Threshold-Optimized Swarm Decomposition Using Grey Wolf Optimizer for the Acoustic-Based Internal Defect Detection of Arc Magnets

Qinyuan Huang 1,2, Qiang Li 1, Maoxia Ran 1, Xin Liu 1, and Ying Zhou 1

1School of Automation and Information Engineering, Sichuan University of Science & Engineering, Zigong 643000, China
2Artificial Intelligence Key Laboratory of Sichuan Province, Zigong 643000, China

Correspondence should be addressed to Qinyuan Huang; qyhuang@suse.edu.cn

Received 11 November 2020; Revised 3 February 2021; Accepted 4 March 2021; Published 15 March 2021

Copyright © 2021 Qinyuan Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The acoustic-based internal defect detection is essential to ensure the quality of arc magnets efficiently. Swarm decomposition (SWD) is conducive to processing acoustic signals, but it is still confronted with threshold optimization problems. Especially, the existing optimization methods for the SWD thresholds are merely available for a single signal with exclusive characteristics, instead of the various signals with similar characteristics. Therefore, a threshold-optimized SWD using grey wolf optimizer (GWO) is proposed to solve these issues and applied to detect the internal defects of arc magnets. In this method, a fitness function is designed to indicate the relationship between the SWD thresholds and the overall decomposition effect of similar signals. The minimum value of it corresponds to the threshold setting yielding the optimal decomposition. GWO is used for searching such a minimum value, and the obtained optimal threshold setting allows SWD to decompose any signal into a series of oscillatory components. The frequency information in the two oscillatory components with the highest energy ratio is extracted as the internal defect features. Random forest is carried out to identify these features. Experimentally, the detection accuracy reaches above 97%, and the detection speed per single arc magnet does not exceed 3.4 seconds. The proposed method cannot only determine the unified threshold setting of SWD for similar signals but also achieve an accurate, rapid detection for the internal defects.

1. Introduction

Arc magnet is a tile-shaped magnetic segment, which is used to form a constant magnetic field within motors [1]. The quality of arc magnets plays a decisive role in the performance and life of motors. However, arc magnets may occasionally suffer from internal defects, due to some unexpected factors during the manufacturing process, such as the improper temperature and pressure. These internal defects will significantly weaken the mechanical integrity and physical properties of an arc magnet, resulting in severe performance degradation or even damage to the motor [2]. Consequently, the internal defect detection of arc magnets is particularly important to ensure the performance and safety of motors. Nonetheless, detecting internal defects is full of challenges. For one thing, internal defects are difficult to identify accurately, owing to their invisibility, complex types, uncertain locations, and varying degrees. For another, the large-scale production of arc magnets always must require a rapid detection efficiency. Meanwhile, a growing number of arc magnet manufacturers tend to prefer low-cost detection approaches. The traditional and existing methods are quite incompetent to enable effective detection. The commonly used detection strategy is to determine whether the internal defects exist through the sensitive human hearing when an arc magnet collides with a mental block, since internal defects are able to change the acoustic and vibration characteristics of an arc magnet being excited [3]. Such a detection method has low efficiency and poor accuracy, and, more importantly, is easily prone to be errors because of...
human factors. Therefore, how to achieve a fast, accurate, low-cost, and automatic detection for internal defects becomes the critical problem to be solved urgently in the manufacture of arc magnets.

Acoustic signals have been widely employed for developing defect detections and fault diagnoses [4, 5]; the reason for this is that they are easily and quickly acquired at low cost. The acoustic signal characteristics of an object are closely linked to its physical structure vibration. The internal defects are bound to change the physical structure of an arc magnet and, further, make an inevitable impact on the corresponding acoustic signal characteristics when the arc magnet vibrates in the air. There is always a characteristic difference in acoustic signals between qualified and defective arc magnets. Besides, instantaneous external forces, such as striking or colliding, provide a convenient and reliable way to excite an arc magnet to produce its high-quality acoustic signal, owing to its high hardness and excellent compressive strength. Analyzing the acoustic signal of arc magnets is expected to realize the internal defect detection. However, the related acoustic signal analysis is also extremely challenging for the internal defect detection. The nonlinearity and instability of acoustic signals can seriously hinder the extraction and identification of the signal features regarding internal defects. This is the reason why the acoustic-based detection for the internal defects of arc magnets is relatively less explored and reported.

Some modern signal processing technologies based on the multiscale adaptive decomposition have already opened the possibility to analyze the acoustic signal of arc magnets. Due to the advantage in decomposing an original signal into several different components adaptively, these methods, for instance, empirical mode decomposition (EMD) [6], local mean decomposition (LMD) [7], ensemble empirical mode decomposition (EEMD) [8], and variational mode decomposition (VMD) [9], are rather beneficial to explore the features and even weak information in nonstationary and nonlinear signals. However, the inability to overcome the adverse influence of the mode aliasing severely limits the decomposition accuracy of both EMD and LMD [10]. Also, EEMD finds it unable to avoid the inefficient computing problem, even though it poses the ability to weaken the mode aliasing significantly [11]. In many recent studies, VMD is confirmed to be superior in both the mode aliasing elimination and the adaptive decomposition, but the difficulty of selecting multiple preset parameters in a wide range greatly restricts its applications [12]. Recently, Apostolidis and Hadjileontiadis proposed a novel approach in the nonlinear, nonstationary signal decomposition, namely, swarm decomposition (SWD) [13], which fosters rules of biological swarms to address signal analysis problems. SWD shows the significant advantages over other signal decomposition technologies in improving the decomposing adaptability and overcoming the mode aliasing. SWD has attracted increasing attention in some studies and has been utilized to cope with different kinds of signals, such as vibrational, biological, and electrical signals. For example, Wan and Peng [14] adopted SWD to analyze the bearing signal and realized a fault diagnosis of the rotating bearing.

By using SWD to process biological signals, Baltatzis et al. [15] developed a detection to determine whether a bullying incident occurred on campus. Alnuaimi et al. [16] employed SWD for analyzing the Doppler signal of the fetal heart, which made it possible to detect the disease early and reduce fetal disease mortality.

Considering the advantages and applications of SWD compared to other signal decomposition methods, it appears that SWD has considerable potential for analyzing the acoustic signal of arc magnets because of the similarity between acoustic signals and vibration signals. Besides, it is well accepted that the optimal decomposition of a signal is most conducive to presenting the characteristics of the signal accurately. However, the performance of SWD is strictly subject to two preset thresholds, including $P_{th}$ and $Std_{th}$, which are termed as the peak threshold and the iteration deviation threshold, respectively. These two thresholds have a wide selection range, but the best performance of SWD commonly corresponds to a specific selection of them. Moreover, such a specific threshold setting is inconsistent in different signals. To realize the best decomposition of SWD, it is necessary to select the optimal threshold setting accurately before performing the decomposition of each signal. A growing number of research studies regarding SWD have come to recognize the importance and necessity of the proper threshold selection depending on optimization algorithms. A few meaningful attempts at the SWD threshold selection using optimization algorithms have emerged. For example, Miao et al. [17] proposed the whale optimization algorithm (WOA) to optimize the thresholds of SWD and then used the SWD with optimized thresholds to achieve the accurate decomposition of the rolling bearing vibration signal in the fault diagnosis. Although the optimization methods for SWD thresholds are not yet common, continuing efforts and progress have been made to resolve parameter optimization problems in the other applications. Samanta and Nataraj [18] adopted particle swarm optimization (PSO) to determine the optimal classifier parameters and achieved the accurate classification of bearing faults. Cerrada et al. [19] applied the genetic algorithm (GA) for the best set selection of gearbox condition parameters on time, frequency, and time-frequency domains and then developed the multiclass fault diagnosis in spur gears. All parameter or threshold optimization processes above need to express the characteristics and rules of parameters as a fitness function; the extremum of it corresponds to the optimal parameter setting. Remarkably, the optimization algorithm based on metaheuristics is responsible for searching the function extremum in the parameter space.

A metaheuristic optimization algorithm is a higher-level heuristic method that can provide an effective solution to an optimization problem, especially, in the cases of imperfect information or limited computation capacity. It is widely applied in solving various complex optimization problems, due to its distinct advantages in simplicity, flexibility, derivation-free mechanism, and local optima avoidance. The common feature of all metaheuristics is that the search process is divided into the exploration and the exploitation. The exploration is to investigate the promising areas of the
search space, whereas the exploitation refers to build the local search capability around the promising regions. Metaheuristics generally have three main classes, including evolutionary algorithms (EA), physics-based algorithms (PA), and swarm intelligence (SI). EA is inspired by the evolution in nature, and some typical algorithms in this branch are GA [20], differential evolution (DE) [21], evolutionary programing (EP) [22], genetic programing (GP) [23], and biogeography-based optimizer (BBO) [24]. PA is derived from imitating physical rules, and some of the most popular algorithms are simulated annealing (SA) [25], gravitational local search (GLS) [26], bigbang-bigcrunch (BBBC) [27], charged system search (CSS) [28], and central force optimization (CFO) [29]. SI refers to the metaheuristic algorithm mimicking the social behavior of swarms of creatures in nature. Recently, in comparison to EA and PA, SI has become the most interesting branch of metaheuristics because of fewer parameters to adjust, less operators, easier implementation, and more comprehensive information about the search space. Some of the widely used algorithms include PSO [30], ant colony optimization (ACO) [31], and artificial bee colony (ABC) [32]. Moreover, some of the typically recent and popular algorithms proposed so far are bat-inspired algorithm (BA) [33], WOA [34], grey wolf optimizer (GWO) [35], grasshopper optimization algorithm (GOA) [36], and fruit fly optimization algorithm (FOA) [37]. Among them, GWO performs an optimization by imitating the hunting mechanism and leadership hierarchy of grey wolves in nature. It has significant merits in the low computational time, the fast convergence, and the little possibility to be trapped at the local minimum. GWO constantly receives relatively wide attention and is used in various engineering fields. For example, Zhang et al. [38] proposed an adaptive stochastic resonance method for machinery fault diagnosis, in which the stochastic resonance system parameters are optimized by GWO algorithm using a redefined signal-to-noise ratio as optimization objective function. Sanjay et al. [39] employed the GWO algorithm for achieving the optimal global solution of the nonconvex discrete distributed generation problem. Verma et al. [40] utilized the GWO algorithm to optimize the parameters of the fractional-order controller for controlling the time-delay and higher-order system. Multiple studies for the parameter optimization show that GWO usually has a high-efficiency optimization performance in unknown, challenging search spaces and is able to provide highly competitive search results compared to other well-known metaheuristics. Thus, adopting GWO would be possible to enable the proper SWD threshold optimization.

