IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 April 2023, accepted 29 April 2023, date of publication 9 May 2023, date of current version 1 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3274696

RESEARCH ARTICLE

A Feature Selection Approach Based on Memory Space Computation Genetic Algorithm Applied in Bearing Fault Diagnosis Model

CHUN-YAO LEE[1], (Member, IEEE), TRUONG-AN LE[2], AND CHUN-LIN HUNG[1]
[1]Department of Electrical Engineering, Chung Yuan Christian University, Taoyuan 320314, Taiwan
[2]Department of Electrical and Electronic Engineering, Thu Dau Mot University, Thu Dau Mot, Binh Duong 75000, Vietnam

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

ABSTRACT The main objective of this study is to propose a motor fault diagnosis model based on machine learning. Compared with the traditional motor fault diagnosis model, the proposed model can reduce the computation time. This model can be divided into three steps: feature extraction, feature selection, and classification. In the feature extraction step, the original signal is extracted by Hilbert-Huang transform (HHT), envelope analysis (EA), and variational mode decomposition (VMD) methods. A feature selection method based on memory space computation genetic algorithm (MSCGA) is proposed and applied in the feature selection step. The advantage of MSCGA is that it eliminates the need to compute data fitness values, saving unnecessary computation time repeatedly. The classifiers use k-nearest neighbor (KNN) and support vector machines (SVM). In order to verify the stability and efficiency of the model, the university of California Irvine (UCI) benchmark dataset, the current signal of motor fault datasets, and case western reserve university (CWRU) were used. The UCI dataset is used to test the efficiency and computation time of the feature selection method. Other datasets are used to compare with traditional motor fault diagnosis models. The simulation results of the proposed model have demonstrated the effectiveness in reducing the computation time without affecting the computation results compared to the traditional motor fault diagnosis model. Furthermore, the performance of MSCGA is proven to be better than that of the other algorithm.

INDEX TERMS Rotating machine, bearing fault diagnosis model, feature selection method, genetic algorithm.

I. INTRODUCTION

In an industrialized society, electrical rotary machinery is widely used in various manufacturing industries and manufacturing plants. Among them, the rolling bearing is the most important core of the rotating machine [1]. The rolling bearing in the rotating machine in the occurrence of the failure ratio is the highest and the most common [2], [3]. The failure of the rotating machinery may cause damage to the manufacturing equipment, and in serious cases may even cause injury to the operator [4]. How to detect the failure of electrical machinery is a solution that needs to be proposed. Therefore, this study's contribution is finding a feature selection method that can be more efficient. The diagnosis model of bearing failure is divided into the following three steps [5], the first step is feature extraction [6], the second step is feature selection [7], and finally, classification. The motor in operation is interspersed with a large number of signals and noises [8]. Measuring the current and vibration signals is the initial signal. The main function of feature extraction is to find the important features from the initial signals and extract them [9]. The important features often include the maximum value, the average value, and the root means square value. The features' quality affects the classification accuracy [10]. There are several feature extraction methods [11], the more famous ones are envelope analysis (EA) [12], wavelet transforms (WT) [13], Hilbert-Huang transform (HHT) [14], variational

The associate editor coordinating the review of this manuscript and approving it for publication was Baoping Cai.

mode decomposition (VMD) [15], and fast Fourier transform (FFT) methods [16]. HHT was proposed by Norden E. Huang *et al,* at Academia Sinica, Taiwan [17], to decompose the analysis data into intrinsic mode functions (IMF) [18], called empirical mode decomposition (EMD) [19]. The IMF is transformed into a Hilbert transform to obtain the instantaneous frequency of the processed data [20]. VMD is one of the latest feature extraction strategies, similar to EMD, which determines the frequency center and bandwidth of each component by iteratively searching for the optimal solution of the variational model in the process of acquiring the decomposed components so that the frequency domain segmentation of the signal and the effective separation of each component can be achieved in an adaptive manner. Among them, HHT, VMD, and EA are applied in this study. After the feature extraction step, the feature set has been initially formed. However, the extracted dataset still has many redundant features, especially in the higher dimensional dataset [21]. These redundant features cause a decrease in the accuracy of the final classification. Therefore, a feature selection step is needed, which is an intermediate step between data extraction and classification. Its function is to pre-process the dataset before it is sent to the classifier so that its features are further selected. The development of feature selection has flourished and meta-inspired algorithms have received attention with the aim of solving the optimization of datasets. The classical algorithms include binary grey wolf optimization (BGWO) [22], ant colony optimization (ACO) [23], binary particle swarm optimization (BPSO) [24], genetic algorithm (GA) [25], and binary differential evolution (BDE) [26]. The GA was proposed by John Hollander of the university of Michigan and is commonly used to solve search algorithms for optimization. It was originally developed by drawing on a number of phenomena in evolutionary biology [27], including genetics, mutation, natural selection, and hybridization [28]. The proposed method in this study: MSCGA has the advantage of reducing the computing time while maintaining the original computing power compared to conventional GA. The main purpose is to avoid finding local optimal in feature selection, to increase the search area, to find the best choice for the whole domain, and to improve the spatial search capability. At the same time, it can shorten the computation time by more than half. The last step in a machine learning-based fault diagnosis model is classification. In this part, the classifier classifies the best features in the selected feature subset, identifies the fault signals, and the accuracy distinguishes the final result. In machine learning, several common classifiers, such as k-nearest neighbor (KNN) [29], discriminate analysis (DA) [30], support vector machines (SVM) [31], decision tree (DT) [32], and naive bayes (NB) [33], and random forest (RF) [34], and so forth. KNN is an example-based learning or inert learning that approximates locally and postpones all calculations until after classification. KNN is an example-based learning or inert learning that approximates locally and postpones all calculations until after classification. KNN is one of the simplest of all machine learning algorithms. SVM is a supervised learning model and associated learning algorithm for analyzing data in classification and regression analysis. Given a set of training instances, each of which is labeled as belonging to one or the other of two classes, it is a non-probabilistic binary linear classifier. KNN and SVM, as the most frequently used classifiers, are considered to have the highest degree of credibility. Therefore, this study uses KNN and SVM as the final classifiers for comparison. The proposed model in this study is a combination of the feature extraction method, the proposed feature selection method, MSCGA, and the final step, classifier.

This paper's main objective is to develop an effective method of bearing diagnosis. Based on this purpose, the main contributions are as follows:

1) Regarding feature extraction, the extracted dataset comprises the original data and the sub-data extracted by HHT, EA, and VMD, respectively. More oriented features can be extracted.
2) In feature selection, MSCGA was applied. Which can effectively shorten the computation time.
3) The proposed feature selection strategy is validated using the UCI datasets to compare the accuracy of the proposed feature selection strategy with other classical feature selection methods. Furthermore, in order to verify the stability, the proposed bearing diagnostic model is compared with the existing diagnostic model using the current signal of motor fault datasets and CWRU dataset.

The rest of the paper is described as follows: the process of feature extraction is discussed in section II. In section III, a detailed introduction of GA is described, the feature selection method based on GA is described, and the features and roles of MSCGA are described. Furthermore, section IV describes the bearing diagnosis model used in this paper. section V describes the experimental results, using UCI datasets, CWRU datasets, and bearing current signal datasets to test the model capability. Finally, the conclusions are described in section VI.

II. THE FEATURE EXTRACTION METHOD

In this study, three feature extraction methods were used, namely HHT, EA and VMD. The feature extraction flow chart is shown in Fig. 1. Feature extraction means extracting different features from the original signal as a basis for analysis. The features of the extraction are including max value (max), mean value (mean), min value (min) mean squared error (mse), root mean square (rms), sum value (sum) and standard deviation (std). The mathematical meanings of the features are shown in the Table 1. Mathematical expressions for every feature are shown in the Table 2.

