# Correlation-Guided Particle Swarm Optimization Approach for Feature Selection in Fault Diagnosis

Ke Chen [ID], *Member, IEEE*, Wenjie Wang [ID], *Member, IEEE*, Fangfang Zhang [ID], *Member, IEEE*, Jing Liang [ID], *Senior Member, IEEE*, and Kunjie Yu [ID], *Member, IEEE*

*Abstract*—A large number of features are involved in fault diagnosis, and it is challenging to identify important and relative features for fault classification. Feature selection selects suitable features from the fault dataset to determine the root cause of the fault. Particle swarm optimization (PSO) has shown promising results in performing feature selection due to its promising search effectiveness and ease of implementation. However, most PSO-based feature selection approaches for fault diagnosis do not adequately take domain-specific a priori knowledge into account. In this study, we propose a correlation-guided PSO feature selection approach for fault diagnosis that focuses on improving the initialisation effectiveness, individual exploration ability, and population diversity. To be more specific, an initialisation strategy based on feature correlation is designed to enhance the quality of the initial population, while a probability individual updating mechanism is proposed to improve the exploitation ability. In addition, a sample shrinkage strategy is developed to enhance the ability to jump out of local optimal. Results on four public fault diagnosis datasets show that the proposed approach can select smaller feature subsets to achieve higher classification accuracy than other state-of-the-art feature selection methods in most cases. Furthermore, the effectiveness of the proposed approach is also verified by examining real-world fault diagnosis problems.

*Index Terms*—Classification, correlation, fault diagnosis, feature selection, particle swarm optimization.

## I. INTRODUCTION

EQUIPMENT health management is the key to ensuring the safe, efficient, and reliable operation of mechanical equipment [1], [2]. Fault diagnosis is one of the important steps of equipment health management, which is crucial to

guarantee the availability and safety of mechanical systems to avoid the occurrence of dangerous situations [3], [4]. Fault diagnosis is essentially a classification task. Its purpose is to classify fault categories according to the information presented by their features. Methods for fault diagnosis can be divided into three classes based on the type of data and how the data is processed: model-based methods, knowledge-based methods, and signal-based methods [5]. The model-based approach determines the type of fault by the residuals of the model's results and the actual results. Knowledge-based methods lee to determine the failure mode based on the knowledge base. Signal-based methods determine the type of fault by analysing the information in various signals. Model-based methods are limited to specific classes of mechanical systems and faults [6]. Knowledge-based methods mainly rely on artificial intelligence methods to directly determine the system fault, which can only be detected within the scope of known knowledge and the structure of the method is more complex. However, the signal-based methods do not rely on precise models or detailed a priori knowledge of a particular system and can therefore be applied to a wide range of different types of systems. It makes direct use of the actual signals generated during the operation of the system, which provide the truest reflection of the current state of the system. Signal-based methods will be the focus of this work.

In signal-based fault diagnosis methods, fault features need to be extracted from multiple perspectives in order to better demonstrate the main characteristics of each fault [7]. Thus, fault features are gradually becoming more dimensional [8]. It is possible for feature selection (FS) to have $2^n$ subsets of features when $n$ represents the number of all available features. This makes exhaustive search techniques almost useless for identifying the root cause of fault problem. Furthermore, when it comes to this data, there is often a large amount of irrelevant and redundant information that can seriously affect fault diagnosis accuracy. Therefore, proper selection of features is the key to fault diagnosis.

FS is an important part of data preprocessing to select a subset of relevant and complementary fault features [9]. FS approaches can be categorized into three types, i.e., filter-based, wrapper-based, and embedded-based approaches [10], [11]. Filter-based approaches select features based on data characteristics without utilizing learning algorithms [12]. Wrapper-based approaches typically use learning algorithms to evaluate selected features until the maximum number of iterations or the maximum number of evaluations is reached

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

2

IEEE/CAA JOURNAL OF AUTOMATICA SINICA

[13]. Embedded-based approaches embed FS into the learning process [14]. The classification accuracy of wrapper-based and embedded-based approaches is generally higher than that of filter-based approaches [15]. With the same number of features, the wrapper-based approach has a shorter computation time than the embedded-based approach. Hence, the purpose of this study is to create a new wrapper-based FS method for fault diagnosis problems.

Evolutionary computation (EC) techniques have received significant attention due to their powerful search ability. Compared to other kinds of evolutionary algorithms, particle swarm optimization (PSO) has faster convergence, provides better exploration of the feature space, and has better fitness in the face of fault diagnosis problems. These significant advantages make PSO as an ideal choice in the feature selection phase of signal-based fault diagnosis methods [16]. For bearing fault diagnosis in [17], the Fisher criterion is used for determining the fitness function as well as the PSO approach based on maximum class separability. The use of F-score and Fisher discriminate analysis before using the PSO approach to eliminate redundant and unnecessary features can improve the effectiveness of fault diagnosis [18]. In [19], a FS approach based PSO was used for fault diagnosis of a wind power conversion system, which applied Euclidean distance to measure the dissimilarity between fault features. Nevertheless, the nonlinear relationships inherent in the fault diagnosis domain are not adequately considered by this approach. Hence, additional research is required to investigate the development of new fault diagnosis methods based on evolutionary computation.

For fault diagnosis problems, there is a non-linear relationship between features and faults under certain operating conditions. For example, in a mechanical system, the peak value of the vibration signal is relatively stable during normal operation. When an unbalance fault occurs, the peak increases, but this increase is not linear. As the degree of imbalance increases, the peak value may first increase slowly and then rise sharply at some stage. By exploiting the non-linear relationship between fault features in fault diagnosis problems, invalid searches can be avoided and feature subset quality can be improved [20]. There is a correlation between fault features and states in [21] that uses the Pearson correlation coefficient to determine the extent to which they are correlated. It utilizes features with high correlation as inputs for the neural network, which reduces the noise features in the fault dataset to a certain extent and shortens the training time for learning classifiers. Most existing strategies, however, work in the same manner as the methods described above, namely, by eliminating all noise directly from the problem search space. If features with low correlation are directly removed, it will have an impact on the search effectiveness of an algorithm. For example, multiple features with low correlation may also produce better results when combined together. At the same time, a similar problem exists in the initialisation phase. Features with low correlation cannot be directly removed altogether in the initialisation phase, and their potential value needs to be determined gradually during the exploration process. As a solution to the above problems, we propose an initialisation strategy, and a probability updating mechanism

using correlation information derived from the fault datasets. By using these strategies, the initial population can be improved in terms of quality, and more potential combinations of features can be identified. The initialisation strategy of feature grouping can better adapt to the characteristics of different types of features and improve the algorithm's search effect in different feature groups. Furthermore, by increasing the flip probability, the algorithm can try different combinations of features more aggressively, increasing the likelihood of finding a better subset of features.

Another challenge for fault diagnosis is the increasing dimensionality of fault features. In the case of a large and complicated search space, the approach will fall into a local optimum, which will limit search performance. When this happens, it generally goes one of two ways, i.e., "jumping out" and "restarting", i.e., searching in other directions based on the current solution, or ignoring the current solution and searching again in a new region. Based on this detection, a mutation operator is applied to the *gbest* value to improve performance by jumping out after a certain amount of time [22]. However, the mutation operator has a certain degree of randomness, and does not discover a more promising search direction within a limited number of iterations. In order for the population to find a more suitable search direction as soon as possible, changing the method of fitness evaluations of the particles is an option, where we yse a different training set. By pulling out or putting in a certain proportion of samples, new information can be introduced and the existing information structure can be changed, thus providing new search clues for the particles, changing the existing search balance, and giving the algorithm a chance to jump out of the local optimum.

In this paper, by employing PSO in the FS approach, an approach is proposed to solve the fault diagnosis problem in this paper. First, the correlation information between features and class labels is fully utilised with the initialisation strategy and probability updating mechanism, making the subsequent search more focused on promising features. Secondly, a sample shrinkage mechanism is proposed in the hope of helping the population to jump out of the local optimum by increasing the diversity of the population. A brief summary of this paper's main contributions is as follows:

1) A FS approach that takes into account the properties of datasets with complex relationships is proposed to improve the classification performance of fault diagnosis problems. Correlation is introduced to fully consider the nonlinear relationship between fault features and fault categories, and to find a more promising subset of features from the candidate features.

2) Based on the correlation information between features, a correlation-guided initialisation strategy and a probability updating mechanism are proposed to improve the quality of the initial population and offspring-generated particles, respectively. The proposed strategies can effectively guide the feature selection process to find more feature combinations with higher quality.

3) An effective sample shrinkage mechanism is designed, which adjusts the search direction of the individual by changing the size of the training set. This strategy can effectively

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

CHEN *et al.*: CORRELATION-GUIDED PARTICLE SWARM OPTIMIZATION APPROACH FOR FEATURE SELECTION IN FAULT DIAGNOSIS 3

overcome the defect of premature convergence and help increase the diversity of the population.

The remainder of this article is organized in the following way. Background and related work are presented in Section II. In Section III, the proposed approaches are described in more detail. Described in Section IV is the experimental design, while described in Section V are the results and discussion. We conduct the application process in Section VI. This article is concluded by Section VII, which concludes the article.

## II. BACKGROUND AND RELATED WORK

This section briefly describes fault diagnosis, binary particle swarm optimization, and the application of feature selection in fault diagnosis.