According to the literature reviews mentioned above, it can be concluded that the combination of SWD and GWO would have intrinsic yet complementary advantages in the acoustic signal decomposition that depends on the threshold optimization. However, it is still unknown and unclear how such a combination in the acoustic signal analysis could be reasonably and properly used for the internal defect detection of arc magnets. Moreover, most of the optimization methods merely focus on a single signal and its exclusive characteristics, resulting in that the optimization results are inapplicable to another signal even if they are quite similar. In parallel, a fast detection for arc magnets also does not allow an individual threshold optimization to be performed for each acoustic signal because of the inevitable time-consuming and complicated calculations during the optimization process. To address these issues, we propose an acoustic signal analysis method for detecting the internal defects of arc magnets based on the optimized SWD, in which GWO is used to determine the optimal \( P_{\text{th}} \) and \( \text{Std}_{\text{th}} \) thresholds of SWD. The fitness function used for GWO is designed to denote the relationship between the SWD thresholds and the overall optimal decomposition performance of similar signals. The experimental results demonstrated that the proposed method realizes a unified optimization of SWD thresholds for the various similar signals and offers the reliable performance to detect the existence of internal defects of arc magnets rapidly and accurately.

2. Materials and Methods

2.1. Swarm Decomposition. SWD has the ability to iteratively extract the oscillation component (OC) from a multicomponent signal through the appropriate parametric swarm filtering (SWF). Each of OCs can be regarded as a real component of the original signal. There are two aspects to SWD, including the swarming model and the swarm-prey hunting. To achieve the swarm-prey hunting, the swarming model contains two interaction forces. The first one is the driving force, and the other one is the cohesive force. For the \( i \)-member at the step \( n \), the driving force, \( F_{\text{dr}}(n,i) \), is defined as

\[
F_{\text{dr}}(n,i) = P_{\text{prey}}(n) - P_i(n - 1),
\]

where \( P_{\text{prey}}(n) \) represents the position of the prey at the step \( n \) and \( P_i(n - 1) \) represents the position of the \( i \)-member at the step \( n - 1 \).

Unlike the driving force, the cohesive force refers to the inductive force between all members. The cohesion force for the \( i \)-member at the step \( n \), \( F_{\text{coh}}(n,i) \), is described as

\[
F_{\text{coh}}(n,i) = \frac{1}{M - 1} \cdot \sum_{j=1,j \neq i}^{N} f[P_i(n - 1) - P_j(n - 1)],
\]

where

\[
f(d) = -\text{sgn}(d) \cdot \ln \left| \frac{d}{d_{\text{cr}}} \right|
\]

where \( M \) represents the number of members in the swarm model, \( d \) refers to the distance between the two members, and \( d_{\text{cr}} \) is defined as the standard distance that keeps the two members not to affect each other, \( \text{sgn}(\cdot) \) denotes the sign
function, $f(\cdot)$ determines the type of the cohesion from the $j$th member to the $i$th member, and $f(d_{ij}) = 0$.

To track the prey, the members need to update their respective positions. Sequentially, the velocity and position of the $i$th member at the step $n$ is described as

$$
V_i(n) = V_i(n-1) + \delta \cdot \left[ F_{ir}(n) + F_{coh}(n, i) \right],
$$

$$
P_i(n) = P_i(n-1) + \delta \cdot V_i(n),
$$

where $\delta$ is used to control the flexibility of the swarm and $V_i(n)$ represents the velocity of the $i$th member at the step $n$. It is worth noting that $P_{ir}(n), P_i[n], \text{ and } V_i[n]$ are sequences of scalar values in order to simplify the data.

When the hunting is completed, the positions of all members in the swarm are summed as the output of SWF, and the corresponding equation is expressed as

$$
y_n = \beta \cdot \sum_{i=1}^{M} P_i(n),
$$

where $\beta$ is the scale factor that impacts the order of $M$ and $\beta = 0.005$ affects the smallest reasonable $M$. Through the above analysis, $\delta$ and $M$ as the decisive factor have an impact on the output of SWF. Therefore, the best $\delta$ and $M$ can be calculated by the following equations with the normalized frequency $\bar{\omega}$:

$$
M(\bar{\omega}) = 33.46\bar{\omega}^{-0.375} - 29.1,
$$

$$
\delta(\bar{\omega}) = -1.5\bar{\omega}^2 + 3.454\bar{\omega} - 0.01.
$$

The core idea of SWD is to perform the filtering-like operations on a multicomponent input signal iteratively. In each iteration, a frequency band with a significant amplitude in the energy spectral density is selected as a candidate for optimizing the extraction mode. The optimal frequency band $\bar{\omega}_m$ can be calculated by the following equation:

$$
\bar{\omega}_m = \text{arg max} \left[ S'X_{it}(\bar{\omega}) > P_{th} \right],
$$

$$
S'X_{it}(\bar{\omega}) = \text{SGF}[SX_{it}(\bar{\omega})],
$$

where $SX_{it}(\cdot)$ represents the fast Fourier transform of $X_{it}(n)$ and $\bar{\omega}$ denotes the normalized signal frequency.

The algorithm framework of SWD is summarized as follows [41]:

Step 1: input the original signal, $X(n)$, and initialize the preset thresholds $P_{th}$, and $\text{StD}_th$, where $\forall n = 0, \ldots, L - 1$.

Step 2: discretize $X(n)$ into $X_{it}(n)$ and assign it to $y_{it}(n)$, where the subscript, $it$, in $X$ is the serial number and $i = 0$ initially.

Step 3: smooth the energy spectrum of $y_{it}(n)$ and calculate its optimal frequency band $\bar{\omega}_m$.

Step 4: the swarm parameters, namely, $M$ and $\delta$, are calculated based on the optimal frequency band $\bar{\omega}_m$.

Step 5: with the swarm parameters $M$ and $\delta$, SWF is performed on the discretized signal $X_{it}(n)$ to obtain $y_{i}(n)$, where $i$ is initially 1.

Step 6: calculate the variance between $y_{i}(n)$ and its initial signal, where the equation is as follows:

$$
\text{StD} = \sum_{n=1}^{N} \frac{|y_{i}(n) - y_{i-1}(n)|^2}{|y_{i-1}(n)|^2}.
$$

If $\text{StD} > \text{StD}_th$, assign $y_{i}(n)$ to the input signal $X_{it}(n)$, return to Step 2 to perform SWF, and set $i = i + 1$. Otherwise, assign $y_{i}[n]$ to $X'_{it}[n]$.

Step 7: update the input signal, according to the calculation equation that is as follows:

$$
X_{it+1}(n) = X_{it}(n) - C_{\bar{\omega}_d}(n),
$$

$$
C_{\bar{\omega}_d}(n) = X_{it}'(n - \tau_{delay}),
$$

$$
\tau_{delay} = \text{arg max}_{\tau} \left[ R_{x_i x_t}(\tau) \right],
$$

where $R_{x_i x_t}(\tau)$ is the correlation number, $\tau$ is the delay, and $-L \leq \tau \leq L - 1$. Through the updated signal, repeat Step 2 to Step 7 until $S'X_{it}(\bar{\omega}) \leq P_{th}, \forall \bar{\omega}$.

Step 8: extract the OCs and the residue signal $r(n)$. Defining $\Omega$ is the set of the recognized OCs, the corresponding calculation equations are as follows:

$$
OC_m = C_{\bar{\omega}_m}(n), \quad \forall m = 1, 2, 3, \ldots, \Omega,
$$

$$
C_{\bar{\omega}}(n) = \sum_{m=1}^{\Omega} C_{\bar{\omega}_m}(n), \quad \forall \bar{\omega} \in \Omega,
$$

$$
\Omega: [\bar{\omega}], \bar{\omega} = \bar{\omega}_{\text{dom}},
$$

$$
r(n) = X_{it}(n)_{i=t=\text{final}}.
$$

2.2. Grey Wolf Optimizer. GWO is a new type of bionic swarm intelligent optimization algorithm proposed by Mirjalili et al. [42]. This algorithm mainly imitates the grey wolf leadership mechanism and hunting mechanism in nature. All grey wolves are divided into four levels, including $w1$, $w2$, $w3$, and $w4$. As a decision maker, $w1$ is responsible for the hunting guidance as well as the decision formulation. As a subordinate of $w1$, $w2$ primarily helps $w1$ in its decision-making. $w3$ is chiefly in charge of patrolling their land, warning the wolf pack in case of emergencies, guarding weak wolves, and hunting the prey. In this hierarchical system, though $w4$ has the lowest status, it can perform as a venting object for other members when they suffer setbacks.