A. HILBERT-HUANG TRANSFORM

HHT consists of two parts: empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA) [35]. This method is potentially viable for nonlinear and nonstationary data

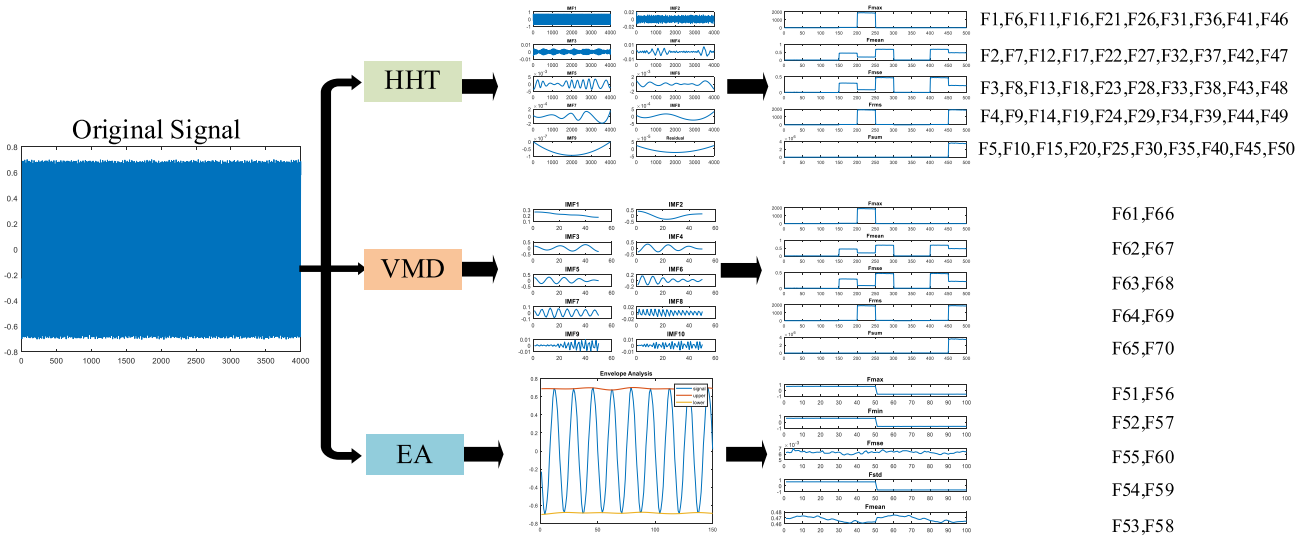

FIGURE 1. Feature extraction flow chart.

TABLE 1. Mathematical meanings of the features.

Feature	Equation	
max value	$x_{\max} = \max(x(n))$	(1)
mean value	$x_{\text{mean}} = \frac{\sum_{n=1}^{N}(x(n))}{N}$	(2)
min value	$x_{\min} = \min(x(n))$	(3)
mean squared error	$x_{\text{mse}} = \frac{1}{N}\sum_{n=1}^{N}(x(n)-\hat{x}(n))^2$	(4)
root mean square	$x_{\text{rms}} = (\frac{\sum_{n=1}^{n}\sqrt{\lvert x(n)\rvert}}{N})^2$	(5)
sum	$x_{\text{sum}} = \text{sum}(x(n))$	(6)
standard deviation	$x_{\text{std}} = \sqrt{\frac{\sum\lvert x(n)-\overline{x(n)}\rvert^2}{n}}$	(7)

Note: $x(n)$ arranged with $n = 1, 2, ..., N$, N is the length of data.

TABLE 2. Mathematical expressions for every feature.

Features	HHT	EA	VMD
max	F1, F11, F21, F31, F41 F6, F16, F26, F36, F46	F51, F56	F61, F66
mean	F2, F12, F22, F32, F42 F7, F17, F27, F37, F47	F53, F58	F65, F70
min	—	F52, F57	—
mse	F3, F13, F23, F33, F43 F8, F18, F28, F38, F48	F55, F60	F63, F68
rms	F4, F14, F24, F34, F44 F9, F19, F29, F39, F49	—	F64, F69
sum	F5, F15, F25, F35, F45 F10, F20, F30, F40, F50	—	F62, F67
std	—	F54, F59	—

analysis, especially for time-frequency-energy representations [36].

1) Empirical mode decomposition (EMD)

Decomposing data into intrinsic mode functions (IMF), such a decomposition process is called empirical mode decomposition (EMD) method. The signal $x(t)$ decomposed process via the EMD are listed as below:

a. Find all the local maximum and local minimum of signal $x(t)$.
b. The upper and lower envelopes are connected from the local maximum and local minimum by the cubic spline line.
c. The mean of the upper envelopes and lower envelopes is designated be m_{ik}, the intermediate component h_{ik} is calculated as follow $h_{ik} = x(t) - m_{ik}$ [37]. Eliminate noise in the signal and meet IMF conditions. Therefore, repeat the process k times (Step *a.* to Step *c.*), the first component is considered to be the input signal.
d. In iteration process, if any intermediate component h_{ik} satisfies the IMF conditions, it is considered the first IMF $c_i(t) = h_{ik}(t)$. The residue is calculated as follow $r_i(t) = x(t) - c_i(t)$.The iterative loop is continued again (Step *(a.)* to Step *(d.)*) with residual is considered as input signal $x(t) - r_i(t)$ for finding the next of IMFs.
e. When termination criteria are met. The stoppage criterion for termination can be either of the following:
   - The residual standard deviation is less than the specified value.
   - No additional IMF can be extracted from the residual signal.

2) Hilbert Transform (HT)

HT is a specific linear operator which acts as a function, $c_i(t)$ of a continuous signal and produces another function of a real variable. The Hilbert spectrum matrix of the signal was formed from the process of analyzing the IMFs by the HT [38]. HT spectrum can be used to analyze the nonlinear and non-stationary signals of the spectrum content over time [39]. Therefore, a spectral matrix is applied to extract local features [40]. Based on the amplitude distribution in the

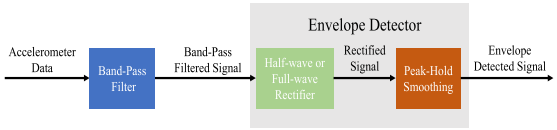

FIGURE 2. EA flow chart.

time-frequency domain of the spectrum matrix, the feature extraction process is as follows: in the time domain, five curves are calculated by five indicators as maximum (T-max), mean (T-mean), mean square error (T-mse), root mean square (T-rms), and energy (T-sum). Afterward, each indicators curve has five statistical values extracted as maximum (max), mean, mean square error (mse), root mean square (rms), and standard deviation (std).

B. ENVELOPE ANALYSIS

This method is mainly used to extract periodic mechanical vibration excitation signals [41], which can detect defects in various motor parts. The flow chart of EA is shown in Fig. 2. There are N samples, and $x(t)$ is the original vibration mode. $X(f)$ is the FFT-transformed signal of the original shock wave.

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt \tag{8}$$

$x(n)$ is the recombined signal passed through the bandpass filter.

$$x(n) = \frac{1}{N} = \sum_{k=0}^{N-1} X(f)e^{-j\omega kn} \tag{9}$$

$x_{env(n)}$ is the envelope signal to be analyzed, as shown follow:

$$x_{env(n)} = \sqrt{x(n)\overline{x(n)}}. \tag{10}$$

$X_{env}(k)$ is the envelope spectrum

$$X_{env}(k) = \sum_{k=0}^{N-1} x_{env}(n)e^{-j\omega kn}. \tag{11}$$

The main advantage of EA is reliable wide-frequency fault detection in certain frequency bands [42].

C. VARIATIONAL MODE DECOMPOSITION

This method is based on the traditional Hilbert transform [15]. Find the optimal solution of the variational model through iteration, and determine each mode's bandwidth and center frequency. Implementing adaptive decomposition in a variational framework [43].

Update in the frequency domain, and finally inverse Fourier transformed to the time domain. Complete adaptive segmentation of the signal band. Effectively avoid blend mode. K mode corresponds to different periods [44]. The VMD function calculates the IMFs in the frequency domain, reconstructing $X(f) = DFT\{x(t)\}$ in terms of $Uk(f) = DFT\{uk(t)\}$. The algorithm extends the signal by mirroring half its length on either side to remove edge effects. The Lagrange multiplier introduced in optimization (signal processing toolbox) has the Fourier transform $\wedge(f)$. The length of the Lagrange multiplier vector is the length of the extended signal. For the $(n+1)$-th iteration, the algorithm performs these steps:

1) Iterate over the k modes of the signal to compute:
   a. The frequency-domain waveforms for each mode using

$$U_k^{n+1}(f) = \frac{X(f) - \sum_{i<k} U_k^{n+1}(f) - \sum_{i>k} U_k^n(f) + \frac{\wedge^n}{2}(f)}{1 + 2\alpha\left\{2\pi\left(f - f_k^n\right)\right\}^2} \tag{12}$$

   b. Where $U_k^{n+1}(f)$ is the Fourier transform of the kth mode calculated in the $(n+1)$-th iteration.
   c. The kth central frequency f_k^{n+1} using

$$f_k^{n+1} = \frac{\int_0^{\infty}\left|U_k^{n+1}(f)\right|^2 fdf}{\int_0^{\infty}\left|U_k^{n+1}(f)\right|^2 df} \approx \frac{\sum f\left|U_k^{n+1}(f)\right|^2}{\sum\left|U_k^{n+1}(f)\right|^2} \tag{13}$$

2) Update the Lagrange multiplier using

$$\wedge^{n+1}(f) = \wedge^n(f) + \tau(X(f)) - \sum_k U_k^{n+1}(f) \tag{14}$$

where τ is the update rate of the Lagrange multiplier.

