

Article

Fault Detection of Wind Turbine Electric Pitch System Based on IGWO-ERF

Mingzhu Tang ^{1,†} , Jiabiao Yi ¹, Huawei Wu ^{2,*} and Zimin Wang ^{3,†}

¹ School of Energy and Power Engineering, Changsha University of Science & Technology, Changsha 410114, China; tmz@csust.edu.cn (M.T.); 20106011037@csust.edu.cn (J.Y.)

² Hubei Key Laboratory of Power System Design and Test for Electrical Vehicle, Hubei University of Arts and Science, Xiangyang 441053, China

³ School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China; worthyman@guet.edu.cn

* Correspondence: whw_xy@hbuas.edu.cn

† Mingzhu Tang and Zimin Wang contributes equally to this work and should be considered co-first authors.

Abstract: It is difficult to optimize the fault model parameters when Extreme Random Forest is used to detect the electric pitch system fault model of the double-fed wind turbine generator set. Therefore, Extreme Random Forest which was optimized by improved grey wolf algorithm (IGWO-ERF) was proposed to solve the problems mentioned above. First, IGWO-ERF imports the Cosine model to nonlinearize the linearly changing convergence factor α to balance the global exploration and local exploitation capabilities of the algorithm. Then, in the later stage of the algorithm iteration, α wolf generates its mirror wolf based on the lens imaging learning strategy to increase the diversity of the population and prevent local optimum of the population. The electric pitch system fault detection method of the wind turbine generator set sets the generator power of the variable pitch system as the main state parameter. First, it uses the Pearson correlation coefficient method to eliminate the features with low correlation with the electric pitch system generator power. Then, the remaining features are ranked by the importance of the RF features. Finally, the top N features are selected to construct the electric pitch system fault data set. The data set is divided into a training set and a test set. The training set is used to train the proposed fault detection model, and the test set is used for testing. Compared with other parameter optimization algorithms, the proposed method has lower FNR and FPR in the electric pitch system fault detection of the wind turbine generator set.

Keywords: wind turbine generator set; electric pitch system; extreme random forest; grey wolf optimization; fault detection



Citation: Tang, M.; Yi, J.; Wu, H.; Wang, Z. Fault Detection of Wind Turbine Electric Pitch System Based on IGWO-ERF. *Sensors* **2021**, *21*, 6215. <https://doi.org/10.3390/s21186215>

Academic Editors: Chunhua Yang and Qiwu Luo

Received: 4 August 2021

Accepted: 25 August 2021

Published: 16 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a renewable energy source, wind energy has many advantages, such as being pollution-free and renewable. It also has a wide distribution and large reserves, which lead to broad application prospects [1]. Because of poor working conditions, it is easy to cause damage to the components of the fan. If there is an accident, such as a shutdown caused by a fault, it will not only affect the normal operation of the wind turbine generator set, but also cost a lot in maintenance [2,3]. Therefore, it is of great significance to accurately detect the fault location of the wind turbine generator set.

As a subsystem of a large wind turbine generator set, the electric pitch system can maintain the rated power output and protect the system by controlling the adjustment of the pitch angle. Since the electric pitch system has a high failure rate, which can result in a long downtime, the fault detection becomes particularly important. Currently, the wind turbine fault detection (FD) methods can be divided into model-based fault detection and supervisory control and data acquisition (SCADA). The model-based fault detection method focuses on the combination of the internal models of the wind turbine,

while the data-based fault detection method processes the original data, and obtains the correlation analysis of the processed data by different methods. Castellani, F. [4] established the relationship between wind turbine power, voltage, and current based on principal component analysis and support vector regression. The results show that the occurrence of fault can be individualized as the change of residual behavior between model estimation and measurement. Zhao, Y.Y. [5] designed an integrated prediction and diagnosis method, which can predict the service life of wind turbines through the SCADA system and diagnose the state of wind turbines when faults occur. Li, M. [6] used multiple Small-World Neural Networks (SWNNs) and the weighted integration method to train and classify the data collected by the system. The experimental results showed that the effective accuracy rate of this integration strategy can exceed 93.8% in a short time. Shi Yiran [7] proposed a pitch angle control system for the variable-speed wind turbine generator set based on an adaptive neural network with FD and fault tolerance. As for the failure of the pitch actuator of wind turbines, Wang [8] developed a multi-scale spatio-temporal convolution deep belief network to perform feature learning and classification tasks in order to address the difficulty of the SCADA system in fault detection due to multivariate time series with spatio-temporal correlation characteristics. In view of the faults of the electric pitch system, the above method integrates multiple learning strategies to analyze and detect a specific feature, but ignores the relationship between multiple sub-components in the electric pitch system, which has certain limitations.

The FD methods based on machine learning include the Artificial Neural Network [9,10], Random Forest [11–13], LightGBM [14], Deep Learning [15,16], and the Large Margin Distribution Machine [17,18]. J. H. Pan [9] designed a data-driven method based on a deep convolutional neural network (DCNN) for the gain error, position deviation, and the faults of sensors and actuators. This method can not only reduce the model training time, but also achieve a high fault recognition accuracy. To solve the problem of low efficiency and accuracy of traditional improvement algorithms during the processing of the large data of wind turbines, Tang, M.Z. [14] proposed an adaptive LightGBM FD model of the gearbox, whose results showed that the gearbox FD method based on the adaptive LightGBM had a low FPR and a low FNR. Considering the shortcomings of traditional fault diagnosis methods, Tang, S.N. [16] summarized the rotating machinery—such as bearing, gear/gearbox, and pumps FD methods based on deep learning—and carried out its prospect forecast. Because the FD method based on machine learning can analyze and process fuzzy faults without establishing an accurate model, it can better deal with the uncertainty and suddenness of random faults in the wind turbine generator set.

Random Forest is often used for classification and regression tasks in FD models. L.J. Wan [19] proposed a high-efficiency rolling bearing FD method based on Spark and improved random forest, which can increase the detection speed and obtain a higher accuracy. Ma, S.L. [20] developed an effective identification system based on wavelet packet technology and random forest by observing that high voltage circuit breakers are prone to mechanical faults in long-term operation. Ma, Suliang [21] focused on the problem that the vibration signal of a high voltage circuit breaker is easy to ignore, and subsequently proposed a hybrid feature transformation method to optimize the fault diagnosis performance of HVCB. Firstly, a random forest binary code is designed. Secondly, the feature depth is compressed by a laminated self-coding neural network, and is finally substituted into the data set for experiments. Qin, S.Y. [22] focused on the input/output and feature selection of the SCADA system, based on the general CMDF model. The random forest algorithm is used to find the best input feature, wavelet analysis is used for noise reduction, and an RLS filter is used to reduce the False Alarm Rate. Yan, Xiangwu [23] designed a fault diagnosis method combining multi-layer neural network and random forest based on the SCADA system. Wang, X.D. [24] focused on the difficult and time-consuming problem of the DOWF short-distance transmission line location, proposing a two terminal fault location method combining ST and RF. In order to do so, the fault features are first extracted from the collected wind turbine current signal through ST

transform, and then the data are processed and transformed into the Random Forest model for diagnosis. The results show that this location method has a high fault identification rate. Mansouri, M. [25] focused on the shortcomings of the traditional random forest algorithm, and proposed a new random forest algorithm based on simplified Gaussian regression for the detection and diagnosis of the wind energy conversion system. Z.H. Xie [26] not only used RF for feature selection, but also proposed an improved cuckoo search (CS) algorithm to optimize the neural network parameters to avoid local optimum, which can significantly improve the accuracy of FD. To solve the problem of unbalanced data types in the wind turbine generator set, M.Z. Tang [10] proposed a cost-sensitive extreme random tree algorithm. Because of the randomness of RF, the actual operation of the model cannot be accurately known, and it is difficult to find a reasonable explanation for the assignment of different parameters. Therefore, the grey wolf optimization algorithm, which has a strong optimization ability, is combined with extreme random forest to optimize the FD model.