In fact, the hunting hierarchy of GWO can be considered as the process to find a globally optimal solution, where $w1$, $w2$, $w3$, and $w4$ are assumed as the first, second, third, and fourth optimal solution, respectively. To hunt the prey successfully, the prey must be surrounded firstly. The surrounding behavior can be described by the following equations [43]:

$$
D = \left| C \cdot \overrightarrow{X}_p(t) - \overrightarrow{X}(t) \right|,
$$

$$
\overrightarrow{X}(t+1) = \overrightarrow{X}_p(t) - \overrightarrow{A} \cdot \overrightarrow{D},
$$

where $t$ represents the current iteration number, $\overrightarrow{X}_p(t)$ and $\overrightarrow{X}(t)$ are the position vectors of the prey and grey wolf,
respectively, and $\mathbf{a}$ and $\mathbf{c}$ are coefficient vectors, and they can be calculated as follows:

$$\mathbf{a} = 2 \cdot \mathbf{d} \cdot \mathbf{r}_1 - \mathbf{d},$$
$$\mathbf{c} = 2 \cdot \mathbf{r}_2,$$

(11)

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are random vectors within $[0, 1]$ and $\mathbf{d}$ decreases linearly from 2 to 0 during the entire iteration.

After the encircle of the prey is realized, the search agent constantly updates its position through the position of $w_1, w_2,$ and $w_3,$ to reduce the encircling circle rapidly. The corresponding mathematical equations are as follows [44]:

$$\begin{align*}
\mathbf{b}_{w1} & = |\mathbf{c}_1 \cdot \mathbf{x}_{w1} - \mathbf{x}|, \\
\mathbf{b}_{w2} & = |\mathbf{c}_2 \cdot \mathbf{x}_{w2} - \mathbf{x}|, \\
\mathbf{b}_{w3} & = |\mathbf{c}_3 \cdot \mathbf{x}_{w3} - \mathbf{x}|, \\
\mathbf{x}_1 & = \mathbf{x}_{w1} - \mathbf{a}_1 \cdot \mathbf{b}_{w1}, \\
\mathbf{x}_2 & = \mathbf{x}_{w2} - \mathbf{a}_2 \cdot \mathbf{b}_{w2}, \\
\mathbf{x}_3 & = \mathbf{x}_{w3} - \mathbf{a}_3 \cdot \mathbf{b}_{w3},
\end{align*}$$

(12)

2.3. Random Forest. RF is a robust and practical decision tree learning method, which trains several weak models to form a strong model, and it belongs to an integrated machine learning method [45]. Taking a binary decision tree as an example, when training a model, it needs to consider how to segment variables and points, and how to measure their quality. Generally, the quality of the variables and points are measured by the impurity of the nodes after the segmentation. The calculation equation can be defined as follows:

$$G(x_l, v_{lk}) = \frac{n_{left} \cdot H(X_{left}) + n_{right} \cdot H(X_{right})}{N_s},$$

(13)

where $x_l$ is a feature of the segmentation vector, $v_{lk}$ is a value of the segmentation variable, $n_{left}, n_{right},$ and $N_s$ are the number of training samples of the left child node, the number of training samples of the right child node, and the number of all training samples in the current stage, respectively, $X_{left}$ and $X_{right}$ are the training sample sets of the left and right child nodes, respectively, and $H(X)$ is a function to measure the impurity of the node.

The machine learning model of RF can be transformed into an optimization problem of finding the smallest segmentation vector $x^*$ and segmentation value vector $v_{k}^*$, as shown below [46]:

$$\arg\min_{x^*, v_{k}^*} G(x, v_k).$$

(14)

2.4. Experimental Sample. The unmagnetized ferrite arc magnets with four different sizes are procured as experimental samples, whose appearances and specifications are given in Figure 1(a) and Table 1, respectively. The set of all collected samples is able to form a robust representativeness of conventional arc magnets because the regular material and general size of arc magnets are covered by samples. According to the sizes, the samples are divided into four categories, which are termed as Type A, B, C, and D, severally. For ease of description, $R, H, T, L,$ and $W,$ in turn, are used to represent the radius, height, thickness, length, and width of an arc magnet. Each category of samples is made up of 240 arc magnets, which is equally split between qualified samples and defective samples. The qualified one refers to the arc magnet without any defects, while the defective one has only internal defects. Both qualified and defective samples in each category are further broken down into two groups as training samples and testing samples. Randomly collected 50-qualified and 50-defective samples from each category act as the training group to develop the detection algorithm. In contrast, the remaining samples in each category serve as the testing group for the performance verification of the detection algorithm. In order to ensure the reliability of experimental samples, the condition of all collected arc magnets has been consistently confirmed by three more experienced inspectors through the manual detection method.

2.5. Experimental Detection System. As exhibited in Figure 1(b), the acoustic-based internal defect detection of arc magnets is a low-cost automation system composed of mechanical and electrical devices. Its running process is made up of three parts in sequence, including exciting an acoustic signal, collecting the acoustic signal, and processing the signal data. Firstly, a stepped structure consisting of two conveyor belts of different heights is used for transporting arc magnets from the upper conveyor to the lower one at a uniform speed. The vertical drop of each arc magnet occurs at this step, and then, the falling arc magnet collides with a metal block placed in the direction of falling to excite an acoustic signal. The difference in height between the two conveyors is set to 20 mm according to the engineering experience to ensure that the collision does not damage the arc magnet while producing a high-quality acoustic signal. Additionally, a pair of photoelectric sensors is also installed in the direction of falling. It will generate a trigger signal to initiate the signal acquisition if its optical path is blocked by a falling arc magnet. At this point, a 0.5-second acoustic signal can be acquired by a microphone having a frequency response ranging from 10 Hz to 20 kHz. The signal conditioning is employed for filtering and amplifying the collected signal. After that, a data acquisition card with a 40 kHz sampling rate and a 12 bit resolution is responsible for converting this signal into digital data stored in a computer.
with 2.5 GHz CPU and 16 GB RAM. Finally, the signal data is analyzed by a detection algorithm developed by a MATLAB program. Such an algorithm is able to provide a result regarding the presence or absence of internal defects in an arc magnet through extracting and identifying features of the acoustic signal. A push rod driven by an air cylinder is in charge of transferring the detected arc magnet to the lower conveyor, and subsequently, the sorting air cylinder is devoted to delivering the arc magnet to the corresponding outlet in accordance with the detection result.

2.6. Experimental Detection Algorithm. The acoustic-based detection for internal defects of arc magnets is to determine whether the internal physical structure is abnormal by the relationship between the structure attribute and its acoustic characteristics. As illustrated in Figure 1(c), the algorithm flow refers to the process of analyzing the acoustic signal of arc magnet, which is divided into the signal decomposition, the feature extraction, and the feature identification, respectively. In the first part, the signal decomposition is in charge of locating the feature information from a large amount of the signal data. The acquired acoustic signal of each sample is processed by a high-pass filter to remove the frequency information below 4 kHz. Then, according to characteristics for acoustic signals of arc magnets, a fitness function, which can reflect the relationship between the threshold setting of both $P_{th}$ and $S_{Th}$ and the SWD performance for the whole samples, is constructed. Moreover, the function minimum is designed to indicate the overall optimal SWD effect. Next, GWO is responsible for searching this value from the selectable threshold range to determine the optimal threshold setting. After that, the acoustic signals of all samples are uniformly decomposed by SWD with the optimal threshold setting so that each signal can be processed as optimally as possible.

In the next part, the feature extraction is responsible for combining all specific signal information into the feature data. The first two OCs with the highest energy ratio, which can be considered to have the greatest characteristic contributions, are selected from the SWD results. Among them, the one with the highest energy ratio is referred to as the first characteristic OC (COC1), and the other is the second characteristic OC (COC2). Then, the frequencies of the maximum peak are, respectively, extracted from COC1 and COC2 as the feature data. Such two obtained data are combined into a feature vector that quantifies whether there are internal defects in the sample generated the corresponding acoustic signal.

In the last part, the feature identification is used to identify the extracted feature data and determine whether the arc magnet corresponding to the acoustic signal is defective. The feature vectors of all training samples in the same category are input to RF to train a classifier model for the corresponding category. Then, this classifier is employed for identifying the feature vector of each testing sample in the category. According to the identification result, whether each sample has internal defects can be determined.

3. Results and Discussion

3.1. Comparison of Acoustic Signals of the Qualified and Defective Samples. Exploring the signal characteristic differences between qualified and defective samples are crucial to the success of the detection. It is generally accepted that both time domain and frequency domain are the most conventional signal analysis methods. To this end, a qualified sample and a defective sample are randomly selected from Type A to analyze their differences in the time and frequency domain. In order to fairly compare the characteristics of signals that are likely to be of different magnitudes, each signal of selected samples is normalized by the proportion of the signal amplitude to its maximum value. As can be seen in Figure 2(a), there are no significant distinctions in the time domain waveforms of qualified and defective samples. Likewise, as exhibited in Figure 2(b), their waveforms in the frequency domain are also quite similar. Directly analyzing the time-frequency characteristics of the entire signal seems to be unable to distinguish between qualified and defective samples easily, but it can be seen that their acoustic signals are obviously composed of multiple frequency components. Moreover, it is worth noting that the characteristics of the acoustic signal of an arc magnet, such as significant peaks, are mainly concentrated in the high-frequency components. For example, several prominent peaks always appear in the high-frequency parts, regardless of whether they belong to a qualified sample or a defective one. In contrast, the low-frequency components of both qualified and defective samples are too low in amplitude and too similar to each other, resulting in that it is hard to establish a robust distinction. Thus, decomposing the signal into multiple frequency components is supposed to play an important role in achieving an effective signal analysis.

