IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 29, 2021, accepted December 10, 2021, date of publication December 13, 2021,
date of current version December 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3135283

Fault Diagnosis of Oil-Immersed Power Transformer Based on Difference-Mutation Brain Storm Optimized Catboost Model

MEI ZHANG[1], WANLI CHEN[1], YU ZHANG[2], FEI LIU[1], DONGSHUN YU[1], CHAOYIN ZHANG[1], AND LI GAO[1]

[1]College of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China
[2]State Grid Xinjiang Electric Power Company Ltd., Hetian Power Supply Company Ltd., Ürümqi, Hetian, Xinjiang 848000, 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 Graduate Innovation Fund Project of Anhui University of technology under Grant 2021CX2081.

ABSTRACT To address the problem of low accuracy of power transformer fault diagnosis, this study proposed a transformer fault diagnosis method based on DBSO-CatBoost model. Based on data feature extraction, this method adopted DBSO (Difference-mutation Brain Storm Optimization) algorithm to optimize CatBoost model and diagnose faults. First, for data preprocessing, the ratio method was introduced to add features to the original data, the SHAP (Shapley Additive Explanations) method was applied for feature extraction, and the KPCA (Kernel Principal Component Analysis) algorithm was employed to reduce the dimension of data. Subsequently, the preprocessed data were inputted into the CatBoost model for training, and the DBSO algorithm was adopted to optimize the parameters of the CatBoost model to yield the optimal model. Lastly, the DBSO-CatBoost model was exploited to diagnose the transformer fault and output the fault type. As indicated from the example results, the accuracy of the transformer fault diagnosis based on DBSO-Catboost model could be 93.71%, 3.958% higher than that of CatBoost model and significantly exceeding that of some common models. Furthermore, compared with other preprocessing methods, the accuracy of fault diagnosis by employing the data preprocessing method proposed in this study was significantly improved.

INDEX TERMS Power transformer, fault diagnosis, catboost model, DBSO algorithm, feature extraction.

I. INTRODUCTION

Transformer is vital equipment of power system, capable of achieving voltage transformation, power distribution and power transmission. Its safe and reliable operation is correlated with the safety and power supply quality of the whole power grid. Accordingly, accurate diagnosis of transformer faults is critical to maintaining the safe operation of power grid and ensuring the quality of power supply [1]–[5].

The causes and types of power transformer faults are difficult to directly detect. Currently, Dissolved Gas Analysis (DGA) has been the most common fault diagnosis method. When the power transformer is overheated and discharged, its insulating oil will emit gases that dissolves in the oil. By analyzing the dissolved gas, the DGA method can determine the operating condition of the transformer. Conventional DGA methods consist of three-ratio method, Rogers ratio method and non-coding ratio method [6]–[9]. The mentioned methods exploit the relative content of dissolved gas to determine the fault type, and the calculations are simple. However, the classification effect of data close to the threshold is relatively poor, and there are common 'missing code' or 'super code' phenomena [10]–[12].

Over the past few years, as artificial intelligence is leaping forward, several intelligent algorithms combined with DGA method are applied to the fault diagnosis of power transformers. On the whole, the mentioned intelligent algorithms fall to non-ensemble learning and ensemble learning. Non-ensemble learning algorithms consist of BP neural network, support vector machine, extreme learning machine and others, each of which exhibits certain advantages, whereas some problems remain unsolved [5], [13], [14], [16].

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

Zhang *et al.* combined the optimized BP neural network with DGA method to increase the accuracy of transformer fault detection to a certain extent, while defects remain (e.g., slow training speed and difficult parameter determination) [17]. Huang Tongxiang *et al.* used a support vector machine for transformer fault diagnosis. Such a machine exhibited strong learning generalization ability, whereas the accuracy is not high when there are many fault types and information is missing [18]. Du Wenxia *et al.* used an extreme learning machine for transformer fault diagnosis, which exhibited the advantages of fast learning speed and high generalization performance. In the diagnosis process, however, hidden layer neurons are prone to redundancy and classification accuracy decline [19].

The ensemble learning algorithm integrates multiple learners and exhibits higher learning performance. Gradient Boosting Decision Tree (GBDT) is a branch of the ensemble learning algorithm, which reduces the total error by decreasing the deviation and raises lower requirements for parameter adjustment and better robustness. GBDT is extensively adopted in transportation, medical, financial and other fields, whereas it has been rarely applied in power system fault diagnosis. Liao Weihan *et al.* built an oil-immersed transformer fault diagnosis model based on GBDT, Li Hejian *et al.* investigated an oil-immersed transformer fault diagnosis method by complying with extreme gradient lifting. As demonstrated from the comparative experiments of two literatures, the accuracy of the transformer fault diagnosis based on GBDT could be higher than that of non-ensemble learning algorithm [20], [21].

CatBoost is a machine learning library based on GBDT framework, which was proposed by Yandex in 2017. CatBoost, as compared with XGBoost, LightGBM and other GBDT algorithms, has been improved in numerous manners. It addresses the problem of gradient deviation in the iteration by complying with orderly principle, orderly enhancement algorithm and greedy strategy. In addition, it is capable of reducing the possibility of over-fitting of the model, increasing the execution speed of the model, improving the robustness of the model, and further increasing the prediction accuracy. On the whole, the performance of CatBoost is determined by the appropriate hyper-parameter set [22]–[25]. At present, the hyper-parameter optimization of the ensemble learning model largely adopts the grid search method, so the parameter set should be traversed. As impacted by the considerable parameters, the efficiency is low, and even the dimension explosion is triggered. Thus, the optimization algorithm should be applied for super-parametric optimization [26]–[28].

Brain Storm Optimization Algorithm (BSO) simulates the process of human creative thinking to tackle down problems, and it exhibits a strong global and local search ability [29]–[32].

Brainstorm optimization algorithm and optimized algorithm have exhibited prominent performance in numerous fields (e.g., medical image registration, image segmentation, engine parameter prediction, data feature selection and multi-objective optimization [33]–[37]). Many scholars have optimized the brainstorming optimization algorithm [37]–[41] to form various variants of brainstorming optimization algorithm, as an attempt to improve the performance of the algorithm [38]–[41]. ZHU H Y *et al.* proposed using k-medians algorithm for clustering, as an attempt to avoid the weaknesses attributed to outliers in k-means clustering, while increasing the algorithm speed [42]. Pourpanah F *et al.* extended BSO to an adaptive algorithm based on multiple groups, thereby improving the mutation effect of BSO, whereas the effect on multi-parameter optimization was insignificant [43]. In this study, the difference-mutation Brain Storm Optimization Algorithm (DBSO) replaced the Gaussian mutation of the BSO algorithm by complying with the BSO algorithm, which could improve the convergence rate and especially apply to the hyper-parametric optimization of the ensemble learning model [44].

As chromatographic technology has been advancing over the past few years, the detection of gas composition and concentration turns out to be rapid and accurate [45], [46]. Accordingly, in this study, the chromatographic technology acted as the vital technology of the transformer fault diagnosis. The chromatographic technology was employed to detect the transformer oil of the respective fault type, and the relevant data information was acquired. A series of preprocessing was performed on the data, and the data characteristics were extracted and normalized. A variety of fault identification models were built and classified for the processed data [6]–[12].

This study proposed a transformer fault diagnosis method based on DBSO-CatBoost. First, the dissolved gas data in transformer insulation oil were preprocessed by feature extraction, dimension reduction and normalization. Subsequently, the CatBoost model optimized by DBSO algorithm was built. Next, the processed data were trained and tested by using DBSO-CatBoost model. Lastly, the running state of the transformer was determined, and the power transformer faults were accurately diagnosed. This study builds various classification and recognition models, compares multiple models, and lastly develops a more suitable classification model for the transformer fault diagnosis. In the end, the whole study is summarized.

II. BASIC THEORY PART

A. CATBOOST MODEL

CatBoost is a machine learning library supporting categorical variables, which complies with the GBDT algorithm framework. It is capable of effectively solving various data migration problems in the original GBDT, while exhibiting the advantages of fewer parameters, high accuracy and good robustness [22], [23].

1) GBDT ALGORITHM

Ensemble learning builds multiple machine learners, trains them to form multiple weak learners, and combines multiple

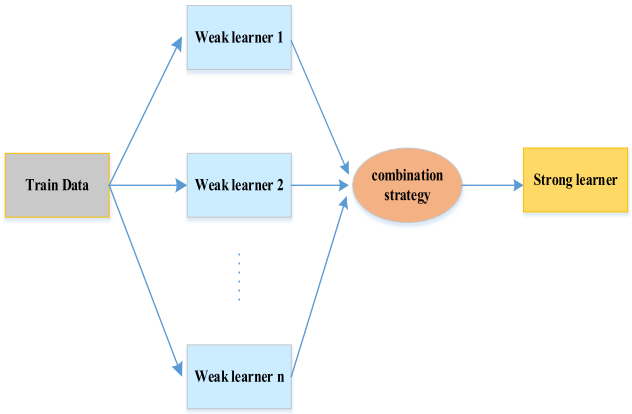

FIGURE 1. Block diagram of ensemble learning principle Boosting.

weak learners via some combination strategies to form a strong learner. Fig. 1 illustrates the principle block diagram of ensemble learning.

The algorithm is a framework algorithm of ensemble learning, with a basic idea to exploit the basic classification weak learner to obtain a strong learner by linear weighting and iterative training.

GBDT algorithm acts as an ensemble learning algorithm based on Boosting algorithm, which combines gradient lifting algorithm and decision tree. The model is an additive model, the learning algorithm is forward step-by-step algorithm, and the basis function is CART tree.

The concrete steps of GBDT algorithm are elucidated below:

Step 1. Initializing the weak learner:

$$f_0(x) = \arg\min_c \sum_{i=1}^{N} L(y_i, c) \tag{1}$$

where L(y_i, c) denotes the loss function; y_i represents the first prediction target; c expresses the parameter with the least square loss function.

Step 2. Calculate the negative gradient of the current loss function as sample residuals:

$$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)} \tag{2}$$

where the number $m = 1, 2, \ldots, M$ of iterations set, M denotes the total number of iterations. The sample is $i = 1, 2, \ldots, N$, N expresses the total number of samples.

Step3. With (x_i, r_{im}) as the training set of the next tree, fitting a CART regression tree, get the leaf node set R_{jm}, Leaf nodes $j = 1, 2, \ldots, J$, J represent the number of leaf nodes in the regression tree.