The grey wolf optimization is a swarm intelligence optimization algorithm proposed by Mirjalili in 2014, which has the advantages of a strong convergence, fewer adjustment parameters, and an easy implementation which are shown in some classical benchmark test function [27]. According to the NFL (No Free Lunch) [28] theorem, the grey wolf optimization has the shortcomings of slow convergence and easy local optimum, like the other optimization algorithm [29]. At present, the improvement of grey wolf optimization mainly includes four aspects. First, change the variation rule of parameter α . In the GWO algorithm, α decreases linearly from two to zero as the number of iterations increases. Long [30] used a logarithmic decay function to dynamically reduce the value of α to enhance the exploitation capability of the algorithm. Luis Rodriguez [31] proposed an improved grey wolf optimization based on fuzzy logic, which had obvious advantages as compared with traditional dynamic algorithms. Second, change the position update equation. Inspired by the PSO algorithm, Long [32] made full use of the location information of the remaining individuals in the wolf pack to update the location, thus enhancing the global survey capability of the algorithm and avoiding local optimum. M. Malik [33] replaced the simple arithmetic average with the weighted sum of the optimal position, finding that the multimodal function optimization effect was better. Third, change the regeneration strategy of the population. Gupta, S. [34] designed a GLF-GWO algorithm in which the alpha wolves update their position through the Levy-flight search mechanism and introduce a greed mechanism to prevent the population from falling into local optimum. Fourth, mix with other algorithms. There are many types of metaheuristic algorithms based on intelligent groups. Various algorithms have different search capabilities. Therefore, mixing with other algorithms is also an improved method. For example, Shaheen [35] proposed to integrate the particle swarm algorithm with the GWO algorithm, which can effectively find the global optimal solution of an optimization problem. Daniel, E. [36] used the cuckoo search algorithm to improve the traditional grey wolf optimization, and applied it to the field of medical image fusion. Compared with the traditional method, there was a significant improvement.

In order to solve the problem that the parameters of the electric pitch system FD of the wind turbine generator set are difficult to optimize, an extreme random forest FD model based on improved grey wolf optimization is proposed. First, the Cosine model is introduced to nonlinearize the linearly changing convergence factor α based on the model. Then, the lens imaging principle in physics is introduced. The mirror wolf of the α wolf is generated and compared with the fitness value of the α wolf, thus avoiding the local optimal solution when the population tends to gather towards the α wolf in the later stage of the algorithm. The grey wolf algorithm optimized based on the above two learning strategies is called the improved grey wolf optimization (Improved Grey Wolf Optimizer, IGWO). IGWO integrates with the extreme random forest electric pitch system FD method of the wind turbine generator set through the fitness function, and defines the accuracy of the binary-classification confusion matrix as the fitness function. IGWO will calculate the

minimum value of the fitness function at each iteration until the accuracy requirements are met, or the maximum number of iterations is reached. Then, the optimal parameters are outputted to the Extreme Random Forest model. The data set is divided into a training set and a test set. The training set is used to train the model, and the test set is used to test the model and output the predicted value. The value is compared with the test value to calculate the FPR and FNR of the electric pitch system FD of the wind turbine generator set for Extreme Random Forest optimized by the improved grey wolf optimization.

2. Materials and Methods

2.1. Grey Wolf Optimization

2.1.1. Algorithms Description

Grey Wolf Optimization is a group metaheuristic algorithm [37] which simulates the predation strategy and hierarchy of wolves in nature, and continuously searches for the optimal value in an iterative method [38]. During the predation process, the distance D between the wolf pack and the prey can be expressed by Equation (1). The wolf pack updates its position according to the distance from the prey, expressed by Equation (2):

$$D = |CX_p(t) - X(t)| \quad (1)$$

$$X(t+1) = X_p(t) - A \cdot D \quad (2)$$

where $X(t)$ is the position vector of the wolf, $X_p(t)$ is the position vector of the prey, t is the current iteration steps, and A and C are coefficient vectors. By adjusting these two vectors, the wolf can reach different positions around the prey. The calculation methods can be represented by Equations (3) and (4):

$$A = 2ar_1 - a \quad (3)$$

$$C = 2 \cdot r_2 \quad (4)$$

where a linearly decreases from 2 to 0 during the iterative process, and r_1 and r_2 are random vectors between (0, 1).

Suppose that α wolf, β wolf, and δ wolf have strong observation abilities to the potential escape position of the prey. The whole predation process is dominated by α , β , and δ wolf. The position of α wolf is the best, followed by β and δ wolf. First, determine the distances between α , β , and δ wolf and the prey according to Equations (5)–(7), and then move to the next position according to Equations (8)–(11). Finally, ω wolf updates their position according to the α , β , and δ wolf. On the basis of the above methods, the optimal solution of the optimization objective can be obtained by continuous iteration until the termination condition is satisfied.

$$D_\alpha = |C_1 \cdot X_\alpha - X| \quad (5)$$

$$D_\beta = |C_2 \cdot X_\beta - X| \quad (6)$$

$$D_\delta = |C_3 \cdot X_\delta - X| \quad (7)$$

$$X_1 = X_\alpha - D_\alpha A_1 \quad (8)$$

$$X_2 = X_\beta - D_\beta A_2 \quad (9)$$

$$X_3 = X_\delta - D_\delta A_3 \quad (10)$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (11)$$

where $D_\alpha, D_\beta, D_\delta$ represent the distances between α , β , and δ wolf and the current candidate wolf, respectively. $X_\alpha, X_\beta, X_\delta$ represent the position vector of the current population of α wolf, β wolf, and δ wolf, X represents the position vector of the grey wolf, and $X(t+1)$ represents the position vector of the next iteration.

2.1.2. Cosine Model-Based Constriction Factor Change Equation

From the Section 2.1. Equation (3), the contraction factor α in GWO changes linearly. In the GWO algorithm, the value of α determines whether the algorithm is in the development or exploration stage in the iterative process. For performance better intelligent algorithms, they should have strong exploration abilities in the initial stage of iteration and find as many global best points as possible. Therefore, the value of α should be large, and the decreasing trend should be steep. In the latter stage of the iteration, it should have strong exploitation capabilities to ensure the quality and convergence speed of the optimal solution, so the value of α should be small, and the decreasing trend should be gentle. The change trend of the cosine function can be seen from the analysis of the change trend. The dynamic cosine function model of contraction factor is shown in Formula (12):

$$a(t) = 1 + \cos(t/Maxiter \cdot \pi) \quad (12)$$

where t is the current number of iterations, and $Maxiter$ is the maximum number of iterations.

2.1.3. Grey Wolf Optimization based on Lens Imaging Learning Strategy

The principle of lens imaging is that an object is refracted by a convex lens, and an image that is opposite to the original object is generated at the other end of the convex lens. Its position and size are determined by the object distance and image distance of the convex lens, as shown in Figure 1.

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \quad (13)$$

where u is the object distance, v is the image distance, and f is the focal length of the lens.

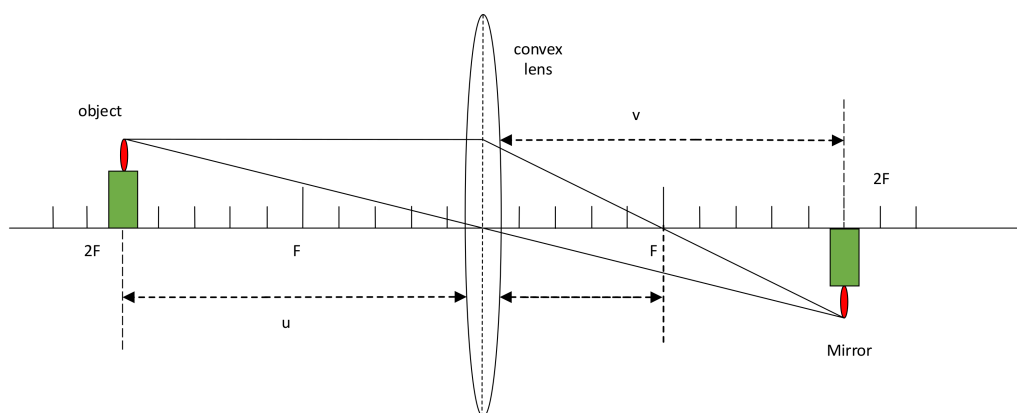


Figure 1. Schematic diagram of lens imaging.