III. THE FEATURE SELECTION METHOD

The purpose of feature selection is to extract features that can effectively improve the final accuracy of the dataset after feature extraction, and delete features that are useless for classification. This section applies the genetic algorithm (GA) and the proposed method, MSCGA.

A. GENETIC ALGORITHM

GA is a feature-based optimization technique that searches for heuristics based on population calculus, imitating the process of natural human evolution. The principle of genetic algorithm comes from the iterative process of manipulating a set of population or organism chromosomes (candidate solutions) to generate new individuals, through genetic functions, such as selection, crossover and mutation (similar to Charles Darwin's theory of biological evolution, inheritance and recombination of chromosomes, survival of the fittest [45]).

The GA is composed of three steps: selection, crossover and mutation. The steps as below

step1 Encoding
Encode each solution in gene strings. Each chromosome is divided into explicit (1) and implicit (0) [46].

step2 Creating a group
Create a primitive population that contains a certain number of sequences.

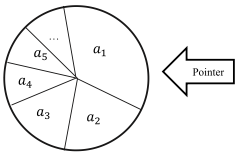

FIGURE 3. Roulette wheel selection.

step1 Implement GA's principle
Implement the following steps until the termination conditions are reached:

a. Calculate the fitness values for every solution in the group. The fitness function determines how close a given solution is to the best solution to the desired problem. The purpose is to determine the suitability of the solution [47].
b. In GA, each chromosome is represented as a sequence of binary codes. These binary codes are called solutions, and these solutions are tested and the best solution to the problem is proposed. A score is determined for each solution, indicating how close the solution is to satisfy the desired solution. The score is calculated by applying the fitness function to the test or from the tested solution.
c. How to determine the appropriate fitness function for a given problem? Each problem has a suitable fitness function. How to decide the fitness function depends on the nature of the problem itself. When using GA, deciding which fitness function is appropriate for the problem is the most challenging part. The computational parameters and basic functions related to the domain in which the problem is formulated can be used as fitness functions for optimization problems.
d. In this model, the KNN is applied. In other words, accuracy is the score that determines whether a model is good enough. Accuracy is defined as follow:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (15)$$

In this study, the metric used for the fitness function is the error rate, where the error rate formula is as follows

$$\text{Fitness value} = \text{Error Rate} = 1 - \text{Accuracy} \quad (16)$$

e. Create new solution by three genetic operations as follows:
Make choices in the group based on solution's fitness values, and choose two solutions as parents. The schematic of the roulette wheel is shown in Fig. 3. Consider a population of size n (the number of solutions is n), $P = \{a_1, a_2, a_3, \ldots, a_n\}$, each solution has fitness values of $f(a_i)$, or each solution selected

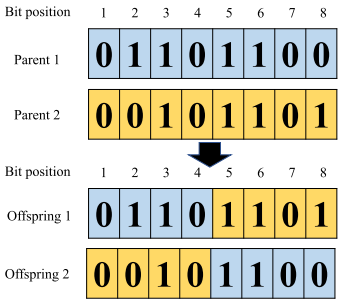

FIGURE 4. Single node crossover.

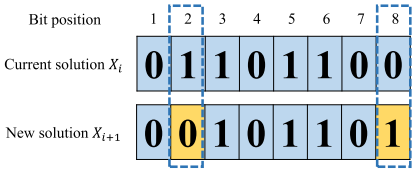

FIGURE 5. Single point mutation.

chance is: [48]

$$P_s(a_i) = \frac{f(a_i)}{\sum_{i=1}^{n} f(a_j)}, j = 1, 2, 3, \ldots, n \quad (17)$$

f. Two parents created two new solutions, and the combination of new solutions comes from the reorganization of the intersection of the two parents. Select one of the nodes, and swap the remaining chromosomes. The schematic diagram of the single-node crossover is shown in Fig. 4.
g. By probabilistic selection, mutate an existing solution. Select several of the chromosomes and flip them. The schematic diagram of single point mutation is shown in Fig. 5.

step4 Termination
When the termination condition is reached, this calculus's subset of optimal fitness values is obtained. The flow chart of GA is shown in Fig. 6. The fitness of the candidate (chromosome) is used for the function to evaluate the target or fitness function. The value given by the fitness value function (objective function) is used for the fitness sorting of a specific chromosome. The applicability of the fitness function depends on the application problem.

B. THE PROPOSED METHOD (MSCGA)

The GA algorithm is widely applied in many research fields, but it still has some disadvantages. In traditional GA, the fitness value is recalculated at each iteration, including duplicate chromosomes. If the number of generations and populations is large enough, it will take a lot of time, and this is where the most time is spent. In addition, the genetic

randomization of chromosomes is so uncertain that it is sometimes difficult to find the optimal fitness value by random search. Sometimes it can get stuck with a local maximum problem, which means that GA does not guarantee that every computation is optimal.

The main idea of this study is to find a way to reduce the computation time efficiently without affecting the original GA computation and to further expand the original population in the hope of solving the local maximum problem. The advantage of MSCGA is that it eliminates the need to compute data fitness values, saving unnecessary computation time repeatedly. Therefore, the computation time can be effectively reduced.

step1 Initial population generator
Create a primitive population that contains a certain number of sequences.

step2 Fitness assignment
Calculate the fitness values for every solution in the group.

step1 Selection
Make choices in the group based on solution's fitness values, and choose two solutions as parents. In the traditional GA selection method, the roulette wheel method is used. In MSCGA, the roulette wheel selection of one of the parents generations is replaced using a method called the ranking roulette wheel. The roulette wheel is ranked according to the fitness value from the largest to the smallest, expecting that the order with the largest fitness value will be increased. Therefore, in the selection phase of MSCGA, a mixed roulette wheel selection is used.
The check rate parameter has been added in the selection stage. The purpose of this check rate parameter is to check the similarity of the two parent generations, if the similarity is too high, the selection is considered meaningless and the selection is skipped. The check rate is set to 50%, which means that if the similarity of the two parent generations is more than 50%, the selection is skipped.

step4 Crossover
Two parents created two new solutions, and the combination of new solutions comes from the reorganization of the intersection of the two parents. Selecting one of the nodes, swap the remaining chromosomes. The check rate parameter is added to the selection stage to reduce the number of selection stages that are canceled due to high similarity. The cross rate (the probability of selecting parental mating at the current time in traditional GA) is eliminated in the crossover stage. This prevents the program from stopping due to the cancellation of two steps.

step5 Mutation
In the mutation step, random interval (RI) mutation is proposed since the traditional GA has the problem of the local optimal due to excessive convergence.

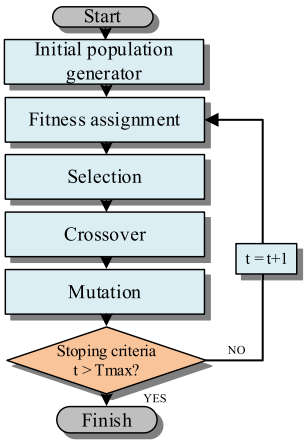

FIGURE 6. GA flow chart.

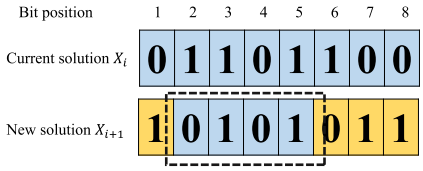

FIGURE 7. Random Interval mutation.

RI selects an interval of the chromosome and mutates the rest of the chromosome. The maximum mutation is performed while retaining some of the original chromosome characteristics, as shown in Fig. 7.

step6 Delete the repeat offspring
A memory space has been added in this step. The purpose of this space is to record the population chromosomes of each iteration in memory before the termination condition is reached. The new offspring from each iteration will be compared in memory. If a chromosome is found to be duplicated in the existing population, the chromosome is deleted. This means that the fitness value of the chromosome does not need to be calculated again in the current iteration, further reduction of computing time.

step7 Compare if any chromosomes have been deleted
In the previous step, chromosomes that overlap with existing populations will be deleted. This means that there are not enough offspring in the current iteration to move to the next iteration. In this step, it is detected that if any offspring is deleted, the number of deleted the offspring will be increased. The new offspring will be added in the next step: the additional crossover.

step8 Additional crossover
In each iteration, duplicate chromosomes are deleted, which means that the population may not be large enough. Deleted chromosomes are replenished in this step, and the memory space is divided into five classes according to their fitness values, from high to low.