Step4. Calculate the minimum loss function for leaf node j:

$$r_{jm} = \underbrace{\arg\min}_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma) \tag{3}$$

where γ denotes the parameter of the respective leaf node.

Step 5. Update the strong learner:

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} \gamma_{jm} I_{jm}(x \in R_{jm}) \tag{4}$$

where I_{jm} represents the jth regression tree.

Step5: Combine weak learner to form strong learner:

$$f(x) = f_M(x) = f_0(x) + \sum_{m=1}^{M} \sum_{j=1}^{J} \gamma_{jm} I_{jm}(x \in R_{jm}) \tag{5}$$

2) CATBOOST ALGORITHM

The prediction model in GBDT algorithm is determined by the target variables of training samples, and there is an over-fitting problem attributed to biased point-state gradient estimation. Catboost algorithm is an improvement based on GBDT framework, which can effectively address the mentioned problems [47].

Compared with other GBDT algorithms (e.g., XGBoost and LightGBM), Catboost has been optimized in numerous aspects. First, CatBoost adopts the 'ordered principle' to avoid the conditional displacement issue inherent in the iteration of GBDT algorithm, while making it possible to exploit the whole data set for training and learning. Second, CatBoost transforms the conventional gradient enhancement algorithm into Ordered Boosting algorithm, thereby solving the inevitable problem of gradient offset in the iteration, improving the generalization ability, reducing the possibility of overfitting and enhancing the robustness of the model. Lastly, CatBoost builds the combination of classification features through greedy strategy, and takes the mentioned combinations as additional features, which makes it easier for the model to capture high-order dependencies and improve the prediction accuracy more significantly. Furthermore, CatBoost selects the forgetting decision tree as the basic prediction period, thereby reducing the possibility of overfitting and increasing the execution speed of the model [24]–[28].

Set the dataset to:

$$D = (X_i, Y_i) \tag{6}$$

where i = 1, 2 . . . n, n is the number of sample groups. The respective group of samples $X_i = (x_i^1, x_i^2 \ldots x_i^m)$, x_i^m is the first feature vector of group i samples. Y_i denotes the label value. The main methods of CatBoost algorithm include:

Multiple rankings are randomly generated for learning, the same class samples are found under the respective feature, and the classification feature conversion value is calculated:

$$\hat{x}_i^k = \frac{\sum_{j=1}^{n} \varphi(x_j^k = x_i^k) Y_j + \alpha p}{\sum_{j=1}^{n} \varphi(x_j^k = x_i^k) + \alpha} \tag{7}$$

where φ denotes the indicator function, which is 1 when $\left\{x_j^k = x_i^k\right\}$ is satisfied; otherwise, it is 0. p is a priori value. α is a priori weight.

The respective group of samples X_i in the training set has a model obtained by training the other training sets without X_i.

The combination of classification features is built in accordance with greedy strategy, and the tree structure is selected. The Ordered Boosting algorithm is adopted to calculate the gradient of X_i, and the gradient is employed to train the weak learner. Besides, the final model is developed by weighting.

B. DBSO ALGORITHM

1) BSO ALGORITHM

Brain Storm Optimization Algorithm (BSO) is an intelligent algorithm proposed by Professor Shi Yuhui in 2011, largely simulating the group behavior in human creative problem solving. It exploits the clustering idea to search the local optimum, while obtaining the global optimum by comparing the local optimum [48]. The mutation idea complicates the algorithm and avoids the algorithm falling into local optimum, which applies to solving the multi-peak high-dimensional function problem.

The BSO algorithm mainly comprises the steps below:

① Initialize the population.

② Individual evaluation and clustering.

③ Selecting cluster centers.

④ New individuals are generated through variation and then updated.

⑤ If the maximum number of iterations is reached, the optimal individual is outputted; otherwise, it is transferred to the second step.

The main part of BSO algorithm is clustering and mutation [49].

BSO employs K-means clustering algorithm to cluster individuals into k categories in accordance with the distance between individuals, while taking the individuals with the optimal fitness function value as the clustering center. To prevent falling into local optimum, the mutation individuals generated by probability replace one of the clustering centers.

BSO variation covers four major ways:

(1) adding random disturbance to a random class center, i.e., the optimal individual of this class, to generate new individuals;

(2) randomly selecting an individual in a random class to add random perturbations for generating novel individuals.

(3) randomly fusing two class centers and adding random perturbation to generate novel individuals;

(4) randomly fusing two random individuals in the two classes, while adding random disturbance to generate new individuals.

2) DBSO ALGOAITHM (DIFFERENCE-MUTATION BRAIN STORM OPTIMIZATION)

For several complex optimization problems, BSO algorithm exhibits slow convergence speed or premature problem. To improve the optimization performance, this study adopted DBSO algorithm to optimize the parameters of CatBoost model.

The DBSO algorithm exhibits the identical overall structure to the classical BSO algorithm, whereas the difference mutation is applied, other than the Gaussian mutation in the fourth step.

The classical BSO algorithm applies Gaussian mutation, and the novel individual generation equation is expressed as:

$$x_{nd} = x_{sd} + \xi \times N(0, 1)_d \tag{8}$$

where x_{nd} denotes the new d-dimensional individual; x_{sd} represents the selected individual; $N(0, 1)_d$ expresses the d-dimensional standard normal distribution; ξ is the coefficient of Gaussian function:

$$\xi = \lg sig(\frac{0.5 \times T - t}{k}) \times R(0, 1) \tag{9}$$

where T and t respectively represent the maximum number of iterations and the current number of iterations. k could adjust the slope of the lg $sig()$ function, and $R(0, 1)$ is a random value from 0 to 1.

In this variation, the requirements can be met at the early stage, whereas the coefficient of variation of Gaussian variation tends to be fixed at the subsequent stage, so it cannot well capture the search characteristics [44]. Thus, DBSO algorithm adopts differential mutation.

In human brain storms, everyone' s ideas at the early stage will be significantly different. Differences in existing ideas should be considered when creating novel ideas. Accordingly, DBSO algorithm determines the mutation step by differential mutation. The specific operation is defined as follows:

$$y = \begin{cases} R \times (H_d - L_d) + L_d, & rand() < p_r \\ x + R \times (x_a + x_b), & other \end{cases} \tag{10}$$

where y denotes the new generated individual; R represents the random number between 0 and 1; L_d and H_d express the upper and lower bounds of the search space; p_r is the open approximation set; rand() denotes the function to generate random numbers; x represents the selected individual; x_a and x_b express two different individuals selected in the contemporary global.

According to Eq. (10), compared with Gaussian variation, the calculation amount of the above differential variation is significantly reduced. Moreover, since the variation could be adaptively adjusted by complying with the dispersion degree of individuals in the group, it could more effectively share information and improve the search efficiency. Thus, compared with the BSO algorithm, DBSO algorithm could better balance local search and global search, and improve the algorithm performance.

III. TRANSFORMER FAULT DIAGNOSIS MODEL BASED ON DBSO-CATBOOST

This study adopted CatBoost model to diagnose transformer faults. As impacted by some parameters of Catboost model under the default value, there would be overfitting or underfitting. If manually adjusted, it would be time-consuming to find the optimal value. Accordingly, DBSO optimization algorithm is adopted to optimize the parameters of Catboost model to improve the performance of diagnosis

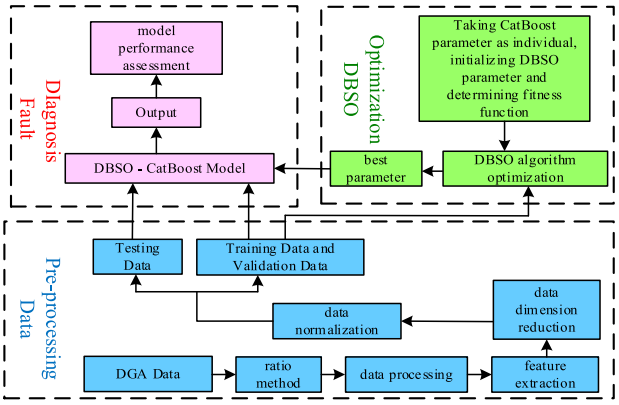

FIGURE 2. Transformer fault diagnosis model based on DBSO-CatBoost.

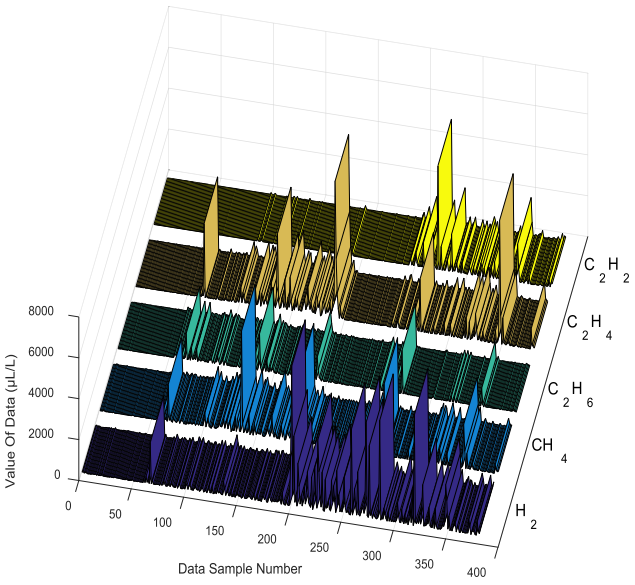

FIGURE 3. 3D view of DGA data.

model. For transformer fault diagnosis, this study built a DBSO-CatBoost model (Fig. 2).

The transformer fault diagnosis model based on DBSO-CatBoost primarily comprises data preprocessing, DBSO optimization and fault diagnosis. Data preprocessing mainly covers feature extraction, dimension reduction and normalization of the collected DGA sample data, as well as sequence division. The DBSO optimization part exploits the DBSO model to optimize several parameters of the CatBoost model to obtain the optimal parameters. The model training and testing part is training and testing CatBoost model, while outputting transformer fault types and assessing the model.

IV. CASE ANALYSIS

A. DATA ACQUISITION

The data in this study was provided by a power grid in the northwest of the State Grid Corporation of China, and H_2, CH_4, C_2H_6, C_2H_4, C_2H_2 were selected as the attributes of the transformer fault diagnosis, including 381 groups of fault data. The three-dimensional view of the data was shown in Fig. 3.

