Fault Diagnosis of Power Transformer Based on SSA–MDS Pretreatment

MEI ZHANG AND WANLI CHEN
College of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China

Corresponding author: Wanli Chen (cwl18815213451@163.com)

This work was supported in part by the Natural Science Foundation of the Higher Education Institute of Anhui Province under Grant KJ2020A0309, and in part by the National Natural Science Foundation of China under Grant 51874010.

ABSTRACT Aiming at the problems of coupling between transformer input characteristics and low accuracy of transformer fault diagnosis, SSA-MDS and other soft technologies are used to analyze the key characteristics of transformer faults, so as to improve the accuracy of transformer fault diagnosis. The SSA algorithm cascade MDS algorithm to process the DGA data is proposed. Subsequently, the TSSA-RF model is introduced to classify the DGA data. The DGA data is first mapped to a high-dimensional space. Next, the optimal feature subset is encoded using the SSA algorithm to reduce irrelevant and redundant features. In this study, the correlation between the optimal feature dimension and the transformer fault diagnosis accuracy is investigated. The expression of the optimal feature subset is obtained by decompiling the SSA operator. The pre-processed data are classified using the RF model, and the TSSA-RF model for classifying the DGA data is found with the highest accuracy through the comparison of different optimization algorithms. After the RF model is optimized using the TSSA algorithm, its accuracy increases by 7.89%, and the accuracy of the TSSA-RF model is obtained as 92.11%. The example results show that compared with the original data, the proposed data processing algorithm improves the diagnostic accuracy of transformer by 11.97% in the RF model. Compared with multiple preprocessing methods, SSA-MDS has the highest accuracy. Compared with the original data, the accuracy of TSSA-RF model increases by 11.64%.

INDEX TERMS Power transformer, fault diagnosis, RF model, TSSA algorithm, feature extraction.

I. INTRODUCTION
Oil-immersed transformer is the key equipment in the operation of the power grid. How to accurately and quickly diagnose the fault location and fault type of the transformer is of great significance to the stable operation of the power grid [1].

When the oil-immersed transformer fails, the discharge phenomenon usually occurs and reacts with the insulation paper, resulting in various gases around it, mainly including H₂, CH₄, C₂H₆, C₂H₄ and C₂H₂ [2]. These gases are partially melted in the surrounding oil, which increases the gas content of transformer oil. Therefore, dissolved gas analysis (DGA) in oil is commonly used to analyze the concentration of various gases in transformer oil. The concentration of H₂, CH₄, C₂H₆, C₂H₄ and C₂H₂ in transformer oil is detected by DGA method, and then the operation state of transformer is evaluated [3].

The traditional methods of transformer fault diagnosis using DGA include three-ratio method, Rogers ratio method, uncoded ratio method and IEC ratio method [4], [5], [6]. However, the diagnostic accuracy of the above method is low, and there is often a lack of code, which cannot diagnose the transformer fault type more accurately.

A large number of examples show that the intelligent algorithm has achieved excellent results in transformer fault diagnosis [7]. With the development of artificial intelligence, a large number of new algorithms are used for power transformer fault diagnosis. Wu et al. used the DBSCAN algorithm to establish the fault diagnosis model of power transformer. Although the results show that this method can solve the code shortage problem of the three-ratio method to a certain extent, the generalization ability of the model
needs to be improved, and there is no construction of data feature engineering, and the high-dimensional mapping between data is not found [8]. Guo C et al. established several neural network model algorithms and compared them one by one to obtain the optimal machine learning model. Although good results have been achieved, the problem of data coupling has not been solved. The diagnostic accuracy under the condition of less data needs to be discussed [9]. De Andrade Lopes S M et al. introduced a new application based on deep neural network, sampling a few oversampling methods to solve the problem of less data sets [10]. Han et al. used random forests to reversely search the transformer features, and the optimized features can improve the performance of the model. However, using the posterior model of random forests will lead to a low score of some main features, while the score of model-independent features is high in the posterior model, which is prone to multiple linear and dimensional explosion problems [11]. H.A. Illias et al. proposed a hybrid support vector machine (SVM) and an improved evolutionary particle swarm optimization (EPSO) algorithm to determine the type of transformer fault. By comparing the results with the actual fault diagnosis, the unoptimized SVM and the previously reported work, the superiority of the improved PSO technology and SVM is evaluated. In this paper, the SVM-MEPSO-TVAC method is used in combination with stepwise regression to establish the feature engineering [12]. The results of this method show that this method can be used as an alternative solution fault diagnosis of the field transformer. Compared with the above literature, this paper maps the data to a higher dimensional space in more ways in feature construction.

The above literature has made outstanding contributions to transformer fault diagnosis, but there are deficiencies in data feature engineering. Especially in data decoupling, how to solve the coupling problem in DGA data is still the focus and difficulty at present.

In order to solve the problem of low accuracy of transformer fault diagnosis caused by data coupling, this paper maps the data to high-dimensional space, and uses SSA-MDS algorithm to extract feature and reduce dimension of algorithm data. After analysis, the optimal feature subset is obtained, and the relationship between the optimal feature subset is obtained. Finally, TSSA-RF model was used to classify and diagnose the preprocessed data.