Definition 1. Mirror point: the sample of M -dimensional space is $X = (x_1, x_2, \dots, x_M)$, $x_i \in [A_i, B_i]$, $i = 1, 2, \dots, M$. According to the principle of lens imaging, it can generate the mirror point $X' = (x'_1, x'_2, \dots, x'_M)$ and $x'_i = A_i + B_i - x_i$.

Definition 2. Base point: The Euclidean distance between the sample point and the mirror point and a certain point O in the M -dimensional space is d and d' . If d/d' is an integer, the point O is called the base point.

Suppose that the optimal solution is $X^* = (x_1^*, x_2^*, \dots, x_M^*)$. According to Definitions 1 and 2, it can be found that the mirror image point of X^* is $X^{*'} = (x_1^{*'}, x_2^{*'}, \dots, x_M^{*'})$. Suppose that the base point o_j is $(a_{j1}, a_{j2}, \dots, a_{jM})$. In this case, the mirror point $X^{*'}$ can be calculated according to Equation (12). Figure 2 shows the reverse learning strategy based on lens imaging in one-dimensional space.

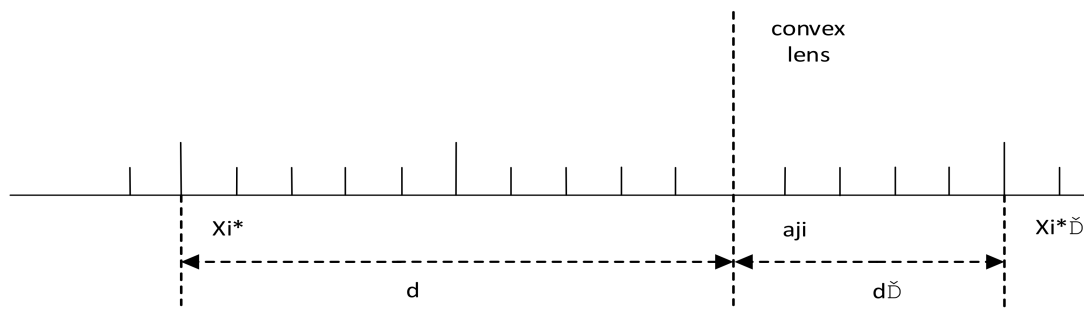


Figure 2. One dimensional inverse learning strategy diagram based on lens imaging principle.

During the iterative process of the grey wolf optimization, the population of wolves tends to converge toward the head wolf and fall into the local optimal solution. The reverse group is generated through the lens imaging learning strategy, which increases the diversity of the group and avoids the local optimum solution, thus improving the local exploitation ability in the later stage of the algorithm.

2.2. Random Forest

Random Forest (RF) is an ensemble model composed of multiple decision trees $\{h(x, \theta_k) | k = 1, 2, 3 \dots n_{tree}\}$, where θ_k is an independent and identically distributed random vector. The basic ideas and steps of the RF model are as follows.

First, RF randomly selects K times from the sample set $D(X, Y)$ through sampling with replacement, and obtains K sample subsets with the same dimensions as the sample set $[D_1, D_2, \dots, D_k]$.

Then, RF uses the CART decision tree as the weak learner. For each sample in the $N \times M$ dimensional sample set $D(X, Y)$, there are M attributes, which are selected based on the Gini coefficient. The criterion for the selection of the Gini coefficient is that each child node achieves the highest purity. In this case, the Gini coefficient is the smallest, the purity is the highest, and the uncertainty is the smallest. The calculation equation of Gini coefficient is as follows:

$$\text{Gini}(D) = \sum_{k=1}^{|y|} \sum_{k' \neq k} p_k p_{k'} = 1 - \sum_{k=1}^{|y|} p_k^2 \quad (14)$$

where p_k indicates that the proportion of the k -th sample in the current sample set $D(X, Y)$. The Gini coefficient of the attribute α is defined as:

$$G_{ini_index}(D, a) = \sum_{v=1}^V \frac{|D^v|}{|D|} G_{ini}(D^v) \quad (15)$$

Therefore, in the candidate attribute set A , the optimal partition attribute minimizes the Gini index after partitioning, namely $a^* = \text{argmin} G_{ini_index}(D, a)$.

Finally, repeat the above steps to build multiple decision trees to form a random forest. Input the prediction data into the constructed RF model where multiple decision trees make decisions at the same time, and make category decisions based on the principle that the minority obeys the majority.

2.3. Extreme Random Forest

Extreme random trees (ERT) is a derivative algorithm of RF. It is a machine learning algorithm proposed by Pierre geurts and other scholars after a lot of experimental research in 2006. Extreme random forest (ERF) is also a classifier integrated by multiple decision trees, but compared with RF, ERF is better in classification accuracy and training time, which is mainly due to its two differences from RF:

- (1) The training set of decision tree is obtained in different ways. RF adopts bagging model, which has put back randomly selected equal dimensional training set, and its training set is random, but there may be duplicate samples in the training set, which can not ensure that all samples are fully utilized, and there may be similarity between two arbitrary training sets, resulting in that the trained classifiers can not play their respective functions. The training set of ERF does not use random sampling, but uses all the original training sets, that is, each decision tree applies the same all training samples, which ensures the utilization of training samples and reduces the final prediction deviation to a certain extent.
- (2) Characteristics are divided in different ways. The decision tree of RF will select an optimal eigenvalue for division based on the principles of information gain, Gini coefficient and mean square deviation. For example, in candidate attribute set A , select the attribute that minimizes the Gini index after division as the optimal division attribute, $\text{argmin}_{G_{ini_index}}(D, a)$. ERF has strong randomness for the acquisition of splitting features and segmentation values. It randomly selects an eigenvalue for division, so that each decision tree presents structural differences

The algorithm principle of ERT is the same as RF except that the acquisition of training set and feature division are different. It scores by integrating multiple decision trees and votes according to the average value of the predicted value of each decision tree. The process is as follows:

- (1) Sample selection. Each decision tree is trained with the original data set.
- (2) Select the partition feature. ERT randomly selects an eigenvalue to divide the decision tree. For $N \times M$ dimensional sample set $D(x, y)$, given sample x_i use m dimensional eigenvector f_i represents the characteristics of the sample. Then, a partition value a_c is randomly selected between at the maximum value of the variable a_{max}^k and minimum variable of a_{min}^k partition. If the value of variable k less than the split value a_c sample ($a < a_c$), then put them in the left leaf node, and if the value of variable k is greater than or equal to the split value a_c sample ($a \geq a_c$), then put them in the right leaf node.
- (3) Build the decision tree. The formation of the decision tree is divided and split according to the division rules in step (2) until it can no longer be split.
- (4) Extreme random tree prediction. Repeat steps (1), (2) and (3) to establish a large number of decision trees and gradually form a forest until the number of iterations is met. Input the prediction data into the constructed forest, calculate the output results of each decision tree for classification or regression, and get the final classification or regression prediction results.

The selection of feature K and threshold a_c in $N \times M$ dimensional sample set $D(x, y)$ according to Formula (16). Calculating the score measurement of each feature, and selecting the highest score as the splitting feature and splitting threshold of the leaf node.

$$\text{Score}_c(k, S) = \frac{2I_c^k(S)}{H_k(S) + H_c(S)} \quad (16)$$

where $I_c^k(S)$ indicates that node S was splitted based on characteristic K and threshold a_c^k , the two subsets have mutual information about categories. $H_k(S)$ represents the splitting entropy of characteristic K . $H_c(S)$ represents the information entropy of node S about the category. Compared with Gini index and information gain, this index introduces $H_c(S)$ and $H_k(S)$ symmetry reduces the influence of class distribution on node splitting. Formula (15) represents the fractional measure of feature K on leaf node S . The final probability of each sample is the probability average of all trees, which is defined as follows:

$$P_{(c|f_i)} = \frac{1}{M} \sum_{t=1}^M P_t(c|f_i) \quad (17)$$

$$\hat{c} = \text{argmax}_c P(c|f_i) \quad (18)$$

where M is the total number of trees, f_i represents eigenvector of sample x_i , P_t is expressed conditional probability that the sample belongs to category C under the condition of in vector f_i . Formula (17) defines the classification probability of samples in the decision tree. Finally, on the extreme random forest, Formula (18) uses the voting principle to determine the category of the sample.