According to Fig. 3, any single feature with large difference in DGA data cannot accurately determine a fault type of transformer, and there were some coupling relationships between the feature attributes of the data, so it was necessary to extract the feature of the data [50], [51].

B. DATA PRE-PROCESSING

1) FEATURE EXTRACTION

According to GB-T 7252 - 2016 Guidelines for Analysis and Judgment of Dissolved Gases in Transformer Oil, the gas production rate of transformer insulating oil is correlated with the fault type of transformer, i.e., the fault type of transformer is correlated with the ratio of the respective gas concentration. Thus, the ratio between the characteristics of transformer fault types and input attributes was related. The common three-ratio method and non-coding method could roughly determine a fault type of transformer independently [6]–[9], so the characteristic variables generated by the interactive ratio method of input attributes exerted the decoupling effect on transformer fault diagnosis data.

The common three-ratio method and non-coding method could roughly determine a certain fault type of transformer separately, whereas the characteristic dimension generated by them cannot completely decouple the data. To achieve better decoupling effect, this study selected to traverse the data attribute ratio. The selected data feature variables were mainly composed of component concentration and its ergodic ratio. The form of ergodic ratio is determined by the formula below. DGA data had five-dimensional attributes, so the interactive ratio of data attributes is expressed below:

$$\frac{N_1}{N_2}, \frac{N_1}{N_3+N_4}, \frac{N_1}{N_3+N_4+N_5}, \frac{N_1}{N_3+N_4+N_5+N_6}.$$

To be specific, N_3, N_4, N_5, N_6 denotes any different attribute of DGA data, N_1, N_2 represents any attribute of DGA data ($N_1 \neq N_2$). Using enumeration algorithm, the new 145-dimensional feature variables were obtained by traversing all the permutations and combinations of four groups, and the original 5-dimensional feature variables were added to 150-dimensional data feature variables.

Since some data in the collected DGA data were zero, the feature attributes added by the ratio method achieved the case of zero division, so abnormal data would be generated.

On the whole, the processing methods of abnormal data comprised Laida criterion filling and fixed value filling. The DGA data was excessively scattered, and the data level difference was significant. Filling with the Layida criterion would eliminate most of the data, so this method was not suitable for DGA data. Here, the fixed-value filling method was used to process the abnormal data.

Each of the 150-dimensional data feature variables had different contributions to the sample, and the addition of some variables sometimes increased the complexity of the model, while affecting the accuracy of the model. Accordingly, the Shapley Additive Interpretation (SHAP) method was used for feature extraction.

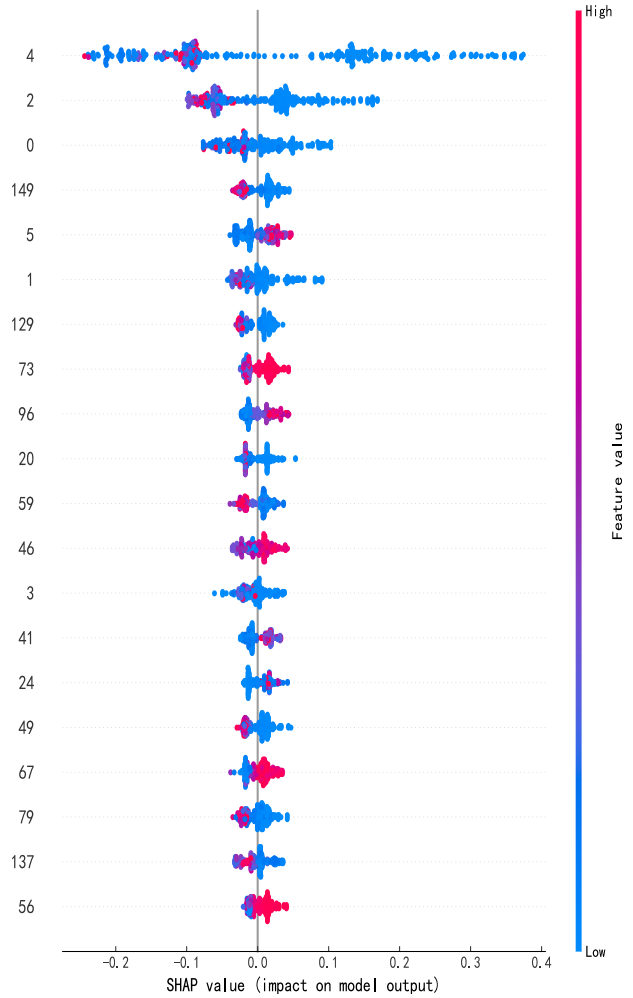

FIGURE 4. The beeswarm graph of SHAP values of each feature.

The SHAP value method builds an additive explanatory model. The core idea was to calculate the marginal contribution of features to the output of the model, and then explain the black box model from the global and local levels. All features were regarded as 'contributors.' For the respective prediction sample, the model produces a predictive value, and the SHAP value was the value assigned to the respective feature in the sample [52].

The SHAP values of the respective feature were calculated for 150-dimensional feature variables, and the feature density scatter plot was made. The respective row in the beeswarm graph represents a feature. Considerable samples were gathered in a wide area, and the abscissa was the SHAP value. A point represents a sample, and the color of the point represents the relative value of the point. The redder the color, the greater the blue would be, and the smaller the color would be. The ordinates in Fig. 4 were sorted by descending order of the average absolute value of SHAP value, and the first 20 characteristics of the intermediate temperature overheating category were taken as beeswarm diagram (Fig. 4).

The ordinate number in Fig. 4 represents the name of the sample property. Numbers 0, 1, 2, 3 and 4 represent H_2, CH_4, C_2H_6, C_2H_4, C_2H_2, respectively. Fig. 4 shows that the average absolute value of SHAP value of C_2H_2 was the largest, C_2H_2 has the greatest impact on the classification of samples. In addition, H_2, CH_4, C_2H_6, C_2H_4 was also very important for sample classification [53].

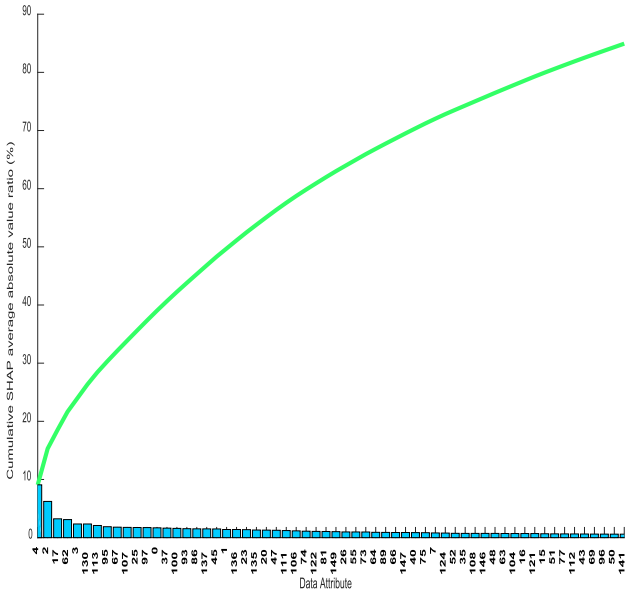

FIGURE 5. Influence of sample characteristics on histogram.

The beeswarm graph only visualizes the SHAP values of all samples in one category, which does not represent the interpretability of the overall model. For the multi-classification situation in this study, the mean of the average absolute value of SHAP in the respective classification was taken to obtain the overall average absolute value of SHAP, and the sample characteristics were used to influence the histogram [54].

In the histogram, the respective column represents a feature. The abscissa was sorted by descending order of the average absolute value of the SHAP value, and the ordinate was the proportion of the average absolute value of the feature SHAP to the sum of the average absolute values of all the features SHAP. The line graph represents the ratio of the average absolute value of SHAP accumulated by the previous features to the sum of the average absolute value of SHAP of all features, while taking the top 60 features (Fig. 5).

According to Fig. 5, the average absolute value of SHAP of C_2H_2 was the largest, which has the greatest impact on data classification. According to the curve in Fig. 5, the average absolute value of the top 60 cumulative SHAP in the figure took up nearly 90%, so the mentioned 60 features were taken as the attributes of the data.

2) DATA DIMENSION REDUCTION

The respective sample has 60 features after feature extraction, a total of 381 samples, so the data dimension was still too large. Data dimension was too large, would increase the complexity of the model, so, the data dimension should be reduced.

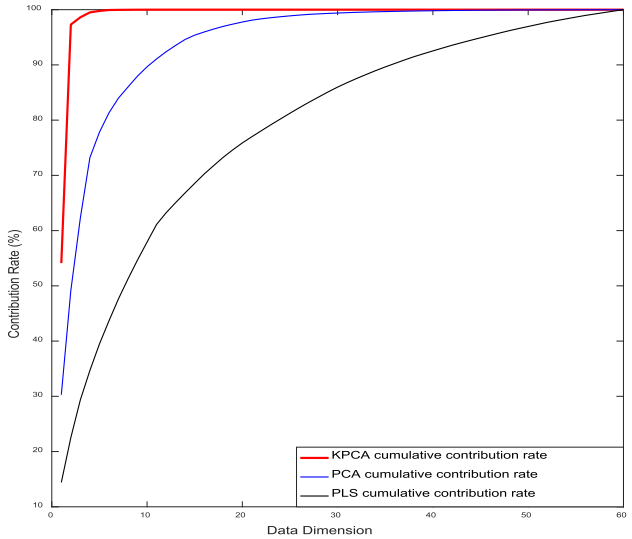

FIGURE 6. Cumulative contribution rate of dimension reduction algorithm.

The common dimensionality reduction algorithms were principal component analysis (PCA), kernel principal component analysis (KPCA) and partial least squares (PLS) [55]–[59].

Principal component analysis (PCA) maps the original variables to a new variable space. In the new variable space, several variables could be used to replace the original variables, and the data content of the original variables could be retained as much as possible. The new variables were orthogonal to each other to eliminate the collinearity of the original variables.

Kernel principal component analysis (KPCA) achieved the nonlinear mapping of data by mapping the original data to a higher dimensional space, and then employed principal component analysis to reduce the linear dimension of data from high dimensions [55]–[57].

PCA, PLS and KPCA were adopted to reduce the dimension of the data, and the results were shown in Fig 6.

