Progress In Electromagnetics Research C, Vol. 127, 83–99, 2022

Hybrid Feature Selection Approach for Power Transformer Fault Diagnosis Based on Whale Optimization Algorithm and Extreme Learning Machine

Zhiyang He[1], Tusongjiang Kari[1, *], Yilihamu Yaermaimait[1], Lin Du[1], Yannan Zhou[2], and Zhichao Liu[1]

Abstract—To further improve fault diagnosis performance, a new hybrid feature selection approach combined with whale optimization algorithm and extreme learning machine is presented in this study. Firstly, three filter methods based on different evaluation metrics are employed to select and rank 25 input features derived from gases concentration values, gases ratio, and energy-weighted dissolved gas analysis. Then, feature fusion approaches are applied to aggregate feature ranks and form a lower-dimension candidate feature subset. Afterwards, the whale optimization-based extreme learning machine model is implemented to optimize parameters and select optimal feature subsets. The accuracy of the model is used to evaluate the fault diagnosis capability of the concerned feature subsets. Finally, novel subsets are determined as the optimal feature subset to establish a fault diagnosis model. According to the experimental results, the average accuracy of the proposed approach is better than that of other conventional methods, which indicates that the optimal feature subset obtained by the proposed method can significantly promote the fault diagnosis accuracy of the power transformer.

Nomenclature			
DGA	dissolved gas analysis	KGM	key gas method
IEC	international electrotechnical commission	SVM	support vector machine
BPNN	back propagation neural network	ANFIS	adaptive neuro fuzzy inference system
ELM	extreme learning machine	MVR	minimum violations ranking
EWDGA	energy weighted dissolved gas analysis	LED	low-energy discharge
PD	partial discharge	LT	thermal fault of low temperature
HED	high-energy discharge	HT	thermal fault of high temperature
MT	thermal fault of medium temperature	HFS	hybrid feature selection method
NC	normal condition	DM	dowdall method
MVS	majority voting strategy	SFS	similar feature subset
WOA-ELM	the whale optimization-based extreme learning machine mode	OFS	optimal feature subset

1. INTRODUCTION

The oil-immersed power transformer is a critical piece of equipment in transmission and substation networks. When failures or malfunctions occur in power transformers, it may lead to not only the

Received 18 September 2022, Accepted 17 November 2022, Scheduled 29 November 2022

* Corresponding author: Tusongjiang Kari (karlsjtu@163.com).
[1] School of Electrical Engineering, Xinjiang University, Urumqi 830047, China. [2] Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool England L69 3GJ, UK.

interruption of electricity supply, but also the collapse of the entire power grid and other serious economic loss. As a result, it is critical to diagnose and evaluate power transformer incipient faults and working condition [1].

Power transformers in operation encounter kinds of stresses, such as electrical, mechanical, thermal, chemical stress, and aging, which lead to the decomposition and degradation of mineral oil and solid insulation. The deterioration process will cause defections and incipient faults, such as partial discharge, arcing, and overheating, and a variety of gases would be consequently generated. The gas types and concentration values are related closely to the type and severity of faults. Besides, these generated gases are dissolved in power transformer oil. Hence, internal insulation of the power transformer or fault condition can be evaluated by measuring types and concentration values of the gases dissolved in power transformer oil.

In recent years, oil-immersed power transformer fault analysis has successfully utilized the dissolved gas analysis (DGA) technologies [2]. Conventional DGA methods, including key gas method (KGM) [3], international electrotechnical commission (IEC) method [4], and improved three ratio method [5], are intensively used worldwide. However, due to flaws including experience and previous knowledge gaps, a lack of coding, and strict borderline testing, the accuracy of the aforementioned approaches is mediocre [6].

With the rise of intelligent algorithms and machine learning theory, various artificial intelligence approaches, including back propagation neural network (BPNN) [7], support vector machine (SVM) [8], extreme learning machine (ELM) [9], adaptive neuro fuzzy inference system (ANFIS) [10], etc., have been well employed in the field of transformer fault diagnosis and made impressive progress. However, the AI methods aforementioned have pros and cons, leaving significant space for further improving the accuracy of fault diagnosis.