2.4. Wind Turbines Pitch System FD of Random Forest based on Improved Grey Wolf Optimization

Because the SCADA system records various historical data of the wind turbine during operation—including normal data and various fault data over a period of time, before the model is trained—operations such as preprocessing and feature selection should be performed on the extracted data to remove redundant variables and to reduce the complexity of model training. The steps in the solid box on the left in Figure 3 represent the preprocessing process. First, data cleaning, then Pearson correlation analysis is used to eliminate redundant variables, then Random Forest is used to sort the importance, and the features of the first N are selected to construct the data set. The steps in the light green solid box on the right in Figure 3 represents the extreme random forest fault detection model. The input of the model includes the parameters optimized by IGWO, training set and test set, and the output includes the evaluation index of FNR, FPR and fitness function ACC. Figure 4 is the fault diagnosis flow chart of wind turbine pitch system. Firstly, the steps of population initialization, parameter initialization and data preprocessing are used as the input of ERF fault detection model, and then the fitness function ACC is output to calculate the fitness function value of each individual of grey wolf population, and assign the top three to α , β , and ω wolf. Then update the position of grey wolf individual according to Formulas (1)–(11). Then, according to the lens imaging reverse learning strategy, the learning strategies were reversed to generate α' wolf and to calculate its fitness value as compared with α wolf's fitness value. If the α' wolf's fitness value was lower than the α wolf, α' Wolf was substituted for α wolf and participated in the population iteration process, A , C were updated by the original grey wolf algorithm, and a was updated according to the Cosine model. After updating the parameters, the X_α and the preprocessed data are substituted into the ERF fault detection model, calculating the fitness of grey wolf population individuals, updating the positions of α , β and δ wolf, and the above steps are repeated until the maximum iteration requirements are met.

2.4.1. Data Preprocessing

The SCADA system records the operating status of each system of the wind turbine generator set. The records mainly include wind speed, wind direction, temperature of each component, power, rotor speed, etc. Since the wind turbine operating environment is relatively harsh, the data recorded by the SCADA system has problems such as null values and abnormal values. The electric pitch system has many internal parameters: there are differences in the nature, dimension, and unit of each parameter. Therefore, after the initial cleaning of the original data extracted by the SCADA system, the Z-score method is needed to scale the data proportionally, and the data of the same parameter is processed into the interval $(-1, 1)$ with the mean $u = 0$ and the variance $\sigma = 1$. The Equation (19) is the mathematical expression of the Z-score, and the variance calculation is shown in Equation (20).

$$z = \frac{x - u}{\sigma} \quad (19)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (20)$$

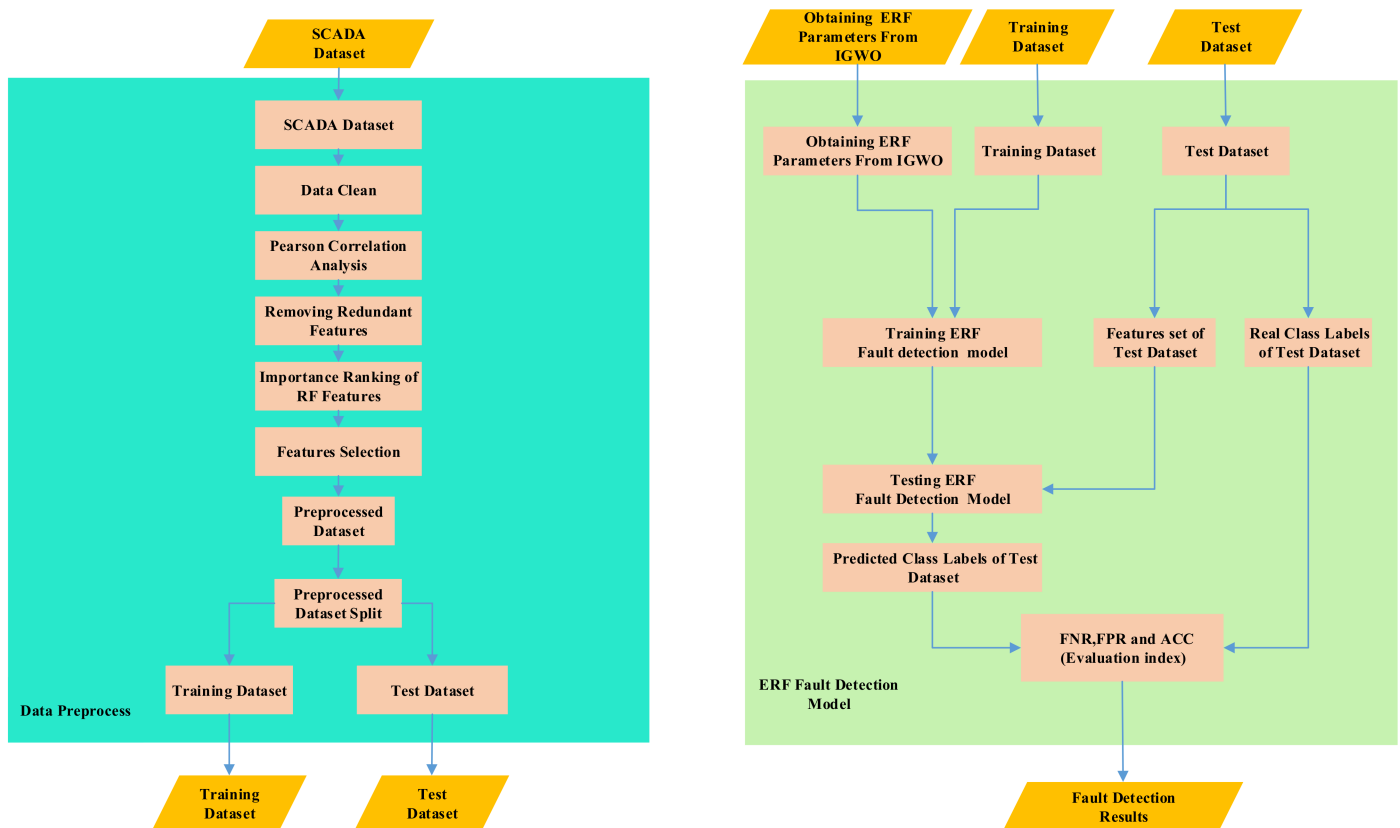


Figure 3. The flow chart of Data Preprocess and ERF Fault Detection Model.

During the optimization of the Extreme Random Forest parameters using the improved grey wolf optimization, the above two parameters (Table 1) are set into a two-dimensional vector $X(t + 1)$ ($n_estimators$, $min_samples_leaf$). The fitness of each individual is calculated during each iteration of the algorithm. If the fitness after iteration is better than the current stage, replace it. Otherwise, discard the current state vector and proceed to the next iteration until the maximum number of iterations is met or the accuracy requirement is met. The pseudo code of the optimized Extreme Random Forest parameters based on improved grey wolf optimization is shown in Algorithm 1.

Algorithm 1. IGWO algorithm optimizes Extreme Random Forest**Input:** $N, Maxiters, k, t$ **Output:** X_α

1. $X_\alpha, X_\beta, X_\delta, X_w \leftarrow X_0$
2. **repeat**
3. Surround, The parameters A and C are calculated according to Formulas (3) and (4)
4. Hunting, X_{t+1} is calculated according to Formulas (3) and (4),
 $X_{t+1}(n_estimators, min_samples_leaf) \leftarrow X_t(X_\alpha, X_\beta, X_\delta, X_w, A, C)$
5. $RF \leftarrow X_{t+1}$, training set
6. $f(X_\alpha) \leftarrow ACC \leftarrow RF \leftarrow$ test set
7. Track, Assign according to the greed criterion,
 $BestX_\alpha, BestX_\beta, BestX_\delta \leftarrow X_\alpha, X_\beta, X_\delta, X_P$
8. $X'_\alpha \leftarrow BestX_\alpha$
9. $ERF \leftarrow (X'_\alpha, \text{training set})$
10. $f(X'_\alpha) \leftarrow ACC \leftarrow RF \leftarrow$ test set
11. **If** $f(X'_\alpha) < f(X_\alpha)$ **then**
12. $BestX_\alpha \leftarrow X'_\alpha$
13. **end if**
14. **until** $Maxiter$
15. **print** X_α
16. $ERF(n_estimators, min_samples_leaf) \leftarrow BestX_\alpha$
17. $ERF(n_estimators, min_samples_leaf) \leftarrow$ training set
18. $Y_pred \leftarrow RF(n_estimators, min_samples_leaf) \leftarrow$ test set
19. $FPR, FNR \leftarrow confusion_matrix(Y_pred, Y_test)$

Table 1. RF parameter: $n_estimators$ and min_sample_leaf .