Fig. 6 shows that the cumulative contribution increases with the increase in the dimension, and no longer increases after reaching 100%. The cumulative contribution of KPCA was obviously higher than other dimension reduction algorithms. When the dimension was 7, the cumulative contribution rate of KPCA was 99.9%, while the cumulative contribution rates of PCA and PLS did not reach 90%. Subsequently, with the increase in the dimension, the cumulative contribution rate of KPCA increased slightly, and the training time of model increased with the increase in the dimension.

According to Fig. 6, KPCA was significantly better than the other algorithms, so this study uses KPCA algorithm to reduce the data dimension to 7 dimensions.

3) DATA NORMALIZATION

The difference of DGA data was large, affecting the processing speed of the model, so the data normalization processing [60]. In this study, the interval value method was used to normalize the data, so that the data was scaled to a specific interval in proportion to avoid the interaction between values. Here, the extreme value method was selected for linear function transformation:

$$X_i(d) = \frac{x_i - \min X_i}{\max X_i - \min X_i} \times [1 - (-1)] + (-1) \quad (11)$$

$X_i(d)(i = 1, 2, \ldots, n)$ denotes the normalized data, and the mapping interval is $[-1, 1]$. X_i represents the original data. $\max X_i$ denotes the maximum value in the data sample. $\min X_i$ expresses the minimum value in the data sample.

The normalized data after dimension reduction could be inputted to train and test the model.

C. FAULT STATE CODING AND SEQUENCE DIVISION

The output result of the diagnosis model was the fault type of the transformer. According to GB-T 7252-2016 transformer oil dissolved gas analysis and judgment guidelines, this study takes low temperature overheating, medium temperature overheating, high temperature overheating, partial discharge, low energy discharge, high energy discharge and normal operation as the output characteristics of the transformer fault diagnosis. In this study, a training set, a validation set and a test set were set at a ratio of 3:1:1. The number of fault state codes and their corresponding sequences was shown in Table 1.

D. COMPARISON OF MULTI-MODEL DIAGNOSIS RESULTS

For the preprocessed data, six models, including extreme learning machine (ELM), support vector machine (SVM), GRNN, random forest (RF), XGboost and Catboost, were used for fault diagnosis to test the performance of various models for transformer fault diagnosis. The diagnosis results were shown in Fig. 7.

According to Fig. 7, the overall accuracy of Catboost was the highest in the six models, and the SVM algorithm was the highest in the single learner. The accuracy of ensemble learning algorithm was higher than that of single learner. The specific accuracy of the respective model for each type was shown in Table 2.

According to Table 2, the overall accuracy of the six models from low to high was GRNN, ELM, SVM, random forest, XGboost, Catboost. The overall accuracy of Catboost algorithm was the best when the empirical parameters were used, but the training time of ensemble learning algorithm was long. If the grid search traversal parameter adjustment method was used, the time required was too long, and the parameter adjustment range was relatively limited. However, single learner classification was not good. Compared with the single learner model, the ensemble learning model exhibited higher fault diagnosis accuracy for oil-immersed transformers.

E. COMPARISON OF PARAMETER OPTIMIZATION ALGORITHMS OF CATBOOST MODEL

The performance of Catboost model was better than other models. The training set of Catboost classification model was

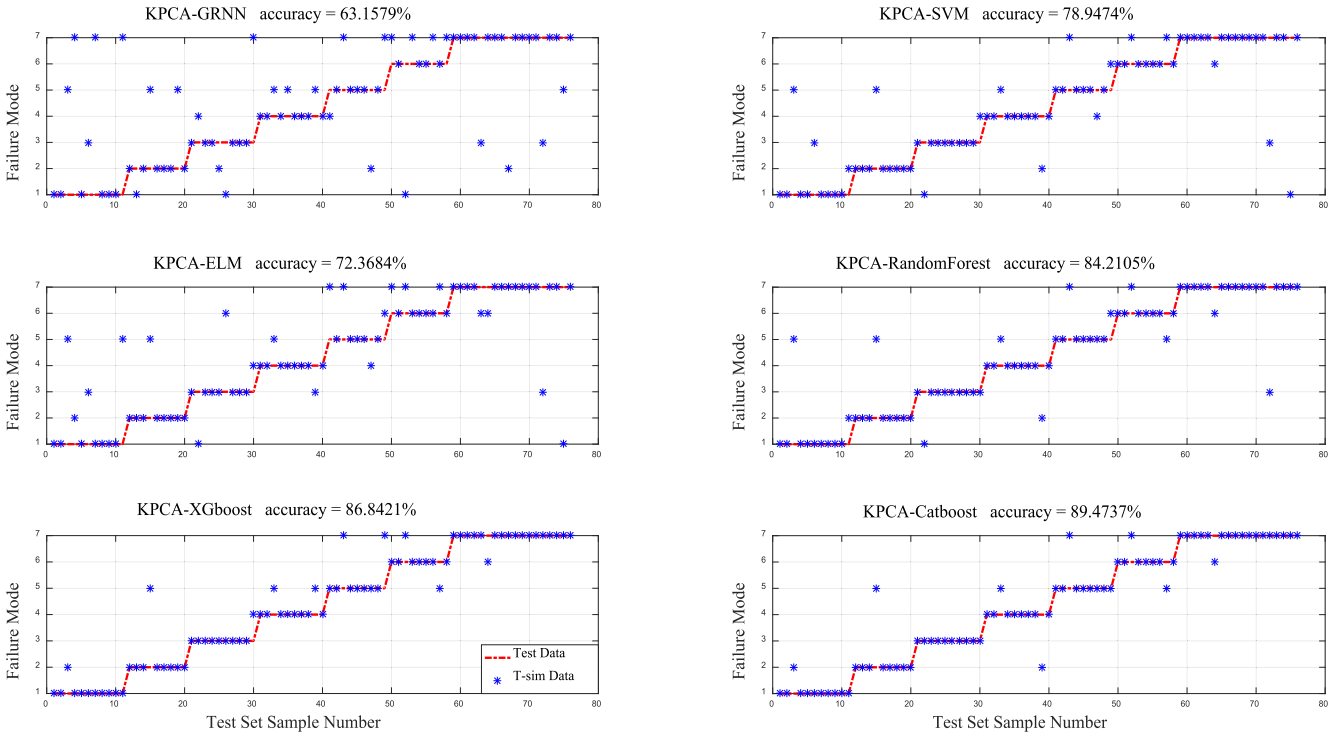

FIGURE 7. Diagnostic results of various model.

TABLE 1. Transformer fault type coding table.

Failure mode	encoding	Number of training sets	Number of validation sets	Number of test sets
normal operation	1	32	11	11
low temperature overheating	2	27	9	9
middle temperature overheating	3	31	10	10
high temperature overheating	4	29	10	10
partial discharge	5	27	9	9
low energy discharge	6	28	9	9
high energy discharge	7	55	18	18

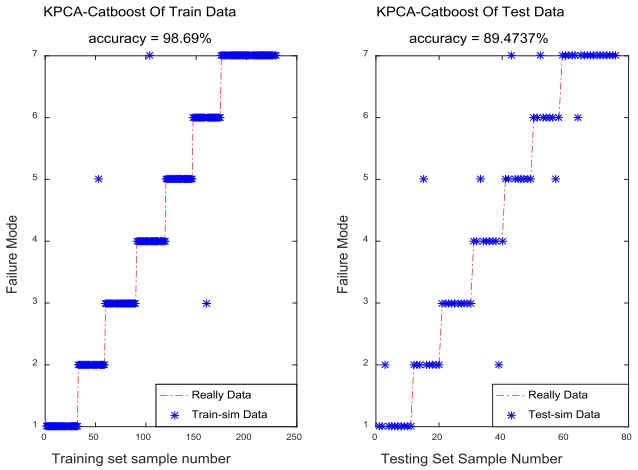

FIGURE 8. Catboost model training set and test set diagnostic results.

analyzed, and the data processed by ratio method combined with KPCA were used as input features. The diagnosis results of Catboost model training set and test set are presented in Fig. 8.

In Fig.8, Catboost model uses default parameters. According to Fig.8, the Catboost model classification results were over-fitting, so the parameters of the Catboost model should be optimized.

If Catboost model adopted manual adjustment of parameters, it would not only take a long time to adjust parameters, but also find the global optimum of parameters. If the grid search method was used for parameter adjustment, the time required was too long and the range of parameter adjustment was limited. Accordingly, the optimization algorithm was used to adjust the parameters of Catboost model.

Catboost was trained by the gradient lifting method. In the respective iteration, the basis for producing a new learner was that the regularization objective function was the smallest, and the regularization parameter L2_leaf_reg was too large

TABLE 2. Accuracy of classification models.

	normal operation	low temperature overheating	middle temperature overheating	high temperature overheating	partial discharge	low energy discharge	high energy discharge	total accuracy
ELM	7/11	8/9	7/10	8/10	5/9	6/9	14/18	55/76
SVM	8/11	8/9	8/10	8/10	6/9	7/9	15/18	60/76
GRNN	6/11	6/9	6/10	7/10	5/9	4/9	14/18	48/76
RandomForest	9/11	8/9	9/10	8/10	7/9	7/9	16/18	64/76
XGboost	10/11	8/9	9/10	8/10	7/9	7/9	17/18	66/76
Catboost	10/11	8/9	10/10	8/10	8/9	7/9	17/18	68/76

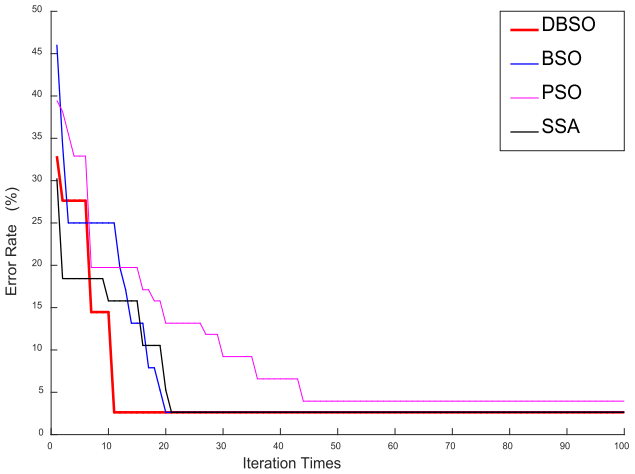

FIGURE 9. Fitness curve of each optimization algorithm.

or too small, which would cause over-fitting or under-fitting of the model. The learning rate parameter learning _ rate was too small, and the gradient descent was too slow. Too large, it may cross the optimal value and produce oscillation. The iteration number parameter iteration was too small would cause under-fitting, resulting in insufficient model solving ability. Too big would cause overfitting, resulting in a decline in generalization ability of the model. In addition, the random strength parameter random _ strength of the model was used to score the split tree, and improper selection would affect the learning ability and classification ability of the model [28]. Thus, this study selects the optimization algorithm to optimize the parameters of the above four Catboost models to improve the performance of the diagnosis model.