3.2. Advantages and Limitations of SWD. To obtain multiple frequency components of the acoustic signal, SWD is selected as the signal decomposition method. To demonstrate the superiority of SWD, two widely used signal decomposition methods, namely, EMD and VMD, are applied to compare with it. In order to perform a fair comparison between algorithms, the same qualified sample in Type A is randomly selected out, and then its signal is separately decomposed by EMD, VMD, and SWD. As seen in Figure 2(c), a total of 6 intrinsic mode functions (IMF) are generated by EMD. The frequency spectrum of the first IMF is extremely similar to that of the original signal, and the remaining IMFs only correspond to the low-frequency components. As a result, EMD is not suitable for the fine decomposition of the frequency components in the arc magnet acoustic signal, especially those in the high-frequency part. In addition, the same signal is processed by VMD. Since the implementation of VMD requires four parameters to be set in advance, two different parameter settings are chosen to analyze and compare the decomposition performance of VMD. These four parameters in each setting are termed as the mode number $K$, the penalty factor $\alpha$, the Lagrange multiplier $\tau$, and the iteration accuracy $\epsilon$.
respectively. It can be seen from Figure 2(d) that a total of three modes are obtained by using the parameter setting at $K = 3$, $\alpha = 100$, $\tau = 0$, and $\varepsilon = 10^{-4}$, but Mode 1 and Mode 2 in the decomposition results are quite similar, indicating that the mode aliasing has occurred. Simultaneously, Figure 2(e) shows the results of decomposing the same signal through VMD with the other parameter setting at $K = 4$, $\alpha = 2000$, $\tau = 0.1$, and $\varepsilon = 10^{-6}$. It can be seen that there are four obtained modes, each of which corresponds to an exclusive frequency component. Accordingly, VMD is indeed superior to EMD in decomposing such a signal. However, the performance of VMD depends on the proper setting of these four parameters within wide ranges. More importantly, the parameter $K$ directly determines the number of modes, namely, the number of signal components that can be expressed, but the exact number of signal components is commonly unknown or uncertain before the decomposition. On the contrary, a better decomposition effect than EMD and VMD is given by SWD. Figure 2(f) shows the results of SWD at $P_{th} = 0.3$ and $\text{Std}_{th} = 0.4$, when decomposing the same signal. With this threshold setting, SWD can produce a total of three OCs, which are all linked to different high-frequency components. Furthermore, the last OC indicates the residue after the decomposition, which is not available in other methods. The results show that high-frequency components representing signal characteristics can be obtained separately by SWD, resulting in that the differences in signal details between qualified and defective samples are conducive to discovery. Compared to VMD, SWD relies on fewer parameters, does...
not need to determine the number of signal components in advance, and provides more information. However, the implementation of SWD is always inseparable from the presets of both \( \text{P}_{\text{th}} \) and \( \text{StD}_{\text{th}} \). As exhibited in Figure 2(g), after changing the threshold setting to \( \text{P}_{\text{th}} = 0.3 \) and \( \text{StD}_{\text{th}} = 0.2 \), the SWD results are completely different from the previous one. For instance, a total of four different OCs can be obtained, in which the high-frequency components are broken down more finely. Hence, it is necessary to make the optimal selection for the threshold setting in order to realize the best decomposition results.

### 3.3. Preprocessing of Original Signals

Since only the high-frequency components are meaningful for detecting the internal defects, removing low-frequency components before starting the signal decomposition can significantly reduce the amount of the data processing, decrease the processing difficulty, and shorten the processing time. The high-pass filter is considered as an effective tool to remove low-frequency components, but its filtering range needs to be determined in advance. As seen in Figure 2(b), the frequency components below the dominant peak are relatively similar. However, removing components near the dominant peak may result in the loss of too much potentially useful information because they are inextricably linked to the dominant peak characteristics despite their weak amplitude. In addition, a similar low-frequency range between different samples may not be the same. To determine the unified range used for a high-pass filter, five qualified and five defective samples are randomly selected from each type. Figure 3 shows their frequency spectra, respectively. By observing the similarity in low-frequency components between all qualified and defective samples, 4 kHz is suitable as the filtering range of a high-pass filter. Below such a frequency, the four types of samples always have the greatest similarity between qualified and defective samples. On the contrary, above this frequency, there is a distinction between qualified and defective samples. It means that the internal defect information of arc magnets carried by acoustic signals would most likely exist in the high-frequency part above 4 kHz. The analysis of the low-frequency part presumably brings about ineffective calculations as well as redundant time consumptions. Therefore, to simplify the signal data and realize an accurate signal analysis, each acoustic signal is preprocessed by a 4 kHz high-pass filter before being decomposed by SWD.

### 3.4. Acoustic Signal Decomposition Based on the Threshold-Optimized SWD

#### 3.4.1. Fitness Function for the Optimization

The signal decomposition performance heavily depends on two preset thresholds, including \( \text{P}_{\text{th}} \) and \( \text{StD}_{\text{th}} \). Inappropriate thresholds can cause either over-decomposition or under-decomposition. The over-decomposition inevitably produces redundancy results, such as completely repetitive or excessively approximate OCs. As shown in Figure 4(a), the threshold setting at \( \text{P}_{\text{th}} = 0.1 \) and \( \text{StD}_{\text{th}} = 0.1 \) leads to the over-decomposition of the previous qualified sample. For example, OC 1 and OC 2 are so extremely similar that they only contribute the same information of signal details but require extra computation to obtain these two components. Conversely, the under-decomposition is bound to result in the inability to break down several components into a set of separate and different OCs. As shown in Figure 4(b), the under-decomposition occurs in the same sample when \( \text{P}_{\text{th}} = 0.5 \) and \( \text{StD}_{\text{th}} = 0.5 \). The significant frequency components of the original signal are integrated into the same OC 2, and they cannot be separately obtained by any individual OC. Consequently, SWD using the optimal threshold should be able to minimize the over-decomposition as well as the under-decomposition.

These two thresholds have a wide selectable range, and any value in their range can be set as a possible threshold. The threshold optimization is to select a proper setting from such a range to avoid the over-decomposition and the under-decomposition. To quantify the relationship between the decomposition performance and the threshold setting, a fitness function, \( f (\text{P}_{\text{th}}, \text{StD}_{\text{th}}) \), is designed as follows:

\[
f (\text{P}_{\text{th}}, \text{StD}_{\text{th}}) = \frac{1}{N_{t}} \sum_{i=1}^{N_{t}} \left[ \left( 1 - C_{i}(x_{i}, y_{i}) \right) + M_{i} \right],
\]

\[
C_{i}(x_{i}, y_{i}) = \frac{\sum_{t=1}^{L} \left[ x_{i}(t) - \bar{x}_{i} \right] \left[ y_{i}(t) - \bar{y}_{i} \right]}{\sqrt{\sum_{t=1}^{L} \left[ x_{i}(t) - \bar{x}_{i} \right]^{2}} \sqrt{\sum_{t=1}^{L} \left[ y_{i}(t) - \bar{y}_{i} \right]^{2}}}
\]

\[
x_{i} = \sum_{j=1}^{k-1} u'_{j}, M_{i} = \sum_{j=1}^{k-1} v'_{j}, v'_{j} = \begin{cases} 1, & \text{if } C_{i}(u'_{j-1}, u'_{j}) > 0.1 \text{ and } j > 1, \\ 0, & \text{if } C_{i}(u'_{j-1}, u'_{j}) \leq 0.1 \text{ or } j \leq 1, \end{cases}
\]

where \( N_{t} \) is the number of training samples, \( y_{i} \) is the original signal of the \( i \)th sample, \( x_{i} \) is the sum of the first \( k - 1 \) OCs obtained from \( y_{i} \) by SWD, \( k \) represents the number of all obtained OCs, \( u'_{j} \) refers to the \( j \)th OC of \( y_{i} \), \( L \) is the length of the data, \( \bar{x}_{i} \) and \( \bar{y}_{i} \), respectively, represent the average of \( x_{i} \) and \( y_{i} \), \( C_{i}(x_{i}, y_{i}) \) indicates the Pearson correlation coefficient [47] between \( x_{i} \) and \( y_{i} \), \( M_{i} \) is the number of times that the Pearson correlation coefficient of adjacent OCs in \( y_{i} \) is
Table 1: The specification of experimental samples.

| Type | Size (mm) | Condition       | Number of the training samples | Number of the test samples |
|------|-----------|-----------------|--------------------------------|---------------------------|
| A    | 50 23 8 36 64 | Qualified       | 50 70                          |                            |
|      |           | Defective       |                                |                            |
| B    | 45 21 7 30 62 | Qualified       | 50 70                          |                            |
|      |           | Defective       |                                |                            |
| C    | 30 10 5 50 21 | Qualified       | 50 70                          |                            |
|      |           | Defective       |                                |                            |
| D    | 20 9 3 35 25 | Qualified       | 50 70                          |                            |
|      |           | Defective       |                                |                            |

Figure 2: Continued.
Figure 2: Comparison of acoustic signals of randomly selected samples in type A. (a) The time-domain diagram of acoustic signals of a qualified sample and a defective sample. (b) The frequency-domain diagram of acoustic signals of a qualified sample and a defective sample. (c) The results of EMD when decomposing a qualified acoustic signal. (d) The results of VMD at $K = 3$, $\alpha = 100$, $\tau = 0$, and $\varepsilon = 10^{-4}$ when decomposing the same qualified acoustic signal. (e) The results of VMD at $K = 4$, $\alpha = 2000$, $\tau = 0.1$, and $\varepsilon = 10^{-6}$ when decomposing the same qualified acoustic signal. (f) The results of SWD at $P_{th} = 0.3$ and $StD_{th} = 0.4$ when decomposing the same qualified acoustic signal. (g) The results of SWD at $P_{th} = 0.3$ and $StD_{th} = 0.2$ when decomposing the same qualified acoustic signal.