### A. Fault Diagnosis as a Classification Problem

A signal-based method is used for fault classification based on features extracted from original signals [23]. The original signals are typically vibration signals, voltage signals, current signals, etc. The features extracted in a certain fault state of a mechanical device can be expressed as a vector $[c, f_1, f_2, \ldots, f_D]$. There are a total of $D$ features that are failure related. A class label $c$ represents a fault state. The number of types of class labels indicates how many fault states there are. For the same mechanical system, there may be several different fault states [24]. A set of fault features $[f_1, f_2, \ldots, f_D]$ obtained after feature extraction of each acquired signal corresponds to a class label indicating that the signal is acquired in the corresponding fault state. In the feature selection phase, after the feature selection method selects a critical subset of features, the training data with class labels are input to the classification model for training. The classification model adjusts its own parameters by learning the relationship between features and fault modes [25]. When faced with new equipment operation data, the classification model is able to determine the failure modes of mechanical equipment based on the evaluation of the new data [26].

### B. Binary Particle Swarm Optimization Algorithm

It has been found that binary particle swarm optimization (BPSO) [27] uses particles to represent candidates for solutions to the problem. There is a previously best position for the ith particle represented as $pbest_i = [pbest_{i1}^t, pbest_{i2}^t, \ldots, pbest_{iD}^t]$, and so far the global best position is $gbes^t = [gbest_1^t, gbest_2^t, \ldots, gbest_D^t]$. Hence, the two best locations should be prioritized for further investigation since they will guide the swarm into a more productive area in the near future [28]. During the update process, the current position $X_i^t = [X_{i1}^t, X_{i2}^t, \ldots, X_{iD}^t]$ of the $i$-th particle consists of $D$ binary vectors, where each identifies whether or not a specific feature has been selected and the current position of the particle. Each particle has a moving velocity $V_i^t = [V_{i1}^t, V_{i2}^t, \ldots, V_{iD}^t]$. These two vectors describe the direction and velocity magnitude that the particle should move in the next update.

The optimization process of BPSO is actually the process in which the particles in the population update their position information and velocity information iteratively, thus finding optimal solutions step by step [29]. There are two types of velocity and position information updated for generation $t+1$ of the $i$-th particle, and the formula for updating them is shown in Eq. 1 and Eq. 2. Due to its principle of operation, PSO has multiple advantages and disadvantages. On the one hand, it is simple and easy to implement, and involves relatively few parameters, which lowers the threshold of use [30]. In addition, PSO converges faster and can approach the optimal solution quickly, and it does not require much memory and CPU, and occupies less resources. However, PSO also has many disadvantages. It has poor local search ability and insufficient search accuracy, which may lead to missing the global optimal solution. During the search process, it is easy to fall into local extreme points [31], which is due to the fact that particle diversity may disappear rapidly, which in turn generates the problem of premature convergence.

$$V_{id}^{t+1} = \omega * V_{id}^t + c_1 * r_1 * (pbest_{id}^t - X_{id}^t) + \\ c_2 * r_2 * (gbest_d^t - X_{id}^t) \tag{1}$$

$$x_d^{t+1} = \begin{cases} 1, & \text{if rand ( )} \leq s(v_d^{t+1}) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $t$ denotes the current iteration number, $d$ indicates the $d$-th dimension of the search space, $\omega$ indicates the inertia weight, and $c_1$ and $c_2$ denote acceleration constants. $r_1$ and $r_2$ are uniformly distributed values in the interval [0, 1]. $s(v_d^{t+1}) = 1/(1 + e^{-v_d^{t+1}})$ is called a transfer function.

### C. Feature Selection in Fault Diagnosis

Signal-based methods involve three major steps: extracting features, selecting features, and classifying faults [32]. As a result of the features extracted from the data, we have an extensive feature space that provides a great deal of information for fault diagnosis. However, classification in high-dimensional feature space is more complex and difficult than that in low-dimensional feature space. FS is therefore needed to reduce feature space dimensions and make fault classification more feasible, thereby improving fault diagnosis accuracy.

Over the past decades, there have been some studies of FS approaches in the field of fault diagnosis. First, we look at *correlation based FS approaches*. By considering the correlation between the fault features and fault classes, it has been possible to select dominant features by examining the Measured Resting Metabolic Rate [33]. Second, we look at *distance based FS approaches*. A new distance assessment approach based on standard deviation is proposed in [34]. The introduction of standard deviation can improve the approach to analyse the looseness within the dataset and compare the differences within the data. Third, we look at *clustering based FS approaches*. The approach in [35] groups features with similar distributions to obtain a subset of features that is low in redundancy and highly representative. The search efficiency of the above approaches decreases greatly when facing the complex nonlinear relationship between fault features and fault states. Moreover, fault diagnosis has uncertainty, and the fault features may be mixed with some environmental noise or outliers, which may affect the approach's direction of search

and reduce the final diagnostic accuracy. An efficient search technique is an important task for the FS problem.

At present, some intelligent fault diagnosis frameworks have already shown promising diagnostic performance in related fields [36]. As a general rule, the PSO has proven to be more productive at specifying the ideal subset of features compared to other approaches. As an example, there is effective methodology for detecting bearing faults through the integration of multiple approaches for extracting and selecting features [37]. It has been shown that BPSO can be improved by using a group initialisation strategy based on the feature weight, a new update mechanism, and an effective FS approach, which are all capable of enhancing the classification accuracy while reducing the size of the data set. As a means of improving fault diagnosis performance, in [38], the maximum information coefficient is first introduced as a correlation measure to obtain a rough feature subset, after which the improved PSO-based approach is used to narrow down the best set of features. Searching at a deeper level is conducted in the search area using particle reset and swarm initialisation strategies. In addition, there are some approaches that combine PSO with other evolutionary approaches. To identify more salient features from datasets, PSO and wheelbarrow differential evolution are used [39]. The ultimate realisation is to improve the accuracy of diagnosing the status of induction motors. With fault diagnosis, different fault classes and features may have different levels of correlation and importance. The above PSO approaches do not adequately consider the complex relationship that exist between fault classes and features, resulting in no deeper exploration of those potential feature subsets.

## III. PROPOSED APPROACH

A new PSO-based FS approach for fault diagnosis is described in this section. Specifically, we first describe the framework of the proposed correlation-guided BPSO (CGBPSO), and then detail the designed correlation-guided initialisation strategy, probability updating mechanism, and sample shrinkage mechanism.

### A. Framework of the Proposed Approach

The flowchart of the proposed CGBPSO method (Algorithm 1) is shown in Fig. 1.The CGBPSO algorithm consists of three main steps: initialisation, updating and evaluation. First, the correlation initialisation strategy in Fig. 1 corresponds to Line 1 in Algorithm 1. That is, the population is initialised according to Algorithm 2 such that features with different correlations are initialised according to their respective different probabilities. Secondly, to find more potential combinations of features with moderate correlations, the velocity and position of the particles are updated (Line 7 to Line 8) by Eq. 1 and Algorithm 3, where the probability of position update is redefined. Thus, the probability update mechanism in Fig. 1 makes it more (less) probable that more (less) relevant features are selected. Thirdly, it is necessary to conduct an evaluation of all of the particles once the updates of their position and velocity have been completed. In order to overcome the problem where the population tends to fall into local
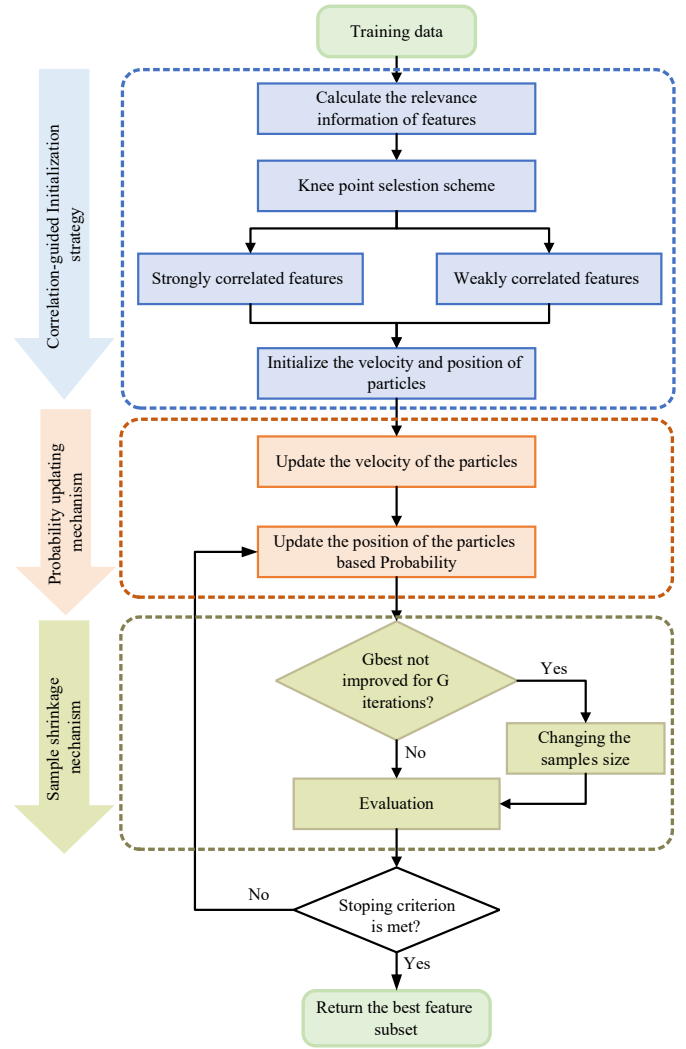


Fig. 1. Flow chart of CGBPSO.

optimum, the sample scaling mechanism in Fig. 1 is used, i.e., the size of the training set used to train the model is modified to change the search direction of the population. When it is found that the *gbest* of the population does not improve in consecutive generations, a portion of the samples is taken out or put in from the training set in order to jump out of the local optimum (Line 9 to Line 19). When the maximum number of iterations is reached, the loop ends and the optimum is returned as the final subset of features selected by the algorithm (Line 20).