Parameter	Meaning	Value Range
$n_estimators$	The number of weak classifiers parameter, which is used to adjust the number of trees	(4200)
min_sample_leaf	Minimum leaf node sample number, which is used to adjust the minimum sample number of leaf nodes of the base classifier	(1300)

2.4.2. Fault Detection Performance Evaluation Index

To verify the effectiveness of the optimized Extreme Random Forest model based on the improved grey wolf, the optimized Extreme Random Forest based on the Particle Swarm Optimization (PSO) [39], Sine and Cosine Optimization Algorithm (SCA) [40], Harris Hawks Optimization (HHO) [41], and Improved Grey Wolf Optimization (IGWO) are applied to the electric pitch system FD model of wind turbines for comparison. The research objects are the high temperature fault of the pitch super capacitor, the high temperature fault of the pitch shaft box, and the main power supply fault of the pitch. For the binary-classification problem of the electric pitch system FD of wind turbines, a confusion matrix is introduced, as shown in Table 2. The false positive rate and the false negative rate of the model are used as evaluation indexes. Equation (21) represents the false positive rate, and Equation (22) indicates the false negative rate.

$$FPR = \frac{FP}{TN + FP} \quad (21)$$

$$FNR = \frac{FN}{TP + FN} \quad (22)$$

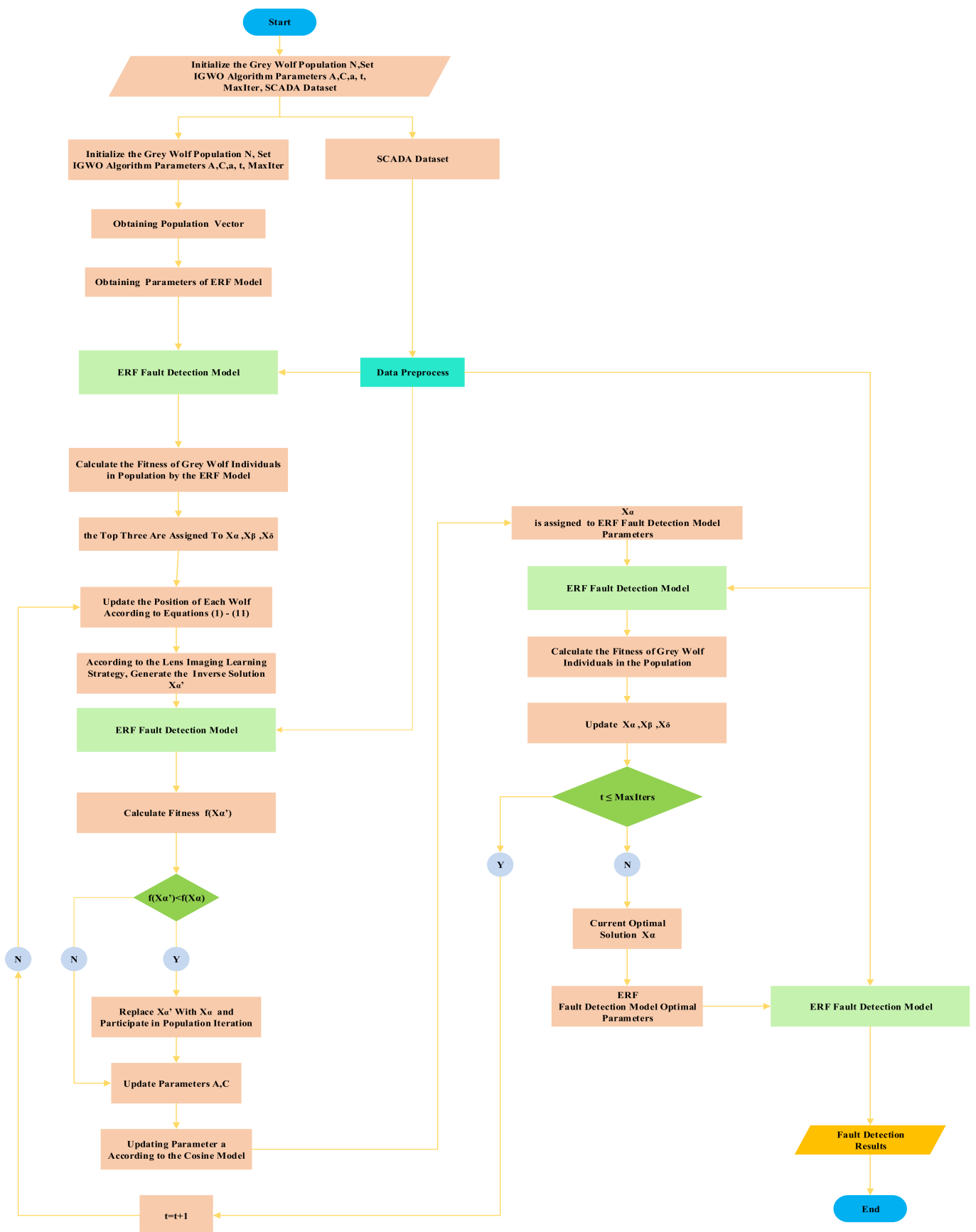


Figure 4. FD flow chart of wind turbine electric pitch system based on IGWO-ERF.

Table 2. Confusion matrix of dichotomies problem.

The Actual Category	The Predict Category	
	Normal	Fault
Normal	TP	FN
Fault	FP	TN

2.5. Experimental Analysis

2.5.1. Data Description

The experiment selects the actual operating data of a wind farm in Inner Mongolia from January to June 2021. There are 33 sets of 1.5 MW double-fed wind turbines in the wind farm, and the data sampling interval is 60 s. Partial operation data is shown in Table 3. Because the working conditions of the wind turbine generator set are easily affected by the surrounding environment—especially the uncertainty of wind speed changes [42]—the electric pitch system should automatically adjust the angle between the blades and the wind direction in accordance with the wind speed, so that the generator has a rated output power and to ensure the safe operation of the entire wind turbine generator set.

Table 3. Partial operation data of fan No. A16 on 1 February 2021.

State Parameter	Time									
	00:00:00	00:01:00	00:02:00	...	08:27:00	08:28:00	08:29:00	...	08:48:00	08:49:00
rotor_speed/(r·m ⁻¹)	17.383	17.338	17.586	...	16.88	17.492	17.538	...	17.567	17.246
converter_motor_speed/(r·m ⁻¹)	1746.6	1751	1744.8	...	1696.1	1757.6	1762.2	...	1765.1	1732.9
...
pitch_ssb_motor_current_2/A	15.686	4.412	9.314	...	19.1176	3.9216	4.4118	...	2.9412	13.725

2.5.2. Sample Feature Selection

First, the collected data is performed with data processing to delete the data whose state quantity is “0” and unchanged. The Z-score [43] is used for normalization processing. The Pearson correlation coefficient is used to conduct the correlation analysis with Converter_power. The results are shown in Table 4. Because the electric pitch system controls the speed of the wind wheel by controlling the angle of the blades, thus controlling the output power of the wind turbine, the output power of the electric pitch system is selected as the main electric of the Pearson correlation analysis.

Table 4. Pearson correlation coefficient between each parameter and Converter_power.