The common parameter optimization algorithms were particle swarm optimization (PSO), sparrow search algorithm (SSA) and so on. In this study, DBSO, BSO, PSO and SSA were used to optimize the four hyperparameters of Catboost model, and the results were compared [61]–[63].

The fitness function curve was made with the error rate of the classification results of the validation set as the fitness value. The fitness curve of the respective optimization algorithm was shown in Fig. 9.

According to Fig. 9, the DBSO algorithm first reached the optimal result, and the number of iterations to achieve the optimal fitness was 11, and the fitness value at the optimal time was the same as that of SSA and BSO algorithms, which was 2.132%. The final fitness value of PSO algorithm was the largest, and the optimization effect was the worst.

The Catboost model optimized by four algorithms was used for fault diagnosis, and the results were shown in Fig 10.

According to Fig.10, the test set accuracy of DBSO-Catboost model, BSO-Catboost model and SSA-Catboost model was the same, which was higher than that of PSO-Catboost.

In summary, although the accuracy of DBSO-Catboost model was the same as that of the other two models, it could find the optimal point faster and the optimization effect was the best.

F. CASE DATA ANALYSIS

Using 381 sets of data collected to build the model, some sample data are listed in Table 3.

The above data is adopted to construct features by ratio method, then feature screening, KPCA dimensionality reduction, normalization, and finally DBSO-Catboost algorithm is applied for prediction. The results are listed in Table 4.

According to Table 3 and Table 4, the proposed model achieves better accuracy than the traditional three-ratio method. According to the samples presented in Table 4, Catboost model and DBSO-Catboost model are used to analyze the confidence of the samples [64], [65].

The confidence of Catboost model is listed in Table 5. The confidence of DBSO-Catboost model is shown in Table 6.

According to Table 5 and Table 6, the confidence of the DBSO-Catboost model in the correct classification of samples is higher than that of the Catboost model, so the model classification method proposed here is effective.

G. PERFORMANCE EVALUATION OF DBSO-CATBOOST MODEL

1) DIAGNOSTIC RESULTS UNDER DIFFERENT PRETREATMENT METHODS

In this study, the ratio method was used to process the data and then the dimension reduction algorithm was used to reduce the dimension of the data. The original five-dimensional data, the data formed by the dimension reduction based on ratio method combined with KPCA, the data formed by the dimension reduction based on ratio method combined with PCA, and the data formed by the dimension reduction based on ratio method combined with PLS were used to form four different data sets with four different data processing methods. The DBSO-Catboost model was used to classify the four data. The classification results of the test set were shown in Fig. 11.

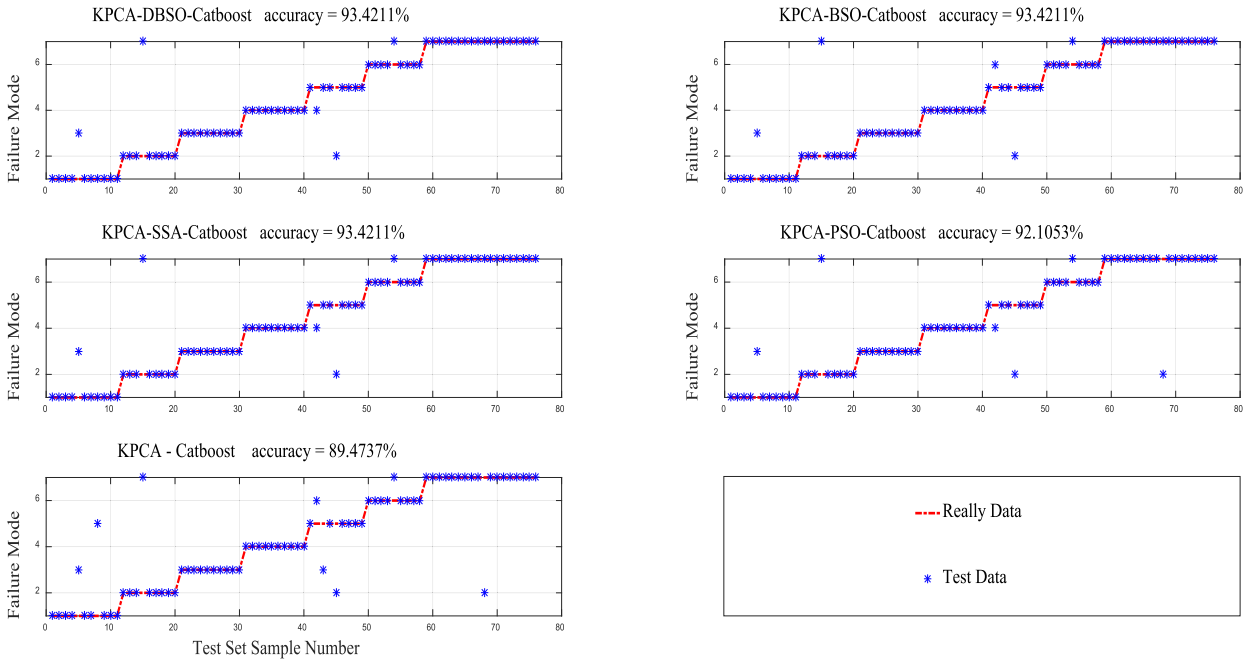

FIGURE 10. Diagnosis results of Catboost model optimized by different algorithms.

TABLE 3. Transformer fault data sample.

Example number	Data/($\mu L \cdot L^{-1}$)					three-ratio method	true fault
	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2		
Data 1	25.6	67.5	69.15	40.42	0.1	low temperature overheating	low temperature overheating
Data 2	211.14	769.25	465.03	1210.65	0	middle temperature overheating	middle temperature overheating
Data 3	83	60	46	48	1.5	coding-free	middle temperature overheating
Data 4	1549	93.2	35	47.3	0	coding-free	partial discharge
Data 5	289	13.4	18.2	49.8	36.3	high energy discharge	low energy discharge

TABLE 4. The processed data samples.

Example number	Sample data using ratio method to construct features and feature selection and KPCA dimension reduction normalized data							KPCA-DBSO-Catboost	true fault
Data 1	-0.482	0.175	-0.171	0.527	0.335	0.648	0.281	low temperature overheating	low temperature overheating
Data 2	-0.632	0.457	-0.449	0.486	0.450	0.681	0.224	middle temperature overheating	middle temperature overheating
Data 3	-0.498	-0.011	-0.322	0.513	0.349	0.667	0.164	middle temperature overheating	middle temperature overheating
Data 4	-0.592	-0.445	-0.402	0.484	0.431	0.705	-0.019	partial discharge	partial discharge
Data 5	-0.523	-0.279	-0.611	0.475	0.039	0.625	0.030	low energy discharge	low energy discharge

According to Fig. 11, when the data were reduced to seven dimensions, the data classification effect of ratio method combined with PLS was the worst, and the classification effect of ratio method combined with KPCA was the best. In the case of DBSO-Catboost model, the accuracy of data after the dimension reduction based on ratio method combined with KPCA was 3.950%, 10.526% and 5.263% higher than that of the ratio method combined with PCA, the ratio method combined with PLS, and the original five-dimensional data, respectively. Accordingly, the classification effect of the data processed by the ratio method and the KPCA dimension reduction algorithm was better than that of the original data.

In addition, when the classification algorithm is implemented, the precision, recall and F1 score of the model are the three main indicators to judge the classification effect of

TABLE 5. Confidence table of Catboost diagnostic model.

Example number	normal operation	low temperature overheating	middle temperature overheating	high temperature overheating	partial discharge	low energy discharge	high energy discharge	KPCA-Catboost	true fault
Data 1	0.060	0.810	0.060	0.000	0.060	0.000	0.010	low temperature overheating	low temperature overheating
Data 2	0.000	0.000	0.990	0.010	0.000	0.000	0.000	middle temperature overheating	middle temperature overheating
Data 3	0.020	0.310	0.520	0.000	0.020	0.020	0.110	middle temperature overheating	middle temperature overheating
Data 4	0.010	0.010	0.010	0.000	0.960	0.010	0.000	partial discharge	partial discharge
Data 5	0.000	0.000	0.000	0.000	0.040	0.440	0.520	high energy discharge	low energy discharge

TABLE 6. Belief table of DBSO-Catboost diagnostic model.

Example number	normal operation	low temperature overheating	middle temperature overheating	high temperature overheating	partial discharge	low energy discharge	high energy discharge	KPCA-DBSO-Catboost	true fault
Data 1	0.010	0.860	0.070	0.000	0.041	0.009	0.010	low temperature overheating	low temperature overheating
Data 2	0.000	0.000	0.990	0.010	0.000	0.000	0.000	middle temperature overheating	middle temperature overheating
Data 3	0.050	0.210	0.590	0.000	0.010	0.010	0.130	middle temperature overheating	middle temperature overheating
Data 4	0.000	0.000	0.000	0.000	0.980	0.020	0.000	partial discharge	partial discharge
Data 5	0.020	0.000	0.000	0.010	0.035	0.485	0.450	low energy discharge	low energy discharge

the model [66].

$$Precision = \frac{TP}{\mathrm{TP} + FP} \tag{12}$$

The recall rate is determined by:

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

F1-Score, also known as the balanced F-fraction method, is calculated by:

$$F1\text{-}Score = \frac{Precision \times Recall}{Precision + Recall} \times 2 \tag{14}$$

where TP is true positive; FP is false positive; FN is false negative.

Taking the normal operation of the category as an example, the true positive represents: it is predicted as the correct number in the normal operation. False positive denotes the number of errors predicted in normal operation. False negative represents: the true value is the number of normal operation and prediction errors.

Macro-F1, i.e., the macro average method, is obtained by substituting the precision rate and recall rate of each transformer state into formula (14), and then the values of seven F1 - Scores are averaged.

Calculate the precision, recall, and F1-Score values of KPCA-DBSO-Catboost in Fig.11. The KPCA-DBSO-Catboost detailed prediction results are shown in Table 7.