---

**Algorithm 1: Overview of CGBPSO**

**Input:** The population size, $N$; Original features, $F = [f_1, f_2, \ldots, f_D]$; The maximum number of iterations, *MaxGen*; The number of iterations of *gbest* stopped updates, *NI*.

**Output:** Selected features.

1: **Initialisation:** Initialise population with $N$ particles using Algorithm 2.
2: Calculate fitness of *Popsize* particles.
3: Update $pbest_N^1 = [pbest_{N1}^1, pbest_{N2}^1, \ldots, pbest_{ND}^1]$ and $gbest^1 = [gbest_1^1, gbest_2^1, \ldots, gbest_D^1]$ based on the fitness values.
4: set NI ← 0.
5: set t ← 1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

CHEN *et al.*: CORRELATION-GUIDED PARTICLE SWARM OPTIMIZATION APPROACH FOR FEATURE SELECTION IN FAULT DIAGNOSIS                5

6:  **while:** $t \leftarrow MaxGen$ **do**
7:    **Update:** Calculate particle's velocity $V^t$ based on Eq. 1.
8:    Update particle's position $X^t$ using Algorithm 3.
9:    **Evolution:** Calculate fitness of particles using Algorithm 4.
10:   Update $pbest_N^t = [pbest_{N1}^t, pbest_{N2}^t, \ldots, pbest_{ND}^t]$ and $gbest^t = [gbest_1^t, gbest_2^t, \ldots, gbest_D^t]$ based on the fitness values.
11:   Curve(t) $\leftarrow gbest$.
12:   **if:** $t \geq 2$ **then**
13:     **if** $Curve(t) = Curve(t-1)$ **then**
14:       NI $\leftarrow$ NI + 1.
15:     **else**
16:       NI $\leftarrow$ 0.
17:     **end**
18:   **end**
19: **end**
20: **return** Position indexes in $gbest$ that have value 1.

---

**Algorithm 2: Correlation-guided Initialisation**

---

**Input:** Original features, $F = [f_1, f_2, \ldots, f_D]$; Population size, N; weight set, U.
**Output:** Initial population, $Pop$.
1:  Use ReliefF to calculate each feature's weight, and store the weights of the all features in the set $U = [cd_1, cd_2, \ldots, cd_D]$.
2:  After sorting all features in $F$ by weight, denote the feature set and weight set after sorting as follows: $F' = [f_1', f_2', \ldots, f_D']$ and $U' = [cd_1', cd_2', \ldots, cd_D']$.
3:  Find the knee point $M$ according to the index and weight of the features.
4:  Divide $F'$ into strongly correlated features subset $SF' = [f_1', f_2', \ldots, f_M']$ and weakly correlated features subset $WF' = [f_{M+1}', f_{M+2}', \ldots, f_D']$.
5:  The weights of the two feature subsets correspond to the sets $SU' = [cd_1', cd_2', \ldots, cd_M']$ and $WU' = [cd_{M+1}', cd_{M+2}', \ldots, cd_D']$, respectively.
6:  Calculate the initialisation probability $P$ according to Eq. 3.
7:  **for** $i = 1$ to $N$ **do**
8:    **for** $j = 1$ to $D$ **do**
9:      **if** $j \leq M$ **then**
10:       **if** $rand \leq P$ **then**
11:         $x_{ij} = 1$;
12:       **else**
13:         $x_{ij} = 0$;
14:       **end**
15:     **else**
16:       **if** $rand > P$ **then**
17:         $x_{ij} = 1$;
18:       **else**
19:         $x_{ij} = 0$;
20:       **end**
21:     **end**
22:     $Pop(i) = x_i$;
23:   **end**
24: **end**
25: **return** Initial population $Pop$.

---

**Algorithm 3: Probability Updating Mechanism**

---

**Input:** Velocity and position of the $i$-th particle in the $t$-th iteration, $V_i^t, X_i^t$; Dimension of features, $D$.
**Output:** Position of the $i$-th particle in the $t$+1-th iteration, $X_i^{t+1}$.
1:  **for** $j = 1$ to $D$ **do**
2:    Determine the velocity $V_{ij}^{t+1}$ by Eq. 1.
3:    Determine the flip probability $TF(j)$ by Eq. 4.
4:    **if** $rand \leq TF(j)$
5:      $X_{ij}^{t+1} = 1 - X_{ij}^t$;
6:    **else**
7:      $X_{ij}^{t+1} = X_{ij}^t$;
8:    **end**
9:  **end**
10: **return** Position of the $i$-th particle in the $t$+1-th iteration, $X_i^{t+1}$.

---

### B. Correlation-guided Initialisation Strategy

As part of the FS process, the correlation between features and class labels is examined. Features with different correlation degrees are chosen to be initialised separately. Algorithm 2 provides the steps to implement the strategy.

First, we calculate the significance of features. ReliefF [40] measures non-linear relationships between variables better than other approaches based on correlation [41]. Other feature weight-based methods rely on data obeying specific distribution assumptions, such as having a normal distribution. When the actual data does not conform to such assumptions, the performance of these methods may be severely affected. ReliefF does not rely on such strict data distribution assumptions, and it can handle datasets with a variety of distributions, including non-normal, skewed distributions, and data with complex mixed distributions. In the field of fault diagnosis, the relationship between fault characteristics and fault type may not be a simple linear relationship. In the case of non-linear data distributions, ReliefF can capture complex dependencies more accurately. It can be used to analyze the correlation between features and labels according to their weights. In Algorithm 2, ReliefF is used to evaluate the significance of original features $F$, and the weights are stored in the set $U$. Note that the larger the weights, the higher of the correlation of a feature to the class label. To facilitate the next step, the feature weights are normalised to [0, 1]. Next, the features are sorted in descending order according to their correlation, and the sorted feature set is denoted as $F'$ and the set of weights is re-denoted as $U'$.

Second, features are grouped together based on correlation. In order to make a more effective distinction between features with different levels of correlation, the weights threshold is set based on the data characteristics. The knee point [42] $M$ (rounded to an integer) is used to divide all the features into two groups (i.e., the strongly correlated feature subset $SF'$ and the weakly correlated feature subset $WF'$). The weights of the two feature subsets correspond to the sets $SU'$ and $WU'$ respectively. The use of the knee point enables the automatic determination of threshold weights and grouping based on the original features' intrinsic characteristics.

Third, the two sets of features are initialized separately. Strongly correlated features and weakly correlated features have different importance for classification tasks. To better preserve the population information, in the initialisation strat-

egy proposed in this study, the correlation of the feature subset is used to determine the probability of initialisation represented by $P$ (Eq. 3). In Algorithm 2 (Line 7 to Line 24), the strongly correlated feature subset $SF'$ and the weakly correlated feature subset $WF'$ are initialised according to the probability $P$, respectively. In the Fig. 2, a simple illustration of the proposed strategy can be given by taking the particle $x_i = (x_{i1}, x_{i1}, \ldots, x_{i10})$ as an example.

$$P = \frac{W_1}{W_1 + W_2} \tag{3}$$

where $W_1$ stands for the average value of $SU'$ and $W_2$ stands for the average value of $WU'$.
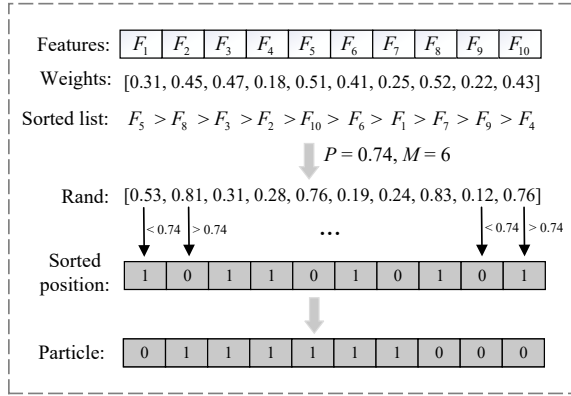


Fig. 2. Correlation-guided initialisation.

### C. Probability Updating Mechanism

The particle's position is updated by Eq.2 in standard BPSO. In the process of calculating the probability of a position updating taking a value of 1, the velocity is taken into account. To better describe the change of the particle, the position value of each dimension of the particle is determined by the flip probability $TF$ in this work.