State Parameter	Pearson Correlation Coefficient	State Parameter	Pearson Correlation Coefficient
rotor_speed/(r·m ⁻¹)	0.83827	generator_winding_temperature_u1/°C	0.67398
converter_motor_speed/(r·m ⁻¹)	0.83846	pitch_ssb_motor_current_1/A	0.47505
...
nacelle_temperature/a	−0.75862	hydraulic_main_sys_pressure/N	−0.35687

The analysis of Table 4 shows that the correlation between different characteristic variables and Converter_power is different. The characteristic variables with an absolute value of a Pearson correlation coefficient greater than 0.6 (such as the bolded part in Table 4) are retained, and redundant variables with a lower correlation are eliminated. The RF is used to conduct the importance ranking of the data set after the Pearson screening. The top N = 8 features are selected as the main influencing factors of the electric pitch system failure. The feature importance ranking of RF is shown in Figure 5.

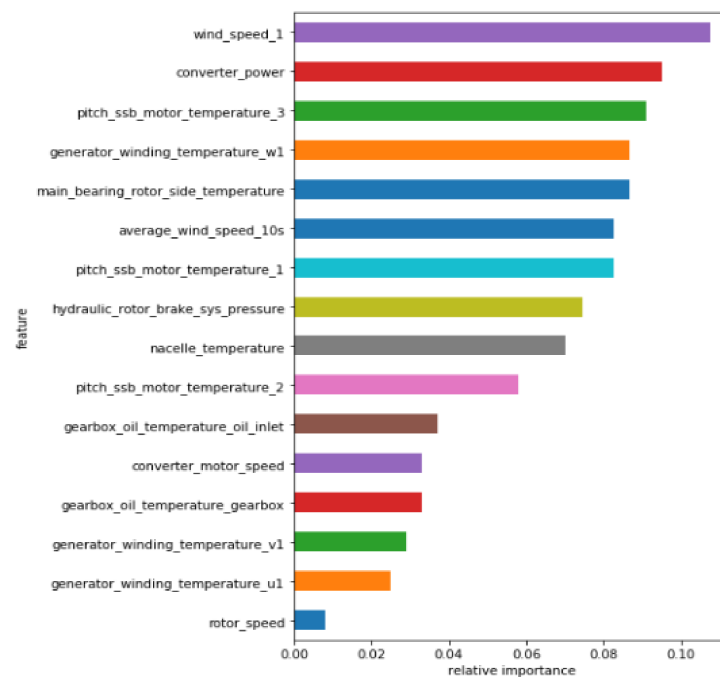
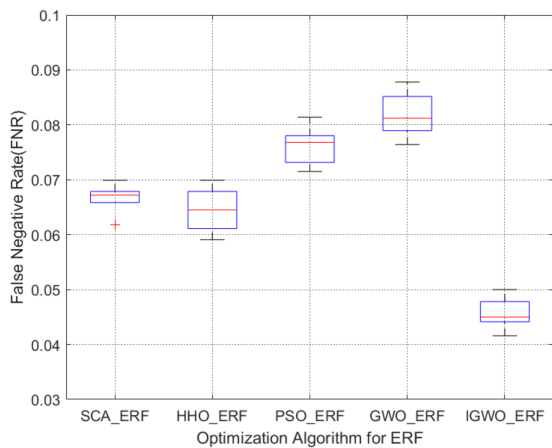


Figure 5. Random forest feature importance ranking diagram.

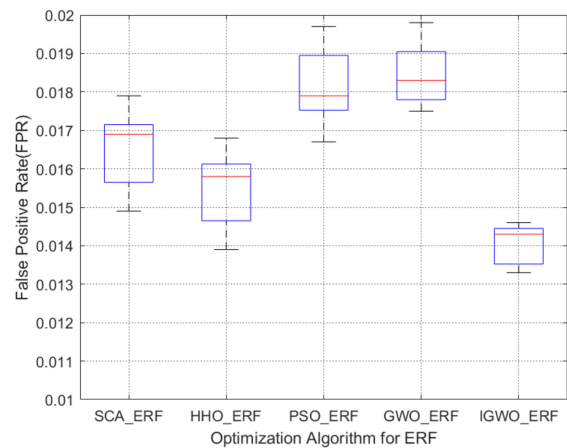
3. Results

To compare IGWO with SCA, HHO, and PSO, these algorithms are substituted into the wind turbine generator set electric pitch system FD. The initial populations of the four optimization algorithms are all set to 30, the maximum number of iterations is set to 500, and the experiments are performed 10 times. The FPR and the FNR of the overall experiment are used to draw a box plot. The high temperature fault of the pitch supercapacitor are shown in Figure 6a,b. The high temperature failures of the pitch shaft box are shown in Figure 6c,d. The main power supply failures of the pitch are shown in Figure 6e,f.

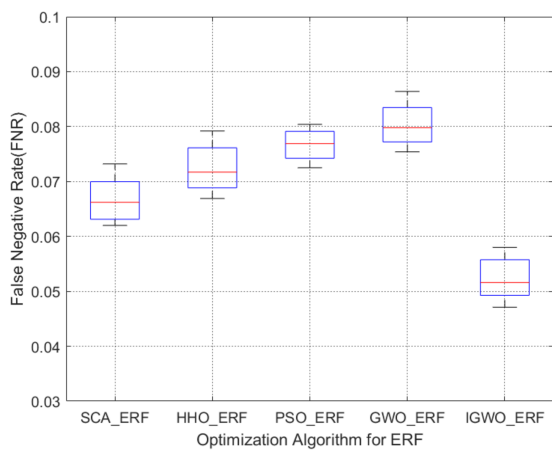
From the box diagrams of FPR and FNR of the three types of faults, it can be known that the FD model of the wind turbine pitch system with four optimization algorithms optimized with Extreme Random Forest parameters has lower false positive rates and false negative rates and, compared with GWO, IGWO has a better optimization effect. The false negative rates of the three types of faults are all smaller than 10%, and the false positive rates are all smaller than 2.5%. For the high temperature fault of the pitch supercapacitor, the false positive rates of SCA, HHO, and PSO optimization algorithms are within 6–8%, and the false negative rates are within 1.5–2%. The false positive rate of IGWO is within 4–5% and the false positive rate is within 1.3–1.5%. The experimental results of the high temperature fault of the rotor axle box show that the Extreme Random Forest effect of PSO optimization is not obvious, the optimization effect is very weak, and the overall optimization performance of IGWO is good. The results of the FD of the main power supply of the pitch change show that, compared with SCA, HHO, and PSO optimization algorithms, the false negative rate of IGWO is reduced by 1–5%, and the false positive rate is reduced by 0.2–0.4%. Based on the above analysis, it is evident that IGWO has a lower false positive rate and false negative rate as compared with SCA, HHO, and PSO, indicating that the wind turbine generator set electric pitch system FD model of Extreme Random Forest, whose parameters are optimized by the IGWO, has a better performance.



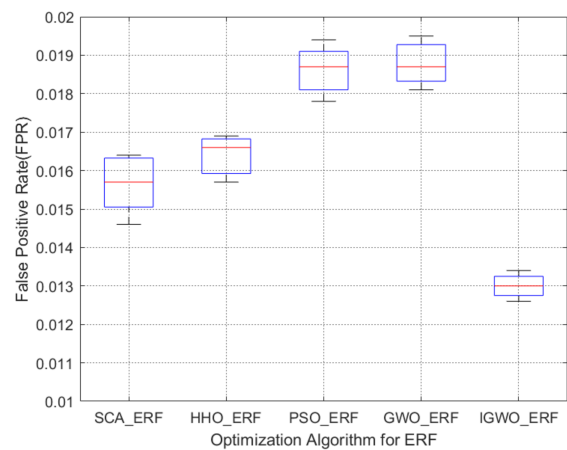
(a)



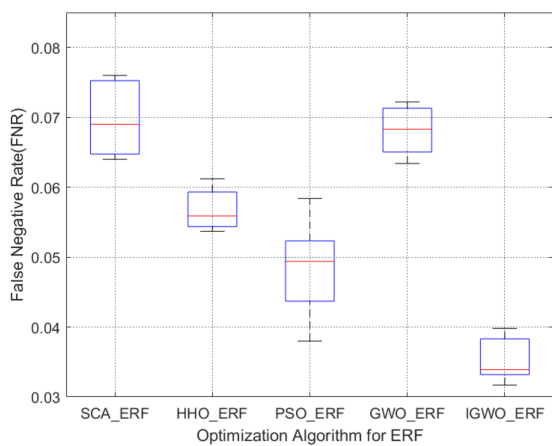
(b)