Previous research on transformer fault diagnosis has found that the input features are typically composed of specific gas concentration values, gas ratio, or relative percentages recommended by IEC or IEEE standards. In order to explore the impact of inputs and identify the ideal feature subset, numerous researches have been carried out since various input qualities lead to varying fault detection accuracies. To find feature subsets in the feature space made up of 46 DGA feature parameters, Yu et al. [11] presented a hybrid feature selection approach based on fuzzy information entropy. Mo et al. [12] used support vector machine and particle swarm optimization to diagnose problems while using two feature selection approaches to rank features and choose partial features based on their order. To diagnose defects in power transformers, a method based on optimizing the kernel parameters and weight parameters of a kernel extreme learning machine was proposed by Li [13].

All fault diagnosis approaches mentioned above in literatures have yielded impressive results. However, there are two significant issues with existing DGA-based fault diagnosis models that must be addressed. The absence of a widely acknowledged feature set for feature selection is the first issue. An oil-immersed transformer fault results in the oil decomposing, which releases fumes. The energy content of the fault controls how concentrated the different gases that are produced are. But the fault energy criteria for fault identification are not considered in most fault diagnosis studies. Therefore, it is expected in this paper to establish a feature set containing energy weighted dissolved gas analysis (EWDGA), which encompasses the amount of energy expended in the faulting process because different gases require different amounts of energy to form faults [14]. The prior best feature subset may be biased, which is the second issue that needs to be resolved. Using feature selection approaches, the distinctness of each feature is evaluated. As a result, a single feature selection approach might not be able to gather and use all the data required to accurately and thoroughly assess the discriminative power of features. Even a mediocre feature subset could be the outcome. As a result, combining multiple feature selection techniques has produced a more solid and trustworthy outcome [15]. Therefore, in order to remove bias from feature selection and increase diagnosis effectiveness, we present a hybrid feature selection method with 2 parts. Relief F, Fisher Score, and Laplacian are employed in the first part to rank features and combine rank orders into a single rank. In the second part, the whale optimization-based extreme learning machine (WOA-ELM) model is adopted to select an optimal feature subset and optimize the classifier's parameters simultaneously. The optimal feature subset is utilized to build a fault diagnosis model.

The rest of this article is organized as follows. Section 2 provides a brief overview of the methodology employed in this research. Section 3 briefly illustrates the fault diagnosis process. Section 4 implements numerical experiments to verify the validity of the developed fault diagnosis approach. The conclusion is drawn in Section 5.

2. METHODOLOGY

2.1. Feature Selection Techniques

In order to make it easier to interpret data, a procedure known as feature selection is applied to select subsets of features from the original, high-dimension set, which lowers features dimensions, reduces computational requirement, eliminates the effect of the curse of dimensionality, and improves the classification and prediction performance [16]. In general, the filter methods have the advantages of being fast, efficient, scalable, and suitable for ample data space. Additionally, a combination of multi-filter techniques can complement one another and offer better approximations of the ideal subset than a single assessment criterion [17]. Multi-criteria filter approaches are suggested and investigated in this study because they can also enhance performance and offer a more dependable and stable feature subset [18]. Before creating model for fault diagnostics, three popular filter techniques, including Relief F [19], Fisher Score [20], and Laplacian [21], are employed to evaluate features importance, rank related features, and supply lower-dimensional and more informative feature subsets.

2.2. Energy-Weighted Theory

In order to find fault features closely related to fault types, many scholars have studied EWDGA. EWDGA is a mechanism for balancing the relative energy needed for each gas's creation with the concentration values of each individual gas [22]. The ensuing Equation (1) can be used to represent the energy-weighted concentration values of fault gas:

$$\textit{energy weighted gas concentration} = \textit{Gas concentration} \times \textit{Weighting factor} \tag{1}$$

To account for the energy content of fault gases, estimation of the relevant weighting factor from thermodynamic decomposition model of the fault gases is crucial. The relative enthalpies of the fault gas formation are applied to calculate the weighting factor. Varied fault gases occur with significantly different enthalpies [23]. The weighting factors, which show the severity of the fault, are generated by enthalpy normalization of the respective fault gas formation enthalpies. Table 1 displays the relative enthalpies and enthalpy of formation [24].

Table 1. Standard enthalpies of formation for gaseous molecules.