Figure 3: Frequency spectra of randomly selected five qualified and five defective samples. (a) Type A. (b) Type B. (c) Type C. (d) Type D.
greater than 0.1, and $v_i^j$ is the specific binary result of the Pearson correlation coefficient between two adjacent OCs, namely, $u_{i+1}'$ and $u_i'$. In such a fitness function, the two thresholds of SWD, including $P_{th}$ and $StD_{th}$, are regarded as the function variables. Since the last OC often acts as the residue of the decomposition, only the first $K - 1$ OCs are used for this function. The increase of $C_i(x_i, y_i)$ means that the obtained components are closer to the original signal so that the smaller $1 - C_i(x_i, y_i)$ indicates the less possibility of under-decomposition. On the contrary, the decrease in $M_i$ accounts for the fact that the obtained components are more irrelevant to each other. The lower $M_i$ signifies that the over-decomposition is less likely to occur. Hence, the combination of $1 - C_i(x_i, y_i)$ and $M_i$ has the ability to simultaneously reflect the degree of over-decomposition and under-decomposition, resulting in that the decomposition performance of SWD for a single signal can be fully expressed. Besides, the average of the decomposition performance of all training samples is utilized to quantify the overall decomposition performance of samples in the same type. As a result, the designed fitness function can establish a sufficient correlation between the threshold setting and the whole or individual decomposition performance of SWD. Any setting regarding $P_{th}$ and $StD_{th}$ is able to produce a unique value of the fitness function; the minimum of it corresponds to the optimal threshold setting.

### 3.4.2. SWD Threshold Optimization with GWO

The threshold optimization is to search the minimum of the function above from the available threshold range. Both thresholds range from 0 to 0.5, and consequently, this function is distributed in a three-dimensional space, which is composed of $P_{th} \in [0, 0.5]$, $StD_{th} \in [0, 0.5]$, and $f(P_{th}, StD_{th})$. To complete the optimization, GWO is used to search for the coordinates of the minimum function value in this space. The GWO algorithm in this work is performed through fixed parameters determined by both empirical values and specific search ranges, including $n_u$, $n_i$, $n_j$, $b_i$, and $b_u$. The first two parameters indicate the number of search agents and the maximum number of iterations, respectively, which are empirically set as $n_u = 10$ and $n_j = 10$. The third parameter, namely, $n_i$, denotes the number of search objects and is set to two because of two dimensions formed by $P_{th}$ and $StD_{th}$. The last two parameters are the lower and upper limit vectors of search ranges, respectively, and $b_i = [0, 0]$ and $b_u = [0.5, 0.5]$. Taking all training samples in Type A as an example, Figure 4(c) shows the searching path and result of the threshold optimization. It can be seen that the minimum value of the function, namely, 0.087, is found at the coordinate of $P_{th} = 0.382$ and $StD_{th} = 0.092$ through ten iterations of the searching location update. To verify the consistency of the optimization results, such a threshold optimization was performed five times, separately. As shown in Figure 4(d), although the starting coordinate of each optimization is different, the final function value searched by the five optimizations is exactly the same. Moreover, each optimization always consumes only ten iterations, during which the searched values converge rapidly. As also shown in Figure 4(d), the searched function value will no longer change after the tenth iteration, even if the number of iterations is increased. Obviously, GWO provides a short searching path, a low number of iterations, a fast optimization speed, and a result that does not trap in the local optimization.

The $P_{th}$ and $StD_{th}$ at the found coordinates can be considered as the optimal threshold setting. Since the function represents the whole decomposition performance of samples in Type A, the determined optimal threshold setting can enable each sample to reach the overall optimal decomposition effect rather than a single optimal effect. As shown in Figure 4(e), the acoustic signal of the previous qualified sample is decomposed by SWD using the obtained optimal threshold setting. Unlike the over-decomposition and under-decomposition, the optimal threshold setting for the overall samples can produce more elaborate results without redundant information. At the same time, the optimal threshold setting for a single sample can be also obtained by using the designed function and this signal. As shown in Figure 4(f), the optimal threshold setting, which is $P_{th} = 0.357$ and $StD_{th} = 0.124$, is different from the one for the overall samples, but both of them produce almost identical decomposition results. It indicates that the optimal decomposition results for the overall samples do approximate to those for an individual sample. It further means that the optimal $P_{th}$ and $StD_{th}$ for the overall samples can be used as the uniform threshold setting to decompose other signals in the same type, instead of carrying out the threshold optimization for each congeneric new signal. Consequently, the threshold setting for performing the SWD processing of each signal in the same type can be determined from a one-time optimization based on the signals of all training samples. Since the optimization takes only one time and the optimal threshold setting is unique, the sensitivity and uncertainty of the thresholds no longer seriously affect the optimal decomposition performance of SWD. Moreover, such a threshold optimization can tremendously contribute to achieving a rapid signal analysis through the least amount of the optimization because any optimization process always imposes additional computational burdens and extra time consumption on data processing. Similarly, the optimal threshold for other types of samples can also be obtained by GWO. Table 2 shows the results of the threshold optimization for four types of samples.

### 3.5. Feature Extraction

#### 3.5.1. OC Selection

After the original signal is decomposed by SWD using the optimal threshold setting, a set of OCs can be obtained, each of which can be regarded as a component. However, the information representing the internal defects is generally only part of the signal decomposition results, rather than all of it. The meaningful results refer to those OCs that are likely to contain the signal characteristics related to internal defects. The energy ratio would be exceptionally useful and beneficial for quantifying how much
signal characteristic information an OC contains because the more abundant characteristic information in a signal component tends to the higher the energy. The energy ratio of each OC can be defined as follows:

\[ E_k = \frac{\sum_{j=1}^{L} |u_k(j)|^2}{\sum_{j=1}^{L} |y(j)|^2}, \]

where \( u_k \) denotes the \( k \)th OC, \( E_k \) acts as its energy ratio, \( y \) indicates the original signal, \( L_k \) is the length of \( u_k \), and \( L_y \) is the length of \( y \).

Within the results of SWD, the OC with the largest energy ratio inevitably has the most signal characteristic information, leading to that it has the highest probability of having the most information regarding internal defects. An example can be found from the energy ratio distribution of OCs, which is derived from a qualified sample and a defective sample, respectively. As seen in Figure 4(g), regardless of qualified or defective samples, the first two OCs with the largest energy ratio almost account for the majority of the whole signal energy. Obviously, the features of internal defects are suitable to be extracted from these two OCs. However, SWD does not ensure that each signal can be decomposed into the same number of OCs. To this end, the two OCs used for the feature extraction are determined as follows:

\[
\begin{align*}
\text{COC1} &= u_{c1}, \\
\text{COC2} &= u_{c2} (N_u > 1), \\
\text{COC1} &= \text{COC2} = u_{c2} (N_u = 1),
\end{align*}
\]

where \( u_{c1} \) and \( u_{c2} \) are the first and second OCs with the largest energy ratio, \( N_u \) is the number of all OCs, and COC1 and COC2 denote the first and second characteristic OCs for the feature extraction. Such a determination method can guarantee that there are always two OCs for the feature extraction regardless of the number of the SWD results.

3.5.2. Acoustic Signal Features of Arc Magnet Internal Defects. Much direct and indirect information in a signal, such as entropy [48], variance [49], and minimum redundancy [50], can be considered as signal characteristics. The vector or matrix composed of their complicated set is often adopted as the feature data, resulting in the huge difficulty and excessive computation of the feature extraction. For a rapid detection, simple yet effective features are vitally important. In this work, the frequencies of the maximum peak in both COC1 and COC2 are directly regarded as features representing the existence of internal defects. The frequency of the maximum peak in COC1 is termed as \( f_1 \), while the one in COC2 corresponds to \( f_2 \). The combination of \([f_1, f_2]\) is the final feature vector for each signal. As depicted in Figure 5, the feature distribution of four types of training samples shows that the extracted features provide a robust distinguishability between qualified and defective samples. The features related to qualified samples are concentrated in a relatively small region, but those belonging to defective samples are randomly distributed in a vast region. Although the minority of features for qualified and defective samples is quite close together, not only do they have no obvious overlap and intersection but also have a clear boundary between each other. Such clustering results of extracted features are conducive to the identification because the feature data of qualified and defective samples are completely different.

3.6. Feature Identification and Detection Results. In this work, RF serves as the classifier to identify the features and determine the detection results. To fairly verify the effectiveness of the feature identification with RF, two frequently used identification methods, namely, support vector machine (SVM) [51] and k-nearest neighbors (KNN) [52], are selected as the classifiers to compare with RF. The classifiers corresponding to these three methods are trained and tested by the same sample signals. The classifier of each type of samples is trained by their own training samples, while the testing samples are employed for the identification using the trained classifier. The recognition accuracy of the detection algorithm for all samples is shown in Table 3. The accuracy includes \( A_D, A_N, \) and \( A_{DN} \), which can be defined as follows:

\[
\begin{align*}
A_D &= \frac{I_D}{N_D}, \\
A_N &= \frac{I_N}{N_N}, \\
A_{DN} &= 100\% \times \frac{(I_D + I_N)}{(N_D + N_N)},
\end{align*}
\]

where \( I_D \) and \( I_N \) indicate the number of defective and qualified samples correctly identified, respectively, \( N_D \) and \( N_N \) are the number of defective and qualified testing samples, respectively, \( A_D \) and \( A_N \) denote the recognition rate of defective and qualified samples, respectively, and \( A_{DN} \) represents the overall recognition rate of qualified and defective samples.