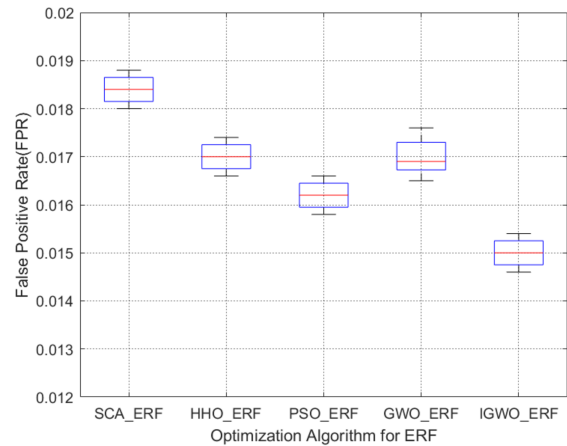
(c)



(d)



(e)



(f)

Figure 6. (a) The FNR box diagram of the supercapacitor fault; (b) the FPR box diagram of the supercapacitor fault; (c) the FNR box diagram of electric pitch axle box fault; (d) the FPR box diagram of electric pitch axle box fault; (e) the FNR box diagram of electric pitch main power; and (f) the FPR box diagram of electric pitch main power.

4. Conclusions

It is difficult to optimize the parameters of a double-fed wind turbine generator set electric pitch system fault model in Extreme Random Forest detection. To solve this problem, an improved double-fed wind turbine generator set electric pitch system FD model that optimizes the parameters of the Extreme Random Forest is proposed. The main contributions of this paper are threefold. First, it introduces the Cosine model. Based on the change of the convergence factor α in the algorithm iteration process, the Cosine model is integrated with the GWO to balance the exploitation and exploration capabilities. Second, it introduces the lens imaging reverse learning strategy. The optimal solution of α wolf will generate its mirror wolf based on the lens imaging learning strategy. This strategy compares the fitness values to avoid the algorithm falling into a local optimal solution at the later stage of the iteration. Third, the above two learning strategies are combined to propose an improved grey wolf optimization, which is combined with Extreme Random Forest and applied to the wind turbine generator set electric pitch system FD model. The improved grey wolf optimization, sine cosine optimization algorithms, Harris hawks optimization algorithm, and particle swarm optimization algorithm are combined with Extreme Random Forest, respectively. The accuracy of the confusion matrix is used as the fitness function, and the false positive rate and false negative rate are used as the evaluation index. The experimental results show that the improved grey wolf optimization has a lower false positive rate and a lower false negative rate, indicating that the wind turbine generator set electric pitch system FD method of Extreme Random Forest optimized by the improved grey wolf optimization has good performance. The reason why the IGWO–RF model has good performance in the fault detection of the wind turbine electric pitch system is that the RF model has a reasonable logical design, and the RF without any optimization algorithm can achieve good results in the field of fault detection. After the IGWO–RF model optimizes Extreme Random Forest, the model performance has been improved, but the qualitative change of magnitude has not necessarily been reached. The original data in the SCADA system is subject to a high degree of processing to use it, and the data preprocessing method is more, which cannot unite a measure. In view of the above problem, we still need to spend more time studying and formulating conforms to the engineering practice of the electric pitch system fault detection model. There is still a long way to go.

Author Contributions: Formal analysis, H.W.; investigation, Z.W.; supervision, M.T.; writing—original draft, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 62173050 and 61403046), the Natural Science Foundation of Hunan Province, China (Grant No. 2019JJ40304), Changsha University of Science and Technology “The Double First Class University Plan” International Cooperation and Development Project in Scientific Research in 2018 (Grant No. 2018IC14), the Research Foundation of the Education Bureau of Hunan Province (Grant No. 19K007), Hunan Provincial Department of Transportation 2018 Science and Technology Progress and Innovation Plan Project (Grant No. 201843), Energy Conservation and Emission Reduction Hunan University Student Innovation and Entrepreneurship Education Center, Innovative Team of Key Technologies of Energy Conservation, Emission Reduction and Intelligent Control for Power-Generating Equipment and System, CSUST, Hubei Superior and Distinctive Discipline Group of Mechatronics and Automobiles(XKQ2021003 and XKQ2021010), Major Fund Project of Technical Innovation in Hubei (Grant No. 2017AAA133), and Guangxi Key Laboratory of Trusted Software (No. 201728), Graduate Scientific Research Innovation Project of Changsha University of Science and Technology (No. 2021-89).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, J.; Fang, L.; Song, D.; Su, M.; Yang, X.; Huang, L.; Joo, Y.H. Review of control strategy of large horizontal-axis wind turbines yaw system. *Wind Energy* **2021**, *24*, 97–115. [[CrossRef](#)]
2. Chen, J.; Wang, F.; Stelson, K.A. A mathematical approach to minimizing the cost of energy for large utility wind turbines. *Appl. Energy* **2018**, *228*, 1413–1422. [[CrossRef](#)]
3. Godwin, J.L.; Matthews, P. Classification and Detection of Wind Turbine Pitch Faults through SCADA Data Analysis. *Int. J. Progn. Health Manag.* **2013**, *4*. [[CrossRef](#)]
4. Castellani, F.; Astolfi, D.; Natili, F. SCADA Data Analysis Methods for Diagnosis of Electrical Faults to Wind Turbine Generators. *Appl. Sci.* **2021**, *11*, 3307. [[CrossRef](#)]
5. Zhao, Y.Y.; Li, D.S.; Dong, A.; Kang, D.H.; Lv, Q.; Shang, L. Fault Prediction and Diagnosis of Wind Turbine Generators Using SCADA Data. *Energies* **2017**, *10*, 1210. [[CrossRef](#)]
6. Li, M.; Wang, S.X. Dynamic Fault Monitoring of Pitch System in Wind Turbines using Selective Ensemble Small-World Neural Networks. *Energies* **2019**, *12*, 3256. [[CrossRef](#)]
7. Yiran, S.; Shoutao, L.; Shuangxin, W. Pitch angle control with fault diagnosis and tolerance for wind turbine generation systems. *Proc. Inst. Mech. Eng.* **2021**, *235*, 1355–1366.
8. Wang, H.; Wang, H.B.; Jiang, G.Q.; Wang, Y.L.; Ren, S. A Multiscale Spatio-Temporal Convolutional Deep Belief Network for Sensor Fault Detection of Wind Turbine. *Sensors* **2020**, *20*, 3580. [[CrossRef](#)]
9. Pan, J.H.; Qu, L.L.; Peng, K.X. Sensor and Actuator Fault Diagnosis for Robot Joint Based on Deep CNN. *Entropy* **2021**, *23*, 751. [[CrossRef](#)] [[PubMed](#)]
10. Wang, J.R.; Li, S.M.; Han, B.K.; An, Z.H.; Bao, H.Q.; Ji, S.S. Generalization of Deep Neural Networks for Imbalanced Fault Classification of Machinery Using Generative Adversarial Networks. *IEEE Access* **2019**, *7*, 111168–111180. [[CrossRef](#)]
11. Wang, H.T.; Feng, D.G.; Liu, K. Fault detection and diagnosis for multiple faults of VAV terminals using self-adaptive model and layered random forest. *Build. Environ.* **2021**, *193*, 107667. [[CrossRef](#)]
12. Fezai, R.; Dhibi, K.; Mansouri, M.; Trabelsi, M.; Hajji, M.; Bouzrara, K.; Nounou, H.; Nounou, M. Effective Random Forest-Based Fault Detection and Diagnosis for Wind Energy Conversion Systems. *IEEE Sens. J.* **2021**, *21*, 6914–6921. [[CrossRef](#)]
13. Tang, M.Z.; Chen, Y.T.; Wu, H.W.; Zhao, Q.; Long, W.; Sheng, V.S.; Yi, J.B. Cost-Sensitive Extremely Randomized Trees Algorithm for Online Fault Detection of Wind Turbine Generators. *Front. Energy Res.* **2021**, *9*. [[CrossRef](#)]
14. Tang, M.Z.; Zhao, Q.; Ding, S.X.; Wu, H.W.; Li, L.L.; Long, W.; Huang, B. An Improved LightGBM Algorithm for Online Fault Detection of Wind Turbine Gearboxes. *Energies* **2020**, *13*, 807. [[CrossRef](#)]
15. Jian, Y.F.; Qing, X.G.; He, L.; Zhao, Y.; Qi, X.; Du, M. Fault diagnosis of motor bearing based on deep learning. *Adv. Mech. Eng.* **2019**, *11*. [[CrossRef](#)]
16. Tang, S.N.; Yuan, S.Q.; Zhu, Y. Deep Learning-Based Intelligent Fault Diagnosis Methods toward Rotating Machinery. *IEEE Access* **2020**, *8*, 9335–9346. [[CrossRef](#)]
17. Tang, M.Z.; Hu, J.H.; Kuang, Z.J.; Wu, H.W.; Zhao, Q.; Peng, S.H. Fault Detection of the Wind Turbine Variable Pitch System Based on Large Margin Distribution Machine Optimized by the State Transition Algorithm. *Math. Probl. Eng.* **2020**, *2020*, 1–9. [[CrossRef](#)]
18. Tang, M.; Ding, S.X.; Yang, C.; Cheng, F.; Shardt, Y.; Long, W.; Liu, D. Cost-sensitive large margin distribution machine for fault detection of wind turbines. *Clust. Comput.* **2019**, *22*, 7525–7537. [[CrossRef](#)]
19. Wan, L.J.; Gong, K.; Zhang, G.; Yuan, X.P.; Li, C.Y.; Deng, X.J. An Efficient Rolling Bearing Fault Diagnosis Method Based on Spark and Improved Random Forest Algorithm. *IEEE Access* **2021**, *9*, 37866–37882. [[CrossRef](#)]
20. Ma, S.L.; Chen, M.X.; Wu, J.W.; Wang, Y.H.; Jia, B.W.; Jiang, Y. Intelligent Fault Diagnosis of HVCB with Feature Space Optimization-Based Random Forest. *Sensors* **2018**, *18*, 1221. [[CrossRef](#)] [[PubMed](#)]
21. Ma, S.; Chen, M.; Wu, J.; Wang, Y.; Jia, B.; Jiang, Y. High-Voltage Circuit Breaker Fault Diagnosis Using a Hybrid Feature Transformation Approach Based on Random Forest and Stacked Autoencoder. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9777–9788. [[CrossRef](#)]
22. Qin, S.Y.; Zhang, M.Z.; Ma, X.J.; Li, M. A new integrated analytics approach for wind turbine fault detection using wavelet, RLS filter and random forest. *Energy Sources Part A Recover. Util. Environ. Eff.* **2019**. [[CrossRef](#)]
23. Yan, X.; Jin, Y.; Xu, Y.; Li, R. Wind Turbine Generator Fault Detection Based on Multi-Layer Neural Network and Random Forest Algorithm. In Proceedings of the 2019 IEEE PES Innovative Smart Grid Technologies Asia, ISGT 2019, Chengdu, China, 21–24 May 2019; pp. 4132–4136.
24. Wang, X.D.; Gao, X.; Liu, Y.M.; Wang, Y.H. Stockwell-transform and random-forest based double-terminal fault diagnosis method for offshore wind farm transmission line. *IET Renew. Power Gener.* **2021**, *15*. [[CrossRef](#)]
25. Mansouri, M.; Fezai, R.; Trabelsi, M.; Nounou, H.; Nounou, M.; Bouzrara, K. Reduced Gaussian process regression based random forest approach for fault diagnosis of wind energy conversion systems. *IET Renew. Power Gener.* **2021**. [[CrossRef](#)]
26. Xie, Z.H.; Yang, X.H.; Li, A.Y.; Ji, Z.C. Fault diagnosis in industrial chemical processes using optimal probabilistic neural network. *Can. J. Chem. Eng.* **2019**, *97*, 2453–2464. [[CrossRef](#)]
27. Long, W.; Wu, T.B.; Xu, M.; Tang, M.Z.; Cai, S.H. Parameters identification of photovoltaic models by using an enhanced adaptive butterfly optimization algorithm. *Energy* **2021**, *229*. [[CrossRef](#)]
28. Macready, W.G.; Wolpert, D.H. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82.