During the search, there are some features with moderate correlations. They may be combined with features with Strong or weak correlations to produce more promising feature subsets. Therefore, in order to fully explore the features with moderate correlation, this section proposes a new transformation function to redefine the probability of position flipping (Eq. 4). Eq. 4 consists of two parts, where $\frac{1}{1+e^{-v_{ij}^{t+1}}}$ can map the velocity to [0, 1], and $\frac{1}{1+(cd_j-\theta)^v}$ can adjust the size of the flip probability using correlation of features. As $cd_j$ gets closer to 0.5 (the more moderate the correlation), the larger the value of $\frac{1}{1+(cd_j-\theta)^v}$. Therefore the value of $TF$ is also larger. The probability that the position vector corresponding to the particle is flipped is also larger. As the $cd_j$ gets further away from 0.5 (the stronger or weaker the correlation), the smaller the value of $\frac{1}{1+(cd_j-\theta)^v}$. Therefore the value of $TF$ is also smaller. The probability of the particle's corresponding position vector being flipped is also smaller. Within a certain range, the shape of the new transformation function satisfies the requirement that the middle output value is large, while the output values on both sides are small. Thus, more promis-

ing regions (feature space with moderate correlation) can be explored efficiently.

$$TF(j) = \lambda * \frac{1}{1+e^{-v_{ij}^{t+1}}} + (1-\lambda)\frac{1}{1+(cd_j-\theta)^v} \tag{4}$$

where $TF(j)$ is the flip probability of the $j$-th dimension feature, and its value range is [0, 1]. Set $\lambda$ to adjust the range of $TF$. $X_{ij}^{t+1}$, and $X_{ij}^{t}$ denote the position values of the $i$-th particle on the $j$-th dimension before and after flipping, respectively. After analysis, the $\theta$ value is taken as 0.5 and the $v$ value is taken as 2. Sensitivity analysis will be performed to determine the value of $\lambda$ in the experimental setting.

In Algorithm 3, the $i$-th particle is taken as an example. Firstly, the velocity $V_{ij}^{t+1}$ and flip probability $TF(j)$ of the particle in the $j$-th dimension needs to calculated as follows by Eq. 2 and Eq. 4. Then, if the $j$-th dimension feature satisfies the random event ($rand \leq TF$), set the new position $x_{ij}^{t+1}$ to $1-V_{ij}^{t}$. Otherwise, the new position is the same as the original position.

### D. Sample Shrinkage Mechanism

If the problem is too complex and the dimensions of the features are too high, PSO is susceptible to falling into local optima. This subsection describes the proposed sample shrinkage mechanism to overcome the deficiency.

There has been a substantial amount of research that has shown that changing the search direction is a very effective way of jumping out of the local optimum [43]. The particle's velocity determines its search direction during the evolutionary process. From Eq. 1, when the values of *gbest* and *pbest* are changed, the search direction of the population will also be changed. The fitness value of the model will differ depending on whether it has been trained with a different set of training data. Thus, it is possible to have an impact on the values of *gbest* and *pbest*. Each change of the fitness value indirectly indicates that the population adjusts the search direction many times until it finds a new search direction. Thus, this work proposes changing the size of the training set to find a better solution by assisting the population in changing the direction of the search.

As shown in Algorithm 4, the proposed sample shrinkage mechanism is explained in detail. Particle evaluation begins with the original training set $T$ at the beginning of the first iteration. Then, samples of size $\rho$ are taken from $T$ in proportion to the share of class labels to obtain a new training set $T'$ (Line 4). The number of iterations which *gbest* stops updating is represented by $NI$. During the evolutionary process, if $NI$ is detected to be greater than a set threshold $G$, it indicates that the population may be trapped in a local optimum. Then the training set $TS$ used in the previous iteration is detected to be consistent with $T$. If it is consistent, $T'$ is used as the *new training set* in this iteration. Otherwise, $T$ serves as the *new training set* (Line 5 to Line 11). When the value of $NI$ does not exceed $G$, $TS$ is still used to train the classification model (Line 12). Following that, classification error rates and ratios of selected features are combined in accordance with certain weights to obtain the final fitness value (Line 14 to Line 18).

Whenever the population falls into a local optimum, the training model from the previous iteration is no longer used. This operation of changing the size of training set allows the particles to search over a wider field.

---

**Algorithm 4: Sample Shrinkage Mechanism**

---

**Input:** All particles; the number of particles; $N$, Dimension of features, $D$; original training set; The training set from the previous iteration; The number of iterations which *gbest* stops updating, *NI*.

**Output:** The fitness of all particles *Fitness(i)* ($i$ = 1, 2,\ldots, $N$).

1:  **for** $i$ = 1 to $N$ **do**
2:    $TS \leftarrow$ the training set from the previous iteration.
3:    $T \leftarrow$ the original training set.
4:    $T' \leftarrow$ take a part out of T.
5:    **if** $NI \geq G$ **then**
6:      **if** $TS \equiv T$ **then**
7:        *new training set* $\leftarrow T'$.
8:      **else**
9:       *new training set* $\leftarrow T$.
10:     **end**
11:   **else**
12:     *new training set* $\leftarrow TS$.
13:   **end**
14:   Training the model with the *new training set* through five-fold cross-validation.
15:   *Error* $\leftarrow$ Calculate the classification error rate.
16:   Calculate the number of selected features in the $i$-th particle $S$.
17:   $Sf \leftarrow S / D$. Feature selection rate.
18:   *Fitness(i)* $\leftarrow$ 0.9 *Error+0.1 *Sf.
19: **end**
20: **return** The fitness of all particles *Fitness(i)* ($i$ = 1,2,\ldots,$N$).

---

### E. Algorithm Complexity

It can be solved by assuming the dataset has $D$ features, and the population has $N$ particles. CGBPSO's complexity is primarily determined by the correlation-guided initialisation strategy, probability updating mechanism and sample shrinkage mechanism. Correlation-guided initialisation strategy consists of computing the ReliefF weights of all features. The computational complexity of this system thus equals $O(D)$. The probability updating mechanism uses the ReliefF weights to compute the probability of the positional update with the probability updating mechanism. Hence, it performs $O(N * D)$ tasks in each generation. The sample Shrinkage Mechanism calculates the fitness value of all particles. Its computational complexity is $O(N)$. The overall complexity of CGBPSO is approximately $O(D+N*D+2N)$.

## IV. EXPERIMENT DESIGN

In the following section, the experimental setup is described, including the dataset composition, comparison approaches, parameter settings, and sensitivity analysis.

### A. Datasets

The following datasets are publicly available and have become a standard for testing different diagnostic approaches. The features of the original vibration signals are extracted in the time domain frequency domain and time frequency,

respectively. Each dataset has different characteristics, and the extracted features are related to its diagnostic objects and operating conditions. The following Table I shows the basic information about the datasets that are being used.

TABLE I
BASIC INFORMATION OF THE DATASETS

| Datasets | #Features | #Instances | #Classes | %Smallest Classes | %Largest Classes |
|---|---|---|---|---|---|
| WHU | 35 | 180 | 4 | 25 | 25 |
| CWRU | 40 | 1687 | 6 | 29 | 14 |
| JNU | 40 | 1020 | 4 | 29 | 24 |
| XJTU | 48 | 1363 | 3 | 39 | 25 |

*1) CWRU:* This comes from the Case Western Reserve University Bearing Data Centre [44].

*2) XJTU:* This comes from the Institute of Design Science and Basic Component at Xi'an Jiaotong University, Shaanxi. China and the Changxing Sumyoung Technology Co., Ltd. (SY), Zhejiang. China [45].

*3) WHU:* This is provided by Wuhan University [46].

*4) JNU:* This is provided by Jiangnan University [47].

### B. Comparative Approaches and Parameter Settings

To test the CGBPSO's capability in a variety of scenarios, we have selected the following state-of-the-art FS approaches as comparison approach:

● Bare bones PSO with Adaptive Chaotic Jump Strategy based Bi-Directional FS (BBPSO) [48],

● Hybrid FS based on Correlation-guided Clustering and PSO (HFS-C-P) [49],

● Surrogate-Assisted PSO based FS with Correlation-Guided Updating Strategy (CUS-SPSO) [50],

● Dynamic Sticky binary PSO based FS (SBPSO-D) [51],

● PSO-based FS with Variable-Length Strategy (VLPSO) [52].

● A Length-Adaptive Genetic Algorithm With Markov Blanket (LAGAM) [53]

● Multiple filtering algorithm with a competitive swarm optimizer (MFCSO) [54]

The maximum number of iterations (max_iter) and population size for the above five approaches are both 100. In the following Table II, you will find a summary of the main parameters that were set in these five comparison approaches. All other parameters of these comparison approaches are set in accordance with the recommendations of the relevant literature. Each approach is run 30 times independently. 20% of the samples in every dataset that are used as test samples and 80% of the samples are used as training samples based on their proportion in each category. When training is complete, a five-fold cross-validation approach is used to evaluate the classification error rate. Based on its performance and clarity, the k-nearest neighbour method is used in this approach, where $k$ is 5. To assess the test accuracy, the subset of evolved features is evaluated on the test set after training.

