IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 25, 2021, accepted April 20, 2021, date of publication April 26, 2021, date of current version May 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3075547

Fault Diagnosis of Wind Turbines Based on a Support Vector Machine Optimized by the Sparrow Search Algorithm

WUMAIER TUERXUN[1,2], **XU CHANG**[3], **GUO HONGYU**[3], **JIN ZHIJIE**[3], **AND ZHOU HUAJIAN**[3]

[1]College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China
[2]College of Hydraulic and Civil Engineering, Xinjiang Agricultural University, Urumqi 830052, China
[3]College of Energy and Electrical Engineering, Hohai University, Nanjing 210098, China

Corresponding author: Xu Chang (zhuifengxu@hhu.edu.cn)

This work was supported in part by the Ministry of Science and Technology of Peoples Republic of China under Grant 2019YFE0104800, and in part by the National Natural Science Foundation of China under Grant U1865101.

ABSTRACT Fault diagnosis technology is key to the safe and stable operation of wind turbines. An effective fault diagnosis technology for wind turbines can quickly identify fault types to reduce the operation and maintenance costs of wind farms and improve power generation efficiency. Currently, most wind farms obtain operation and maintenance data via supervisory control and data acquisition (SCADA) systems, which contain rich information related to the operation characteristics of wind turbines. However, few SCADA systems provide fault diagnosis functionality. Support vector machines (SVMs) are a popular intelligence method in the fault diagnosis of wind turbines. SVM parameter selection is key for accurate model classification. The sparrow search algorithm (SSA) is a novel and highly efficient optimization method used to optimize the penalty factor and kernel function parameter of SVM in this paper and to construct the SSA-SVM wind turbine fault diagnosis model. Data are acquired from a wind farm SCADA system and form a faulting set after preprocessing and feature selection. Experiments show that the SSA-SVM diagnostic model effectively improves the accuracy of wind turbine fault diagnosis compared with the GS-SVM, GA-SVM and PSO-SVM models and has fast convergence speed and strong optimization ability. Moreover, the SSA-SVM diagnostic model can be used to diagnose faults in practical engineering applications.

INDEX TERMS Sparrow search algorithm (SSA), fault diagnosis, wind turbines, support vector machine (SVM), parameter optimization.

I. INTRODUCTION

In recent years, with the deteriorating global ecological environment and the gradual depletion of fossil fuels, countries around the world have increased their research efforts related to renewable energy [1], [2]. As a clean and environmentally friendly renewable energy, wind energy does not produce pollution, does not contribute to global warming, and has no known emissions or hazardous waste [3]. Therefore, wind turbine energy generation is gradually replacing traditional power sources [4], [5]. With the gradual development of wind energy resources, the proportion of wind power in the power grid is increasing, and large-scale wind turbines are frequently put into operation. According to the relevant wind power capacity statistics released by the World Wind Energy Association (WWEA) at the beginning of 2020, the total installed capacity of wind turbines in the world in 2019 reached 650.8 GW [6].

Wind farms are generally located in remote mountains or offshore areas, with inconvenient transportation and a dispersed arrangement of a great number of wind turbines, away from the control center, in a harsh working environment subject to constantly changing uncontrolled factors, all of which will increase the incidence of wind turbine failures. One European research institute has calculated statistics on wind turbine faults [7], as shown in Figure 1. The failure rate of the electrical system is the highest, followed by the electric control system and sensors. The faults that take the most time to repair are faults caused by the blades, gearbox and generator. The use of fault diagnosis technology effectively reduces

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

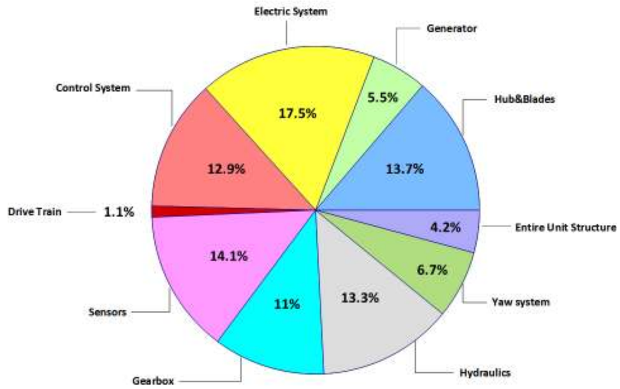

FIGURE 1. Failure rate for wind turbine unit components.

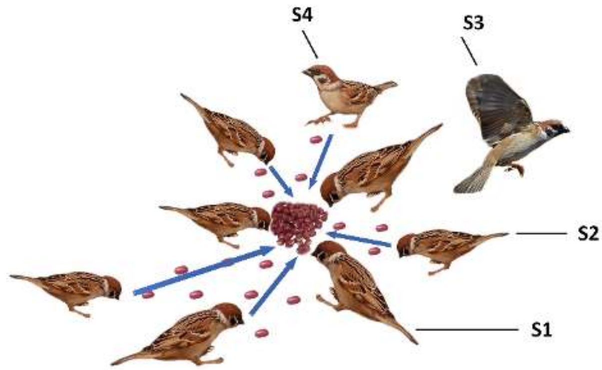

FIGURE 2. Schematic diagram of the foraging of a family sparrow group (S1-Producer, S2-Scrounger, S3-Leaver, S4-Spectator).

the incidence of faults, and fault diagnosis technology plays an important role in the safety and normal operation of wind turbines [8], [9].