It can be seen that RF achieves the best results compared to the other two classifiers, regardless of the types of samples. In the results of RF, the overall recognition rate for Type A, B, and C reaches 100%, but that of Type D is 97.86%. Although three qualified samples in Type D are not correctly identified, the recognition rate of all defective samples is still 100%. By contrast, due to the lowest false detection rate, such an algorithm using RF has a significant value of utilization.

In addition to the recognition accuracy, the running time is also another important performance of the detection algorithm. According to the designed detection algorithm, it consists of the time cost of the preparation and detection. The former refers to the time consumption of processing all training sample signals, including the acoustic signal acquisition, the threshold optimization, the feature extraction, and the classifier training. The latter is on behalf of the time consumed by a single testing sample from starting the signal acquisition to producing the detection result by the well-trained classifier. Table 4 shows the running time of the detection algorithm in units of seconds. \( T_s \) denotes the time
Figure 4: Results of the threshold optimization and signal decomposition of SWD for type A. (a) The over-decomposition results of SWD at $P_{th} = 0.1$ and $StD_{th} = 0.1$ when decomposing the same signal. (b) The under-decomposition results of SWD at $P_{th} = 0.5$ and $StD_{th} = 0.5$ when decomposing the same signal. (c) The search path of the SWD threshold optimization using GWO for the whole training samples. (d) Five optimizations of the SWD thresholds for the whole training samples. (e) The overall optimal decomposition results of SWD at $P_{th} = 0.382$ and $StD_{th} = 0.092$. (f) The individual optimal decomposition results of SWD at $P_{th} = 0.357$ and $StD_{th} = 0.124$. (g) The energy ratio distribution of OCs from a qualified sample and a defective sample.

Table 2: Optimization results of SWD thresholds for four types of samples.

| Type | $P_{th}$ | $StD_{th}$ |
|------|----------|------------|
| A    | 0.382    | 0.092      |
| B    | 0.435    | 0.361      |
| C    | 0.391    | 0.393      |
| D    | 0.324    | 0.196      |

The consumption of collecting all training sample signals of each type of samples. As aforementioned, the number of training samples in each type is 50-qualified and 50-defective samples, and the duration of each signal is 0.5 seconds, results in $T_a = 0.5 \times 100$. $T_g$ indicates the time consumption of the SWD threshold optimization using GWO. Since each iteration during the optimization needs to perform the signal decomposition with the corresponding thresholds, $T_d$ is the time consumed by both a series of the signal decomposition with SWD for all training samples and multiple threshold searches with GWO. $T_e$ is the sum of time costs regarding the feature extraction and the classifier training. On the contrary, $t_{max}$, $t_{min}$, and $t_{ave}$ are separately the maximum, minimum, and average time consumption for detecting a single testing sample. Any of them is made up of four time costs, including the signal acquisition with 0.5 seconds, the signal decomposition with the threshold-optimized SWD, the feature extraction, and the internal defect identification based on the trained classifier. Significantly, $t_{ave}$ representing the average performance can be regarded as the primary detection speed, whereas $T_{a1}$, $T_{a2}$, and $T_{e}$ occurring before the detection always belong to the time consumption generated by the offline processing. It can be seen from these running time results that $T_e$ is the longest time cost. It also means that the optimization process is always the most time-consuming, but the unified thresholds obtained by the proposed
method eliminate the need for the threshold optimization during the detection of each signal, which greatly reduces the time consumption. Although the time cost in different types of samples is quite distinct from each other, the maximum detection time for a single sample does not exceed 3.4 seconds. Therefore, the proposed detection algorithm is able to achieve an acceptable detection speed with a relatively short running time.

Besides, since the acoustic signal differences in Type D between qualified and defective samples are relatively similar, the extraction and identification of features are difficult to process, resulting in a lower recognition rate and more extended time consumption. More notably, the running time of the algorithm could be further shortened by either optimizing the algorithm code or improving the computer hardware configuration.

3.7. Evaluation and Validation

3.7.1. Influence of the Filtering Range on the Detection Performance. To confirm the rationality of the used filtering range for the high-pass filter, its influence on the detection performance, including the accuracy and speed, is obtained by adjusting the filtering range. The overall recognition rate of all testing samples in each type is regarded as the accuracy, while their time consumption from preprocessing the signal to identifying the features is defined as the speed. The filtering range is adjusted from 0 Hz to 6 kHz, and each filtering range corresponds to an exclusive accuracy and speed. It can be found in Figure 6(a) that the speed decreases obviously synchronously with the increase of the filtering range because the larger filtering range means fewer signal data to be processed. However, the decrease in the accuracy occurs after 4 kHz, and the highest accuracy is always stable within 4 kHz. According to these results, the filtering range exceeding 4 kHz is bound to lose some beneficial signal information for the extraction or identification of features, resulting in an obvious decrease in the accuracy. It also suggests that the acoustic signal characteristics representing the presence or absence of the internal defects are distributed within the relatively high-frequency range instead of the low-frequency part below 4 kHz. Therefore, for the sake of both the highest overall recognition rate and the lowest time consumption, the selected 4 kHz is the best filtering range for all samples in each type.

3.7.2. Impact of Training Sample Quantities on the Detection Accuracy. The training sample quantities have a primary impact on the identification performance of the RF classifier. Too many training samples will cause more calculations and longer training time, while too few training samples will lead to a lack of feature information. To assess the most appropriate number of training samples, seven training groups with different amounts of training samples are used for establishing an RF classifier for each type. Each group is equally divided between the qualified and defective samples. All training samples in each group are randomly collected from the same type. The previous testing samples are applied to obtain the corresponding detection accuracy. The accuracy results are utilized to verify the best number of training samples. It can be seen from Figure 6(b) that testing samples in Type B obtain the stable highest detection accuracy in any training group, but other types of samples get the highest accuracy only at the beginning of the training group made up of 50 qualified and 50 defective samples. In the training group where the number of training samples is lower than that of 50 qualified and 50 defective samples, the accuracy of Type B, C, and D increases with the number of samples. The accuracy of training groups exceeding this number tends to be stable, neither increasing nor decreasing. Therefore, the best setting for training sample quantities is the combination of 50 qualified and 50 defective samples.

3.7.3. Influence of the Threshold Optimization on the Detection Accuracy. To verify the importance and validity of the threshold optimization for SWD, all testing samples in each type are decomposed by the nonoptimized thresholds and then used for the internal defect detection. The non-optimized thresholds refer to \( P_{\text{th}} \) and \( \text{StD}_{\text{th}} \) that are available within the threshold range except the optimal threshold setting determined by the optimization. Figure 6(c) shows the detection accuracy of all testing samples under the optimized and nonoptimized thresholds. It can be clearly seen that the detection accuracy corresponding to the nonoptimized thresholds is always lower than that of optimized thresholds regardless of the sample type. It also means that only the optimal thresholds can make SWD produce the best signal decomposition, resulting in the highest detection accuracy. Furthermore, the accurate decomposition results would be quite helpful in exposing even relatively weak signal characteristics in detail. Thus, it is necessary to optimize the SWD thresholds properly, and adopting the proposed method helps to realize the optimized SWD.

3.7.4. Influence of Different Parameters on the Optimization Performance of GWO. As aforementioned, the number of search agents and the maximum number of iterations, namely, \( n_{\text{a}} \) and \( n_{\text{i}} \), respectively, can be regarded as the main controlling parameters of GWO. However, the optimization performance of GWO might be sensitive to the values of its main controlling parameters, whereas the execution of GWO in this work is based on fixed parameters. To prove the validity of the used parameters, different values for \( n_{\text{a}} \) and \( n_{\text{i}} \) are employed for implementing the threshold optimization for Type A. Then, the detection accuracy for all testing samples is applied to evaluate the influence of different parameter values on the optimization performance of GWO. As depicted in Figure 6(d), these two parameters have a relatively wide range of values for achieving the optimal accuracy. For instance, any position of the area consisting of \( n_{\text{a}} \in [10, 16] \) and \( n_{\text{i}} \in [10, 20] \) always corresponds to the best accuracy. It signifies that the performance of GWO in the SWD threshold optimization is not extremely sensitive to changes in the controlling parameter values. The fixed parameter values we used, namely, \( n_{\text{a}} = 10 \) and \( n_{\text{i}} = 10 \), are
located within the most suitable value range, and more importantly, they are the smallest value in the range. It also means that our parameter values are able to perform the optimization more efficiently compared with other values because small controlling parameter value in GWO usually corresponds to less calculation and low complexity.

| Type | RF  | SVM  | KNN   |
|------|-----|------|-------|
|      | $A_D$ | $A_N$ | $A_{DN}$ | $A_D$ | $A_N$ | $A_{DN}$ | $A_D$ | $A_N$ | $A_{DN}$ |
| A    | 70/70 | 70/70 | 100% | 70/70 | 62/70 | 94.29% | 70/70 | 68/70 | 98.57% |
| B    | 70/70 | 70/70 | 100% | 70/70 | 50/70 | 85.71% | 70/70 | 70/70 | 100% |
| C    | 70/70 | 70/70 | 100% | 70/70 | 62/70 | 94.29% | 70/70 | 70/70 | 100% |
| D    | 70/70 | 67/70 | 97.86% | 68/70 | 62/70 | 92.86% | 70/70 | 66/70 | 97.14% |

Table 4: The running time of the detection algorithm in units of seconds.

| Type | Preparation | Detection |
|------|-------------|-----------|
|      | $T_a$ | $T_o$ | $T_c$ | $t_{max}$ | $t_{min}$ | $t_{ave}$ |
| A    | 0.5 × 100 | 11615 | 1.389 | 2.472 | 2.441 | 2.457 |
| B    | 0.5 × 100 | 11597 | 1.362 | 2.541 | 2.513 | 2.525 |
| C    | 0.5 × 100 | 11611 | 1.396 | 2.251 | 2.218 | 2.235 |
| D    | 0.5 × 100 | 11609 | 1.401 | 3.396 | 3.367 | 3.389 |

**Figure 5:** Results of extracted features. (a) Type A. (b) Type B. (c) Type C. (d) Type D.
Therefore, the used controlling parameter values for GWO in this work are reasonable and effective. In addition, it is worth noting that the increase of $n$ and $n_1$ is generally beneficial to obtain higher accuracy, but the best one cannot be obtained when these two parameters get too large. The reason is that too large parameters complicate the decision-making mechanism, which is not conducive to the correct search.