Each rank is given a weight. The highest rank has a weight of 1, the next rank has a weight of 1/2, …, the lowest rank has a weight of 1/5. Whenever a chromosome is deleted, two ranks are selected by using roulette wheel selection from memory. The two parents are selected again using roulette wheel selection to generate a new offspring at the crossover between the two ranks. This method maximizes population diversity while maintaining population fitness values.

step9 If the offspring meet the original number
Repeating the additional crossover step is expected to bring the number of chromosomes up to the original number that was not deleted. Once the number of chromosomes is checked at this step, the next step is performed: again mutation.

step10 Again mutation
When the number of chromosomes has reached the number of the offspring that would have resulted from the original iteration, it is expected that the chromosome diversity can be increased further. Therefore, we perform an again mutation. Again mutation works by selecting chromosomes with high fitness values in the top 20% of memory for mutation. (The original mutation rate was 0.01 while the again mutation rate was 0.05). New offspring are mutated and added to the existing population.

step11 Fitness assignment
Calculate the fitness values for every solution in the group. Every solution's fitness value in the group should reach the maximum of this iteration.

step12 If stop criteria t > T max
Preset number of iterations, the number of t is 100. When the current iteration number reaches 100, the calculate is stopped.

The method MSCGA was proposed in this research. Compared with the traditional GA, MSCGA is more powerful in feature selection and effectively reduces calculation time. The flowchart of the proposed method is shown as Fig. 8. The bearing fault diagnosis model has three parts: feature extraction, feature selection and classification. In the feature extraction part, HHT, EA and VMD are used to extract features from the initial signal. HHT extracts 50 features, EA extracts 10 features, and 10 features are extracted by VMD; Total of 70 features. In the feature selection part MSCGA is applied. In this part, the optimal feature subset is generated. Support vector machine (SVM) and k-nearest neighbor (KNN) are used in the classification part. Identify the feature of healthy motor bearings. And faulty motor bearings with different degrees of damage.

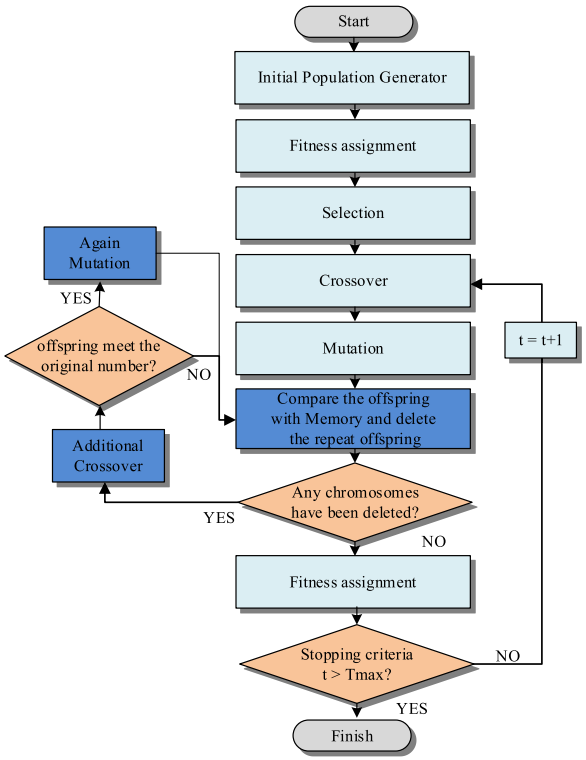

FIGURE 8. MSCGA method flow chart.

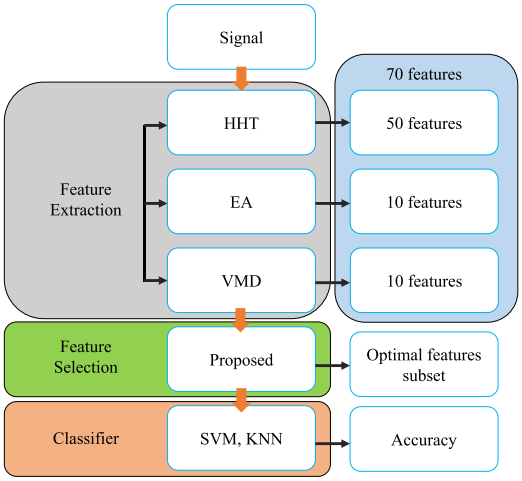

FIGURE 9. Flow chart of bearing fault diagnosis model.

IV. THE BEARING FAULT DIAGNOSIS MODEL

The bearing fault diagnosis model consists of three stages: feature extraction, feature selection, and classification. Fig. 9 shows the flowchart of the bearing fault diagnosis model. During the feature extraction stage, three feature selection methods are applied: HHT, VMD, and EA. A total of 70 features are extracted in this stage. Among them, 50 features are from HHT, 10 features are from VMD and 10 features are from EA. In the feature selection stage, an efficient feature selection method: MSCGA is proposed. The proposed method is able to save time by avoiding unnecessary recalculation of fitness values. At the same time, it is possible to retain the parents' features to a greater extent. Two main traditional classifications are applied in the classification

TABLE 3. Parameters of the feature selection method used.

MSCGA	GA	ACO	BDE
number of chromosomes:10	number of chromosomes:10	initial tau:1	number of vectors:10
number of generations:10	number of generations:10	initial beta:1	maximum number of iterations:100
check rate 50%	crossover rate:0.8	coefficient control tau:1	crossover rate: 0.9
mutation rate: 0.01	mutation rate: 0.01	beta:1	—
again mutation rate: 0.05	—	phermone: 0.2	—
		coefficient: 0.5	

stage: SVM and KNN. The training data set is set to 70% and the testing data set is set to 30%. SVM is a widely used classifier, which is convenient for solving small sample and nonlinear problems. The basic concept of SVM is to treat the input signal as a high dimensional vector and use lower dimensional hyperplane to separate the points, which is a so-called linear classifier. KNN consider the input signal as "*i*", find out which type of the k data closest to "*i*" is mostly, and predict the type of "*i*". The advantage is high precision. Both of the above can efficiently complete data classification.

V. EXPERIMENTAL VERIFICATION AND RESULTS

In the study, the personal computer (4 CPUs Intel(R) Xeon(R)CPU E3-1230 v3,32GB of RAM) was utilized. Experimental equipment includes three-phase squirrel-cage induction motor, three-phase power supply, three-purpose meter and oscilloscope. The software used for experiment is MATLAB R2017a version. In the study, three different datasets are applied to verify the motor diagnostic robustness of the proposed method and traditional methods.

A. PARAMETER SETTING

MSCGA is a kind of wrapper feature selection method, and the model also requires a classifier to predict the fitness values. The value computed by the classifier can evaluate the accuracy of the dataset. In this model, KNN and SVM are used to evaluate the fitness of the dataset. The parameters of the two classifications are set as follows. In KNN, k value is set to 3 and kfold is set to 10. In SVM, kfold is set to 10 and kernel is selected as radial basis function. The proposed method and other data selection methods used for comparison are GA, ACO, and BDE. The parameter settings are shown in Table 3.

B. CASE STUDY 1: UCI BENCHMARK DATASETS

1) The UCI benchmark was adopted to test the data selection ability of the proposed method. Three traditional methods are compared with the proposed feature selection method, GA, ACO and BDE.
2) In this case, a total of 11 datasets are used, namely BreastEW, CongressEW, Exactly, HeartEW, Ionosphere, KrVsKpEW, Sonar, SpectEW, Vote, Waveform and Wine. The data structure information is listed Table 4. In this phase, three characteristics were compared: fitness average error, fitness best performance and computation time. The average error and lowest error are presented in Table 5. When using the proposed method, eight data sets have the best performance in Average Error data compared with GA, ACO, and BDE. They are BreastEW, CongressEW, Exactly, Ionosphere, KrVsKpEW, Sonar, Waveform, and Wine. There are 7 data sets with the best performance in the Lowest Error data. In the Lowest Error data, there are 7 data sets with the best performance: BreastEW, Exactly, Ionosphere, KrVsKpEW, Sonar, SpectEW and Waveform. The data compared in Fig. 10 are the computation times of 100 iterations. The units are seconds. Of all 11 datasets used, MSCGA has the shortest computation time. In some data sets it is even possible to reduce the computation time by 50% compared to GA using a repetitive offspring removal technique. This can confirm that the performance of the proposed feature selection method is superior to that of the traditional methods.