Fault diagnosis technology can help operation and maintenance personnel identify abnormalities in time and address them accordingly, thereby preventing faults, which could improve the service life of the wind turbine and ensure that the wind turbine works in safe, stable and economic manner. Currently, most wind farms rely on supervisory control and data acquisition (SCADA) systems to obtain operation and maintenance data, which can provide abundant data for fault diagnosis [10], [11]. In fault diagnosis technologies, Artificial neural networks (ANNs) [12], [13], gray models (GMs) [14] and support vector machines (SVMs) [15], [16] are widely used classification algorithms for fault diagnosis; SVMs, which have the advantages of low sample demand, strong generalization ability, and high diagnostic accuracy, and are the main focus of the literature [17].

In the application of SVM, the selection of the kernel function and the parameter settings have a considerable impact on the classification effect. The RBF kernel function has been proven to produce good results for nonlinear problems [18]. Therefore, the selection of the penalty factor C and the kernel parameter g is important. Common parameter optimization methods include grid search (GS) [19], genetic algorithm (GA) [20], particle swarm optimization (PSO) [21], [22] etc. The GS method is similar to the inch-by-inch search and is applicable only to low-dimensional data with a slow search speed. The GA is an effective global optimization algorithm, but its convergence rate is slow, the encoding and decoding processes are complicated, and the parameter selection often relies on experience. The PSO algorithm easily becomes trapped into local extreme points, has a slow convergence rate and has low accuracy in late iterations.

To address the above problems, we propose a SVM fault diagnosis model for wind turbines optimized by the sparrow search algorithm (SSA-SVM). This paper makes the following contributions:

1) The SSA is used to optimize the penalty factor and kernel function parameter of the SVM to form an SSA-SVM fault diagnosis model.
2) The data collected from wind farm SCADA systems is preprocessed via cleaning and screening. Then, the characteristic quantities related to the faults of the generator and converter are extracted, and the fault sample sets are constituted using the random forest method.
3) Some fault sample sets are used as test data. Models such as SVM, GA-SVM, PSO-SVM, SSA-SVM and other diagnostic models are applied to diagnose the faults and classify the results. Finally, the performance of the various methods is compared.

The rest of this article is organized as follows. In Section 2, the fault diagnosis method based on SSA-SVM is introduced. In Section 3, we verify the feasibility of the SSA-SVM model through experimental comparison. Finally, Section 4 presents the main conclusions.

II. SSA-SVM ALGORITHM

A. SSA

The SSA, a new swarm intelligence optimization algorithm proposed by Jiankai Xue in 2020, is inspired by the foraging and antipredation behaviors of sparrows [23]. The SSA is not restricted by the differentiability, derivability and continuity of the objective function. This algorithm has the advantages of strong global search capability, good stability and fast convergence rate. The SSA is a novel well-organized metaheuristic algorithm that can be used to solve optimization problems in various fields. Zhu and Yousefi [24] used the adaptive SSA to optimize the relevant parameters of an exchange membrane fuel cell stack. Liu and Rodriguez [25] used the adaptive SSA to optimize the energy load and cost of residential buildings. Yuan [26] *et al.* used the SSA to optimize the parameters of the DMPPT control system of the photovoltaic power station.

Sparrows are gregarious birds with strong memories. Generally, two different types of sparrows exist in a flock, producers and the scroungers [27], as shown in Figure 2. Producer *S1* has a high energy reserve, and scrounger *S2* has a relatively low energy reserve. In the flock, producers are mainly responsible for finding a foraging area and providing directions for the whole flock, while scroungers use producers to obtain food. Generally, individual behaviors of producers and scroungers can be transformed [28], [29].

During the foraging process, individuals monitor the behavior of other individuals. Some sparrows take charge of observing the surroundings, while the rest hunt for food, and sparrow may alternate among these behaviors at any time [30], [31]. In Figure 2, sparrow *S4* observes the surroundings, while the remaining sparrows continue to feed and keep an eye on sparrow *S4*. If *S4* gives a warning signal, the whole flock will immediately flee from danger and fly to another safe area for food. Sparrows at the edge are more vulnerable to predators, so they constantly adjust their position and move towards the center. *S3* in Figure 2 represents a sparrow at the most dangerous edge of the foraging area who is most likely to fly elsewhere.

The bionics principle of the SSA is as follows: the behavior of a sparrow in the foraging process can be abstracted as a producer-scrounger model with a scouting and warning mechanism. The producer has high adaptive ability and a wide search range to guide the flock to search and forage. To improve their fitness, the scroungers follow the producer to forage. At the same time, some of the scroungers monitor the producer for food competition or foraging to improve their own predation rate. When the entire flock is threatened by a predator or becomes aware of danger, it will engage in antipredation actions immediately.

According to the above description, we can establish the mathematical model of the SSA. Suppose there are N sparrows in a D-dimensional search space and that the position of the i-th sparrow in the D-dimensional search space is $X_i = [x_{i1}, x_{i2}, \ldots, x_{id}, \ldots, x_{iD}]$.The position of the flock X is composed of N sparrows is expressed as follows:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1d} & \ldots & x_{1D} \\ x_{21} & x_{22} & \ldots & x_{2d} & \ldots & x_{2D} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \ldots & x_{id} & \ldots & x_{iD} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \ldots & x_{Nd} & \ldots & x_{ND} \end{bmatrix}, \quad i = 1, 2, \ldots, N \tag{1}$$

where, x_{id} represents the position of the i-th sparrow in the dimension D. The fitness values F_X of all sparrows can be expressed as follows:

$$F_X = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_i \\ \vdots \\ f_N \end{bmatrix} = \begin{bmatrix} f[x_{11} & x_{12} & \ldots & x_{1d} & \ldots & x_{1D}] \\ f[x_{21} & x_{22} & \ldots & x_{2d} & \ldots & x_{2D}] \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f[x_{i1} & x_{i2} & \ldots & x_{id} & \ldots & x_{iD}] \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f[x_{N1} & x_{N2} & \ldots & x_{Nd} & \ldots & x_{ND}] \end{bmatrix}, \quad i = 1, 2, \ldots, N \tag{2}$$