According to Table 7 and formula 12-14, the precision, recall and F1-Score of KPCA-DBSO-Catboost method can be calculated. The details are shown in Table 8.

Therefore, combined with formula 12-14, Table 7 and Table 8, the F1-Score is 93.42%, and the Macro-F1 value is 92.63% by adding the F1-Score values of each class and dividing them into 7. The Macro-F1 value of

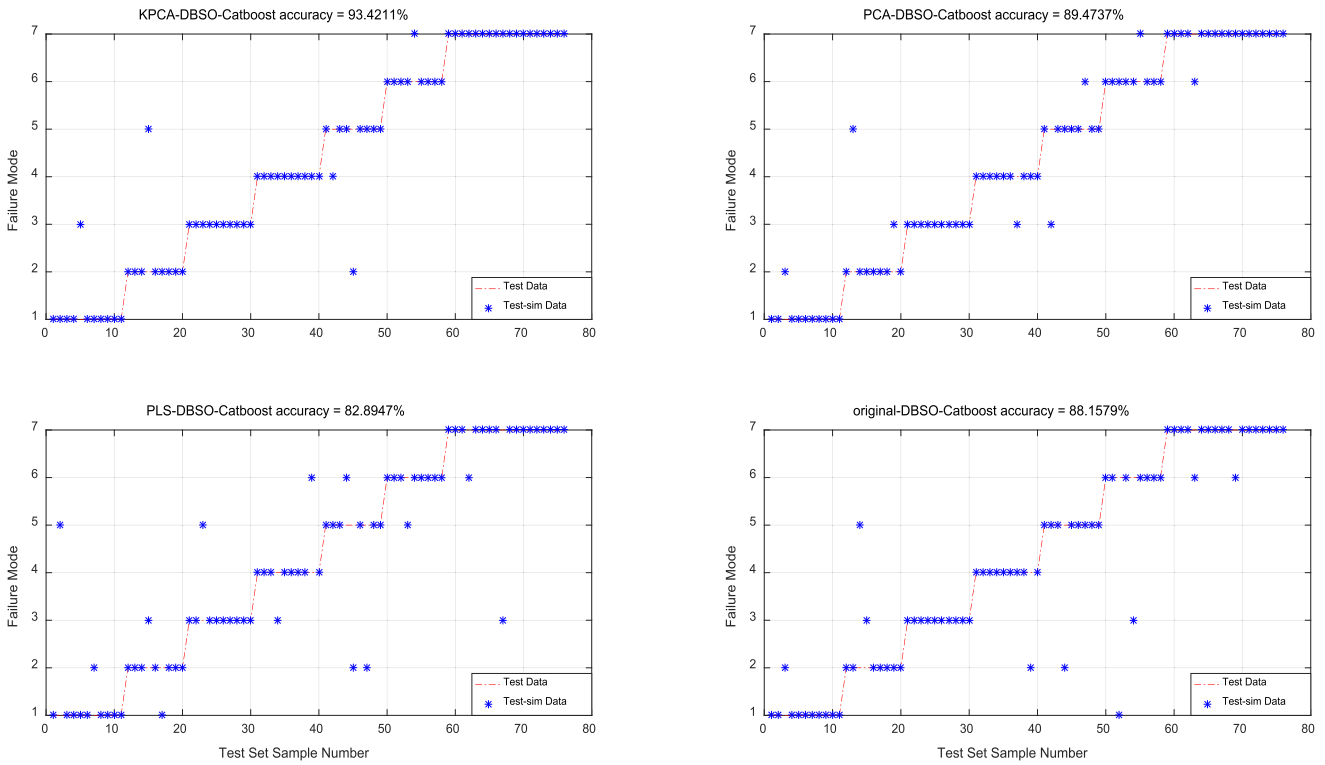

FIGURE 11. Diagnostic results of different pretreatments.

TABLE 7. Detailed diagnosis results.

failure mode	testing set	TP	FP	FN
normal operation	11	10	0	1
low temperature overheating	9	8	1	1
middle temperature overheating	10	10	1	0
high temperature overheating	10	10	1	0
partial discharge	9	7	1	2
low energy discharge	9	8	0	1
high energy discharge	18	18	1	0
grand total	76	71	5	5

TABLE 8. Detailed information table of each index.

failure mode	Precision/%	Recall/%	F1-score/%
normal operation	100	90.91	95.24
low temperature overheating	88.89	88.89	88.89
middle temperature overheating	90.91	100	95.24
high temperature overheating	90.91	100	95.24
partial discharge	87.50	77.78	82.35
low energy discharge	100	88.89	94.12
high energy discharge	94.74	100	97.30
grand total	93.42	93.42	93.42

Orginal-DBSO-Catboost model in Fig. 11 can be calculated by the same method, and it is found that it is less than 90%. This shows that the KPCA-DBSO-Catboost model of transformer fault diagnosis classification method is effective.

2) COMPARISON OF DIAGNOSTIC RESULTS OF DIFFERENT MODELS

DBSO optimization algorithm is employed to classify ELM, SVM, GRNN, Random Forest, XGboost and Catboost. After the data is processed by ratio method and KPCA, the optimal classification model is built.

The optimization algorithm is adopted to optimize the initial weights and thresholds of ELM model. The penalty factor C and kernel function parameter g of SVM model are optimized by optimization algorithm. The smoothing factor of GRNN model is optimized by using the optimization algorithm. Optimization algorithm is used to optimize decision tree tree and split feature number of Random Forest [67]. The regression tree k, learning rate η, maximum regression tree depth (max_depth), regularization coefficient λ, min_chile_weight and minimum splitting gradient descent δ of XGboost model are optimized by optimization algorithm. The optimization algorithm is adopted to optimize the regularization coefficient L2_leaf_reg parameter of Catboost model, the random strength random_strength parameter of splitting tree score, the iteration number iteration parameter and the learning rate learning_rate parameter. The population size is set to 20, and the number of iterations is set to 100. The classification diagnosis results of the respective model are presented in Fig. 12. The detailed information of each model classification is listed in Table 9.

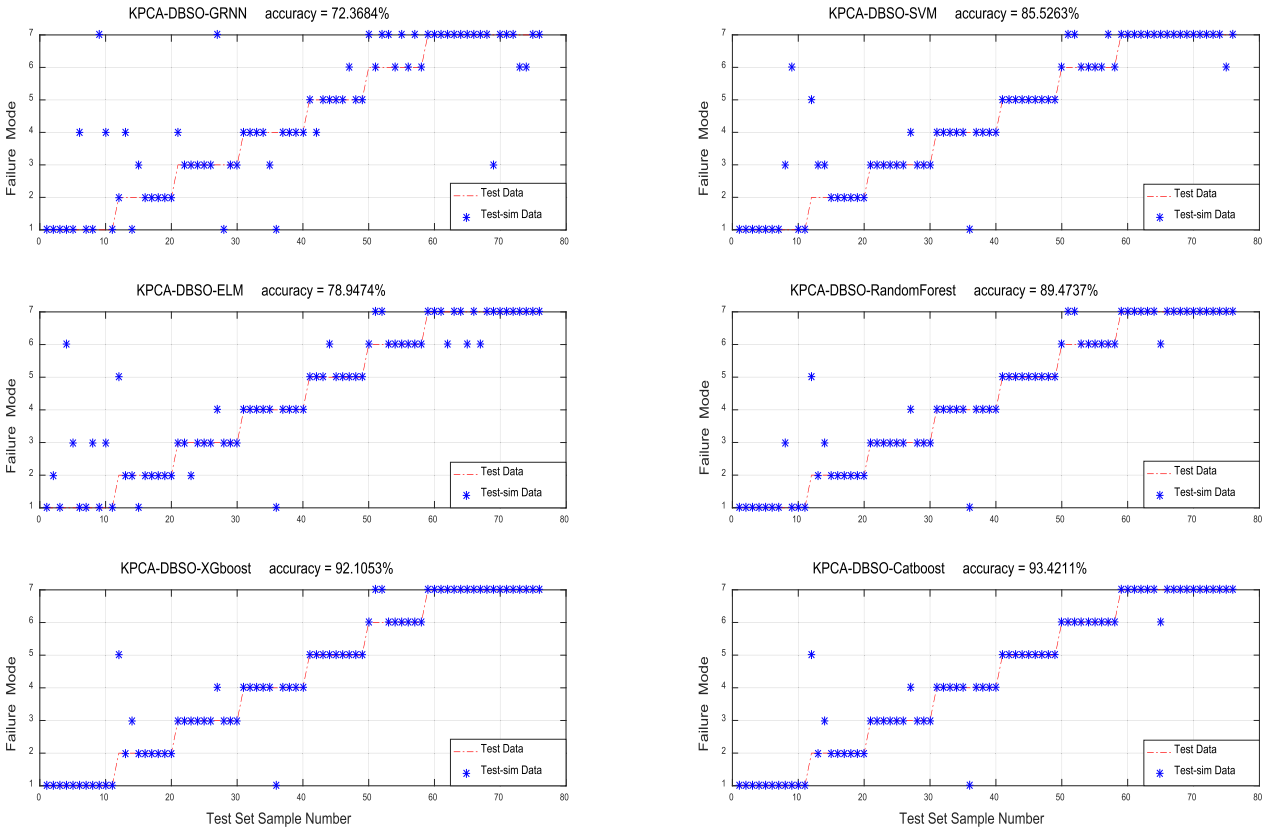

FIGURE 12. Classification and diagnosis results of each model.

TABLE 9. Detailed information table of diagnosis results of each model.

	KPCA-DBSO-GRNN	KPCA-DBSO-SVM	KPCA-DBSO-ELM	KPCA-DBSO-RandomForest	KPCA-DBSO-XGboost	KPCA-DBSO-Catboost
normal operation	8/11	9/11	6/11	10/11	11/11	11/11
low temperature overheating	6/9	6/9	7/9	7/9	7/9	7/9
middle temperature overheating	7/10	9/10	8/10	9/10	9/10	9/10
high temperature overheating	8/10	9/10	9/10	9/10	9/10	9/10
partial discharge	7/9	9/9	8/9	9/9	9/9	9/9
low energy discharge	4/9	6/9	7/9	7/9	7/9	9/9
high energy discharge	15/18	17/18	15/18	17/18	18/18	17/18
grand total	55/76	65/76	60/76	68/76	70/76	71/76

Analysis of Fig. 12 and Table 9 indicates that the DBSO-Catboost model has the optimal classification effect. After optimization by the optimization algorithm, the diagnostic effect of the classification model is improved. After the optimization algorithm, the classification effect of the respective model is significantly improved.