The main contributions of this paper are:

1. Based on the idea of IEC method and non-coding method, DGA data is mapped to high-dimensional space by interactive ratio method. Provide ideas for data decoupling.

2. Combined with the idea of Wrapper, the heuristic Wrapper algorithm is used to encode the optimal feature subset of data, analyze the optimal feature dimension of data, and fully decouple the text data.

3. Combined with TSSA-RF model, the diagnostic model of DGA data is constructed. Compared with other algorithms, the accuracy of TSSA-RF model can reach 92.11%, which greatly improves the accuracy of the model on the basis of the original data.

II. ALGORITHM DESCRIPTION

A. SSA (SPARROW SEARCH ALGORITHM)

Sparrow Search Algorithm (SSA) play three roles in sparrow population, including finder, participant and alert. The finder provides foraging direction for other individuals in the population by finding food, and the participant forages in accordance with the direction information offered by finder [13].

Discoverers usually have high energy reserves, undertake search tasks among populations, have high risk of variation, and find regions with abundant food in various regions, indicating the direction for team participants. In the model, the richness of energy reserves depends on the fitness function value of the corresponding region of the individual, and the better fitness value corresponds to the higher energy reserves.

When there are predators in this area, individuals will issue alerts to remind their peers. When the alert intensity is greater than the threshold, the discoverer will lead the population to other safe places foraging.

The roles between discoverers and participants can be changed, not fixed, but the proportion of discoverers in the population is constant. If a discoverer is converted into an adder, then there must be an adder converted into a discoverer in the whole population.

B. TSSA ALGORITHM

During the initialization of the population, the sparrow search algorithm adopts the random generation method, thus leading to the uneven distribution of the sparrow population and having an effect on the later iteration optimization. When chaotic mapping exhibits a series of characteristics (e.g., randomness, ergodicity and regularity), TSSA (Tent Chaotic Map Sparrow Search Algorithm) algorithm combines the randomness, ergodicity and regularity of Tent mapping to optimize the search ability of SSA algorithm. [14], [15].

Chaos is a common nonlinear phenomenon in nature. Chaos traversal has regularity and does not repeat, which is faster than random traversal and has more traversal states. In order to improve the optimization performance of the basic SSA optimization algorithm, the chaotic search mechanism is introduced in the remote step of the SSA optimization algorithm through algorithm mixing to strengthen the local optimization efficiency and global optimization performance of the particle swarm. The particles are divided into two categories, and they perform different evolutionary mechanisms to achieve collaborative optimization, so as to construct a new chaotic hybrid sparrow search optimization algorithm.

The definition of Tent mapping is as follows:

\[
Z_{n+1} = f(x_n) = \begin{cases} 
\frac{Z_n}{\alpha} & \text{if } Z_n \in [0, \alpha) \\
\frac{1-Z_n}{1-\alpha} & \text{if } Z_n \in [\alpha, 1) 
\end{cases}
\]  

(1)
Among them, $\alpha$ is the chaotic parameter of $[0, 1)$, $Z_i$ is the initial value and $Z_{i+1}$ is the value after Tent mapping. When $\alpha$ takes 0.5, the system is in a short period state, and the Tent mapping expression is:

$$Z_{i+1} = \begin{cases} 2 \times Z_i & 0 \leq Z_i \leq \frac{1}{2} \\ 2 \times (1 - Z_i) & \frac{1}{2} < Z_i \leq 1 \end{cases}$$

(2)

where $Z_i$ denotes the initial value; $Z_{i+1}$ represents the value after Tent mapping.

The Tent map is shown in Fig. 1.

The basic idea of Tent chaotic optimization algorithm is to map the chaotic sequence generated by chaotic mapping to the value range of optimization variables through linear transformation, and then search. Tent chaos optimization algorithm is divided into 'coarse search' and local 'fine search' by adding a small disturbance around the optimal solution. Because Tent chaotic search can traverse all states in the space, when the space is large, the traversal time is longer; when the search starting point or control parameters are not properly selected, it takes a long time to search to achieve the convergence requirements.

The specific optimization steps of the TSSA algorithm are elucidated below:

Step 1: Use Eq. 2 to generate chaotic variables $Z^d$ according to initial particles $X^d$.

Step 2: The chaotic variables are transferred to the solution space of the problem to be solved:

$$X^d_{new} = X^d_{min} + (X^d_{max} - X^d_{min}) \times Z^d$$

(3)

where $X^d_{max}$ and $X^d_{min}$ denote the maximum and minimum of the $d$th dimensional variable $X^d_{new}$.

Step 3: Chaotic disturbance is conducted on individuals using Eq. (3):

$$X'_{new} = \frac{(X + X^d_{new})}{2}$$

(4)

where $X$ denotes the individual requiring chaos disturbance; $X^d_{new}$ represents the amount of chaos disturbance generated; $X'_{new}$ expresses the individual after chaos disturbance.

C. RF MODEL (RANDOM FOREST MODEL)

RF model is a strong classifier formed by multiple CART trees as weak classification and bagging as integration strategy. The Bagging strategy generates many parallel classifiers by random sampling, and determines the final result by a few obeying the majority principle [16].