In SSA, the fitness value F_X represents the energy reserve. Producers with higher energy reserves will have priority in obtaining food during the search process. Generally, producers, which account for 10% $\sim$ 20% of a flock, are responsible for finding food and have a larger foraging search range than that of scroungers. Meanwhile, producers should constantly update their position via the following expressions:

$$X_{id}^{t+1} = \begin{cases} X_{id}^t \cdot \exp\left(\dfrac{-i}{\alpha \cdot T}\right), & R_2 < ST \\ X_{id}^t + Q \cdot L, & R_2 \geq ST \end{cases} \tag{3}$$

where t current number of iterations; T represents the maximum number of iterations; $\alpha \in (0, 1]$ is a uniform random number; Q is a random number that follows a normal distribution; L represents the matrix of size $1 \times d$ with all elements being 1; and $R_2 \in [0, 1]$ and $ST \in [0.5, 1]$ represent the warning value and the safe value, respectively. When $R_2 < ST$, there are no predators around the foraging area, and producers can perform extensive search operations; when $R_2 \geq ST$, the detecting sparrow in the flock has identified a predator and immediately sends an alert to the other sparrows. The flock then engages in antipredation behavior, adjusts its search strategy, and rapidly moves towards a safe area.

In the foraging process, in addition to the producer, all sparrows act as scroungers that find the best feeding area for foraging. Scroungers update their positions according to the following formula:

$$x_{id}^{t+1} = \begin{cases} Q \cdot exp\left(\dfrac{\boldsymbol{xw}_d^t - x_{id}^t}{i^2}\right), & i > n/2 \\ \boldsymbol{xb}_d^{t+1} + \left|x_{id}^t - \boldsymbol{xb}_d^{t+1}\right| \cdot \boldsymbol{A}^+ \cdot \boldsymbol{L}, & i \leq n/2 \end{cases} \tag{4}$$

where, $\boldsymbol{xw}_d^t$ represents the worst position of a sparrow in the d-th dimension in the t-th iteration of the flock and $\boldsymbol{xb}_d^{t+1}$ denotes the optimal position of a sparrow in the d-th dimension in the $t+1$-th iteration of the flock. A denotes a $1 \times d$ matrix where each element is randomly assigned to 1 or -1, such that $A^+ = A^T(AA^T)^{-1}$. When $i > n/2$, the i-th scrounger does not obtain any food and is in a state of starvation with low fitness. This sparrow has a high probability of flying to another location to forage and obtain higher energy. When $i \leq n/2$, the i-th scrounger finds a random location near the current optimal location xb for foraging.

When danger is detected, the sparrows at the edge of the flock will quickly move to a safe area to obtain a better position, while the sparrows in the middle of the flock will walk randomly to get closer to other sparrows, according to the following mathematical expression:

$$x_{id}^{t+1} = \begin{cases} \boldsymbol{xb}_d^t + \beta \left(x_{id}^t - \boldsymbol{xb}_d^t\right), & f_i \neq f_g \\ x_{id}^t + K \cdot \left(\dfrac{x_{id}^t - \boldsymbol{xw}_d^t}{|f_i - f_w| + e}\right), & f_i = f_g \end{cases} \tag{5}$$

where, β represents the step size control parameter, which is a normally distributed random number with a mean of 0 and variance of 1; $K \in [-1, 1]$ is a random number that indicates the direction of sparrow movement and is also a

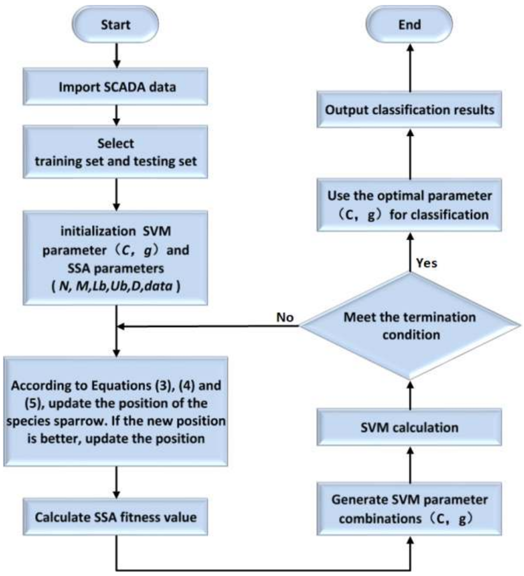

FIGURE 3. SSA-SVM flow chart.

step size control parameter; and e is a small constant to avoid the case in which the denominator is 0. f_i denotes the fitness value of the i-th sparrow, while f_g and f_w are the optimal and worst fitness values of the current sparrow flock, respectively. When $f_i \neq f_g$, the sparrow is at the edge of the flock and is vulnerable to predators; when $f_i = f_g$, the sparrow is in the middle of the flock. Once a sparrow becomes aware of the threat from predators, it will move closer to other sparrows and adjust its search strategy to avoid being attacked.