3.7.5. Comparison of Different Threshold Optimization Methods. Although the necessity of the threshold optimization is incontrovertible, there are many ways to achieve the optimization. In relevant studies, some methods have been proven to have excellent optimization capabilities. To verify the validity of GWO in the threshold optimization, a total of other five optimization algorithms are applied to compare with GWO because they are widely used for many cases for determining the optimal value. These algorithms are all derived from metaheuristics similar to GWO and are composed of some typically popular and recent methods, including GA, PSO, ACO, WOA, and GOA. To make a comparison under the same conditions, they are arranged in the same numerical space and the same random initial location to search for the minimum of the same fitness function based on the identical training samples in Type A. After that, all these algorithms can find the same optimal thresholds as obtained by GWO, resulting in the same optimization accuracy. However, their optimization efficiency, such as iterations and their time cost, is completely different from that of GWO. As listed in Table 5, compared with the other five optimization algorithms, GWO obviously has the lowest number of iterations as well as the least time consumption. As aforementioned, the time consumption covers both a series of the signal decomposition with SWD for all training samples and multiple threshold searches with GWO. According to these results, the combination of GWO and the designed fitness function not only is suitable as a threshold optimization method for SWD but also provides a fast and accurate optimization result. In comparison to other metaheuristic optimization algorithms with the same fitness function, the metaheuristic strategy of GWO has relatively high efficiency and is further more suitable for the determination of the SWD threshold performance reflected by the acoustic signal characteristics of arc magnets. It can also confirm the rationality of choosing GWO in this work.

3.7.6. Potential Applications of the Proposed Method. This work focuses on achieving the acoustic-based detection for internal defects of arc magnets via the threshold-optimized SWD, and consequently, the proposed method is also likely applicable to other cases where the acoustic signal analysis is used to determine defects or faults. Not only that, it has the potential to be applied to decompose other signals with similar characteristics to acoustic signals. To evaluate such a possibility, the vibration signal of a rotating machine is imitated, and then, our method is performed to decompose the simulated vibration signal. In general, there are similarities between acoustic signals and vibration signals, but the vibration signals measured in industrial environments are subject to multiple components, various excitations, and low signal-to-noise ratio (SNR) [53]. To establish a signal model as realistic as possible, the simulated vibration signal consists of five frequently encountered components, which are described as the following equation [54]:

\[
\begin{align*}
    x(t) &= \sum_{i=1}^{M} A_i \cos(2\pi f_{i0} t + \alpha_i) + \sum_{i=1}^{N} B_i \cos(2\pi G f_{i0} t + \beta_i) \\
    &+ \sum_{i=1}^{P} C_i S_i(t - iT - \tau_i) + \sum_{i=1}^{Q} D_i O_i S_i(t - O_k) + n(t) .
\end{align*}
\]

(19)

For such a simulated vibration signal $x(t)$, the first term indicates the rotor vibration components, where $f_{i0}$ is the fundamental rotating frequency and $A_i$ and $\alpha_i$, respectively, stand for the amplitude and initial phase of the $i$th harmonic. The second term represents the normal gear meshing vibration components, where $G$ denotes the number of gear tooth, and $B_i$ and $\beta_i$ are, respectively, the amplitude and initial phase of the $i$th meshing harmonic. The third term is the impulse components caused by mechanical faults in a gear or a rolling element bearing, where $C_i$ specifies the amplitude of the $i$th impulse, $T$ is the time interval between two adjacent impulses, $\tau_i$ is a random variable with zero-mean, which simulates the roller slip effect and typically accounts for 1-2% of fault period [55], and $S_i(t)$ denotes the $i$th impulse with the unit amplitude, which can be expressed by an exponentially decaying sinusoid as [56]

\[
S_i(t) = e^{-\theta t} \sin(2\pi f_i t),
\]

(20)

where $f_{i0}$ denotes the resonance frequency excited by the $i$th impact impulse and $\theta_i$ is the decay rate of the corresponding impulse. The fourth term represents some random external shock components from occasional strikes on the machine casing, or electromagnetic interferences during the data acquisition, where two random variables, $D_i$ and $O_i$, indicate the amplitude and occurrence time of the occasional impulses, respectively. The last term is the white Gaussian noise components with the SNR of 0 dB.

The parameters for the simulated vibration signal are listed in Table 6. The waveform and frequency spectrum of such a signal and all its components are shown in Figure 7(a). It can be seen that the fault-related component is unable to be visualized from neither the time domain nor the frequency domain in the final vibration signal because of the noises and undesirable components. Therefore, the mechanical fault diagnosis using a vibration signal analysis is to eliminate these interferences, such that the fault features can be confidently recognized. For this purpose, the simulated signal is analyzed by our method, in which the optimal threshold setting is determined as $P_{th} = 0.215$ and $\text{StD}_{th} = 0.490$, and then, the threshold-optimized SWD is applied to provide the appropriate decomposition of the vibration signal. As illustrated in Figure 7(b), a total of four
OCs are obtained. As aforementioned, the last OC belongs to the residue after the decomposition, which is almost unavailable in the signal analysis. However, the frequency-domain spectra of the first three OCs exhibit some significant information that contributes to the fault diagnosis. In the first OC, the rotor vibration frequency $f_o$ and two gear meshing vibration frequencies $Gf_o$ and $2Gf_o$ exactly correspond to the frequencies of the first-three
Figure 7: The application of the vibrational signal decomposition using the proposed method. (a) The simulation for the vibrational signal of a rotating machine. (b) The decomposition results of the simulated vibrational signal.
maximum peaks in turn. In the second OC, the fault impulse frequency \( f_{r1} \) is precisely the frequency of the maximum peak. Similarly, the external shock frequency \( f_{r2} \) is completely equal to the frequency of the maximum peak in the third OC. As a result, the weak fault component that is difficult to be visualized under the severe interference can be conveniently and accurately identified by our method. Thus, it is obvious that the proposed method is also competent to analyze some vibration signals precisely, resulting in that our method is expected to be extended to more potential applications, not just the acoustic-based detection of arc magnet internal defects.

4. Conclusions

This paper developed an SWD algorithm with GWO for the acoustic signal analysis and applied it to rapidly detect the presence or absence of internal defects of arc magnets at low cost. The main contributions of this paper are outlined as follows:

1. The use of a high-pass filter allows the acoustic signal data to be effectively reduced, and the meaningful signal characteristics can be fully retained. It is conducive to weakening the complexity and workload of the signal decomposition.

2. With the correlation between oscillation components, a fitness function to reflect the association between the SWD threshold setting and the overall decomposition effect in similar signals can be established. The optimal threshold setting corresponds to the extremum of the function, and the GWO algorithm enables the fewer iterations to carry out a rapid, accurate search for the fitness function extremum. More importantly, GWO can highly avoid the blind search and escape the traps of local minima. Through such a fitness function and its extremum search process, the threshold optimizations that are supposed to perform multiple times are simplified to the one-time optimization.

3. The unified threshold setting determined by the optimization ensures the consistent decomposition for all signals, and each original signal is expressed by a set of OCs. In addition, the first two OCs with the highest energy ratio have the most obvious information linked to the internal defects. The information regarding the frequency of the maximum peak in these two OCs can be combined as the acoustic signal features associated with the internal defects of arc magnets.

4. Not only the identification difficulty but also the dependence of the classifier on a large number of training samples are significantly decreased owing to the advantages in concise features with clearly distinguishable clustering distributions. Furthermore, the simple structure of RF is suitable for the rapid feature classification.

In addition to contributions and advantages, the proposed method has some limitations that are worthy of attention. On the one hand, performing the proper optimization for the unified SWD threshold setting and training the correct feature classifier can hardly be done without a certain number of precisely known signals. On the other hand, the recognition rate and running time of the proposed method are not ultimate in the ideal detection performance. Consequently, we will endeavour to raise the accuracy and speed of the detection through enhancing the optimization efficiency and improving the feature extraction in the future work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant no. 61701330), Scientific Research Fund of Sichuan Provincial Education Department of China (Grant no. 18ZB0428), Open Fund for Artificial Intelligence Key Laboratory of Sichuan Province of China (Grant no. 2016RZJ01), Talent Introduction Project of Sichuan University of Science and Engineering (Grant no. 2016RCL29), and College Students’ Innovation and Entrepreneurship Training Program of Sichuan University of Science and Engineering (Grant no. cx2020156).

References

[1] Q. Huang, Y. Yin, and G. Yin, “Automatic classification of magnetic tiles internal defects based on acoustic resonance analysis,” *Mechanical Systems and Signal Processing*, vol. 60-61, pp. 45–58, 2015.