Because RF model training can run in parallel, RF model has a faster training speed. Due to the random sampling, the variance of the trained model is small and the generalization ability is strong. Due to the random selection of decision tree node partition features, RF model can still maintain efficient training model under high dimensional features. Because the RF model is easy to produce over-fitting phenomenon on the data with large differences, and the characteristics of more value division are easy to produce adverse effects on RF decision-making, it is necessary to optimize the super parameters of the RF model to avoid the RF model falling into a serious over-fitting state.

III. TRANSFORMER FAULT DIAGNOSIS MODEL BASED ON TSSA-RF

The RF model is greatly affected by parameters, and the selection of super parameters directly affects the classification accuracy and generalization ability of RF [17]. In this paper, the RF model is used to diagnose the transformer fault, and the TSSA algorithm is used to optimize the number of decision trees, the maximum depth of the decision tree, the minimum separation sample number, the minimum leaf node sample number and the maximum separation feature number of the RF model to improve the performance of the diagnosis model. To solve the problem of transformer fault diagnosis, a TSSA-RF model is built in this study, as depicted in Fig. 2.
TSSA model is used to search for several hyper-parameters of the RF model. In the fault diagnosis part, the RF model is trained and tested, and then the transformer fault type is output for model evaluation.

IV. OPTIMAL FEATURE SUBSET UNDER SSA CODING

A. DATA ACQUISITION

In this paper, the operation status of oil-immersed transformer in a northwest power grid of China State Grid Corporation in recent five years is selected, and the concentrations of \( H_2, CH_4, C_2H_6, C_2H_4 \) and \( C_2H_2 \) are selected as the attribute of transformer fault diagnosis, which involve 381 groups of fault data. Some sample data are listed in Table 1.

### TABLE 1. Partial data samples.

| Data(μL·L⁻¹) | \( H_2 \) | \( CH_4 \) | \( C_2H_6 \) | \( C_2H_4 \) | \( C_2H_2 \) | True Faults |
|--------------|-----------|-----------|-----------|-----------|-----------|-------------|
| N            | 9.1       | 11.7      | 5.2       | 1.8       | 0.9       | N           |
| T₁           | 97        | 110       | 34        | 85        | 0         | T₁          |
| T₂           | 170       | 300       | 44        | 580       | 9         | T₂          |
| D₂           | 83        | 60        | 46        | 48        | 2         | D₂          |
|              | 179       | 413       | 27        | 544       | 473       |              |

In accordance with the dissolved gas analysis and judgment guide in GB-T 7252-2016 transformer oil, this study adopts seven fault types as the output characteristics of transformer fault diagnosis (i.e., low temperature overheating (\( T_1 \)), medium temperature overheating (\( T_2 \)), high temperature overheating (\( T_3 \)), partial discharge(PD), low energy discharge (\( D_1 \)), high energy discharge (\( D_2 \)), Normal operation (N)). In this study, the ratio of the training set and test set is set at 4:1. The number of samples of fault type data, training set and test set is shown in Table 2.

### TABLE 2. Transformer fault type coding table.

| Fault type | State code | \( N \) | \( T_1 \) | \( T_2 \) | \( T_3 \) | PD | \( D_1 \) | \( D_2 \) |
|------------|------------|-----|-----|-----|-----|---|-----|-----|
| N          | 1          | 43  | 36  | 41  | 39  | 36 | 37  | 73  |
| \( T_1 \)  | 2          | 11  | 9   | 10  | 10  | 9  | 9   | 9   |
| \( T_2 \)  | 3          | 10  | 10  | 10  |     |     |     |     |
| \( T_3 \)  | 4          | 9   |     |     |     |     |     |     |
| PD         | 5          |     |     |     |     |     |     |     |
| \( D_1 \)  | 6          |     |     |     |     |     |     |     |
| \( D_2 \)  | 7          |     |     |     |     |     |     |     |

B. DATA DIMENSION RAISING

Since Duval triangle method and IEC method can be used to judge the fault type of transformer, and Duval triangle method and IEC method are both generated by the characteristic ratios between DGA data, the ratio characteristics between gas concentrations of DGA data are helpful for transformer fault diagnosis. The commonly used DGA detection gases are \( H_2, CH_4, C_2H_6, C_2H_4 \) and \( C_2H_2 \), so the collected DGA data has five-dimensional attributes, and the interactive ratio forms of five-dimensional data attributes are as follows:

\[
\frac{N_1}{N_2}, \frac{N_1}{N_3 + N_4}, \frac{N_1}{N_3 + N_4 + N_5}, \frac{N_1}{N_3 + N_4 + N_5 + N_6},
\]

where \( N_3, N_4, N_5 \) and \( N_6 \) are any different attributes of DGA data, and \( N_1 \) and \( N_2 \) are any attributes of DGA data \( (N_1 \neq N_2) \). The transformer features are mapped to 150 dimensions by adding 145 dimension feature variables in four ratio forms and adding the original 5-dimension feature.