### C. Sensitivity Analysis of Parameters

CGBPSO has three key parameters: 1) sample size taken

TABLE II
PARAMETER SETTING

| Approaches | Parameter values |
|---|---|
| BBPSO | Number of features per domain $T = 3$, Number of iterations to compute the lifted average fitness value $W = 10$. |
| HFS-C-P | $c_1 = c_2 = 0.5$, selected threshold $\lambda = 0.5$. |
| CUS-SPSO | $c_1 = c_2 = 1.49445$, $\omega = 0.9 - 0.5*(\text{iter/MaxIter})$, the number of runs for the correlation-guided probability strategy $n_c = 2$, two constants (A and B) used to determine $A = 0.15$, $B = 0.05$. |
| SBPSO-D | The fix number of steps $ustkS^L = 1*\text{MaxIter}/100$, and $ustkS^U = 10*\text{MaxIter}/100$, $i_s^L = 0/n$, and $i_s^U = 10/n$, the ratio $\alpha = 2$. |
| VLPSO | $c = 1.49445$, $w = 0.9\text{-}0.5*(\text{iter/Maxiter})$, the max iterations for renew exemplars $\alpha = 7$, the number of divisions $\eta = 12$, $\theta = 9$. |
| LAGAM | The population size $p = |F|/20$, the rates of crossover $p_c = 0.8$, mutation $p_m = 0.2$, length changing $p_l = 0.3$, length changing oper-ator $\alpha = 0.1$, $\beta = 0.4$. |
| MFCSO | Random numbers $r_1, r_2, r_3 \in [0, 1]$, weight value of task 1, 2, 3 and 4, $g_1 = 0.1$, $g_2 = g_3 = g_4 = 0.45$, transfer probability $p_{trans} = 0.5$. |

from the training set ($\rho$). 2) the weight of the transfer function to calculate the selected threshold ($\lambda$). 3) the threshold ($G$) which *gbest* stops updating.

In addition to determining the size of the scaled sample set, $\rho$ also determines the direction of the population search. This section sets the value of $\rho$ is to {1%, 5%, 10%, 15%, 20%, 25%, 30%}. Table III illustrates the average classification accuracy for different $\rho$ values for the FS problems. It is remarkably clear that classification accuracy increases continuously with an increase in $\rho$ value, but when $\rho$ value reaches a certain value, increasing $\rho$ value cannot further improve the classification accuracy, and even a decrease in accuracy occurs. From the experimental results, the good value of $\rho$ is 10%, since the best classification results are achieved on all the datasets. Therefore, the ratio $\rho$ of the scaling factor to all samples in the CGBPSO is set to 10%.

TABLE III
THE RESULTS OF AVERAGE CLASSIFICATION ACCURACY

| $\rho$ values | 1% | 5% | 10% | 15% | 20% | 25% | 30% |
|---|---|---|---|---|---|---|---|
| WHU | 99.91 | 100 | **100** | 100 | 100 | 100 | 99.81 |
| JNU | 98.13 | 98.13 | **98.21** | 98.11 | 98.20 | 98.05 | 97.90 |
| CWRU | 88.07 | 88.12 | **88.70** | 87.79 | 86.91 | 86.79 | 87.27 |
| XJTU | 97.63 | 97.58 | **98.06** | 97.45 | 97.17 | 97.25 | 97.14 |

The size of $\lambda$ determines the impact of the feature weight value on the updated position value. $G$ is an important parameter that determines whether to resize the training set. For the two parameters $\lambda$ and $G$ in the experiment, 11 different values of $\lambda$ are set to {0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99}, and 5 different values of $G$ are set to {5, 6, 7, 8, 9}. The CWRU dataset is chosen for the experiments. This is because the medium number of features in the CWRU dataset provides a more moderate scenario for parameter validation. Meanwhile, rich samples and multiple types of labels can cover more fault types and change scenarios, thus testing the adaptability and robustness of the parameters in a more comprehensive way. Finally, 55 different combinations are obtained. Based on data from the CWRU dataset, Fig. 3 illustrates the average training accuracy of each combination on the CWRU dataset. Based on Fig. 3, it can be seen that the combination of the parameters {$\lambda = 0.9$, $G = 8$} yields the best performance. We have the same observations on other datasets as well. This lead to the CGBPSO model being able
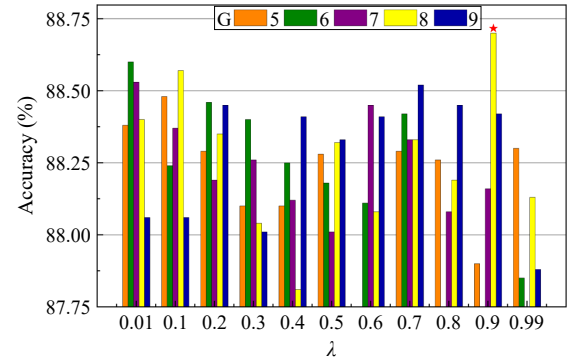


Fig. 3. Average results of 55 combinations of two parameters (calculate the weight of the transfer function and the threshold which *gbest* stops updating).

to use 0.9 and 8 as the values of both parameters $\lambda$ and $G$, respectively.

*D. Effect of CGBPSO With Different Classifiers*

In the proposed FS approach, KNN is used as a classifier to evaluate the performance of a selected subset of features. In order to investigate the effect of different classifiers in the proposed method, we use decision tree (dt) and naive bayesian (nb) instead of KNN in the proposed method, which are referred to as CGBPSO-dt and CGBPSO-nb, respectively.The parameter settings of CGBPSO-dt and CGBPSO-nb are the same as those of CGBPSO. Table IV shows the results of CGBPSO-dt, CGBPSO-nb and CGBPSO in terms of the number of feature selections (AvgNF), running time (Time), average classification accuracy (AvgAcc) and the corresponding standard deviation (Std).

From the Table IV, it can be seen that CGBPSO has higher accuracy values than CGBPSO-dt and CGBPSO-nb on the three datasets. The highest classification accuracy is achieved by CGBPSO-dt on the JNU dataset. This suggests that the selection of suitable classifiers can further contribute to the higher accuracy of the proposed FS approach. In terms of the number of features selected, CGBPSO selects the least number of features on the WHU and JNU datasets. In terms of training time, CGBPSO is much shorter than CGBPSO-dt and CGBPSO-nb. The shorter runtime reduces the cost of the fault diagnosis method. In summary, the proposed FS approach with KNN can better evolve a subset of features with higher classification.

TABLE IV
COMPARED RESULTS OF THE PROPOSED CGBPSO
WITH DIFFERENT CLASSIFIERS

| Datasets | Approaches | Time | AvgNF | AvgAcc±Std |
|---|---|---|---|---|
| | CGBPSO-dt | 282.05 | 6.03 | 98.15±2.11(+) |
| WHU | CGBPSO-nb | 383.41 | 5.60 | 98.80±2.02(+) |
| | CGBPSO | 13.96 | 4.27 | **100.00±0.00** |
| | CGBPSO-dt | 402.44 | 7.50 | **98.92±0.86(≈)** |
| JNU | CGBPSO-nb | 806.22 | 7.00 | 98.92±1.01(≈) |
| | CGBPSO | 73.14 | 6.90 | 98.21±1.06 |
| | CGBPSO-dt | 752.66 | 10.90 | 88.27±1.65(≈) |
| CWRU | CGBPSO-nb | 671.38 | 11.57 | 88.59±1.63(≈) |
| | CGBPSO | 220.34 | 11.7 | **88.70±1.54** |
| | CGBPSO-dt | 433.82 | 9.40 | 97.80±0.88(≈) |
| XJTU | CGBPSO-nb | 792.99 | 8.77 | 95.12±2.45(+) |
| | CGBPSO | 127.00 | 9.87 | **98.06±0.77** |

## V. RESULTS AND DISCUSSIONS

This subsection conducts several experiments to verify the effectiveness and efficiency of CGBPSO. Two metrics are used in this study, namely classification accuracy and feature subset size.

### A. The performance of CGBPSO

All six approaches are illustrated in Table V, where the average results are in bold, based on 30 independent classifications of the four datasets. We evaluate the performance of 30 approaches using the Wilcoxon rank-sum test at 0.05. The results indicated that, as indicated by the percentages of "+", "−", and "≈", the proposed CGBPSO is significantly superior, inferior, and similar to the compared approaches, respectively. Friedman's test determines the comprehensive performance of the proposed approach. As shown in the results below, the word "Ranking" is used as a way to indicate the order in which they are approached. The first thing to notice is that the CGBPSO approach has better average classification accuracy across all the datasets than any of the comparison approaches, as presented in Table V below. CGBPSO achieved a classification accuracy of 100% in all 30 independent runs on the WHU dataset, and 90.24% for LAGAM, with CGBPSO showing a significant improvement over LAGAM by 9.86%. CGBPSO achieves the greatest improvement on the JNU dataset compared to the BBPSO approach. The average classification accuracies of CGBPSO and BBPSO are 98.21% and 85.24%, respectively, which is an improvement of 12.97%. As determined by the Friedman test, CGBPSO performed the best, followed by HFS-C-P, SBPSO-D, VLPSO, CUS-SPSO, BBPSO, LAGAM and MFCSO. It is possible that the lower accuracy values obtained for the CWRU dataset are due to the fact that simply extracting features in the time domain, frequency domain, and time-frequency domain may not accurately capture the features in the CWRU dataset that are most relevant to fault diagnosis. A significance test revealed that the CGBPSO approach is significantly more accurate than the other approaches compared in most cases. CGBPSO based on feature correlation has been shown to be effective in this

study.

A comparison of the number of selected features between CGBPSO and the other five approaches is presented in Table VI. Based on all of the datasets, the CGBPSO reaches fewer features than the CUS-SPSO, SBPSO-D and LAGAM. JNU has fewer features selected by CGBPSO than CUS-SPSO, and its classification accuracy is higher. Similarly, VLPSO and MFCSO select fewer featuresthan CGBPSO. During the evolution of VLPSO, the length variation mechanism cuts down the number of selected features by reducing the variation in length. There may, however, be a negative impact on the classification performance of VLPSO if features are quickly removed. As demonstrated in Table VI, CGBPSO achieves much higher average classification accuracy than VLPSO over all datasets. MFCSO uses multiple filtering methods to generate multiple tasks. This strict filtering mechanism results in the direct removal of a large number of features. However, the classification accuracy of CGBPSO is higher than that of MFCSO on all datasets. Overall, although CGBPSO does not achieve the smallest feature set, it is able to achieve a good balance between improving the classification process as well as reducing the number of features.