B. SVM

SVM is a machine learning algorithm based on statistical learning theory that has the advantages of fewer training samples, shorter training time and better classification effect for small clusters [32]–[34]. The core concept is that the data in the input space are mapped to a high-dimensional space by means of a kernel function, where linear classification methods can be used. Then, we construct an optimal classification hyperplane in the high-dimensional space and classify the data to maximize the inter-category interval to improve the generalization ability and confidence of the classification. The optimal hyperplane $\omega^T.x + b = 0$ can be produced by nonlinear mapping. For a linear-divisible sample set, the optimization problem can be expressed as:

$$\min_{\omega,b} \phi(\boldsymbol{\omega}, b) = \frac{1}{2}\|\boldsymbol{\omega}\|^2$$
$$s.t.\ y_i(\boldsymbol{\omega}^T x_i + b) \geq 1, \quad i = 1, 2, \ldots, n \tag{6}$$

For non-separable linear samples, SVM uses nonlinear mapping ϕ to map the data samples onto high-dimensional space for classification. The kernel function $K(x_i, y_i)$ is used to solve the problem of high-dimensional space. The kernel function satisfies Mercer's theorem, which corresponds to the inner product of the transformation space, i.e., $K(x_i, y_i) = \phi(x_i) \cdot \phi(y_i)$. The linear classification of nonlinear samples can be achieved by selecting an appropriate kernel function $K(x_i, y_i)$. Meanwhile, a relaxation variable $\xi_i(\xi_i > 0)$ is introduced to weight the classification plane, and a penalty factor C is introduced to weight the penalty degree of the relaxation variable ξ_i. Then Equation (6) can be converted to:

TABLE 1. Description of the faults.

State Codes	State Types	State Codes	State Types
1	Generator overheating	4	Excitation fault, DC link over-voltage
2	Inverter cooling system fault	5	Feeder fault
3	Power failure, under-voltage on L1 line	6	Normal

$$min\ \phi(\boldsymbol{\omega}) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{n}\xi_i$$
$$s.t.\ y_i(\boldsymbol{\omega}^T x_i + b) \geq 1 - \xi_i\ ,\ \xi_i \geq 0, \quad i = 1, 2, \ldots, n \tag{7}$$

The Lagrange multiplier α is introduced to transform the optimal hyperplane problem constructed by the constraints in Equation (7) into a dual quadratic programming problem:

$$\min L(\boldsymbol{\alpha}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i \cdot x_j) - \sum_{i=1}^{n}\alpha_i$$
$$s.t.\sum_{i=1}^{N}\alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \tag{8}$$

By solving equation (8), the optimal classification decision function is finally obtained as:

$$f(x) = \mathrm{sgn}\left[\sum_{i=1}^{n}\alpha_i y_i K\left(x_i \cdot x_j\right) + b\right] \tag{9}$$

The selection of different kernel functions has a considerable influence on the classification ability and application range of kernel functions. Common kernel functions include linear kernel functions, polynomial kernel functions, radial basis kernel functions and sigmoid kernel functions. To obtain the appropriate kernel functions, these kernel functions are tested with relevant data. The results show that the classification achieves the best effect when the radial basis kernel function is used. Moreover, the radial basis kernel function has the advantages of fewer parameters, nonlinear mapping and fast convergence speed, so it is extensively used in practice. Therefore, the radial basis function is selected as the kernel function. Then $K(x_i \cdot x_j)$ can be expressed as:

$$K(x_i \cdot x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2g^2}\right) \tag{10}$$

The penalty factor C and kernel function parameter g must be given in the SVM model, because the classification performance of the SVM is closely related to both C and g. If parameters are given artificially, improper selection will

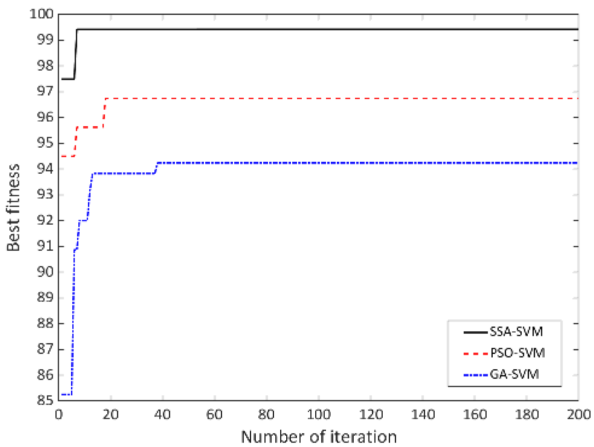

FIGURE 4. Relationship between the number of iterations and the fitness of the three diagnostic models.

easily lead to poor classification results. Therefore, penalty factor C and kernel function parameter g must be optimized. Then the optimized parameters can be used to establish the SVM model and classify samples. In this way, the SVM model achieves high classification accuracy.

C. SSA-SVM

The selection of the SVM penalty factor C and kernel function parameter g has a considerable impact on the classification results. Meanwhile, the SSA algorithm has a strong global searching capability, and it is suitable for optimizing the penalty factor C and kernel function parameter g of the SVM to obtain a better combination of parameters. By using a certain number of sparrows for global optimization, we can obtain the optimal parameter combination. Then, the optimal C and g obtained by the SSA optimization algorithm are used to establish the SVM diagnostic model and obtain the results. The process of using the SSA algorithm to optimize the SVM is shown in Figure 3.

The steps to optimize the relevant parameters of the sparrow search algorithm are as follows:

Step 1: Input SCADA data. First, preprocess the original SCADA data, select the characters and normalize them, and import the data into the algorithm.

Step 2: Select the training set and test set. Divide the imported data according to the research situation. In this paper, 60% of the imported data is selected as the training set, and 40% is used as the test set.