C. FEATURE EXTRACTION BASED ON SSA ALGORITHM

The addition of some redundant features will cause multiple collinearity problems and increase the complexity of the prediction model, while some noise features will have a negative impact on the model. These will increase the complexity of the model, resulting in over-fitting of the model, and then affect the diagnostic effect of the model. Therefore, it is necessary to filter the input features.

In this paper, the Wrapper method is used to filter features, eliminate irrelevant features, and obtain the optimal feature subset. The principle is that the model is used to train and evaluate the feature subset and the target (label) set, and the binary optimization algorithm is constructed. The accuracy of training is used as the standard to measure the quality of the feature subset, and the optimal algorithm is used to select the best feature subset.

The feature extraction of Wrapper method is divided into complete search method and heuristic search method. The complete search method can obtain very good results, but the search space is large and the time complexity is high when there are more features. Heuristic search algorithm can get better results with less search times. In the case of more features, heuristic search method is usually used instead of complete search algorithm.

In this paper, 150-dimensional features are obtained after interactive ratio dimension-up, and 145-dimensional features are needed to preserve the 5-dimensional features of the original data. If the full search method is used to do exhaustive search, the search space is \( 2^{145} - 1 \) possibilities, and the computational complexity of all traversal is too large. And heuristic search can get better features in less search times, so this paper uses heuristic search instead of complete search method, and SSA algorithm is selected as heuristic Wrapper search algorithm.

Since some irrelevant features increase the model complexity and cause poor model diagnosis effect, the binary optimization algorithm is adopted to screen the features.
and remove irrelevant features to obtain the optimal feature subset.

In this study, the SSA algorithm serves as the optimization algorithm to build the binary SSA algorithm for feature extraction of the dataset after dimension raising.

Definition of independent variables: a row of 0/1 binary column vectors with a length of N are set as the independent variables of the optimization algorithm, where N denotes the number of features, the number of newly increased features \( N = 145 \), 0 represents the number of features not selected, and 1 represents the number of features selected.

First, the training set of the dataset is divided into training samples and validation samples using the four-fold cross-validation method. Second, the fitness function is established, the RF model serves as the prediction model, and the training samples are employed to train the RF model. Next, the trained RF model is adopted to classify the validation samples, and the error rate of the predicted value and the actual value of the validation sample serves as the fitness function value. Lastly, the 0/1 binary column vector with a length of N is used as the independent variable, the error rate value is employed as the fitness function value, and the SSA algorithm is applied as the optimization algorithm; on that basis, the binary SSA optimization algorithm is built to search for the optimal feature subset.

As cross-validation randomly divides training sets and validation sets, the selection of validation sets of different training sets is found to have an impact on the accuracy of the model. To ensure the stability and reliability of the model, 10 types of training sets and validation sets are divided through 4-fold cross-validation at the fitness function, and 10 types of datasets are classified by RF model. The average error rate of the 10 types of datasets is taken, and the overall accuracy value serves as the fitness function value.

The number of 1 in binary code is set to make the selected features a fixed number. Subsequently, the effect of the number of feature subsets on model performance is investigated.

The adaptive encoder constructed by the SSA algorithm is used for feature extraction of data. Fig. 2 presents the correlation between the dimension of extracted feature subset and the accuracy of the validation set.

As depicted in Fig. 3, with the number of feature subsets at 10, the accuracy rate of the validation set is the highest and the model performance is the best. When the number of feature subsets is 0, it means that the data is original data, average accuracy under validation set is 72.5%. As depicted in the figure, when the number of novel features is small, the accuracy rate grows faster. When only one novel feature number is selected, the accuracy rate is 5.79% higher than the original data. When the number of features is higher than 10, the newly increased bad features reduce the model performance and lead to the decline of model accuracy. When the new feature data is 0 dimension, the accuracy of RF model is 72.5%, the number of optimal feature subsets is 10, and the accuracy rate is 84.74%, 12.24% higher than the original data.

Fig. 4 plots the fitness function curve of the SSA optimization algorithm based on the 10 dimensions of feature subsets.

As depicted in Fig. 4, with the number of iterations at 17, the fitness function value does not change, the fitness value is obtained as 15.26%, and the average accuracy rate of the verification set is 84.74%.

**D. COMPARISON OF DIFFERENT ALGORITHMS**

Due to strong coupling between various features, all types of coding algorithms have different performances. Thus, different coding algorithms select different feature subsets. To obtain the optimal feature subset, in this study, Particle Swarm Optimization (PSO), Sparrow Search Algorithm (SSA), TSSA and WOA algorithms are used to construct...
adaptive encoders. The population size of each algorithm is set at 50, and the number of iterations is set at 100. Subsequently, 10-dimensional feature subsets are selected using four algorithms.

The error rate of cross validation set is used as the fitness function value of the optimization algorithm. The optimization results of various optimization algorithms are shown in Fig. 5.

As depicted in Fig. 5, the SSA algorithm has the lowest fitness value and the best model performance. Moreover, the TSSA algorithm is second only to the SSA algorithm and superior to WOA and PSO algorithms. Accordingly, the SSA algorithm is selected as the optimal algorithm for constructing an adaptive encoder.