### B. Analysis on the Three Strategies

This subsection investigates the contributions of the designed strategies in the CGBPSO approach.

*1) Correlation-guided Initialisation Strategy:* The randomly initialised CGBPSO approach is chosen as the comparison approach, denoted as CGBPSO-W. To analyze the effect of the strategy, Fig. 4 provides the training curves of two strategies CGBPSO and CGBOPSO-W. On the majority of datasets, CGBPSO has the highest fitness. CGBPSO usually generates better solutions than CGBPSO-W at the beginning. Since CGBPSO-W is primarily focused on exploration, instead of focusing on specific regions, it finds more promising deposits than CGBPSO-W. The fitness value of CGBPSO can improve over CGBPSO-W once it focuses more on exploitation.

*2) Probability Updating Mechanism:* By comparing CGBPSO using the proposed strategy with CGBPSO-C using the random update strategy, this experiment evaluates the effectiveness of the probability updating mechanism in the CGBPSO. The Table VII records the average number of selected features (AvgNF), average classification accuracy (AvgAcc), and standard deviation (Std) of AvgAcc. The proposed CGBPSO approach achieves a higher average classification accuracy on all the datasets in comparison with the existing CGBPSO approach, which can be seen in Table VII. Based on this evidence, it is clear that the proposed update strategy effectively uses the correlation information of features to create more combinations of features that have moderate correlation, thus effectively enhancing the efficiency of the search process for particles.

*3) Sample Shrinkage Mechanism:* We compare the results using the proposed mechanism (CGBPSO) with those without (CGBPSO-S). As shown in Table VIII, CGBPSO obtains higher classification accuracy than CGBPSO-S on three out of four datasets, including WHU, JUN and XJTU.On the three

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

10                                                                                                                                          IEEE/CAA JOURNAL OF AUTOMATICA SINICA

TABLE V
CLASSIFICATION ACCURACIES OF THE EIGHT APPROACHES

| Datasets | BBPSO | HFS-C-P | CUS-SPSO | SBPSO-D | VLPSO | LAGAM | MFCSO | CGBPSO |
|---|---|---|---|---|---|---|---|---|
| WHU | 92.50±5.23(+) | **100.00±0.00(≈)** | 98.10±4.62(+) | **100.00±0.00(≈)** | 95.12±5.68(+) | 90.24±3.86(+) | 93.56±5.09(+) | **100.00±0.00** |
| JNU | 88.36±5.81(+) | 98.11±2.88(≈) | 92.02±6.18(+) | 91.40±6.04(+) | 97.52±1.017(+) | 96.01±4.05(+) | 85.24±9.81(+) | **98.21±1.06** |
| CWRU | 88.07±0.99(+) | 83.61±2.02(+) | 88.25±1.22(+) | 88.43±0.98(+) | 83.79±10.50(+) | 87.81±1.27(+) | 84.58±1.55(+) | **88.70±1.54** |
| XJTU | 97.73±1.17(+) | 98.02±0.89(≈) | 96.48±1.38(+) | 97.38±1.17(≈) | 98.03±0.86(≈) | 95.22±1.42(+) | 97.13±1.79(+) | **98.06±0.77** |
| +\≈\− | 4\0\0 | 1\3\0 | 4\0\0 | 2\2\0 | 3\1\0 | 4\0\0 | 4\0\0 | |
| Ranking | 6 | 2 | 5 | 3 | 4 | 7 | 8 | 1 |

TABLE VI
NUMBER OF SELECTED FEATURES OF THE EIGHT APPROACHES

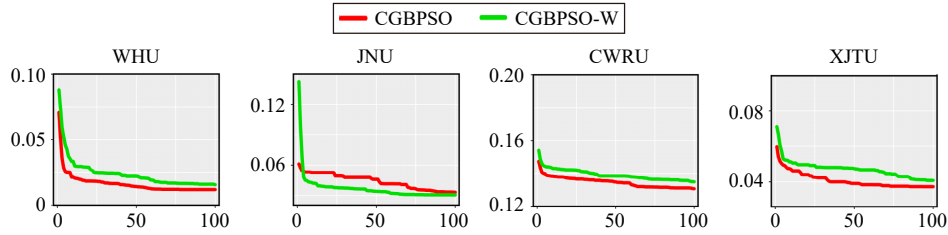| Datasets | BBPSO | HFS-C-P | CUS-SPSO | SBPSO-D | VLPSO | LAGAM | MFCSO | CGBPSO |
|---|---|---|---|---|---|---|---|---|
| WHU | 2.97 | 2.00 | 9.33 | 6.23 | 2.53 | 23.63 | **1.63** | 4.27 |
| JNU | 9.57 | 4.30 | 16.43 | 13.27 | 4.83 | 18.17 | **3.30** | 6.90 |
| CWRU | 6.87 | 13.03 | 14.40 | 11.80 | **6.37** | 29.47 | 6.47 | 11.70 |
| XJTU | 5.20 | 5.00 | 14.33 | 12.50 | **4.87** | 31.33 | 4.98 | 9.87 |



Fig. 4.    Convergence curves of CGBPSO and CGBPSO-W.

TABLE VII
COMPARED RESULTS OF CGBPSO AND CGBPSO-C

| Datasets | Approaches | AvgNF | AvgAcc±Std |
|---|---|---|---|
| WHU | CGBPSO-C | 1.47 | 99.58±2.28(≈) |
| | CGBPSO | 4.27 | **100.00±0.00** |
| JNU | CGBPSO-C | 4.67 | 98.04±1.99(≈) |
| | CGBPSO | 6.67 | **98.21±1.06** |
| CWRU | CGBPSO-C | 7.00 | 86.84±1.82(+) |
| | CGBPSO | 11.7 | **88.70±1.54** |
| XJTU | CGBPSO-C | 5.24 | 97.90±1.04(≈) |
| | CGBPSO | 9.17 | **98.06±0.77** |

TABLE VIII
COMPARED RESULTS OF CGBPSO AND CGBPSO-S

| Datasets | Approaches | AvgNF | AvgAcc±Std |
|---|---|---|---|
| WHU | CGBPSO-S | 5.57 | 99.91±0.51(≈) |
| | CGBPSO | 4.27 | **100.00±0.00** |
| JNU | CGBPSO-S | 7.03 | 98.14±1.36(≈) |
| | CGBPSO | 6.67 | **98.21±1.06** |
| CWRU | CGBPSO-S | 10.13 | **88.10±1.36(≈)** |
| | CGBPSO | 11.7 | 87.70±1.54 |
| XJTU | CGBPSO-S | 9.37 | 97.59±1.26(+) |
| | CGBPSO | 9.17 | **98.06±0.77** |

datasets, compared with CGBPSO-S, the average classification accuracy of CGBPSO is increased by 0.09%, 0.07% and 0.47% respectively, and the feature selection number is reduced by 1.3, 0.36 and 0.2 respectively.

## VI.    APPLICATION

As a means of verifying the effectiveness of the proposed fault diagnosis approach, this section applies the proposed CGBPSO approach to a real electromechanical system.

### A.  Problem Description

The experiments carried out in this section are carried out using the experimental platform for electromechanical system fault diagnosis, known as PT500Mini. This experimental plat-

form can be seen in Fig. 5. This test bench consists of a three-phase asynchronous motor, a frequency converter, a bearing, a gearbox for driving the motor and a powder magnetic brake for stopping it. The shaft is driven by a motor that runs at a certain speed and uses a powder magnetic brake to simulate the load. Vibration signals under five kinds of states will be collected using the experimental equipment through the acceleration sensors (magnetic suction and screw) placed on the bearing upper end.