TABLE 4. UCI dataset information.

	Datasets	Samples	Features
1	BreastEW	568	30
2	CongressEW	436	16
3	Exactly	1000	13
4	HeartEW	270	13
5	Ionosphere	351	34
6	KrVsKpEW	3196	36
7	Sonar	208	60
8	SpectEW	267	22
9	Vote	300	16
10	Waveform	5000	40
11	Wine	178	13

C. CASE STUDY 2: CURRENT SIGNAL OF MOTOR FAULT DATASETS MEASURED BY INDUCTION MOTOR

1) The motor fault dataset was adopted to test the classification ability of the bearing fault diagnosis model. The motor fault dataset was measured by a three-phase squirrel-cage induction motor. Motor specifications are as follows: rated voltage 220V/380V, rated horsepower 2hp, rated current 5.58A/3.23A, efficiency 83.5%, rated speed 1715 RPM and the number of poles is 4. The signal is measured under four conditions: health, bearing damage, stator fault, and rotor bar damage. Data miner (NI PXI-1033) is used for data current collection. The detailed status is shown in Table 6. The measurement conditions are as follows: load rate 0%, torque 0 N/m; load rate 25%, torque 2 N/m; load rate 50%, torque 4

TABLE 5. UCI datasets result of the calculated data.

Datasets	GA		ACO		BDE		The proposed method	
	average error (%)	lowest error (%)	average error (%)	lowest error (%)	average error (%)	lowest error (%)	average error (%)	lowest error (%)
BreastEW	0.0465	0.0423	0.0470	0.044	0.0479	**0.0405**	**0.0456**	**0.0405**
CongressEW	0.0357	**0.030**	0.0406	0.0369	0.037	**0.0276**	**0.0323**	0.0323
Exactly	0.0212	**0**	0.2629	0.0250	0.2106	**0**	**0.0060**	**0**
HeartEW	**0.1379**	**0.1259**	0.2101	0.1778	0.1663	0.1296	0.1407	0.1407
Ionosphere	0.1122	0.1026	0.1198	0.1054	0.1289	0.1140	**0.1087**	**0.0826**
KrVsKpEW	0.022	0.0156	0.0552	0.0438	0.0390	0.0307	**0.0219**	**0.0153**
Sonar	0.1600	0.1298	0.1769	0.1587	0.1821	0.1587	**0.1585**	**0.1250**
SpectEW	**0.1571**	0.1298	0.1607	0.1498	0.1558	0.1386	0.1585	**0.1250**
Vote	**0.0419**	**0.0367**	0.4190	0.0433	0.0463	0.0400	0.0443	0.0400
Waveform	**0.1183**	0.1148	0.1527	0.1432	0.1231	0.1150	**0.1183**	**0.1142**
Wine	0.0241	**0.0112**	0.0258	0.0169	0.0237	**0.0112**	**0.0237**	0.0225

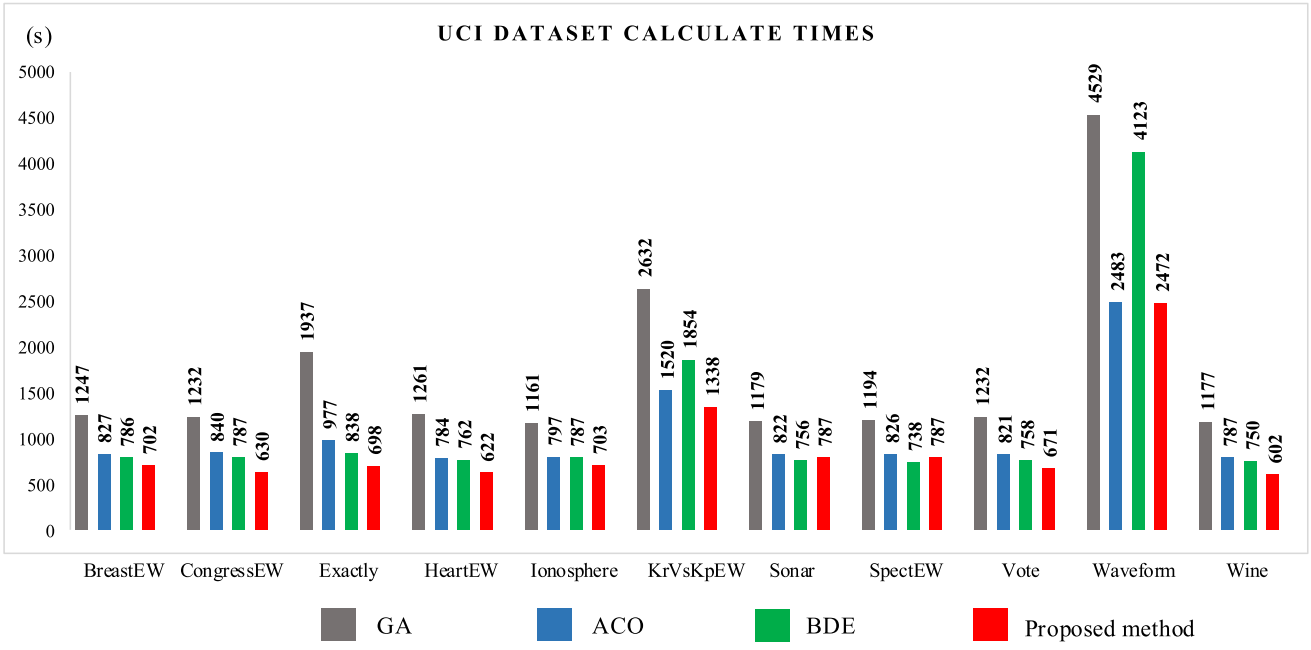

FIGURE 10. UCI datasets average computation time in different feature selection method.

FIGURE 11. Rotor drilling condition.

FIGURE 12. Bearing damage condition.

N/m; load rate 75%, torque 6 N /m and load rate 100%, torque 8 N/m. The measurement equipment is shown in Fig. 11, Fig. 12, Fig. 13.

Original signal is measures by the experiment induction motors in current signal form. In this case, the proposed feature selection method is compared with others three feature selection method same as case study 1.

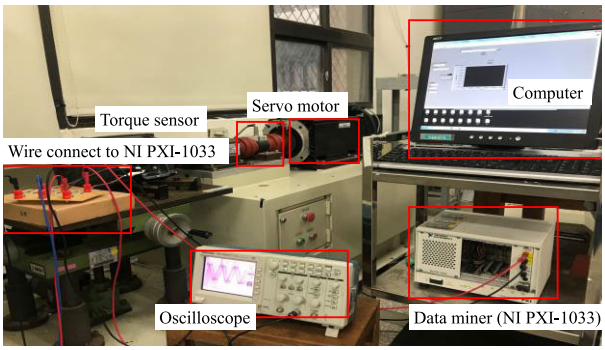

FIGURE 13. Measuring equipment: computer, oscilloscope, data miner (NI PXI-1033), torque sensor, servo motor.

TABLE 6. Test data set information.

Item	Damage Description	Order
Health	—	1
Bearing damage	B0.53: 0.53*0.53	2
	B1.24: 1.24*0.53	3
	B1.96: 1.96*0.53	4
Stator fault	S1: 2 winding short circuit	5
	S2: 4 winding short circuit	6
Rotor bar damage	R1: 2 bars between 1 hole deep 10mm	7
	R2: 3 bars between 2 holes deep 10mm	8

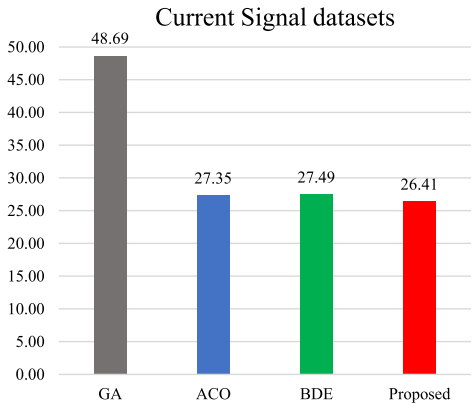

FIGURE 14. Current signal average computation time.