E. OPTIMAL FEATURE ANALYSIS

The optimal feature subset is selected using the SSA algorithm, and the number of feature subsets is set at 10. By decomplicating the SSA operator, the expression of the optimal feature subset is obtained, and the optimal feature subset encoded by the SSA algorithm is expressed as:

\[
\begin{align*}
C_2H_4 & \quad H_2 \\
C_2H_6 + C_2H_2 & \quad CH_4 + C_2H_4 + C_2H_2 \\
C_2H_4 & \quad CH_4 + C_2H_2 \\
C_2H_6 & \quad CH_4 \\
C_2H_2 + CH_4 + C_2H_4 & \quad CH_4 + H_2 \\
C_2H_2 + CH_4 & \quad C_2H_4 + H_2 \\
\end{align*}
\]

The Pearson correlation coefficient between the optimal feature subsets is shown in Fig. 6.

As depicted in Fig. 6. The deeper the color is, the higher the absolute value of the value is. The lighter the color is, the closer the data value is to 0. The deeper the blue is, the closer the data value is to \(-1\). The deeper the red is, the greater the positive value is, and the closer the data value is to 1. There is a high correlation between some extracted features, and there is a multicollinearity problem among the data, which leads to the spatial instability of the solution. Moreover, high-dimensional data leads to a large amount of RF data, resulting in a long training time. Accordingly, the dimension of the data should be reduced.

F. MDS DIMENSION REDUCTION

MDS (Multidimensional Scaling) algorithm: It is a method to reduce the dimension of multidimensional problems to simplify the complexity of problems and solve problems in an effective time. For data, the distance between samples in the original space is required to be maintained in low-dimensional space.

The data dimension increases due to the data dimension raising. If the data dimension is too large, the complexity of the model will increase. Principal Component Analysis (PCA), LLE (Locally Linear Embedding) and MDS algorithms are used to reduce the dimension of the data. The training set of the data set is divided into training samples and verification samples by 4 fold cross validation. RF is used as the prediction model, and the training samples are used to train the RF model. The trained RF model is used to classify the validation samples, and the accuracy of the predicted value and the actual value of the validation sample is calculated. The selection of different training sets and validation sets has an impact on the accuracy of the model. In order to ensure the stability and reliability of the model, four fold cross validation is used to divide 10 training sets and validation sets at the fitness function, and the RF model is used to classify the 10 data sets. The overall accuracy is the average accuracy of the 10 data sets, and the average overall accuracy is the accuracy under this dimension.

Fig. 7 presents the correlation between dimension and accuracy rate achieved by different algorithms.
The MDS algorithm reaches an accuracy of 84.47\% under 6 dimensions, and the accuracy increases slightly when the dimension is higher than 6. To obtain a lower dimension while ensuring the accuracy of the model, the MDS algorithm is adopted to reduce the data dimension to 6 dimensions. Combined with Fig. 3 and Fig. 7, the average accuracy of the RF model increased by 11.97\% after SSA-MDS treatment.

Details of the relationship between dimension and accuracy of each dimension reduction algorithm are shown in Table 3.

**Table 3. Relationship between the dimension of each dimension reduction algorithm and the accuracy of verification set.**

| Dim | 1   | 2   | 3   | 4   | 5   | 6   | 7   |
|-----|-----|-----|-----|-----|-----|-----|-----|
| MDS | 59.26 | 61.18 | 68.29 | 78.95 | 81.84 | 84.61 | 84.47 |
| PCA | 34.74 | 55.26 | 63.03 | 65.92 | 68.82 | 70.66 | 74.3 |
| LLE | 29.21 | 45.39 | 54.34 | 68.16 | 74.08 | 75.92 | 75.53 |

**G. DATA NORMALIZATION**

DGA data vary significantly, so the data after SSA-MDS processing is normalized. In this study, the interval value method is employed to normalize the data, so as to scale the data in a specific range and avoid the mutual influence between the values. Here, normalization of data by linear transformation:

\[
X_i(d) = \frac{X_i - \min X_i}{\max X_i - \min X_i} \times [1 - (-1)] + (-1)
\]

where \(X_i(d)\) denotes the data after normalization, and the mapping interval is \([-1, 1]\); \(i = 1, 2 \ldots n\); \(X_i\) represents the original data; \(\max X_i\) is the maximum value in this feature; \(\min X_i\) is the minimum value in this feature.

**TABLE 4. Data samples after processing.**

| Data samples after processing | True faults |
|------------------------------|-------------|
| 0.997 0.116 -0.174 -0.248 -0.072 0.682 | N            |
| 0.972 0.096 -0.168 -0.229 -0.063 0.702 | T_1          |
| 0.863 0.026 -0.219 -0.225 0.064 0.623 | T_3          |
| 0.970 0.314 -0.582 -0.233 -0.117 0.179 | T_2          |
| 0.824 0.030 -0.089 -0.353 0.252 -0.374 | D_2          |

Table 4 lists the results of the data samples after data processing.

**V. EXPERIMENTAL TEST AND ANALYSIS**

**A. COMPARISON OF DIFFERENT DIAGNOSTIC MODELS**

SSA adaptive encoder and MDS algorithm are employed to preprocess the data, and the normalized data features are the input to classify the data using the classification model. Two classifiers, Extreme Learning Machine (ELM) and Random Forest (RF), are selected to respectively diagnose the data. Fig. 8 presents the results.