Gas type	Enthalpy of formation	Relative enthalpy (weighting factor)
H_2	97.9	2.76
CH_4	35.42	1
C_2H_6	57.8	1.63
C_2H_4	93.5	2.64
C_2H_2	267.9	7.56

Note: The relative values refer to the enthalpy of production of other fault gases, using the enthalpy of production of CH_4 as a reference.

2.3. Ranking Aggregation Theory

Different feature selection approaches may lead to various order lists for the identical feature set, so it is necessary to combine different orders rationally and effectively. Since ranking aggregation methods can merge the collection of rankings over a set of alternatives in a single order, this issue has attracted much attention in various fields. The minimum violations ranking (MVR) method, proposed by Yamamoto, is a well acknowledged and adopted ranking aggregation method [25]. This method computes aggregated order list by giving a degree of a coincidence so that the degree of coincidence between the input rankings is optimal. The brief steps of MVR are described as follows:

a. a constant C_{ij} is predefined as Equation (2)

$$C_{ij} = (Number\ of\ lists\ where\ i\ is\ ranked\ before\ j) - (Number\ of\ lists\ where\ i\ is\ ranked\ after\ j) \quad (2)$$

b. It aims to produce a ranking of the things that reduce this number using the constants matrix C. The decision variables X_{ij} that decide whether the thing should be ranked above item j are defined in order to achieve this purpose. Equation is utilized to calculate X_{ij} (3)

$$x_{ij} = \begin{cases} 1, & c_{ij} > 0 \\ 0, & c_{ij} \leq 0 \end{cases} \quad (3)$$

c. the column sums of X_{ij} are calculated and sorted in ascending order to obtain the final ranking. The larger the summation is, the more important the feature will be.

2.4. WOA-ELM Theory

ELM is a feed forward neural network with one-layer hidden layer that has linear activation function at the output. The connection weights between the input layer and hidden layer, as well as the threshold of the neurons in the hidden layer, are generated at random by the ELM algorithm [26]. Due to its ability to find the singular global optimal solution with a predetermined number of neurons in the hidden layer, ELM has the benefits of rapid learning speed and strong generalization performance [27]. The general optimization process of WOA is shown in Figure 1.

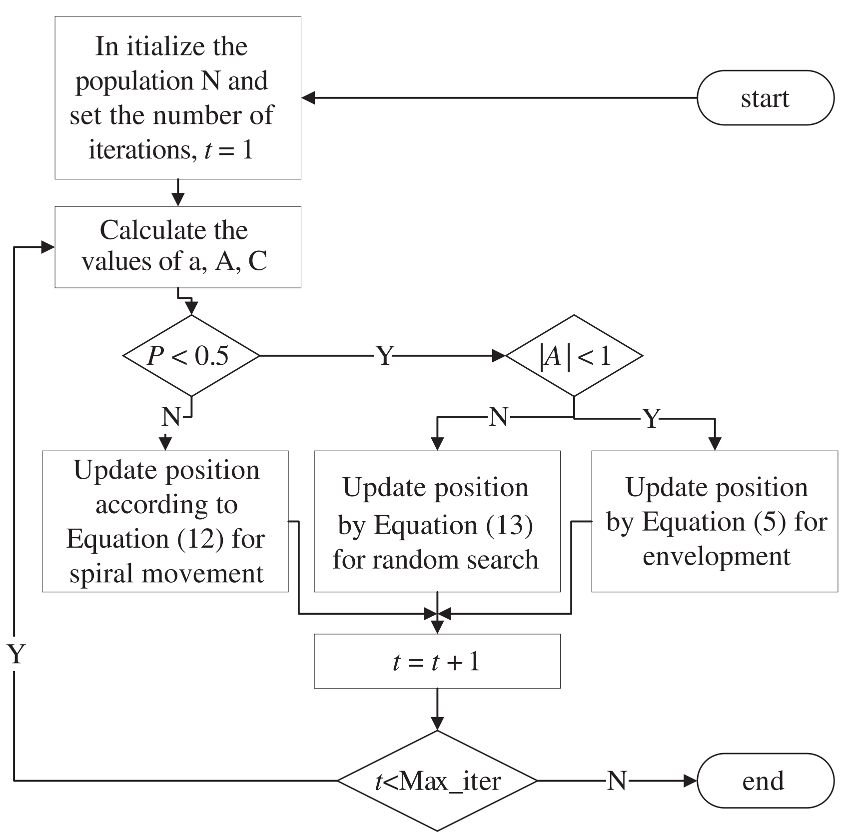

Figure 1. Flowchart for the WOA.

