problem that needs improvement in this study. Besides that, a growing trend applying various deep learning algorithms and automatic feature extraction to bearing fault classification should be considered in the future.

REFERENCES

[1] S. Kumar, D. Mukherjee, P. K. Guchhait, R. Banerjee, A. K. Srivastava, D. N. Vishwakarma, and R. K. Saket, "A comprehensive review of condition based prognostic maintenance (CBPM) for induction motor," *IEEE Access*, vol. 7, pp. 90690–90704, Jul. 2019.

[2] J. Huang, J. Tang, M. Zhang, X. Zhang, and T. Ha, "An improved EMD based on cubic spline interpolation of extremum centers," *J. Vibroeng.*, vol. 17, no. 5, pp. 2393–2409, Aug. 2015.

[3] J. Zarei, M. A. Tajeddini, and H. R. Karimi, "Vibration analysis for bearing fault detection and classification using an intelligent filter," *Mechatronics*, vol. 24, no. 2, pp. 151–157, Mar. 2014.

[4] N. E. Huang, *Hilbert-Huang Transform and Its Applications*, vol. 16. Singapore: World Scientific, 2014.

[5] Y. Gao, F. Villecco, M. Li, and W. Song, "Multi-scale permutation entropy based on improved LMD and HMM for rolling bearing diagnosis," *Entropy*, vol. 19, no. 4, p. 176, Apr. 2017.

[6] M. Van and H.-J. Kang, "Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection," *IET Sci., Meas. Technol.*, vol. 9, no. 6, pp. 671–680, Sep. 2015.

[7] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, "Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 13, no. 5, pp. 971–989, Sep. 2016.

[8] Y.-J. Gong, J.-J. Li, Y. Zhou, Y. Li, H. Shu-Hung Chung, Y.-H. Shi, and J. Zhang, "Genetic learning particle swarm optimization," *IEEE Trans. Cybern.*, vol. 46, no. 10, pp. 2277–2290, Oct. 2016.

[9] J. Too, A. Abdullah, N. Mohd Saad, N. Mohd Ali, and W. Tee, "A new competitive binary grey wolf optimizer to solve the feature selection problem in EMG signals classification," *Computers*, vol. 7, no. 4, p. 58, Nov. 2018.

[10] A. W. Mohamed, H. Z. Sabry, and M. Khorshid, "An alternative differential evolution algorithm for global optimization," *J. Adv. Res.*, vol. 3, no. 2, pp. 149–165, Apr. 2012.

[11] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, Aug. 2016.

[12] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proc. IEEE Int. Conf. Syst., Man, Cybern., Comput. Cybern. Simulation*, Oct. 1997, pp. 4104–4108.

[13] D. Whitley, "A genetic algorithm tutorial," *Statist. Comput.*, vol. 4, no. 2, pp. 65–85, Jun. 1994.

[14] 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.

[15] T. Li, H. Dong, and J. Sun, "Binary differential evolution based on individual entropy for feature subset optimization," *IEEE Access*, vol. 7, pp. 24109–24121, Feb. 2019.

[16] C.-S. Yang, L.-Y. Chuang, J.-C. Li, and C.-H. Yang, "Chaotic maps in binary particle swarm optimization for feature selection," in *Proc. IEEE Conf. SMCIA*, Jun. 2008, pp. 107–112.

[17] W. Song, C. Cattani, and C.-H. Chi, "Multifractional Brownian motion and quantum-behaved particle swarm optimization for short term power load forecasting: An integrated approach," *Energy*, vol. 194, Mar. 2020, Art. no. 116847.

[18] W. Song, X. X. Chen, C. Cattani, and E. Zio, "Multifractional Brownian motion and quantum-behaved partial swarm optimization for bearing degradation forecasting," *Complex.*, vol. 2020, Jan. 2020, Art. no. 8543131.

[19] B. H. Nguyen, B. Xue, P. Andreae, and M. Zhang, "A new binary particle swarm optimization approach: Momentum and dynamic balance between exploration and exploitation," *IEEE Trans. Cybern.*, early access, Oct. 11, 2020, doi: 10.1109/TCYB.2019.2944141.

[20] J. Long, Z. Sun, C. Li, Y. Hong, Y. Bai, and S. Zhang, "A novel sparse echo autoencoder network for data-driven fault diagnosis of delta 3-D printers," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 3, pp. 683–692, Mar. 2020.

[21] J. Long, S. Zhang, and C. Li, "Evolving deep echo state networks for intelligent fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4928–4937, Jul. 2020.

[22] J. Ren, "ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging," *Knowl.-Based Syst.*, vol. 26, pp. 144–153, Feb. 2012.

[23] J. Kim, "Discrete wavelet transform-based feature extraction of experimental voltage signal for Li-ion cell consistency," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1150–1161, Mar. 2016.

[24] J. Seshadrinath, B. Singh, and B. K. Panigrahi, "Investigation of vibration signatures for multiple fault diagnosis in variable frequency drives using complex wavelets," *IEEE Trans. Power Electron.*, vol. 29, no. 2, pp. 936–945, Feb. 2014.

[25] P. Zheng and J. Huang, "Discrete wavelet transform and data expansion reduction in homomorphic encrypted domain," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2455–2468, Jun. 2013.

[26] W. Fan, Q. Zhou, J. Li, and Z. Zhu, "A wavelet-based statistical approach for monitoring and diagnosis of compound faults with application to rolling bearings," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 4, pp. 1563–1572, Oct. 2018.