According to Fig. 8, the RF model has a higher accuracy rate than that of the ELM, and the performance of the RF model is better.

**B. IMPACT OF DIFFERENT OPTIMIZATION ALGORITHMS ON THE RF MODEL**

Since the hyper-parameters of the RF model have significant influence on the model. Therefore, TSSA algorithm is used to optimize the number of decision trees (n_estimators) (the search range is [1-1000], and is an integer), the maximum depth (max_depth) (the search range is [1-500], and is an integer), the minimum number of separated samples (min_samples_split) (the search range is \((0 – 1]\), floating-point decimal), the minimum number of leaf nodes (min_samples_leaf) (the search range is \((0 – 0.5]\), floating-point decimal), and the maximum number of separated features (max_features) (the search range is [1], [2], [3], [4], [5], [6], and is an integer) of the RF model. The results are compared...
with SSA, WTO and PSO algorithms. The population size of the TSSA, SSA, WTO, and PSO algorithms is set at 50, and the iterations are set at 100. The data of the training set is employed for 4-fold cross-validation, and the average value is taken after running 20 times. Subsequently, the error rate of the validation set is taken as the fitness value. The analysis is carried out on the server of Windows 10 standard 64 bit operating system with Intel (R) Core (TM) CPU i9-12900K, 3.2GHz, 32GB RAM and Nvidia GeForce GTX 3080 card. The fitness curve of each algorithm is plotted in Fig. 9.

As depicted in Fig. 9, the TSSA algorithm has the lowest error rate and the best optimization effect. The WOA algorithm has the slowest convergence speed. The PSO algorithm has the highest error rate and relatively weak global search ability. Thus, the TSSA algorithm exhibits stronger local search ability and global search ability than the other three algorithms.

The TSA-RF model is employed for a diagnostic analysis on the data of the test set, and the diagnostic results of the TSSA-RF model are compared with those of the PSO-RF, WTO-RF and SSA-RF models. Fig. 10 presents the overall classification results, and Table 5 lists the detailed diagnosis results.

As depicted in Fig. 10 and Table 5, the TSSA-RF model exhibits the highest accuracy and the best model performance in the test set. As depicted in Figs 10 and 8, the accuracy of the RF model optimized by the TSSA increases by 7.89%.

During the implementation of classification algorithms, model accuracy, recall rate and F1-score are the three main indicators to judge the effect of model classification.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

The equation of recall rate is:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

F1-Score is also called balanced F-score method, and its equation is written as:

$$F1 - \text{Score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 2 \quad (8)$$

where TP is predicted as a correct positive sample.
FP is predicted as a wrong positive sample.
FN is predicted as a wrong negative sample.
Macro-F1 is termed Macro average method. It is obtained by substituting the accuracy rate and recall rate of each transformer state into Eq. (8) and then by averaging the 7 F1-Score values.

The accuracy, recall rate and F1-score of the TSA-RF in Fig. 10 are obtained, as listed in Table 6.
Accordingly, it can be seen from Table 6 that the call rate of the TSSA-RF model is 92.38%. Plus F1-Score values of each class and dividing by 7, the value of Macro-F1 is 92%. With the use of the same method, the Macro-F1 values of the SSA-RF model, the PSO-RF model and the WOA-RF model in Fig. 10 are all lower than 92%, which reveals that the TSSA algorithm is suitable for parameter optimization of the RF.

The average Precision, average callback rate and Macro-F1 index of each model are shown in Table 7.

According to Table 7, the Macro-F1 index of SSA algorithm is higher than that of WOA algorithm, and the Macro-F1 index of TSSA is the highest. SSA algorithm is superior to WOA algorithm in the local solution, and is better than WOA algorithm on the verification set. Among them, TSSA algorithm has the highest accuracy and Macro-F1 index, and the model diagnosis effect is the best.

A diagnostic analysis is conducted on the data of the test set using the TSSA-ELM model, the PSO-ELM model, the WTO-ELM model, and the SSA-ELM model. The overall classification results are presented in Fig. 11.

As depicted in Fig. 12, compared with the original data, the accuracy of the data processed by the SSA-MDS in the TSSA-RF model grows by 13.16%. The SSA-MDS algorithm in this study exhibits a higher accuracy in the TSSA-RF model than the other two feature processing methods.

According to Fig. 12, the DGA data encoded by SSA-MDS have the highest accuracy among the four preprocessing methods. Compared with the original data, the diagnosis of RF model after SSA-MDS encoding increased by 13.16%. The accuracy of each fault category of each preprocessing method is shown in Fig. 13. Different colors in Figure 13 represent different pretreatment methods. In order to make the image more intuitive, the transparency of the canvas is set to 50%, and the seven corners of the positive seven corners of the graph represent seven fault types. The closer to the axis end point, the higher the diagnostic accuracy, the larger the graph area, and the better the comprehensive diagnostic effect of the representative model. There is no color in the graph, which is the mixed effect of two or more colors in the graph.