Step 3: Initialize SVM parameters C, g and the parameters of the SSA, including sparrow flock size N, maximum number of iterations M, upper limit of independent variables Ub, lower limit of independent variables Lb, dimension D and sample data.

Step 4: Update the sparrow's location. Calculate the sparrow's position to obtain the current new position using equations (3), (4) and (5), and update the position if the new position is better.

TABLE 2. Feature variable list and its description.

Feature Codes	Feature Description	Feature Codes	Feature Description
F1	Average wind speed (m/s)	F16	Nacelle ambient temperature1 (□)
F2	Average rotation (r/min)	F17	Nacelle ambient temperature2 (□)
F3	Average power (kw)	F18	Nacelle temperature (□)
F4	Average reactive power (kva)	F19	Nacelle cabinet temperature (□)
F5	Average blade angle (°)	F20	Main carrier temperature (□)
F6	Spinner temperature (□)	F21	Rectifier cabinet temperature (□)
F7	Front bearing temperature (□)	F22	Yaw inverter cabinet temperature (□)
F8	Rear bearing temperature (□)	F23	Fan inverter cabinet temperature (□)
F9	Pitch cabinet blade a temperature (□)	F24	Ambient temperature (□)
F10	Pitch cabinet blade b temperature (□)	F25	Tower temperature (□)
F11	Pitch cabinet blade b temperature (□)	F26	Control cabinet temperature (□)
F12	Rotor temperature1 (□)	F27	Transformer temperature (□)
F13	Rotor temperature2 (□)	F28	Inverter cabinet average temperature (□)
F14	Stator temperature1 (□)	F29	Standard error of inverter cabinet average temperature
F15	Stator temperature2 (□)		

Step 5: Calculate the fitness value. The classification accuracy of the SVM fault model is the current fitness value of each sparrow, and the maximum fitness value is updated in real time. If the current fitness value of a sparrow is greater than the saved fitness value, the original fitness value is updated; otherwise, the original fitness value remains unchanged. Therefore, the saved fitness value is the optimal value, and the combination of parameters (C, g) corresponding to the current optimal value is saved.

Step 6: Save the corresponding SVM parameter combination (C, g) according to the optimal fitness value.

Step 7: Calculate the optimal parameter combination (C, g) using the SVM method.

Step 8: Judge the termination conditions. If the termination conditions are satisfied, output the optimal parameter combination (C, g) and its corresponding classification results. If the conditions are not satisfied, increase the iteration number by 1 and return to Step 4.

III. EXPERIMENT

A. FEATURE EXTRACTION FOR SCADA DATA

Currently, most wind farms obtain operation and maintenance data from a SCADA system. To verify the performance of the SSA-SVM diagnostic model based on wind turbine faults,

TABLE 3. Samples of six state characteristics of wind turbines.

State Codes	Sample Numbers	Feature Codes F1	F2	...	F28	F29
1	1	0.3289	0.2216	...	0.7273	0.0329
	2	0.3589	0.2238	...	0.5597	0.0209
	⋮	⋮	⋮	⋮	⋮	⋮
	150	0.2862	0.0868	...	0.6306	0.0181
2	1	0.1105	0.0083	...	0.4362	0.0041
	2	0.0984	0.0033	...	0.4843	0
	⋮	⋮	⋮	⋮	⋮	⋮
	150	0.0014	0.0043	...	0.0106	0
3	1	0.0067	0.0105	...	0.0274	0.0091
	2	0.1099	0.0078	...	0.2683	0.0093
	⋮	⋮	⋮	⋮	⋮	⋮
	150	0.0058	0.0013	...	0.0152	0.0088
4	1	0.0067	0.0045	...	0.0202	0
	2	0.0039	0.0039	...	0.0138	0
	⋮	⋮	⋮	⋮	⋮	⋮
	150	0.0090	0.0049	...	0.0140	0
5	1	0.0766	0.0025	...	0.3832	0.0146
	2	0.0620	0.0000	...	0.2995	0.0117
	⋮	⋮	⋮	⋮	⋮	⋮
	150	0.0142	0	...	0.0540	0.0097
6	1	0.1416	0	...	0.2748	0.0124
	2	0.1305	0	...	0.2673	0.0118
	⋮	⋮	⋮	⋮	⋮	⋮
	150	0.1412	0.1413	...	0.0735	0.0089

this paper uses SCADA data consisting of 61 features of a wind farm in Inner Mongolia for 365 consecutive days from April 6, 2018, to April 5, 2019, as fault analysis data. Six states of wind turbines, such as generator overheating, inverter cooling system fault, power failure, generator excitation fault, feeder fault and normal state, were selected by preprocessing these data through elimination, cleaning and screening, as shown in Table 1.

For the six states of wind turbines, 61 features of SCADA data were ranked according to importance using random forests, and 29 feature quantities were extracted, as shown in Table 2.

For each state, 90 samples were selected for training, and 60 samples were selected for testing, for a total of 540 training samples and 360 test samples. The normalized eigenvalues of some samples for generator overheating, inverter cooling system fault, power failure, generator excitation fault, feeder fault and normal state are shown in Table 3.

B. COMPARISON OF DIAGNOSTIC MODEL PERFORMANCE

To further illustrate the performance of the model, SSA-SVM was compared with SVM, GA-SVM and PSO-SVM diagnostic models for optimization of the fitness values, and the results are shown in Figure 4.