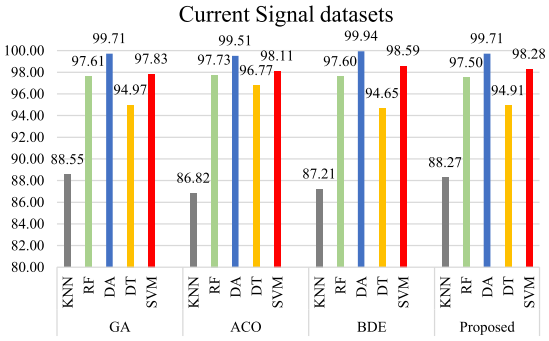

FIGURE 15. Current signal average classification results.

2) In this case, the test data used are divided into eight categories: one healthy data, three damaged bearings, two stators short-circuit conditions, and two rotor bars bore damages. Table 6 shows the detailed information.
3) In this experiment, the original dataset was combined into a dataset of 70 samples by extracting 50, 10 and 10 samples from HHT, EA and VMD individually. GA, ACO, BDE and MSCGA are used for data selection. Each experiment was repeated 30 times and the average computation time was obtained as shown in Fig. 14. It can be seen that the average time for GA is the longest, and among the four data selection methods, the average time for MSCGA can be the shortest.
4) The average classification results of the 30 experiments are recorded in Fig. 15. Although the best performance cannot be obtained by averaging 30 experiments, the best results can be obtained by using each classifier.
5) The original 70 features were selected by different feature selection methods, as shown in Table 7, with 35 features selected by GA, 32 features selected by ACO, 43 features selected by BDE, and 38 features selected by MSCGA. The BDE filtered the most features, but the final accuracy was not the highest, which proves that the BDE has the worst filtering ability among the four feature selection methods used, and it is difficult to retain the valid features effectively. BDE is not a good feature selection method.
6) In the result section, five conditions are counted, namely, no selected, GA, ACO, BDE, and the proposed method. The classification methods are KNN, RF, DA, DT and SVM. Among them, the proposed methods perform best in four of five classifiers—the proposed method, namely, RF, DA, DT and SVM. The results are shown in Fig. 16.
7) To verify the capability of the model, two values of precision and recall are added to demonstrate that the proposed method has a better ability to predict the percentage correctly compared to the original method. Precision is a metric that calculates the percentage of correct predictions for the positive class. Recall calculates the percentage of correct predictions for the positive class out of all positive predictions that could be made. Precision is defined as follow:

$$\text{Precision} = \frac{\text{True Positives}}{(\text{True Positives} + \text{False Positives})} \quad (18)$$

Precision is calculated as the ratio of correctly predicted positive examples divided by the total number of positive examples that were predicted. The result is a value between 0.0 for no precision and 1.0 for full or perfect

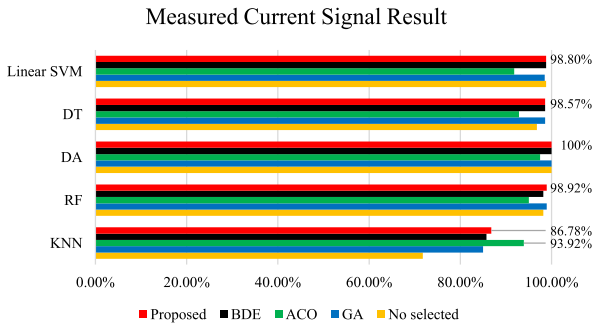

FIGURE 16. Current signal result.

TABLE 7. Current signal selected cases.

	Number	Selected features
GA	35	F1, F5, F7, F10, F12, F13, F14, F17, F19, F21, F22, F23, F24, F28, F29, F30, F36, F39, F42, F45, F46, F47, F51, F52, F57, F58, F60, F62, F64, F65, F66, F67, F68, F69, F70
ACO	32	F1, F2, F3, F5, F8, F10, F11, F13, F16, F17, F18, F21, F24, F25, F27, F28, F30, F32, F33, F43, F48, F49, F52, F53, F54, F58, F59, F52, F63, F64, F66, F67
BDE	43	F2, F3, F4, F5, F6, F7, F8, F9, F10, F13, F14, F15, F16, F17, F18, F19, F20, F21, F22, F23, F25, F26, F27, F28, F29, F30, F32, F34, F35, F36, F37, F38, F41, F42, F43, F44, F46, F49, F52, F55, F56, F58, F59, F60, F62, F63, F64, F65, F66, F67, F68, F69, F70
MSCGA	38	F2, F3, F4, F5, F6, F7, F8, F11, F13, F16, F21, F22, F23, F24, F25, F27, F28, F29, F32, F33, F34, F36, F37, F38, F41, F42, F44, F45, F46, F47, F48, F61, F63, F65, F66, F68, F69, F70

precision. Recall is defined as follow:

$$\text{Recall} = \frac{\text{True Positives}}{(\text{Ture Positives} + \text{False Negatives})} \tag{19}$$

Recall is calculated as the ratio of correctly predicted positive examples divided by the total number of positive examples that could be predicted. The result is a value between 0.0 for no recall and 1.0 for full or perfect recall.

8) In the GA method, the precision value is 0.8422 and the recall value is 0.8334. In MSCGA, the precision value is 0.8439 and the recall value is 0.8376. While in MSCGA, the precision value is 0.8439 and the recall value is 0.8376. There is a significant improvement in comparison. This proves that MSCGA is more effective than GA in this dataset to improve the data selection effect.
9) The conclusion can be found in the experimental results of current signal of motor fault dataset. MSCGA can reduce the computation time by almost half compared to the original method: GA. It can effectively reduce the computation time while maintaining the original accuracy standard. Even though the final accuracy is not the highest among the four feature selection methods, it is only the second-highest. And while reducing computation time, it further improves the accuracy and capability of feature selection.

TABLE 8. CWRU datasets selected cases.

	Number	Selected features
GA	40	F1, F2, F4, F6, F7, F8, F10, F11, F12, F14, F15, F16, F17, F18, F19, F20, F21, F23, F24, F25, F26, F30, F31, F32, F33, F34, F36, F37, F42, F43, F45, F52, F53, F54, F57, F59, F60, F61, F63, F69
ACO	30	F5, F8, F10, F11, F14, F16, F19, F26, F28, F29, F32, F33, F36, F37, F39, F41, F42, F44, F48, F53, F56, F57, F59, F60, F61, F67, F68, F69, F70
BDE	64	F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F11, F13, F14, F15, F17, F18, F19, F20, F21, F22, F23, F24, F25, F26, F27, F28, F29, F30, F31, F32, F33, F34, F35, F36, F38, F40, F41, F42, F43, F44, F45, F46, F47, F48, F51, F52, F53, F54, F55, F56, F57, F58, F59, F60, F61, F62, F63, F64, F65, F66, F67, F68, F69, F70
MSCGA	35	F1, F4, F5, F11, F13, F15, F16, F17, F18, F21, F23, F25, F26, F27, F29, F30, F31, F32, F34, F36, F37, F39, F43, F47, F48, F49, F52, F54, F55, F57, F58, F59, F60, F66, F67

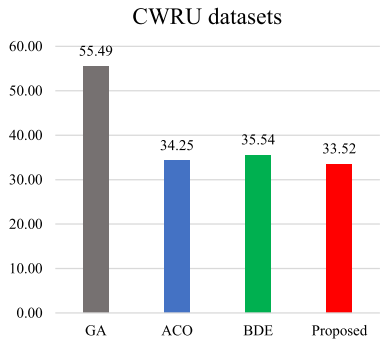

FIGURE 17. CWRU datasets average computation time.

D. CASE STUDY 3: CASE WESTERN RESERVE UNIVERSITY DATASET

1) CWRU dataset was collected for normal bearings, single-point drive end, and fan end defects. Data was collected at 12,000 samples/second and at 48,000 samples/second for drive-end bearing experiments. All fan end-bearing data was collected at 12,000 samples/second. It provides ball test vibration signals for normal motors and faulty bearings. Use 2 hp reliance electric motor as test object. Acceleration data are measured at both ends of the motor bearing. Motor bearing is broken by electro-discharge machining, and the hole diameter is from 0.007 inches to 0.040 inches. Motor fault locations include inner raceway, outer raceway and rolling element (i.e. ball). Relocate to the test motor and

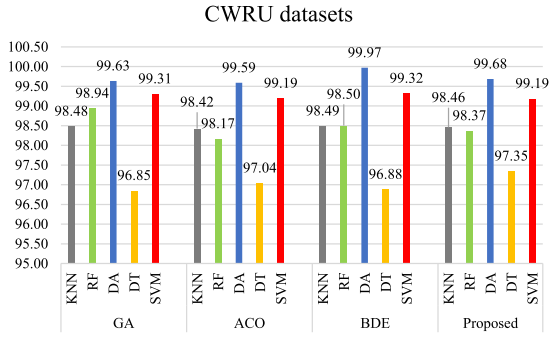

FIGURE 18. CWRU dataset result average classification results for 30 times.