According to the observation in Fig. 13, the characteristics of SSA-MDS coding are more accurate than other preprocessing methods in low-energy discharge and high-energy discharge states. The characteristics of No+IEC+Rogers coding are more accurate than those of other methods under high temperature overheating.

**TABLE 7.** Indicators of each model.

| Model      | Precision rate/% | Recalling rate/% | Macro-F1%  |
|------------|------------------|-----------------|------------|
| TSSA-RF    | 91.86            | 92.38           | 92.00      |
| SSA-RF     | 90.38            | 91.59           | 90.87      |
| WOA-RF     | 90.43            | 90.95           | 90.51      |
| PSO-RF     | 88.83            | 88.70           | 88.57      |

**FIGURE 11.** Classification results of various optimization algorithms in the ELM model.

**FIGURE 12.** TSSA-RF classification results under different preprocessing.

**C. IMPACT OF DIFFERENT FEATURE PROCESSING METHODS ON MODEL ACCURACY**

Impacted by significant differences in the concentration of transformer gas, directly using the concentrations of \( H_2, CH_4, C_2H_6, C_2H_4 \) and \( C_2H_2 \) gases as the input of the classifier will decrease the diagnostic accuracy. Thus, non-coding ratio (e.g., the characteristics of 9 dimensions), IEC ratio method (e.g., the characteristics of 3 dimensions), Rogers ratio method (e.g., the characteristics of 4 dimensions) have been extensively used to build novel features, so as to increase the diagnostic accuracy [18], [19], [20], [21].

In this study, the features constructed using the non-coding ratio method and the IEC ratio and the Rogers ratio are compared with those of the SSA-MDS preprocessing algorithm proposed in this study. To ensure the comparison of experiments, the DGA data is not process in another set of experiments, and the original data are retained. The data of the four different pretreatment methods are classified using the TSSA-RF model, and the classification results of each model are shown in Fig. 12.
To ensure the stability of the calculated results, the four models run 20 times, and all the results are statistically studied. Table 8 lists all the results.

**TABLE 8. Table of comprehensive accuracy rate.**

| Method               | Highest accuracy rate | Lowest accuracy rate | Average accuracy rate |
|----------------------|-----------------------|----------------------|-----------------------|
| SSA-MDS              | 96.05%                | 88.16%               | 92.50%                |
| Nocode               | 88.16%                | 78.69%               | 83.95%                |
| Nocode+IEC+Rogers    | 89.47%                | 82.89%               | 86.97%                |
| Original data        | 86.84%                | 73.68%               | 80.86%                |

According to Table 8, the accuracy of TSSA-RF model is improved by 11.64 % by comparing the original data with the data processing algorithm proposed in this paper. Combined with Fig. 3 and Fig. 7, in the verification set, when the new feature data in Fig. 3 is 0 dimension, the accuracy of RF model is 72.5 %. After SSA-MDS processing, the accuracy of RF model is improved by 11.97 %. As depicted in Figs 10 and 8, the accuracy of the RF model optimized by the TSSA increases by 7.89%.

**VI. CONCLUSION**

A coupling phenomenon exists among features of transformer DGA original data. To investigate the correlation between features of DGA data, the DGA data is mapped to a high-dimensional space, the SSA algorithm is used to obtain the optimal feature subset, and then the effect of newly increased feature variables on the accuracy of transformer fault diagnosis is analyzed. The MDS algorithm is introduced to process the data for solving the multicollinearity problem and reducing the model complexity. According to the DGA data collected by the State Grid, the proposed model is trained and tested using the concentrations of $H_2$, $CH_4$, $C_2H_6$, $C_2H_4$ and $C_2H_2$ in the obtained DGA data. The conclusions are drawn below:

1. The effect of various ratio features on the accuracy of transformer fault diagnosis is analyzed, and the SSA algorithm is adopted to obtain the correlation between the dimension of optimal subset and the accuracy of transformer fault diagnosis. Lastly, it is concluded that, when the feature subset has 10 dimensions, the diagnosis of transformer faults has the highest accuracy. After the decompiling of the SSA operator, the expression of the optimal feature subset is obtained.

2. Compared with different optimization algorithms, the proposed TSSA algorithm has the optimal optimization effect on the hyperparameter optimization of the RF model, and the TSSA-RF model has an accuracy 7.89% higher than that of the RF model.

3. In this paper, SSA algorithm is used to cascade MDS algorithm to process the original data. The experimental results show that the proposed SSA-MDS preprocessing algorithm improves the diagnostic accuracy of RF model by 11.97% compared with the original data. Compared with the two data preprocessing methods, the highest accuracy of the proposed algorithm is 92.5% in TSSA-RF model, 8.55% higher than Nocode, 5.53% higher than Nocode+IEC+Rogers, and 11.64% higher than the original data.

**REFERENCES**

[1] I. B. M. Taha, S. Ibrahim, and D.-E.-A. Mansour, “Power transformer fault diagnosis based on DGA using a convolutional neural network with noise in measurements,” *IEEE Access*, vol. 9, pp. 111162–111170, 2021.

[2] S. S. M. Ghoneim, T. A. Farrag, A. A. Rashed, E.-S.-M. El-Kenawy, and A. Ibrahim, “Adaptive dynamic meta-heuristics for feature selection and classification in diagnostic accuracy of transformer faults,” *IEEE Access*, vol. 9, pp. 78324–78340, 2021.