[27] A. B. Patil, J. A. Gaikwad, and J. V. Kulkarni, "Bearing fault diagnosis using discrete wavelet transform and artificial neural network," in *Proc. 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. (iCATccT)*, 2016, pp. 399–405.

[28] C.-Y. Lee and Y.-H. Hsieh, "Bearing damage detection of BLDC motors based on current envelope analysis," *Meas. Sci. Rev.*, vol. 12, no. 6, pp. 290–295, Jan. 2012.

[29] 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.
[30] J. Liu, Z. Xu, L. Zhou, W. Yu, and Y. Shao, "A statistical feature investigation of the spalling propagation assessment for a ball bearing," *Mechanism Mach. Theory*, vol. 131, pp. 336–350, Jan. 2019.
[31] H. Helmi and A. Forouzantabar, "Rolling bearing fault detection of electric motor using time domain and frequency domain features extraction and ANFIS," *IET Electr. Power Appl.*, vol. 13, no. 5, pp. 662–669, May 2019.
[32] J. Liu, Y. Mei, and X. Li, "An analysis of the inertia weight parameter for binary particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 20, no. 5, pp. 666–681, Oct. 2016.
[33] J. Too, A. Abdullah, N. Mohd Saad, and W. Tee, "EMG feature selection and classification using a pbest-guide binary particle swarm optimization," *Computation*, vol. 7, no. 1, p. 12, Feb. 2019.
[34] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms," *Appl. Soft Comput.*, vol. 18, pp. 261–276, May 2014.
[35] J. Li, J. Zhang, C. Jiang, and M. Zhou, "Composite particle swarm optimizer with historical memory for function optimization," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2350–2363, Oct. 2015.
[36] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *J. Biomed. Informat.*, vol. 85, pp. 189–203, Sep. 2018.
[37] O. Cordón, S. Damas, and J. Santamaría, "Feature-based image registration by means of the CHC evolutionary algorithm," *Image Vis. Comput.*, vol. 24, no. 5, pp. 525–533, May 2006.
[38] H. Banka and S. Dara, "A Hamming distance based binary particle swarm optimization (HDBPSO) algorithm for high dimensional feature selection, classification and validation," *Pattern Recognit. Lett.*, vol. 52, pp. 94–100, Jan. 2015.
[39] B. Tran, B. Xue, and M. Zhang, "Variable-length particle swarm optimization for feature selection on high-dimensional classification," *IEEE Trans. Evol. Comput.*, vol. 23, no. 3, pp. 473–487, Jun. 2019.
[40] 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.
[41] E. L. Lema-Condo, F. L. Bueno-Palomeque, S. E. Castro-Villalobos, E. F. Ordonez-Morales, and L. J. Serpa-Andrade, "Comparison of wavelet transform symlets (2-10) and daubechies (2-10) for an electroencephalographic signal analysis," in *Proc. IEEE 26th Int. Conf. Electron., Electr. Eng. Comput. (INTERCON)*, Aug. 2017, pp. 1–4.
[42] D. Chen, S. Wan, and F. S. Bao, "Epileptic focus localization using discrete wavelet transform based on interictal intracranial EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 5, pp. 413–425, May 2017.
[43] S. Haroun, A. N. Seghir, and S. Touati, "Multiple features extraction and selection for detection and classification of stator winding faults," *IET Electr. Power Appl.*, vol. 12, no. 3, pp. 339–346, Mar. 2018.
[44] *UCI Machine Learning Repository*. Accessed: Sep. 5, 2019. [Online]. Available: http://archive.ics.uci.edu/ml
[45] H. Jia, Z. Xing, and W. Song, "A new hybrid seagull optimization algorithm for feature selection," *IEEE Access*, vol. 7, pp. 49614–49631, Apr. 2019.
[46] K. Chen, F.-Y. Zhou, and X.-F. Yuan, "Hybrid particle swarm optimization with spiral-shaped mechanism for feature selection," *Expert Syst. Appl.*, vol. 128, pp. 140–156, Aug. 2019.
[47] M. Mafarja, I. Aljarah, H. Faris, A. I. Hammouri, A. M. Al-Zoubi, and S. Mirjalili, "Binary grasshopper optimisation algorithm approaches for feature selection problems," *Expert Syst. Appl.*, vol. 117, pp. 267–286, Mar. 2019.
[48] S. Arora and P. Anand, "Binary butterfly optimization approaches for feature selection," *Expert Syst. Appl.*, vol. 116, pp. 147–160, Feb. 2019.
[49] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017.

CHUN-YAO LEE (Member, IEEE) received the Ph.D. degree in electrical engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2007. From 2001 to 2008, he was a Distribution System Designer with the Engineering Division, Taipei Government. From 2004 to 2005, he was a Visiting Scholar with the University of Washington, Seattle, under the sponsorship of the Ministry of Science and Technology, R.O.C. In 2008, he joined as a Faculty Member with Chung Yuan Christian University. His research interests include power distribution, optimization algorithms, and damage diagnosis.

TRUONG-AN LE received the M.S. degree in electronic engineering from the University of Transport and Communications, Vietnam, in 2012. He is currently pursuing the Ph.D. degree with the Electrical Engineering Department, Chung Yuan Christian University, Taiwan. His research interests include fault detection and optimization algorithms.