Figure 4 shows that the SSA-SVM diagnostic model can quickly identify the optimal value after a few iterations. The PSO-SVM diagnostic model reached the optimal value by the 18th iteration, while the GA-SVM diagnostic model achieved the optimal value by the 42nd iteration. The SSA algorithm had the fastest optimization speed and highest diagnostic error rate after optimizing the SVM parameters, followed by the PSO-SVM algorithm and GA-SVM algorithm. The reason the original SVM is not represented in the iteration number versus fitness plot is that the original SVM did not incorporate an intelligent optimization algorithm and did not ensure fitness.

To further intuitively reflect the fault recognition ability of the SVM, GS-SVM, GA-SVM, PSO-SVM, and SSA-SVM models, the confusion matrix of each model was calculated, as shown in Figure 5. A total of 333 of 360 test samples were correctly classified by the SVM model, for an accuracy of 92.5%. However, the SSA-SVM model accurately classified 357 samples, and its accuracy reached 99.2%.

To more directly illustrate the diagnostic performance of these models, performance indicators, namely, the precision, recall, F1-score and accuracy, were calculated, and a comparison of the results is shown in Table 4. The SSA-SVM diagnostic model outperforms the other four diagnostic models.

TABLE 4. Performance indicators of fault diagnostic model.

Diagnostic Model	Evaluation Metrics	Fault Categories 1	2	3	4	5	6
SVM	Precision	100%	100%	100%	87.0%	76.9%	100%
	Recall	70.0%	85.0%	100%	100%	100%	100%
	F1-score	82.4%	91.9%	100%	93.0%	86.9%	100%
	Accuracy	92.5%					
GS-SVM	Precision	100%	100%	100%	88.2%	82.2%	100%
	Recall	78.3%	86.7%	100%	100%	100%	100%
	F1-score	87.8%	92.9%	100%	93.7%	90.2%	100%
	Accuracy	94.2%					
GA-SVM	Precision	100%	100%	100%	95.2%	82.2%	100%
	Recall	78.3%	95.0%	100%	100%	100%	100%
	F1-score	87.8%	97.4%	100%	97.5%	90.2%	100%
	Accuracy	95.6%					
PSO-SVM	Precision	100%	100%	100%	88.2%	100%	100%
	Recall	100%	86.7%	100%	100%	100%	100%
	F1-score	100%	92.9%	100%	93.7%	100%	100%
	Accuracy	97.8%					
SSA-SVM	Precision	100%	100%	100%	95.2%	100%	100%
	Recall	100%	95.0%	100%	100%	100%	100%
	F1-score	100%	97.4%	100%	97.5%	100%	100%
	Accuracy	99.2%					

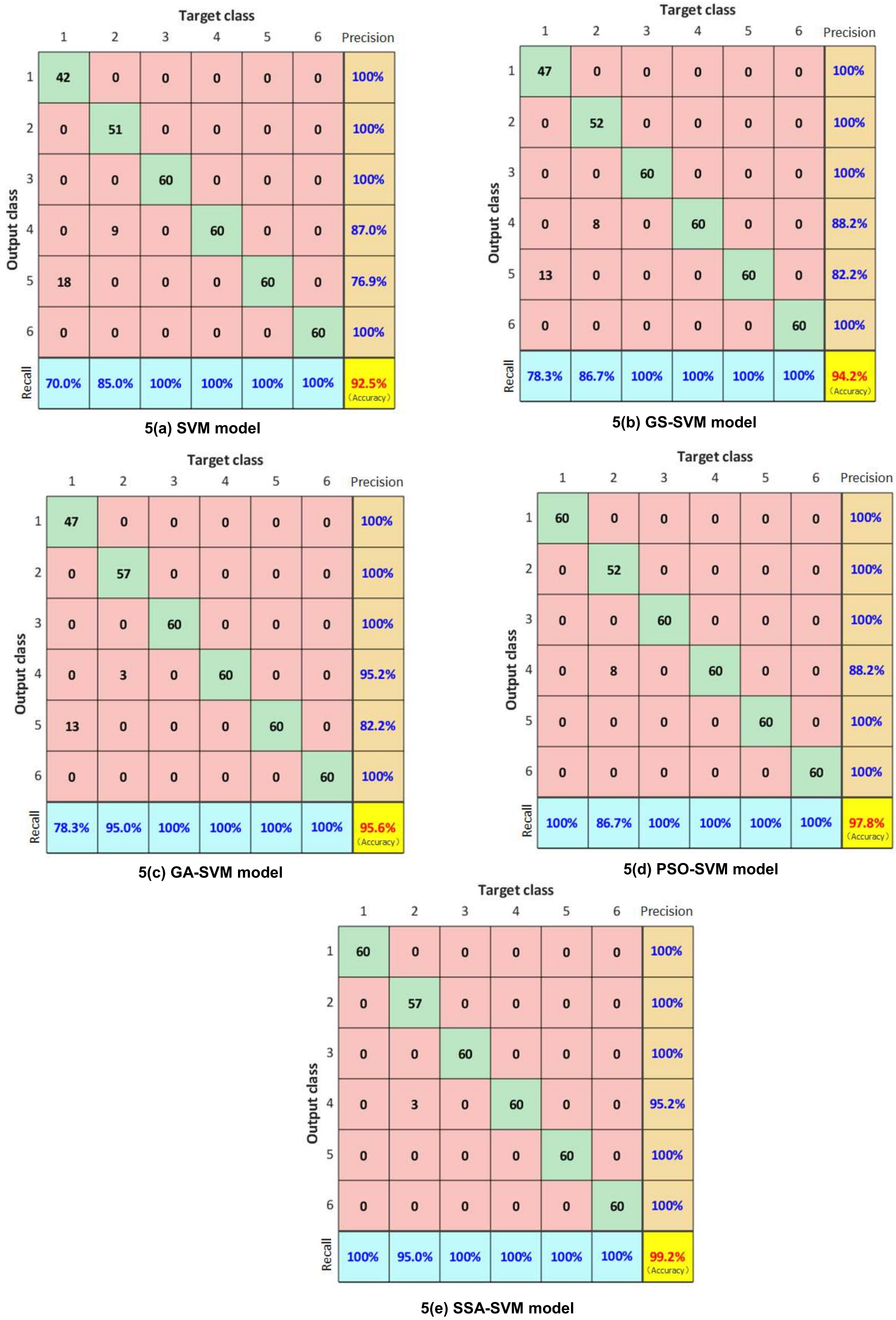

FIGURE 5. Confusion matrices of the five diagnostic models.