[2] L. Xie, M. Yin, Q. Huang et al., “Internal defect inspection in magnetic tile by using acoustic resonance technology,” *Journal of Sound and Vibration*, vol. 383, pp. 108–123, 2016.

[3] Q. Huang, L. Xie, G. Yin, M. Ran, X. Liu, and J. Zheng, “Acoustic signal analysis for detecting defects inside an arc magnet using a combination of variational mode decomposition and beetle antennae search,” *ISA Transactions*, vol. 102, pp. 347–364, 2020.

[4] A. Glowacz, “Fault diagnosis of single-phase induction motor based on acoustic signals,” *Mechanical Systems and Signal Processing*, vol. 117, pp. 65–80, 2019.

[5] X. Liu, D. Pei, G. Lodewijks, Z. Zhao, and J. Mei, “Acoustic signal based fault detection on belt conveyor idlers using machine learning,” *Advanced Powder Technology*, vol. 31, no. 7, pp. 2689–2698, 2020.

[6] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal
20 Shock and Vibration

[7] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *Journal of the Royal Society Interface*, vol. 2, no. 5, pp. 443–454, 2005.

[8] M. Santhosh, C. Venkaiyah, and D. M. Vinod Kumar, "Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction," *Energy Conversion and Management*, vol. 168, pp. 482–493, 2018.

[9] N. Zhao, Z. Mao, D. Wei et al., "Fault diagnosis of diesel engine valve clearance based on variational mode decomposition and random forest," *Applied Sciences*, vol. 10, no. 3, Article ID 1124, 2020.

[10] J. Ma, J. Wu, and X. Wang, "Incipient fault feature extraction of rolling bearings based on the MVMD and Teager energy operator," *ISA Transactions*, vol. 80, pp. 97–111, 2018.

[11] W. Wang and X Chen, "Multiscale modeling of fiber optic gyroscope temperature drift based on improved ensemble empirical mode decomposition," *Applied Optics*, vol. 57, no. 28, pp. 8443–8450, 2018.

[12] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, pp. 35–56, 2018.

[13] G. K. Apostolidis and L. J. Hadjileontiadis, "Swarm decomposition: a novel signal analysis using swarm intelligence," *Signal Processing*, vol. 132, pp. 40–50, 2017.

[14] S. Wan and B. Peng, "An integrated approach based on swarm decomposition, morphology envelope dispersion entropy, and random forest for multi-fault recognition of rolling bearing," *Entropy*, vol. 21, no. 4, Article ID 354, 2019.

[15] V. Baltatzis, K. M. Bintsi, G. K. Apostolidis, and L. J. Hadjileontiadis, "Bullying incidences identification within an immersive environment using HD EEG-based analysis: a swarm decomposition and deep learning approach," *Scientific Reports*, vol. 7, Article ID 17292, 2017.

[16] S. Alnuaimi, S. Jimaa, Y. Kimura et al., "Fetal cardiac timing events estimation from Doppler ultrasound signals using swarm decomposition," *Frontiers in Physiology*, vol. 10, Article ID 789, 2019.

[17] Y. Miao, M. Zhao, V. Makis, and J. Lin, "Optimal swarm decomposition with whale optimization algorithm for weak feature extraction from multicomponent modulation signal," *Mechanical Systems and Signal Processing*, vol. 122, pp. 673–691, 2019.

[18] B. Samanta and C. Nataraj, "Use of particle swarm optimization for machinery fault detection," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 2, pp. 308–316, 2009.

[19] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 87–103, 2016.

[20] R. L. Kadri and F. F. Boctor, "An efficient genetic algorithm to solve the resource-constrained project scheduling problem with transfer times: the single mode case," *European Journal of Operational Research*, vol. 265, no. 2, pp. 454–462, 2018.

[21] S. Das and P. N. Suganthan, "Differential evolution: a survey of the state-of-the-art," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011.

[22] Y. He, J. Zhou, and S. Y. Yuen, "Composing photomosaic images using clustering based evolutionary programming," *Multimedia Tools and Applications*, vol. 78, no. 18, pp. 25919–25936, 2019.

[23] C. Ong, J. Huang, and G. Tseng, "Building credit scoring models using genetic programming," *Expert Systems with Applications*, vol. 29, no. 1, pp. 41–47, 2005.

[24] D. Simon, "Biogeography-based optimization," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 702–713, 2008.

[25] A. Suppapitnarm, K. A. Seffen, G. T. Parks, and P. J. Clarkson, "A simulated annealing algorithm for multiobjective optimization," *Engineering Optimization*, vol. 33, no. 1, pp. 59–85, 2000.

[26] J. Ji, S. Gao, S. Wang, Y. Tang, H. Yu, and Y. Todo, "Self-adaptive gravitational search algorithm with a modified chaotic local search," *IEEE Access*, vol. 5, pp. 17881–17895, 2017.

[27] O. K. Erol and I. Eksin, "A new optimization method: big bang-big crunch," *Advances in Engineering Software*, vol. 37, no. 2, pp. 106–111, 2006.

[28] A. Kaveh and S. Talatahari, "Charged system search for optimal design of frame structures," *Applied Soft Computing*, vol. 12, no. 1, pp. 382–393, 2012.

[29] Y. Liu and P. Tian, "A multi-start central force optimization for global optimization," *Applied Soft Computing*, vol. 27, pp. 92–98, 2015.

[30] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: an overview," *Soft Computing*, vol. 22, no. 2, pp. 387–408, 2018.

[31] B. C. Mohan and R. Baskaran, "A survey: ant colony optimization based recent research and implementation on several engineering domain," *Expert Systems with Applications*, vol. 39, no. 4, pp. 4618–4627, 2012.

[32] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, "A comprehensive survey: artificial bee colony (ABC) algorithm and applications," *Artificial Intelligence Review*, vol. 42, no. 1, pp. 21–57, 2014.

[33] M. A. Al-Betar, M. A. Awadallah, H. Faris, X.-S. Yang, A. Tajudin Khader, and O. A. Alomari, "Bat-inspired algorithms with natural selection mechanisms for global optimization," *Neurocomputing*, vol. 273, pp. 448–465, 2018.

[34] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.

[35] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, "Grey wolf optimizer: a review of recent variants and applications," *Neural Computing and Applications*, vol. 30, no. 2, pp. 413–435, 2018.

[36] S. Z. Mirjalili, S. Mirjalili, S. Saremi, H. Faris, and I. Aljarah, "Grasshopper optimization algorithm for multi-objective optimization problems," *Applied Intelligence*, vol. 48, no. 4, pp. 805–820, 2018.

[37] L. Wang, Y. Xiong, S. Li, and Y.-R. Zeng, "New fruit fly optimization algorithm with joint search strategies for function optimization problems," *Knowledge-Based Systems*, vol. 176, pp. 77–96, 2019.

[38] X. Zhang, Q. Miao, Z. Liu, and Z. He, "An adaptive stochastic resonance method based on grey wolf optimizer algorithm and its application to machinery fault diagnosis," *ISA Transactions*, vol. 71, pp. 206–214, 2017.

[39] R. Sanjay, T. Jayabarathi, T. Raghunathan, V. Ramesh, and N. Mithulanantham, "Optimal allocation of distributed generation using hybrid grey wolf optimizer," *IEEE Access*, vol. 5, pp. 14807–14818, 2017.

[40] S. K. Verma, S. Yadav, and S. K. Nagar, "Optimization of fractional order PID controller using grey wolf optimizer," *Journal of Control, Automation and Electrical Systems*, vol. 28, no. 3, pp. 314–322, 2017.
[41] E. Dokur, "Swarm decomposition technique based hybrid model for very short-term solar PV power generation forecast," *Elektronika Ir Elektrotechnika*, vol. 26, no. 3, pp. 79–83, 2020.

[42] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

[43] A. Donyaii, A. Sarraf, and H. Ahmadi, "Water reservoir multiobjective optimal operation using grey wolf optimizer," *Shock and Vibration*, vol. 2020, Article ID 8870464, 10 pages, 2020.

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

[45] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.

[46] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.

[47] H. Zhou, Z. Deng, Y. Xia, and M. Fu, "A new sampling method in particle filter based on Pearson correlation coefficient," *Neurocomputing*, vol. 216, pp. 208–215, 2016.

[48] H. Li, J. Sun, H. Ma, Z. Tian, and Y. Li, "A novel method based upon modified composite spectrum and relative entropy for degradation feature extraction of hydraulic pump," *Mechanical Systems and Signal Processing*, vol. 114, pp. 399–412, 2019.

[49] B. Li, D.-S. Huang, C. Wang, and K.-H. Liu, "Feature extraction using constrained maximum variance mapping," *Pattern Recognition*, vol. 41, no. 11, pp. 3287–3294, 2008.

[50] M. Toğacar, B. Ergen, and Z. Comert, "Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 23–39, 2020.

[51] S. Pan, T. Han, A. C. C. Tan, and T. R. Lin, "Fault diagnosis system of induction motors based on multiscale entropy and support vector machine with mutual information algorithm," *Shock and Vibration*, vol. 2016, Article ID 5836717, 12 pages, 2016.

[52] X. Du, "Fault detection using bispectral features and one-class classifiers," *Journal of Process Control*, vol. 83, pp. 1–10, 2019.

[53] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, and Z. Zhu, "A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines," *Mechanical Systems and Signal Processing*, vol. 116, pp. 668–692, 2019.

[54] M. Zhao and X. Jia, "A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 94, pp. 129–147, 2017.

[55] Y. Miao, M. Zhao, and J. Lin, "Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition," *ISA Transactions*, vol. 84, pp. 82–95, 2019.

[56] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 25, pp. 485–520, 2011.