There is a sampling frequency of 10 kHz and a rotation speed of 1200 revolutions per minute. A total of 1178 samples are collected. The bearing type used in the experiment is a deep groove ball bearing (deep groove ball bearing) SKF6205, which has an inner diameter of 25 mm and an outer

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

CHEN *et al*.: CORRELATION-GUIDED PARTICLE SWARM OPTIMIZATION APPROACH FOR FEATURE SELECTION IN FAULT DIAGNOSIS 11
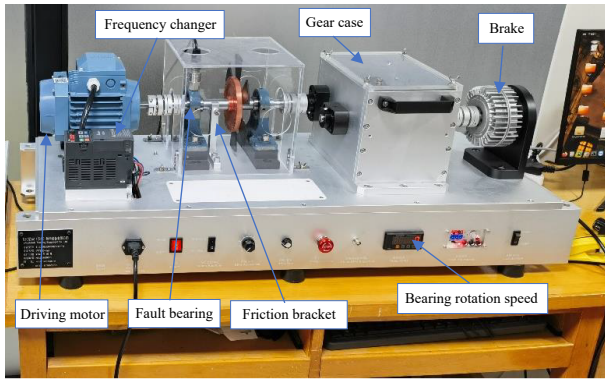


Fig. 5.    Electromechanical system fault diagnosis experimental platform.

diameter of 52 mm, as well as a width of 15 mm and a weight of 0.13 kg. The bearing can be in five states, which are as follows: bearing inner ring fault, bearing outer ring fault, bearing comprehensive fault, rolling element fault, and normal condition. Fig. 6 shows four of these fault states. With the collected time series signals, a 27-dimensional feature dataset (namely SDU) is constructed.
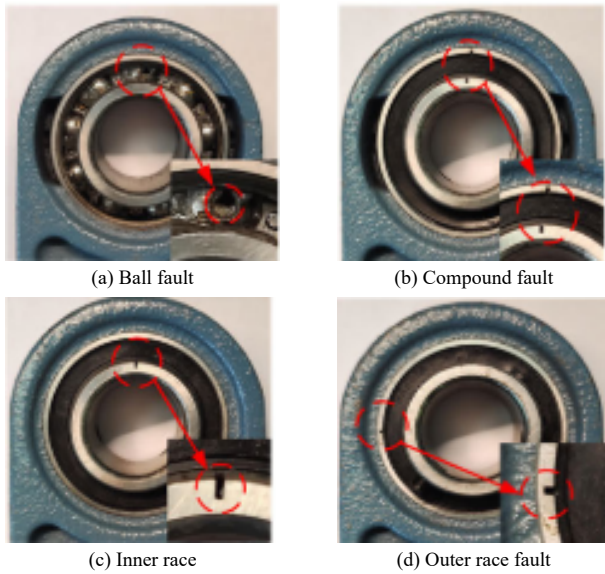


(a) Ball fault                    (b) Compound fault

(c) Inner race                    (d) Outer race fault

Fig. 6.    The different fault bearings.

### B.  Result and Analysis

Table IX provides experimental results on the SDU dataset. On the basis of Table IX, we are able to see that the CGBPSO approach takes a shorter amount of time to run than the rest of the approaches that have been compared. A significant increase in classifier accuracy can be attributed to the CGBPSO approach when it comes to the detection of significance as compared to all other approaches that were compared.

By using the proposed CGBPSO approach, 14.8% of the original number of features are selected, which greatly reduces the original dataset's size. The proposed CGBPSO approach selects no fewer features than the HFS-C-P approach. As a result of the fast partial feature removal, HFS-C-P has the smallest subset of features. In conclusion, it can

TABLE IX
AVERAGE TEST RESULTS BY THE SIX APPROACHES ON
SDU DATASETS

| Datasets | Approaches | AvgAcc±Std | AvgNF | Time |
|---|---|---|---|---|
| SDU | BBPSO | 91.02±1.27(+) | 5.00 | 112.59 |
| | HFS-C-P | 91.78±1.40(+) | **2.80** | 227.11 |
| | CUS-SPSO | 90.82±1.42(+) | 7.63 | 86.33 |
| | SBPSO-D | 91.96±1.51(+) | 5.33 | **76.92** |
| | VLPSO | 85.53±17.66(+) | 5.43 | 174.36 |
| | CGBPSO | **92.80±1.68** | 4.00 | 85.03 |

be seen that the CGBPSO only needs to use a shorter running time to obtain a subset of features with better discriminative ability in fault diagnosis.

### VII.  CONCLUSIONS

This paper presented a PSO-based feature selection approach for fault diagnosis, called CGBPSO. To compensate for particle oscillations, the initialisation and updating phases relied on correlation information of the features. By changing the size of the training set, more productive regions were explored by adjusting the search direction of populations that fall into local optimums.

On all datasets, the proposed CGBPSO approach had higher classification accuracy than the state-of-the-art methods and was able to evolve a small subset of features in a shorter period of time. This was due to the effectiveness of correlation information in the feature selection process. The results demonstrated that the correlation-guided initialisation strategy and probability updating mechanism could effectively improve population convergence and search abilities while continuously improving their quality. Moreover, the proposed mechanism could successfully maintain population diversity during search operations. Based on experimental results on real-world fault diagnosis problems, we found that the CGBPSO approach could evolve a feature subset with a classification accuracy of 92.8%, and the size of feature subset was only 14.8% of the original dataset. In summary, CGBPSO could be successfully applied to PSO-based fault diagnosis approaches to achieve a good fault feature subset.

In our future work, we will study various types of feature extraction algorithms in depth, compare and analyse their advantages and shortcomings, and construct a more complete and efficient feature extraction system to extract the key information that has been missed. This will enable us to further explore the performance of the proposed method on more complex fault diagnosis problems.

### REFERENCES

[1] X. Chen, X. Li, S. Yu, Y. Lei, N. Li, and B. Yang, "Dynamic vision enabled contactless cross-domain machine fault diagnosis with neuromorphic computing," *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 3, pp. 788–790, 2024.

[2] J. Xu, H. Kong, K. Li, and X. Ding, "Generative zero-shot compound fault diagnosis based on semantic alignment," *IEEE Trans. Instrumentation and Measurement*, vol. 73, pp. 1–13, 2024.

[3] C. Sun, H. Yin, Y. Li, and Y. Chai, "A novel rolling bearing vibration impulsive signals detection approach based on dictionary learning,"

*IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 6, pp. 1188–1198, 2021.

[4] X. Yang, T. Ye, X. Yuan, W. Zhu, X. Mei, and F. Zhou, "A novel data augmentation method based on denoising diffusion probabilistic model for fault diagnosis under imbalanced data," *IEEE Trans. Industrial Informatics*, 2024.

[5] Z. Xu, Y. Ma, Z. Pan, and X. Zheng, "Deep spiking residual shrinkage network for bearing fault diagnosis," *IEEE Trans. Cybernetics*, vol. 54, no. 3, pp. 1608–1613, 2024.

[6] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[7] X. Yang, X. Yuan, L. Dong, X. Mei, and K. Chen, "Manifold assistant multi-modal multi-objective differential evolution algorithm and its application in actual rolling bearing fault diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108040, 2024.

[8] L. Ma, N. Li, P. Zhu, K. Tang, A. Khan, F. Wang, and G. Yu, "A novel fuzzy neural network architecture search framework for defect recognition with uncertainties," *IEEE Trans. Fuzzy Systems*, 2024.

[9] C. Zhang, L. Zhu, D. Shi, J. Zheng, H. Chen, and B. Yu, "Semisupervised feature selection with soft label learning," *IEEE/CAA Journal of Automatica Sinica*, pp. 1–13, 2022.

[10] Y. Wang, Z. Zhang, and Y. Lin, "Multi-cluster feature selection based on isometric mapping," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 3, pp. 570–572, 2022.

[11] Y. Gong, J. Zhou, Q. Wu, M. Zhou, and J. Wen, "A length-adaptive nondominated sorting genetic algorithm for bi-objective high-dimensional feature selection," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 9, pp. 1834–1844, 2023.

[12] B. Tran, B. Xue, and M. Zhang, "A new representation in pso for discretization-based feature selection," *IEEE Trans. Cybernetics*, vol. 48, no. 6, pp. 1733–1746, 2017.

[13] K. Chen, F. Zhou, and B. Xue, "Particle swarm optimization for feature selection with adaptive mechanism and new updating strategy," in *Australasian Joint Conf. on Artificial Intelligence*. Springer, 2018, pp. 419–431.

[14] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Computing Curveys (CSUR)*, vol. 50, no. 6, pp. 1–45, 2017.

[15] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data Classification: Algorithms and Applications*, p. 37, 2014.

[16] L. C. Echevarr´ıa, O. L. Santiago, J. A. H. Fajardo, A. J. S. Neto, and D. J. S´anchez, "A variant of the particle swarm optimization for the improvement of fault diagnosis in industrial systems via faults estimation," *Engineering Applications of Artificial Intelligence*, vol. 28, pp. 36–51, 2014.

[17] R. Ziani, A. Felkaoui, and R. Zegadi, "Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized fisher's criterion," *Journal of Intelligent Manufacturing*, vol. 28, pp. 405–417, 2017.

[18] S. Nezamivand Chegini, A. Bagheri, and F. Najafi, "A new intelligent fault diagnosis method for bearing in different speeds based on the fdafscore algorithm, binary particle swarm optimization, and support vector machine," *Soft Computing*, vol. 24, pp. 10 005–10 023, 2020.

[19] M. Mansouri, K. Dhibi, H. Nounou, and M. Nounou, "An effective fault diagnosis technique for wind energy conversion systems based on an improved particle swarm optimization," *Sustainability*, vol. 14, p. 18, 2022.

[20] M. M. Mafarja and S. Mirjalili, "Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection," *Soft Computing*, vol. 23, no. 15, pp. 6249–6265, 2019.

[21] B.-S. Peng, H. Xia, Y.-K. Liu, B. Yang, D. Guo, and S.-M. Zhu, "Research on intelligent fault diagnosis method for nuclear power plant based on correlation analysis and deep belief network," *Progress in Nuclear Energy*, vol. 108, pp. 419–427, 2018.

[22] H. B. Nguyen, B. Xue, P. Andreae, and M. Zhang, "Particle swarm optimisation with genetic operators for feature selection," in *2017 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2017, pp. 286–293.

[23] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Trans. Industrial Informatics*, vol. 9, no. 4, pp. 2226–2238, 2013.

[24] X. Yang, X. Yuan, T. Ye, W. Zhu, F. Zhou, and J. Jin, "Psnn-tada: Prototype and stochastic neural network based twice adversarial domain adaptation for fault diagnosis under varying working conditions," *IEEE Trans. Instrumentation and Measurement*, 2023.