When the SSA-SVM model was used for fault diagnosis, the accuracy reached 99.2%, which is sufficient for the fault diagnosis task of wind turbines.

IV. CONCLUSION

The global energy shortage has led to the rapid development of clean energy both domestically and abroad. In recent years, wind energy, a green and renewable energy source, has become an essential and important force to address energy pollution. With the gradual development of wind energy resources, wind power is playing an increasingly important role in power systems. With the increasing number of wind farms, troubleshooting wind turbines has become a vital issue that requires prompt solutions. SVM is a common method used for the fault diagnosis of wind turbines.

To address the problem that improper selection of the penalty factor C and kernel function parameter g may adversely affect the SVM classification results, this paper proposes an SSA-optimized SVM wind turbine fault diagnosis model.

1) The SSA algorithm is used to optimize the penalty factor C and kernel function parameter g to construct the SSA-SVM diagnostic model. In the wind turbine fault diagnosis test, both the algorithm iteration speed and classification accuracy are superior to those of GS-SVM, GA-SVM and PSO-SVM. Finally, the classification accuracy on the test set reaches 99.2%.
2) SSA-SVM can be used in fault diagnosis of wind turbines. The model has high accuracy rate, strong optimization ability, a fast convergence rate and good convergence performance and has high practical value and wide application prospects.
3) In the typical fault diagnosis of wind turbines, although the classification accuracy of SSA-SVM is high, there are some cases of complicated faults, such as long operation time and difficulty in distinguishing the properties of faults. Exploring more methods that have better performance in extracting and selecting fault features to improve the classification accuracy and operation speed represents a major direction of future research.

REFERENCES

[1] S. Yu, Y. Zheng, and L. Li, "A comprehensive evaluation of the development and utilization of China's regional renewable energy," *Energy Policy*, vol. 127, pp. 73–86, Apr. 2019.

[2] T. Pukšec, P. Leahy, A. Foley, N. Markovska, and N. Duić, "Sustainable development of energy, water and environment systems 2016," *Renew. Sust. Energ. Rev.*, vol. 82, no. 2, pp. 1685–1690, 2018.

[3] M. L. Wymore, J. E. Van Dam, H. Ceylan, and D. Qiao, "A survey of health monitoring systems for wind turbines," *Renew. Sustain. Energy Rev.*, vol. 52, pp. 976–990, Dec. 2015.

[4] M. S. Nazir, N. Ali, M. Bilal, and H. M. N. Iqbal, "Potential environmental impacts of wind energy development: A global perspective," *Current Opinion Environ. Sci. Health*, vol. 13, pp. 85–90, Feb. 2020.

[5] M. Pettersson, K. Ek, K. Söderholm, and P. Söderholm, "Wind power planning and permitting: Comparative perspectives from the Nordic countries," *Renew. Sustain. Energy Rev.*, vol. 14, no. 9, pp. 3116–3123, 2010.

[6] WWEA. *World Wind Capacity at 650.8 GW, Corona Crisis Will Slow Down Markets in 2020, Renewables to be Core of Economic Stimulus Programmes*. [Online]. Available: https://wwindea.org/blog/2020/04/16/world-wind-capacity-at-650-gw/

[7] E. Byon, L. Ntaimo, C. Singh, and Y. Ding, *Wind Energy Facility Reliability and Maintenance*. Berlin, Germany: Springer, 2014, pp. 639–672.

[8] J. Wang, Y. Liang, Y. Zheng, R. X. Gao, and F. Zhang, "An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples," *Renew. Energy*, vol. 145, pp. 642–650, Jan. 2020.

[9] B. Hahn, M. Durstewitz, and K. Rohrig, *Reliability of Wind Turbines*. Berlin, Germany: Springer, 2007, pp. 200–332.

[10] W. Yang, R. Court, and J. Jiang, "Wind turbine condition monitoring by the approach of SCADA data analysis," *Renew. Energy*, vol. 53, pp. 365–376, May 2013.

[11] Y. Liu, Z. Wu, and X. Wang, "Research on fault diagnosis of wind turbine based on SCADA data," *IEEE Access*, vol. 8, pp. 185557–185569, 2020.

[12] X. You and W. Zhang, "Fault diagnosis of frequency converter in wind power system based on SOM neural network," *Procedia Eng.*, vol. 29, pp. 3132–3136, Jan. 2012.

[13] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: A data-mining approach," *Renew. Energy*, vol. 48, pp. 110–116, Dec. 2012.

[14] Y. B. Luo, P. Yu, B. Song, Z. H. Peng, and X. M. Lin, "Prediction of the gas dissolved in power transformer oil by the grey model," *Proc. CSEE*, vol. 21, no. 3, pp. 65–69, 2001.