In order to verify the performance of the model, the PSO-RF model established by using the feature of non-coding ratio is compared with the model proposed in this paper [21]. The experimental results are shown in Fig.13.

According to fig.13, the transformer fault diagnosis model proposed in this paper can significantly improve the classification accuracy.

3) DIAGNOSTIC RESULTS UNDER DIFFERENT SEQUENCE SETS

The training set, test set and verification set of KPCA-DBSO-Catboost model were obtained by random disruption of 381 sets of raw data. Different sequence set partitions would affect the diagnosis results of the model. To ensure the reliability of the results, nine different sequence

TABLE 10. Diagnostic results of different data sets (%).

Different sequence sets	1	2	3	4	5	6	7	8	9	average accurate
KPCA-DBSO-Catboost	93.42%	92.11%	93.42%	94.74%	94.74%	94.74%	96.05%	90.79%	93.42%	93.71%
KPCA-Catboost	89.47%	90.79%	88.16%	92.11%	94.74%	90.79%	92.11%	86.84%	93.42%	89.76%

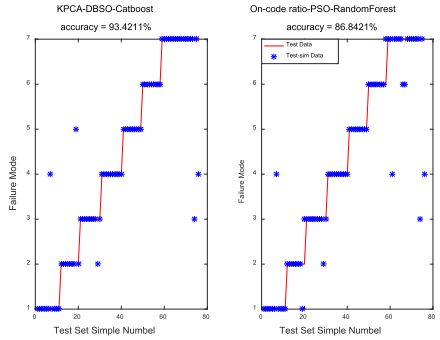

FIGURE 13. Comparison diagrams of different algorithms.

sets were used to train and diagnose the KPCA-DBSO-Catboost model and KPCA-Catboost model, respectively. Table 10 lists the results.

According to Table 10, compared with KPCA-Catboost model, the accuracy of KPCA-DBSO-Catboost model increases by 3.958 %, and the performance of KPCA-DBSO-Catboost model was not lower than that of KPCA-Catboost model under the same data.

V. CONCLUSION

To address the problems of power transformer fault diagnosis under conventional methods (i.e., low accuracy and poor classification effect), this study proposed a power transformer fault diagnosis method by complying with DBSO-Catboost model. With the collected transformer DGA data as an example, the proposed model was trained and tested. The conclusions were drawn below:

(1) Compared with the existing ELM, SVM, GRNN, RF and XGboost models, Catboost model exhibited the maximal accuracy, reaching 89.474%.

(2) When Catboost model was being optimized, the accuracy of DBSO, SSA and BSO algorithms was identical to the optimal fitness value, whereas the number of iterations was the least, with the best effect.

(3) Compared with the conventional feature selection method, the traversal ratio method in this study covered a wider range of effective information. In different dimensionality reduction algorithms, the cumulative contribution rate of KPCA was obviously higher than that of other dimensionality reduction algorithms. As indicated from the example results, after several data preprocessing methods proposed in this study, and then using the DBSO-Catboost model, after the data were processed by the ratio method and dimension reduction algorithm, the classification effect was better than that of the original data. To be specific, the data classification effect after KPCA dimension reduction was the best. When DBSO-Catboost model was adopted, the data accuracy of ratio method combined with KPCA dimension reduction was 3.950%, 10.526% and 5.263% higher than that of ratio method combined with PCA, ratio method combined with PLS and original five-dimensional data, respectively.

(4) Catboost model with default parameters could exhibit the high accuracy of 89.474%. After DBSO algorithm was adopted to optimize Catboost model, DBSO-Catboost model exhibited the accuracy of 93.567%, and the accuracy was improved by 3.958%.

REFERENCES

[1] X. Yanwei, L. Ping, J. Shui, L. Ying, C. Jiagan, "Research on transformer fault diagnosis technology in power network environment," *Transformer*, vol. 56, no. 4, pp. 46–50, 2019, doi: 10.19487/j.cnki.1001-8425.2019.04.012.
[2] N. A. Muhamad, B. T. Phung, T. R. Blackburn, and K. X. Lai, "Comparative study and analysis of DGA methods for transformer mineral oil," in *Proc. IEEE Lausanne Power Tech*, Jul. 2007, pp. 45–50.
[3] S. S. M. Ghoneim and I. B. M. Taha, "A new approach of DGA interpretation technique for transformer fault diagnosis," *Int. J. Electr. Power Energy Syst.*, vol. 81, pp. 265–274, Oct. 2016.
[4] S. A. Khan, M. D. Equbal, and T. Islam, "A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 1, pp. 590–596, Feb. 2015.
[5] H. Cui, X. Zhang, G. Zhang, and J. Tang, "Pd-doped MoS_2 monolayer: A promising candidate for DGA in transformer oil based on DFT method," *Appl. Surf. Sci.*, vol. 470, pp. 1035–1042, Mar. 2019.
[6] S. Bin, L. Zhixiong, and 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.
[7] 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.
[8] 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.
[9] 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.
[10] H. Cui, P. Jia, X. Peng, and P. Li, "Adsorption and sensing of CO and C_2H_2 by S-defected SnS_2 monolayer for DGA in transformer oil: A DFT study," *Mater. Chem. Phys.*, vol. 249, Jul. 2020, Art. no. 123006.
[11] 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.
[12] S. Zhang, Y. Bai, G. Wu, and Q. Yao, "The forecasting model for time series of transformer DGA data based on WNN-GNN-SVM combined algorithm," in *Proc. 1st Int. Conf. Electr. Mater. Power Equip. (ICEMPE)*, May 2017, pp. 292–295.
[13] S. Bustamante, M. Manana, A. Arroyo, P. Castro, A. Laso, and R. Martinez, "Dissolved gas analysis equipment for online monitoring of transformer oil: A review," *Sensors*, vol. 19, no. 19, p. 4057, Sep. 2019.
[14] M. Duval and L. Lamarre, "The new Duval pentagons available for DGA diagnosis in transformers filled with mineral and ester oils," in *Proc. IEEE Electr. Insul. Conf. (EIC)*, Jun. 2017, pp. 279–281.
[15] F. Jawad and S. Milad, "Assessment of computational intelligence and conventional dissolved gas analysis methods for transformer fault diagnosis," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 5, pp. 1798–1806, Oct. 2018.

[16] Z. Kefei, G. Jiang, N. Dexin, Y. Fang, and X. Zhihuai, "Study on oil-immersed transformer model based on chemical reaction optimization neural network and fusion DGA algorithm," *High Voltage Eng.*, vol. 42, no. 4, pp. 1275–1281, 2016, doi: 10.13336/j.1003-6520.hve.20160405005.

[17] T. Huang, H. Liu, G. Xiang, and X. Yang, "Analysis and research on transformer operating state based on SVM," *J. Chongqing Univ. Technol., Natural Sci.*, vol. 31, no. 7, pp. 162–168, 2017.

[18] D. Wenxia, Z. Xiuping, D. Hailian, and L. Feng, "Power transformer fault diagnosis based on extreme learning machine," *J. Shandong Univ. Sci. Technol., Natural Sci.*, vol. 36, no. 5, pp. 29–36, 2017.

[19] W. Liao, C. Guo, Y. Jin, and G. Xiao, "Oil-immersed transformer fault diagnosis method based on four-stage preprocessing and GBDT," *Power Syst. Technol.*, vol. 43, no. 6, pp. 2195–2230, 2019.

[20] Z. Youwen, F. 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.

[21] L. Hejian, X. Xiaowei, W. Ke, Z. Yongjun, W. Shizhe, and L. Kezhen, "Transformer fault diagnosis model based on particle swarm optimization random forest," *J. Kunming Univ. Sci. Technol. (Natural Sci. Ed.)*, vol. 46, no. 3, pp. 94–101, 2021.

[22] G. Huang, L. Wu, X. Ma, W. Zhang, J. Fan, X. Yu, W. Zeng, and H. Zhou, "Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions," *J. Hydrol.*, vol. 574, pp. 1029–1041, Jul. 2019.

[23] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: An interdisciplinary review," *J. Big Data*, vol. 7, no. 1, pp. 1–45, Dec. 2020.

[24] S. B. Jabeur, C. Gharib, S. Mefteh-Wali, and W. B. Arfi, "CatBoost model and artificial intelligence techniques for corporate failure prediction," *Technol. Forecasting Social Change*, vol. 166, May 2021, Art. no. 120658.

[25] J. Hancock and T. M. Khoshgoftaar, "Medicare fraud detection using CatBoost," in *Proc. IEEE 21st Int. Conf. Inf. Reuse Integr. Data Sci. (IRI)*, Aug. 2020, pp. 97–103.

[26] V. A. Dev and M. R. Eden, "Gradient boosted decision trees for lithology classification," *Comput. Aided Chem. Eng.*, vol. 47, pp. 113–118, Jan. 2019.

[27] L. Diao, D. Niu, Z. Zang, and C. Chen, "Short-term weather forecast based on wavelet denoising and catboost," in *Proc. Chin. Control Conf. (CCC)*, Jul. 2019, pp. 3760–3764.

[28] Z. Tao and F. Bo, "Loan risk prediction method based on CLPSO-CatBoost," *Comput. Syst. Appl.*, vol. 30, no. 4 pp. 222–226, 2021.

[29] Y. Shi, "Brain storm optimization algorithm," in *Proc. Int. Conf. Swarm Intell.* Berlin, Germany: Springer, 2011, pp. 303–309.

[30] H. Duan and P. Qiao, "Pigeon-inspired optimization: A new swarm intelligence optimizer for air robot path planning," *Int. J. Intell. Comput. Cybern.*, vol. 7, no. 1, pp. 24–37, Mar. 2014.

[31] Y. Shi, "An optimization algorithm based on brainstorming process," in *Emerging Research on Swarm Intelligence and Algorithm Optimization*. Hershey, PA, USA: IGI Global, 2015, pp. 1–35.

[32] H. Duan, S. Li, and Y. Shi, "Predator–prey brain storm optimization for DC brushless motor," *IEEE Trans. Magn.*, vol. 49, no. 10, pp. 5336–5340, Oct. 2013.