[3] S. Bin, L. Zhixiong, L. Enwen, and W. Guoli, “A preliminary study of the three-ratio code-deficient in DGA,” *Power Automat. Equip.*, vol. 35, no. 12, pp. 60–65, 2015.

[4] S. I. Ibrahim, S. S. M. Ghoneim, and I. B. M. Taha, “DGALab: An extensible software implementation for DGA,” *IET Gener., Transmiss. Distrib.*, vol. 12, no. 18, pp. 4117–4124, Oct. 2018, doi: 10.1049/iet-gtd.2018.5564.

[5] H. Malik and S. Mishra, “Application of gene expression programming (GEP) in power transformers fault diagnosis using DGA,” *IEEE Trans. Ind. Appl.*, vol. 52, no. 6, pp. 4556–4565, Nov. 2016.

[6] H. Cui, P. Jia, X. Peng, and P. Li, “Adsorption and sensing of CO and C$_2$H$_2$ by S-defected Sn$_2$S$_2$ monolayer for DGA in transformer oil: A DFT study,” *Mater. Chem. Phys.*, vol. 249, Jul. 2020, Art. no. 123006.

[7] D. Saravanan, A. Hasan, A. Singh, H. Mansoor, and R. N. Shaw, “Fault prediction of transformer using machine learning and DGA,” in *Proc. IEEE Int. Conf. Comput., Power Commun. Technol. (GUCON)*, Oct. 2020, pp. 1–5.

[8] W. Guoqiang, Y. Jianfeng, G. Minyuian, Z. Xuesong, W. Kai, L. Haoyan, and S. Shanshan, “Power transformer fault diagnosis based on DBSCAN,” *J. Wuhan Univ., Eng. Ed.*, vol. 54, no. 12, pp. 1172–1179, 2021, doi: 10.14188/j.1671-8844.2021-12-012.
C. Guo, M. Dong, and Z. Wu, “Fault diagnosis of power transformers based on comprehensive machine learning of dissolved gas analysis,” in Proc. IEEE 20th Int. Conf. Dielectr. Liquids (ICDL), Jun. 2019, pp. 1–4.

S. M. D. A. Lopes, R. A. Flauzino, and R. A. C. Alufaim, “Incipient fault diagnosis in power transformers by data-driven models with over-sampled dataset,” Electr. Power Syst. Res., vol. 201, Dec. 2021, Art. no. 107519.

A. Guoqing, S. Zhenwen, M. Shifeng, H. Xiaohui, D. Zhenbin, and Z. Chunlin, “WOA-RF transformer fault diagnosis based on RF feature optimization,” High Voltage App., vol. 58, no. 2, pp. 171–178, 2022, doi: 10.13296/j.1001-1609.hva.2022.02.023.

H. A. Illias and W. Zhao Liang, “Identification of transformer fault based on dissolved gas analysis using hybrid support vector machine-modified evolutionary particle swarm optimisation,” PLoS ONE, vol. 13, no. 1, Jan. 2018, Art. no. e0191366.

L. Lina, N. Xinyuan, and S. Yuefei, “Improved sparrow search algorithm for job shop scheduling problem,” Comput. Appl. Res., vol. 38, no. 12, pp. 3634–3639, 2021.

S. Arora and P. Anand, “Chaotic grasshopper optimization algorithm for global optimization,” Neural Comput. Appl., vol. 31, no. 8, pp. 4385–4405, Aug. 2019.

Z. Weikang, L. Sheng, and R. Chunhui, “Hybrid strategy improved sparrow search algorithm,” Comput. Eng. Appl., vol. 57, no. 24, pp. 74–82, 2021.

L. Wei, L. Bo, C. Fei, and M. Teng, “Fault diagnosis method of CNC turret based on PSO-RF and time sequence link,” J. Jilin Univ., Eng. Ed., vol. 52, no. 2, pp. 392–399, 2022.

Z. Wengang, T. Libin, C. Fuyong, and Y. Jiafeng, “Prediction of TBM tunneling speed based on four hyperparameter optimization algorithms and random forest model,” J. Basic Eng. Sci., vol. 29, no. 5, pp. 1186–1200, 2021, doi: 10.16058/j.issn.1005-0930.2021.05.009.

Z. Youwen, P. Bin, C. Page, L. Weihan, and G. Xin, “Fault diagnosis method of oil-immersed transformer based on XGBoost optimized by genetic algorithm,” Power Automat. Equip., vol. 41, no. 2, pp. 200–206, 2021.

D. Yafeng, D. Jinfeng, F. Hua, and Z. Juncheng, “Transformer fault diagnosis based on SSA-AdaBoost-RF,” Control Eng., vol. 29, no. 2, pp. 280–286, 2022.

WANLI CHEN was born in Taizhou, Zhejiang, China, in 1996. He received the bachelor’s degree in electrical engineering from the Xinjiang Institute of Technology, in 2020. Now, he is a Graduate Student with the Anhui University of Science and Technology. His research interests include intelligent control, the Internet of Things (IoT) Technology, and data analysis.