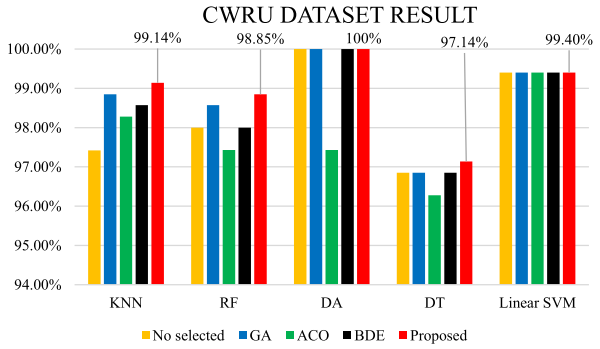

FIGURE 19. The best performance of CWRU result.

record the vibration signal of load 0 ∼ 3 hp (1720 ∼ 1797 RPM). CWRU dataset was adopted to test the classification ability of the bearing fault diagnosis model. In this case, the proposed feature selection method is also compared by the three feature selection methods in case study 1.

2) In this case, the test data used were divided into ten categories: one health data, three innerace, three outerace, and three ball. Innerace, outerace, and ball data were individually classified as diameter 0.0007 inches, 0.0014 inches, and 0.0028 inches.
3) Same as case study 2. The number of samples selected by each data selection method from the original data set of 70 samples and their numbers are organized in the Table 8. In the average results of 30 times experiments, GA is still the feature selection method with the longest computation time. CWRU datasets average computation time is shown in Fig. 17. For the proposed feature selection method, MSCGA still obtains the shortest computation time in comparison.
4) DT had the best classification results among the average results of the 30 experiments of CWRU dataset. The other average classification results are not the best in comparison. However, the highest accuracy can be obtained for each classification. The CWRU dataset result average classification results for 30 times is shown in Fig. 18.
5) Table 8 shows the set of feature selections with the highest accuracy rate for the current final classification result: GA filtered 40 features, ACO filtered 30 features, BDE filtered 64 features, MSCGA filtered 35 features, and BDE still filtered the highest number of features. BDE still has the highest number of features, but the final accuracy is not necessarily the highest. The experimental results of the two datasets are combined. The efficiency of BDE in feature selection is the worst.
6) The same five classifiers are used, KNN, RF, DA, DT, and SVM. Among them, the proposed method can get the best performance among all classifiers. The detailed statistics of the results are shown in Fig. 19.
7) In the experimental results of the CWRU dataset, the precision value of GA is 0.9048, and the recall value is 0.9065, while in MSCGA, the precision value is 0.9122, and the recall value is 0.9108. Both values have been significantly improved. The proposed method was verified to improve the feature selection capability.
8) In the CWRU dataset experiments, it can be concluded that the MSCGA can still obtain nearly half of the computation time compared to GA, and still maintain the final accuracy. In the DT classifier, even the best accuracy can be obtained. Display MSCGA has more feature selection advantages than GA.

VI. CONCLUSION

The contribution of this study is to propose a new bearing fault diagnosis model. In the feature extraction step, the methods used in the model are HHT, EA, and VMD to extract 70 features from the original signals. In the feature selection step, a new feature selection method, MSCGA, is proposed, which is further improved based on GA and can significantly reduce the computation time while maintaining the original feature selection accuracy. Finally, SVM, KNN, RF, DA and DT are used as data subset classifiers under different feature selection methods. Three different datasets are used to test and verify the capability and robustness of the model. In the first case, the data and the UCI benchmark dataset were used to validate the performance of the feature selection method. The experimental results demonstrate that 8 of the 11 UCI data selected achieve the best astringency with the lowest average error. The proposed method achieves the shortest computation time in all experimental comparisons. In all experimental comparisons. The proposed method achieves the shortest computation time in all experimental comparisons, even by nearly half. In the second case, a bearing data set measured by an induction motor is applied to test a diagnostic model. From the results, it is clear that although the best case is not the best performer compared to the other methods, the best results can be obtained in terms of the final average accuracy. The computation time is also the shortest among all the examples. Lastly, the CWRU benchmark

dataset is applied to validate the model's capabilities in the third case. In this case, the proposed models all achieve the highest average correct recognition rate. The SVM method can achieve 99.4% accuracy, while the KNN method achieves 99.14% accuracy. The best results can be achieved in terms of operation time. According to the final average experimental results, the SVM classifier is the most informative classifier. Finally, the proposed model is more effective in selecting features to improve the final accuracy, which can effectively improve the accuracy results. In addition, the computation time can be significantly reduced while improving the classification results. In turn, the efficiency of the work can be improved. The improvement of precision and recall values also proves that MSCGA is more capable of feature selection than GA.

REFERENCES

[1] C. Lee, T. Le, and Y. Lin, "A feature selection approach hybrid grey wolf and heap-based optimizer applied in bearing fault diagnosis," *IEEE Access*, vol. 10, pp. 56691–56705, 2022.

[2] B. Cai, Z. Wang, H. Zhu, Y. Liu, K. Hao, Z. Yang, Y. Ren, Q. Feng, and Z. Liu, "Artificial intelligence enhanced two-stage hybrid fault prognosis methodology of PMSM," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 7262–7273, Oct. 2022.

[3] B. Cai, Y. Liu, and M. Xie, "A dynamic-Bayesian-network-based fault diagnosis methodology considering transient and intermittent faults," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 1, pp. 276–285, Jan. 2017.

[4] C.-Y. Lee and T.-A. Le, "Identifying faults of rolling element based on persistence spectrum and convolutional neural network with ResNet structure," *IEEE Access*, vol. 9, pp. 78241–78252, 2021.

[5] X. Kong, B. Cai, Y. Liu, H. Zhu, C. Yang, C. Gao, Y. Liu, Z. Liu, and R. Ji, "Fault diagnosis methodology of redundant closed-loop feedback control systems: Subsea blowout preventer system as a case study," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 3, pp. 1618–1629, Mar. 2023.

[6] Z. Huo, Y. Zhang, L. Shu, and M. Gallimore, "A new bearing fault diagnosis method based on fine-to-coarse multiscale permutation entropy, Laplacian score and SVM," *IEEE Access*, vol. 7, pp. 17050–17066, 2019.

[7] Y. Tang, Y.-Q. Zhang, and Z. Huang, "Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 4, no. 3, pp. 365–381, Jul. 2007.

[8] S. He, R. Zhang, Q. Wang, Y. Chen, T. Yang, Z. Feng, Y. Zhang, M. Shao, and Y. Li, "A P300-based threshold-free brain switch and its application in wheelchair control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 715–725, Jun. 2017.

[9] M. Peker, A. Arslan, B. Sen, F. V. Çelebi, and A. But, "A novel hybrid method for determining the depth of anesthesia level: Combining ReliefF feature selection and random forest algorithm (ReliefF+RF)," in *Proc. Int. Symp. Innov. Intell. Syst. Appl. (INISTA)*, Sep. 2015, pp. 1–8.

[10] A. A. Raweh, M. Nassef, and A. Badr, "A hybridized feature selection and extraction approach for enhancing cancer prediction based on DNA methylation," *IEEE Access*, vol. 6, pp. 15212–15223, 2018.

[11] S. Wang, C. Gao, Q. Zhang, V. Dakulagi, H. Zeng, G. Zheng, J. Bai, Y. Song, J. Cai, and B. Zong, "Research and experiment of radar signal support vector clustering sorting based on feature extraction and feature selection," *IEEE Access*, vol. 8, pp. 93322–93334, 2020.

[12] C. Lee and T. Le, "Intelligence bearing fault diagnosis model using multiple feature extraction and binary particle swarm optimization with extended memory," *IEEE Access*, vol. 8, pp. 198343–198356, 2020.

[13] F. Xu, X. Song, K. Tsui, F. Yang, and Z. Huang, "Bearing performance degradation assessment based on ensemble empirical mode decomposition and affinity propagation clustering," *IEEE Access*, vol. 7, pp. 54623–54637, 2019.

[14] D. Donnelly, "The fast Fourier and Hilbert–Huang transforms: A comparison," in *Proc. Multiconf. 'a Eng. Syst. Appl.'*, Oct. 2006, pp. 84–88.