[25] L. Ma, H. Kang, G. Yu, Q. Li, and Q. He, "Single-domain generalized predictor for neural architecture search system," *IEEE Trans. Computers*, 2024.

[26] L. Mao, Z. Liu, D. Low, W. Pan, Q. He, L. Jackson, and Q. Wu, "Evaluation method for feature selection in proton exchange membrane fuel cell fault diagnosis," *IEEE Trans. Industrial Electronics*, vol. 69, no. 5, pp. 5277–5286, 2022.

[27] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *1997 IEEE Int. Conf. on Systems, Man, and cybernetics. Computational Cybernetics and Simulation*, vol. 5. IEEE, 1997, pp. 4104–4108.

[28] T. Thaher, H. Chantar, J. Too, M. Mafarja, H. Turabieh, and E. H. Houssein, "Boolean particle swarm optimization with various evolutionary population dynamics approaches for feature selection problems," *Expert Systems with Applications*, vol. 195, p. 116550, 2022.

[29] P. Hu, J.-S. Pan, S.-C. Chu, and C. Sun, "Multi-surrogate assisted binary particle swarm optimization algorithm and its application for feature selection," *Applied soft computing*, vol. 121, p. 108736, 2022.

[30] Z.-J. Wang, Z.-H. Zhan, S. Kwong, H. Jin, and J. Zhang, "Adaptive granularity learning distributed particle swarm optimization for largescale optimization," *IEEE transactions on cybernetics*, vol. 51, no. 3, pp. 1175–1188, 2020.

[31] Y. Zhang, Y.-H. Wang, D.-W. Gong, and X.-Y. Sun, "Clustering-guided particle swarm feature selection algorithm for high-dimensional imbalanced data with missing values," *IEEE Trans. Evolutionary Computation*, vol. 26, no. 4, pp. 616–630, 2021.

[32] Q. Ni, Z. Zhan, X. Li, Z. Zhao, and L. L. Lai, "A real-time fault diagnosis method for grid-side overcurrent in train traction system using signal time series feature pattern recognition," *IEEE Trans. Industrial Electronics*, vol. 71, no. 4, pp. 4210–4218, 2024.

[33] Y. Li, Y. Yang, G. Li, M. Xu, and W. Huang, "A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mrmr feature selection," *Mechanical Systems and Signal Processing*, vol. 91, pp. 295–312, 2017.

[34] Q. Lu, R. Yang, M. Zhong, and Y. Wang, "An improved fault diagnosis method of rotating machinery using sensitive features and rls-bp neural network," *IEEE Trans. Instrumentation and Measurement*, vol. 69, no. 4, pp. 1585–1593, 2019.

[35] L. Du, Z. Xu, H. Chen, and D. Chen, "Feature selection-based low voltage ac arc fault diagnosis method," *IEEE Trans. Instrumentation and Measurement*, 2023.

[36] Y. Liu, Y. Yu, L. Guo, H. Gao, and Y. Tan, "Automatically designing network-based deep transfer learning architectures based on genetic algorithm for in-situ tool condition monitoring," *IEEE Trans. Industrial Electronics*, vol. 69, no. 9, pp. 9483–9493, 2022.

[37] C.-Y. Lee and T.-A. Le, "An enhanced binary particle swarm optimization for optimal feature selection in bearing fault diagnosis of electrical machines," *IEEE Access*, vol. 9, pp. 102 671–102 686, 2021.

[38] X. Li and J. Ren, "Micq-ipso: An effective two-stage hybrid feature selection algorithm for high-dimensional data," *Neurocomputing*, vol. 501, pp. 328–342, 2022.

[39] B. Nayana and P. Geethanjali, "Improved identification of various conditions of induction motor bearing faults," *IEEE Trans. Instrumentation and Measurement*, vol. 69, no. 5, pp. 1908–1919, 2019.

[40] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of relieff and rrelieff," *Machine Learning*, vol. 53, pp. 23–69,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

CHEN *et al.*: CORRELATION-GUIDED PARTICLE SWARM OPTIMIZATION APPROACH FOR FEATURE SELECTION IN FAULT DIAGNOSIS 13

2003.

[41] Z. Huang, C. Yang, X. Zhou, and T. Huang, "A hybrid feature selection method based on binary state transition algorithm and relieff," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, p. 5, 2018.

[42] F. Zhang, Y. Mei, S. Nguyen, and M. Zhang, "Evolving scheduling heuristics via genetic programming with feature selection in dynamic flexible job-shop scheduling," *ieee transactions on cybernetics*, vol. 51, no. 4, pp. 1797–1811, 2020.

[43] M. Clerc and J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.

[44] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64, pp. 100–131, 2015.

[45] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Trans. Reliability*, vol. 69, no. 1, pp. 401–412, 2018.

[46] D. Liu, Z. Xiao, X. Hu, C. Zhang, and O. Malik, "Feature extraction of rotor fault based on eemd and curve code," *Measurement*, vol. 135, pp. 712–724, 2019.

[47] K. Li, X. Ping, H. Wang, P. Chen, and Y. Cao, "Sequential fuzzy diagnosis method for motor roller bearing in variable operating conditions based on vibration analysis," *Sensors*, vol. 13, no. 6, pp. 8013–8041, 2013.

[48] J.-Q. Yang, Q.-T. Yang, K.-J. Du, C.-H. Chen, H. Wang, S.-W. Jeon, J. Zhang, and Z.-H. Zhan, "Bi-directional feature fixation-based particle swarm optimization for large-scale feature selection," *IEEE Trans. Big Data*, vol. 9, no. 3, pp. 1004–1017, 2023.

[49] X.-F. Song, Y. Zhang, D.-W. Gong, and X.-Z. Gao, "A fast hybrid feature selection based on correlation-guided clustering and particle swarm optimization for high-dimensional data," *IEEE Trans. Cybernetics*, vol. 52, no. 9, pp. 9573–9586, 2022.

[50] K. Chen, B. Xue, M. Zhang, and F. Zhou, "Correlation-guided updating strategy for feature selection in classification with surrogate-assisted particle swarm optimization," *IEEE Trans. Evolutionary Computation*, vol. 26, no. 5, pp. 1015–1029, 2022.

[51] 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. Cybernetics*, vol. 51, no. 2, pp. 589–603, 2019.

[52] B. Tran, B. Xue, and M. Zhang, "Variable-length particle swarm optimization for feature selection on high-dimensional classification," *IEEE Trans. Evolutionary Computation*, vol. 23, no. 3, pp. 473–487, 2018.

[53] J. Zhou, Q. Wu, M. Zhou, J. Wen, Y. Al-Turki, and A. Abusorrah, "Lagam: A length-adaptive genetic algorithm with markov blanket for high-dimensional feature selection in classification," *IEEE Trans. Cybernetics*, vol. 53, no. 11, pp. 6858–6869, 2022.

[54] L. Li, M. Xuan, Q. Lin, M. Jiang, Z. Ming, and K. C. Tan, "An evolutionary multitasking algorithm with multiple filtering for highdimensional feature selection," *IEEE Trans. Evolutionary Computation*, vol. 27, no. 4, pp. 802–816, 2023.
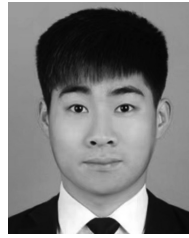
**Ke Chen** (Member, IEEE) received his B.Sc. degree from Weifang University, Weifang, China, in 2014, his M.Sc. degree from Yanshan University, Qinhuangdao, China, in 2017, and his Ph.D degree from Shandong University, Jinan, China, in 2021. He is currently a Associate Professor in the School of Electrical Engineering at Zhengzhou University. He has over 20 journal and conference papers. His current research interests include evolutionary computation, feature selection, fault diagnosis.



**Wenjie Wang** received her B.E. degree from Henan University of Science and Technology, China in 2022. Now, she is currently pursuing her M.S. degree at Zhengzhou University, China. Her current research interests include evolutionary computation, feature selection, and fault diagnosis.



**Fangfang Zhang** (Member, IEEE) received her B.Sc. and M.Sc. degrees from Shenzhen University, China, and her Ph.D. degree in Computer Science from Victoria University of Wellington, New Zealand, in 2014, 2017, and 2021, respectively. She is currently a postdoctoral research fellow in the Centre for Data Science and Artifical Intelligence & School of Engineering and Computer Science, Victoria University of Wellington, New Zealand. She is an Associate Editor of Expert Systems With Applications, Swarm and Evolutionary Computation.



**Jing Liang** (Senior Member, IEEE) received her B.E. degree from Harbin Institute of Technology, China in 2003, and her Ph.D. degree from Nanyang Technological University, Singapore in 2009. She is currently a Professor at the School of Electrical and Information Engineering, Zhengzhou University, China and the School of Electrical Engineering and Automation, Henan Institute of Technology, China. Dr. Liang serves as an associate editor for IEEE Transactions on Evolutionary Computation and the Swarm and Evolutionary Computation.



**Kunjie Yu** (Member, IEEE) received his Ph.D. degree in Control Science and engineering from the East China University of Science and Technology, Shanghai, China, in 2017. Currently, he is an associate professor with the School of Electrical and Information Engineering, Zhengzhou University. His current research interests include evolutional computation, constrained optimization, dynamic optimization and their applications in complex systems. He serves as an Associate Editor of Swarm and Evolutionary Computation, and Expert Systems with Applications.