[33] C. Guogang, Z. Xinyu, C. Ying, C. Cong, and K. Deqing, "Medical image registration method based on improved brainstorming optimization algorithm," *Data Acquisition Process.*, vol. 35, no. 4, pp. 730–738, 2020.

[34] G. Sovatzidi, M. Savelonas, C. Dimitra-Christina Koutsiou, and D. K. Iakovidis, "Image segmentation based on determinative brain storm optimization," in *Proc. 15th Int. Workshop Semantic Social Media Adaptation Personalization*, Oct. 2020, pp. 1–6.

[35] D. Qing, L. Benwei, Y. Siqi, and Q. Renjun, "Performance parameters prediction of turboshaft engine acceleration process based on BSO-ELM," *Syst. Eng. Electron. Technol.*, vol. 43, no. 8, pp. 2181–2188, 2021.

[36] T. Haonan and Z. Hui, "'Two-order high-dimensional data feature selection based on BSO-OS algorithm," *Comput. Eng. Des.*, vol. 41, no. 3, pp. 695–700, 2020.

[37] Z. Lixin, Y. Kang, and L. Yuxi, "Multi-objective optimization of TA7 EDM based on BSO algorithm," *Mech. Strength*, vol. 40, no. 3, pp. 639–646, 2018.

[38] A. Hassanein, M. El-Abd, I. Damaj, and H. Ur Rehman, "Parallel hardware implementation of the brain storm optimization algorithm using FPGAs," *Microprocessors Microsyst.*, vol. 74, Apr. 2020, Art. no. 103005.

[39] M. S. Kumar and B. Indrani, "Brain storm optimization based association rule mining model for intelligent phishing URLs websites detection," in *Proc. 4th Int. Conf. Comput. Methodol. Commun. (ICCMC)*, Mar. 2020, pp. 640–646.

[40] F. Pourpanah, C. P. Lim, X. Wang, C. J. Tan, M. Seera, and Y. Shi, "A hybrid model of fuzzy min–max and brain storm optimization for feature selection and data classification," *Neurocomputing*, vol. 333, pp. 440–451, Mar. 2019.

[41] D. Azuma, Y. Fukuyama, T. Jintsugawa, H. Fujimoto, and T. Matsui, "Improved brain storm optimization with differential evolution strategies for load adjustment distribution state estimation using correntropy," *IFAC-PapersOnLine*, vol. 52, no. 4, pp. 449–454, 2019.

[42] H. Zhu and Y. Shi, "Brain storm optimization algorithms with k-medians clustering algorithms," in *Proc. 7th Int. Conf. Adv. Comput. Intell. (ICACI)*, Mar. 2015, pp. 107–110.

[43] F. Pourpanah, R. Wang, X. Wang, Y. Shi, and D. Yazdani, "MBSO: A multi-population brain storm optimization for multimodal dynamic optimization problems," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2019, pp. 673–679.

[44] M. Weiqiang, G. Yongqi, and Z. Miao, "Brainstorming optimization algorithm based on global optimum and difference mutation," *Syst. Eng. Electron. Technol.*, pp. 1–11, Dec. 2021. [Online]. Available: http://kns.cnki.net/kcms/detail/11.2422.TN.20210531.0826.002.html

[45] S.-X. Zhang, R. Peng, R. Jiang, X.-S. Chai, and D. G. Barnes, "A high-throughput headspace gas chromatographic technique for the determination of nitrite content in water samples," *J. Chromatogr. A*, vol. 1538, pp. 104–107, Feb. 2018.

[46] W.-Q. Xie, K.-X. Yu, and Y.-X. Gong, "Determination of iodate in iodized edible salt based on a headspace gas chromatographic technique," *J. Chromatogr. A*, vol. 1584, pp. 187–191, Jan. 2019.

[47] M. Al-Sarem, F. Saeed, W. Boulila, and A. H. Emara, "Feature selection and classification using CatBoost method for improving the performance of predicting Parkinson's disease," in *Advances on Smart and Soft Computing*. Singapore: Springer, 2021, pp. 189–199.

[48] L. Shijie, D. Chen, C. Mingzhi, Z. Fan, and C. Jinghui, "Summarization of new swarm intelligence optimization algorithms," *Comput. Eng. Appl.*, vol. 54, no. 12, pp. 1–9, 2018.

[49] Y. Yuting, S. Yuhui, and X. Shunren, "Brainstorming optimization algorithm based on discussion mechanism," *J. Zhejiang Univ. (Eng. Ed.)*, vol. 47, no. 10, pp. 1705–1711, 2013.

[50] L. Dianyang, Z. Yujie, F. Jian, and W. Shanyuan, "Multidimensional diagnosis of transformer fault samples and reliability analysis of results," *J. Elect. Technol.*, pp. 1–9, Dec. 2021, doi: 10.19595/j.cnki.1000-6753.tces.210070.

[51] W. Jing, X. Suan, H. Kaixing, W. Ce, L. Yong, and C. Xing, "Transformer fault diagnosis model based on DGA feature selection and GA-SVM," *Transformer*, vol. 57, no. 12, pp. 36–40, 2020.

[52] C. S. Kumar, M. N. S. Choudary, V. B. Bommineni, G. Tarun, and T. Anjali, "Dimensionality reduction based on SHAP analysis: A simple and trustworthy approach," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Jul. 2020, pp. 558–560.

[53] L. Antwarg, R. M. Miller, B. Shapira, and L. Rokach, "Explaining anomalies detected by autoencoders using SHAP," 2019, *arXiv:1903.02407*.

[54] Y. Bi, D. Xiang, Z. Ge, F. Li, C. Jia, and J. Song, "An interpretable prediction model for identifying N7-methylguanosine sites based on XGBoost and SHAP," *Mol. Therapy, Nucleic Acids*, vol. 22, pp. 362–372, Dec. 2020.

[55] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100378.

[56] J. H. Mita, C. G. Babu, and M. G. Shankar, "Performance analysis of dimensionality reduction using PCA, KPCA and LLE for ECG signals," in *Proc. IOP Conf. Mater. Sci. Eng.*, Mar. 2021, vol. 1084, no. 1, Art. no. 012005.

[57] S. Neffati, K. B. Abdellafou, I. Jaffel, O. Taouali, and K. Bouzrara, "An improved machine learning technique based on downsized KPCA for Alzheimer's disease classification," *Int. J. Imag. Syst. Technol.*, vol. 29, no. 2, pp. 121–131, Jun. 2019.

[58] Z. Rongjian and P. Feng, "Prediction model of glutamic acid fermentation product concentration based on PLS-LSSVM," *J. Chem. Eng.*, vol. 68, no. 3, pp. 976–983, 2017.

[59] L. Hongbin and S. Liu, "Prediction of effluent quality of wastewater treatment system by relevance vector machine," *China J. Papermaking*, vol. 34, no. 2, pp. 53–59, 2019.

[60] C. Tie, L. Changqin, Z. Xin, and C. Weidong, "Transformer fault diagnosis model based on KPCA-WPA-SVM," *Elect. Meas. Instrum.*, vol. 58, no. 4, pp. 158–164, 2021, doi: 10.19753/j.issn1001-1390.2021.04.023.

[61] H. Nguyen, H. Moayedi, L. K. Foong, H. A. H. Al Najjar, W. A. W. Jusoh, A. S. A. Rashid, and J. Jamali, "Optimizing ANN models with PSO for predicting short building seismic response," *Eng. Comput.*, vol. 36, no. 3, pp. 823–837, Jul. 2020.
[62] M. A. Hossain, H. R. Pota, S. Squartini, and A. F. Abdou, "Modified PSO algorithm for real-time energy management in grid-connected micro-grids," *Renew. Energy*, vol. 136, pp. 746–757, Jun. 2019.
[63] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: Sparrow search algorithm," *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 22–34, Jan. 2020.
[64] H. Xinbo, M. Yutao, and Z. Yongcan, "Transformer fault diagnosis method based on information fusion and M-RVM," *Power Automat. Equip.*, vol. 40, no. 12, pp. 218–225, 2020, doi: 10.16081/j.epae.202009040.
[65] L. Yunpeng, H. Jiahui, X. Ziqiang, L. Yijin, W. Quan, Y. Ning, and H. Shuai, "Transformer fault diagnosis method combined with AdaBoost and cost sensitive," *J. North China Electr. Power Univ., Natural Sci. Ed.*, pp. 1–9, Dec. 2021. [Online]. Available: http://kns.cnki.net/kcms/detail/13.1212.TM.20210719.1104.002.html
[66] H. Lingling and Z. Yongli, "Transformer fault diagnosis based on DCAE-SVM," *Comput. Appl. Softw.*, vol. 38, no. 5, pp. 49–53, 2021.
[67] W. Xue and H. Tao, "Transformer fault diagnosis based on Bayesian optimized random forest," *Electr. Meas. Instrum.*, vol. 58, no. 6, pp. 167–173, 2021.

MEI ZHANG was born in Suzhou, Anhui, China, in 1979. She received the bachelor's and master's degrees in electrical engineering from the Anhui University of Science and Technology, in 2002 and 2005, respectively. She is currently an Associate Professor with the School of Electrical and Information Engineering, Anhui University of Science and Technology. Her research interests include intelligent control, Internet of Things (IoT) technology, and embedded systems.

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. He is currently pursuing the master's degree with the Anhui University of Science and Technology. His research interests include intelligent control Internet of Things technology, Internet of Things (IoT) technology, and data analysis.

YU ZHANG received the bachelor's degree in electrical engineering and automation from the Xinjiang Institute of Technology. He is currently working at Xinjiang Hotan Power Supply Company of State Grid, focusing on the maintenance of dispatching automation system.

FEI LIU received the B.E. degree in electrical engineering and automation from the Xuchang College. He is currently pursuing the M.E. degree in energy dynamics with the Anhui University of Technology, focusing on data analysis and embedded system.

DONGSHUN YU received the B.E. degree in electrical engineering and automation from Fuyang Normal University. He is currently pursuing the M.E. degree in energy power with the Anhui University of Technology, focusing on process control system.

CHAOYIN ZHANG received the B.E. degree in electrical engineering and automation from the Chaohu College. He is currently pursuing the M.E. degree in energy dynamics with the Anhui University of Technology, focusing on data analysis and photoelectric information processing.

LI GAO received the B.E. degree in electrical engineering and automation from the Anhui University of Science and Technology, where he is currently pursuing the M.E. degree in energy dynamics, focusing on mining information processing.