[15] C. Zhang, Y. Zhang, C. Hu, Z. Liu, L. Cheng, and Y. Zhou, "A novel intelligent fault diagnosis method based on variational mode decomposition and ensemble deep belief network," *IEEE Access*, vol. 8, pp. 36293–36312, 2020.

[16] E. O. Brigham and R. E. Morrow, "The fast Fourier transform," *IEEE Spectr.*, vol. S-4, no. 12, pp. 63–70, Dec. 1967.

[17] K. Wallis, G. Akers, P. Collins, R. Davis, A. Frazier, M. Oxley, and A. Terzuoli, "Complex empirical mode decomposition, Hilbert–Huang transform, and Fourier transform applied to moving objects," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 4395–4398.

[18] S. Admasie, S. B. A. Bukhari, T. Gush, R. Haider, and C. H. Kim, "Intelligent islanding detection of multi-distributed generation using artificial neural network based on intrinsic mode function feature," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 3, pp. 511–520, 2020.

[19] M. Yang, K. Huang, Y. F. Yang, L. Lu, Z. Feng, and H. P. Tsai, "Hyperspectral image classification using fast and adaptive bidimensional empirical mode decomposition with minimum noise fraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1950–1954, Dec. 2016.

[20] C. Li, G. Yu, B. Fu, H. Hu, X. Zhu, and Q. Zhu, "Fault separation and detection for compound bearing-gear fault condition based on decomposition of marginal Hilbert spectrum," *IEEE Access*, vol. 7, pp. 110518–110530, 2019.

[21] A. S. Ashour, M. K. A. Nour, K. Polat, Y. Guo, W. Alsaggaf, and A. El-Attar, "A novel framework of two successive feature selection levels using weight-based procedure for voice-loss detection in Parkinson's disease," *IEEE Access*, vol. 8, pp. 76193–76203, 2020.

[22] Q. Al-Tashi, S. J. A. Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary optimization using hybrid grey wolf optimization for feature selection," *IEEE Access*, vol. 7, pp. 39496–39508, 2019.

[23] F. Dahan, K. E. Hindi, A. Ghoneim, and H. Alsalman, "An enhanced ant colony optimization based algorithm to solve QoS-aware web service composition," *IEEE Access*, vol. 9, pp. 34098–34111, 2021.

[24] N. Jin and Y. Rahmat-Samii, "Advances in particle swarm optimization for antenna designs: Real-number, binary, single-objective and multiobjective implementations," *IEEE Trans. Antennas Propag.*, vol. 55, no. 3, pp. 556–567, Mar. 2007.

[25] Y.-N. Guo, D.-W. Gong, and Z.-G. Xue, "Hybrid optimization method based on genetic algorithm and cultural algorithm," in *Proc. 6th World Congr. Intell. Control Autom.*, 2006, pp. 3471–3475.

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

[27] Z. Chai, X. Yang, Z. Liu, Y. Lei, W. Zheng, M. Ji, and J. Zhao, "Correlation analysis-based neural network self-organizing genetic evolutionary algorithm," *IEEE Access*, vol. 7, pp. 135099–135117, 2019.

[28] Q. Fan, P. Wang, J. Liu, H. Zhang, and X. Gao, "Study on mutation operator based on BP algorithm in genetic algorithm with floating coding," in *Proc. 6th World Congr. Intell. Control Autom.*, 2006, pp. 3260–3264.

[29] D. Jiang, W. Zang, R. Sun, Z. Wang, and X. Liu, "Adaptive density peaks clustering based on K-nearest neighbor and Gini coefficient," *IEEE Access*, vol. 8, pp. 113900–113917, 2020.

[30] J. Ye, R. Janardan, C. H. Park, and H. Park, "An optimization criterion for generalized discriminant analysis on undersampled problems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 982–994, Aug. 2004.

[31] Z. Ebrahimpour, W. Wan, A. S. Khoojine, and L. Hou, "Twin hyper-ellipsoidal support vector machine for binary classification," *IEEE Access*, vol. 8, pp. 87341–87353, 2020.

[32] R. C. Barros, M. P. Basgalupp, A. A. Freitas, and A. C. P. L. F. de Carvalho, "Evolutionary design of decision-tree algorithms tailored to microarray gene expression data sets," *IEEE Trans. Evol. Comput.*, vol. 18, no. 6, pp. 873–892, Dec. 2014.

[33] T. Kim and J. Lee, "Exponential loss minimization for learning weighted naive Bayes classifiers," *IEEE Access*, vol. 10, pp. 22724–22736, 2022.

[34] L. Dong, H. Du, F. Mao, N. Han, X. Li, G. Zhou, D. Zhu, J. Zheng, M. Zhang, L. Xing, and T. Liu, "Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning technique—Subtropical area for example," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 113–128, 2020.

[35] P. Fuentealba, A. Illanes, and F. Ortmeier, "Cardiotocographic signal feature extraction through CEEMDAN and time-varying autoregressive spectral-based analysis for fetal welfare assessment," *IEEE Access*, vol. 7, pp. 159754–159772, 2019.

[36] P. Li, J. Gao, D. Xu, C. Wang, and X. Yang, "Hilbert–Huang transform with adaptive waveform matching extension and its application in power quality disturbance detection for microgrid," *J. Mod. Power Syst. Clean Energy*, vol. 4, no. 1, pp. 19–27, Jan. 2016.
[37] W. Lu, X. Du, J. Ding, and X. Wang, "Modal parameter identification based on fast Fourier transform and Hilbert Huang transform," in *Proc. 2nd Int. Conf. Consum. Electron., Commun. Netw. (CECNet)*, Apr. 2012, pp. 2703–2706.
[38] A. Ahmadpour, A. Dejamkhooy, and H. Shayeghi, "Fault diagnosis of HTS–SLIM based on 3D finite element method and Hilbert–Huang transform," *IEEE Access*, vol. 10, pp. 35736–35749, 2022.
[39] X. Shan, S. Huo, L. Yang, J. Cao, J. Zou, L. Chen, P. G. Sarrigiannis, and Y. Zhao, "A revised Hilbert–Huang transformation to track non-stationary association of electroencephalography signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 841–851, 2021.
[40] M. Dehghani, M. Ghiasi, T. Niknam, A. Kavousi-Fard, and S. Padmanaban, "False data injection attack detection based on Hilbert–Huang transform in AC smart islands," *IEEE Access*, vol. 8, pp. 179002–179017, 2020.
[41] J. Wang, L. Qiao, Y. Ye, and Y. Chen, "Fractional envelope analysis for rolling element bearing weak fault feature extraction," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 2, pp. 353–360, Apr. 2017.
[42] Q. Mao, Y. Zhang, X. Zhang, G. Zhang, H. Fan, and K. Mushayi, "Accurate fault location method of the mechanical transmission system of shearer ranging arm," *IEEE Access*, vol. 8, pp. 202260–202273, 2020.
[43] S. A. Abufana, Y. Dalveren, A. Aghnaiya, and A. Kara, "Variational mode decomposition-based threat classification for fiber optic distributed acoustic sensing," *IEEE Access*, vol. 8, pp. 100152–100158, 2020.
[44] D. Zhang and Z. Feng, "Application of variational mode decomposition based demodulation analysis in gearbox fault diagnosis," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, May 2016, pp. 1–6.
[45] L. G. Fonseca, A. C. C. Lemonge, and H. J. C. Barbosa, "A study on fitness inheritance for enhanced efficiency in real-coded genetic algorithms," in *Proc. IEEE Congr. Evol. Comput.*, Jun. 2012, pp. 1–8.
[46] M. A. Alanezi, H. R. E. H. Bouchekara, T. A. Apalara, M. S. Shahriar, Y. A. Sha'aban, M. S. Javaid, and M. A. Khodja, "Dynamic target search using multi-UAVs based on motion-encoded genetic algorithm with multiple parents," *IEEE Access*, vol. 10, pp. 77922–77939, 2022.
[47] C. Tao, X. Wang, F. Gao, and M. Wang, "Fault diagnosis of photovoltaic array based on deep belief network optimized by genetic algorithm," *Chin. J. Electr. Eng.*, vol. 6, no. 3, pp. 106–114, Sep. 2020.
[48] S. Ernawati, E. R. Yulia, and F. Samudi, "Implementation of the Naïve Bayes algorithm with feature selection using genetic algorithm for sentiment review analysis of fashion online companies," in *Proc. 6th Int. Conf. Cyber IT Service Manage. (CITSM)*, Aug. 2018, pp. 1–5.

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

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

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