29. Wen, L.; Jianjun, J.; Ximing, L.; Tiebin, W.; Ming, X.; Shaohong, C. Pinhole-imaging-based learning butterfly optimization algorithm for global optimization and feature selection. *Appl. Soft Comput. J.* **2021**, *103*, 107146.
30. Long, W.; Jiao, J.; Liang, X.; Tang, M. Inspired grey wolf optimizer for solving large-scale function optimization problems. *Appl. Math. Model.* **2018**, *60*, 112–126. [[CrossRef](#)]
31. Luis Rodríguez, O.C.; Soria, J. Grey Wolf Optimizer with dynamic adaptation of parameters using fuzzy logic. *IEEE Congr. Evol. Comput.* **2016**, 3116–3123. [[CrossRef](#)]
32. Long, W.; Jiao, J.J.; Liang, X.M.; Tang, M.Z. An exploration-enhanced grey wolf optimizer to solve high-dimensional numerical optimization. *Eng. Appl. Artif. Intell.* **2018**, *68*, 63–80. [[CrossRef](#)]
33. Malik, M.; Mohideen, E.R.; Ali, L. Weighted distance Grey wolf optimizer for global optimization problems. In Proceedings of the 2015 IEEE International Conference on Computational Intelligence and Computing Research, Tamilnadu, India, 10–12 December 2015.
34. Gupta, S.; Deep, K. Enhanced leadership-inspired grey wolf optimizer for global optimization problems. *Eng. Comput.* **2020**, *36*, 1777–1800. [[CrossRef](#)]
35. Shaheen, M.A.M.; Hasanien, H.M.; Alkuhayli, A. A novel hybrid GWO-PSO optimization technique for optimal reactive power dispatch problem solution. *Ain Shams Eng. J.* **2021**, *12*, 621–630. [[CrossRef](#)]
36. Daniel, E.; Anitha, J.; Gnanaraj, J. Optimum laplacian wavelet mask based medical image using hybrid cuckoo search—Grey wolf optimization algorithm. *Knowl. Based Syst.* **2017**, *131*, 58–69. [[CrossRef](#)]
37. Xie, Q.Y.; Guo, Z.Q.; Liu, D.F.; Chen, Z.S.; Shen, Z.L.; Wang, X.L. Optimization of heliostat field distribution based on improved Gray Wolf optimization algorithm. *Renew. Energy* **2021**, *176*, 447–458. [[CrossRef](#)]
38. Helaleh, A.H.; Alizadeh, M. Performance prediction model of Miscible Surfactant-CO₂ displacement in porous media using support vector machine regression with parameters selected by Ant colony optimization. *J. Nat. Gas Sci. Eng.* **2016**, *30*, 388–404. [[CrossRef](#)]
39. Jakubik, J.; Binding, A.; Feuerriegel, S. Directed particle swarm optimization with Gaussian-process-based function forecasting. *Eur. J. Oper. Res.* **2021**, *295*, 157–169. [[CrossRef](#)]
40. Mirjalili, S. SCA: A Sine Cosine Algorithm for Solving Optimization Problems. *Knowl. Based Syst.* **2016**, *96*, 120–133. [[CrossRef](#)]
41. Heidari, A.A.; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Futur. Gener. Comput. Syst.* **2019**, *97*, 849–872. [[CrossRef](#)]
42. Song, D.R.; Chang, Q.; Zheng, S.Y.; Yang, S.; Yang, J.; Joo, Y.H. Adaptive Model Predictive Control for Yaw System of Variable-speed Wind Turbines. *J. Mod. Power Syst. Clean Energy* **2021**, *9*, 219–224. [[CrossRef](#)]
43. Henderi, H. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *IJIS: Int. J. Informatics Inf. Syst.* **2021**, *4*, 13–20. [[CrossRef](#)]