[15] C. Hu, M. Huang, Q. Yang, and W. Yan, "On the use of EEMD and SVM based approach for bearing fault diagnosis of wind turbine gearbox," in *Proc. Chin. Control Decis. Conf. (CCDC)*, 2016, pp. 3538–3543.

[16] J. Saari, D. Strömbergsson, J. Lundberg, and A. Thomson, "Detection and identification of windmill bearing faults using a one-class support vector machine (SVM)," *Measurement*, vol. 137, pp. 287–301, Apr. 2019.

[17] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[18] J. Wu and H. Yang, "Linear regression-based efficient SVM learning for large-scale classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2357–2369, Oct. 2015.

[19] B. Li, M. Liu, and P. Yang, "Application of EWT and GS-SVM in mechanical fault diagnosis of circuit breakers," *Harbin Gongcheng Daxue Xuebao/J. Harbin Eng. Univ.*, vol. 39, no. 8, pp. 1422–1430, 2018.

[20] P. M. Shi, K. Liang, N. Zhao, and S. J. An, "Fault diagnosis of wind turbine gearbox bearing based on fractal dimension and GA-SVM," *Jiliang Xuebao/Acta Metrol. Sinica*, vol. 39, pp. 61–65.

[21] X. Zhang, P. Han, L. Xu, F. Zhang, Y. Wang, and L. Gao, "Research on bearing fault diagnosis of wind turbine gearbox based on 1DCNN-PSO-SVM," *IEEE Access*, vol. 8, pp. 192248–192258, 2020.

[22] D. Bai, J. He, X. Wang, and W. Tian, "Combination model for forecasting wind speed based on adaptive PSO-SVM," *Taiyangneng Xuebao/Acta Energiae Solaris Sinica*, vol. 36, no. 4, pp. 792–797, 2015.

[23] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: Sparrow search algorithm," *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 22–34, Jan. 2020.

[24] Y. Zhu and N. Yousefi, "Optimal parameter identification of PEMFC stacks using adaptive sparrow search algorithm," *Int. J. Hydrogen Energy*, vol. 46, no. 14, pp. 9541–9552, Feb. 2021.

[25] B. Liu and D. Rodriguez, "Renewable energy systems optimization by a new multi-objective optimization technique: A residential building," *J. Building Eng.*, vol. 35, Mar. 2021, Art. no. 102094.

[26] J. Yuan, Z. Zhao, Y. Liu, B. He, L. Wang, B. Xie, and Y. Gao, "DMPPT control of photovoltaic microgrid based on improved sparrow search algorithm," *IEEE Access*, vol. 9, pp. 16623–16629, 2021.

[27] C. J. Barnard and R. M. Sibly, "Producers and scroungers: A general model and its application to captive flocks of house sparrows," *Animal Behav.*, vol. 29, no. 2, pp. 543–550, May 1981.

[28] C. Johnson, J. W. A. Grant, and L.-A. Giraldeau, "The effect of handling time on interference among house sparrows foraging at different seed densities," *Behaviour*, vol. 138, no. 5, pp. 597–614, 2001.

[29] A. Liker and Z. Barta, "The effects of dominance on social foraging tactic use in house sparrows," *Animal Behav.*, vol. 77, no. 2, pp. 337–342, 2015.

[30] L. M. Bautista, J. C. Alonso, and J. A. Alonso, "Foraging site displacement in common crane flocks," *Animal Behav.*, vol. 56, no. 5, pp. 1237–1243, Nov. 1998.

[31] Á. Z. Lendvai, Z. Barta, A. Liker, and V. Bókony, "The effect of energy reserves on social foraging: Hungry sparrows scrounge more," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 271, no. 1556, pp. 2467–2472, Dec. 2004.

[32] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995, pp. 25–314.
[33] X. Zhu, J. Xiong, and Q. Liang, "Fault diagnosis of rotation machinery based on support vector machine optimized by quantum genetic algorithm," *IEEE Access*, vol. 6, pp. 33583–33588, 2018.
[34] Y. Xue, D. Dou, and J. Yang, "Multi-fault diagnosis of rotating machinery based on deep convolution neural network and support vector machine," *Measurement*, vol. 156, pp. 1–7, May 2020.

WUMAIER TUERXUN received the M.E. degree in energy and power engineering from Hohai University, Nanjing, China, in 2008, where he is currently pursuing the Ph.D. degree with the College of Water Conservancy and Hydropower Engineering. His research interests include optimal operation and fault diagnosis of wind turbines.

XU CHANG received the M.E. degree in engineering thermal physics from Northeastern University, Shenyang, China, in 2000, and the Ph.D. degree in power machinery and engineering from Southeast University, Nanjing, China, in 2005. He is currently a Professor with the College of Energy and Electrical Engineering, Hohai University, Nanjing. His current research interest includes renewable energy development and technology.

GUO HONGYU received the B.E. degree in engineering from the College of Energy and Electrical Engineering, Hohai University, Nanjing, China, in 2019, where he is currently pursuing the master's degree. His research interests include data clustering, condition monitoring, and fault diagnosis of wind turbines.

JIN ZHIJIE received the B.E. degree in engineering from the Changsha University of Science and Technology, in 2018. He is currently pursuing the master's degree with the College of Energy and Electrical Engineering, Hohai University, Nanjing, China. His research interests include condition monitoring and fault diagnosis of wind turbines.

ZHOU HUAJIAN received the B.E. degree in engineering from the College of Energy and Electrical Engineering, Hohai University, Nanjing, China, in 2020, where he is currently pursuing the master's degree. His research interests include condition monitoring and fault diagnosis of wind turbines.