The WOA algorithm can be divided into 3 main phases: search and foraging, contraction and encirclement, spiral update of position. The search foraging phase can be represented by a mathematical model as

$$\mathbf{D} = |\mathbf{C}\mathbf{X}_{\mathrm{rand}}(t) - \mathbf{X}(t)| \tag{4}$$

$$\mathbf{X}(t+1) = \mathbf{X}_{\mathrm{rand}}\ (t) - \mathbf{A}\,|\mathbf{C} - \mathbf{X}_{\mathrm{rand}}\ (t) - \mathbf{X}(t)| \tag{5}$$

where X_{rand} is the individual whale selected from the contemporary whale population; $X(t)$ is the current location of the individual whale; and A and C are coefficient vectors defined as

$$A = 2ar_1 - a \tag{6}$$

$$C = 2r_2 \tag{7}$$

where a is the control parameter; r_1 and r_2 take values in the range $[0,1]$ and decrease linearly from 2 to 0 as the number of iterations t increases

$$a = 2 - \frac{2t}{\mathrm{Max\_iter}} \tag{8}$$

where *Max_iter* is the maximum value of the number of iterations set by the optimization. When $|A| \geq 1$, the whale will then perform a random search for food in the current state; otherwise, the whale will move towards the optimal position. The next process in the whale search is to carry out the foraging process, which is mainly divided into shrinking envelope and spiral update position [28]. The encircling process is represented in mathematical model as shown in Equations (9) and (10).

$$\mathbf{D} = |\mathbf{C}\mathbf{X}_{\mathrm{best}}(t) - \mathbf{X}(t)| \tag{9}$$

$$\mathbf{X}(t+1) = \mathbf{X}_{\mathrm{best}}\ (t) - \mathbf{A}\,|\mathbf{C} - \mathbf{X}_{\mathrm{best}}\ (t) - \mathbf{X}(t)| \tag{10}$$

where $X_{\mathrm{best}}(t)$ is the location of the best individual in the whale population; $A|C - X_{best}(t) - X(t)|$ is the set bracketing step.

The spiral update position process is represented as

$$\mathbf{D} = |\mathbf{X}_{\mathrm{best}}(t) - \mathbf{X}(t)| \tag{11}$$

$$\mathbf{X}(t+1) = D'e^{b1}\cos(2\pi l) + \mathbf{X}_{\mathrm{best}} \tag{12}$$

where $D\prime$ is the distance between the starting and final positions of the whale movement; b is a constant; and l is a random number between $[-1,1]$. The stage in which the whale algorithm works is simultaneously influenced by the probability factor. When $p \geq 0.5$, the WOA enters the spiral update position stage; when $p < 0.5$, the other 2 stages of the WOA are judged using $|A|$, and the mathematical model is as follows

$$\mathbf{X}(t+1) = \begin{cases} \mathbf{X}_{\mathrm{best}}(t) - \mathbf{AD}, & |\mathbf{A}| \ldots 1 \\ D'e^{bl}\cos(2\pi l) + \mathbf{X}_{\mathrm{best}}, & |\mathbf{A}| < 1 \end{cases} \tag{13}$$

Finally, WOA is applied to optimize network input weights and implied layer thresholds in ELM to improve the accuracy of the model [29].

3. POWER TRANSFORMER FAULT DIAGNOSIS BASED ON HYBRID FEATURE SELECTION

To improve the accuracy of the power transformer fault diagnosis, a new fault diagnosis model is developed utilizing hybrid feature selection and WOA-ELM. The diagnosis process of this method is described in Figure 2. The details of the proposed method are described as below:

a. Collect and sort out power transformer fault samples from the published literatures and power transformer factories. Each DGA sample consists of 5 kinds of characteristic gases, such as hydrogen (H_2), methane (CH_4), acetylene (C_2H_6), ethylene (C_2H_4), and ethane (C_2H_2).

b. 25 fault features based on gas concentration values, gas ratios, and EWDGA derived from conventional methods are utilized to build a feature set which is listed in Table 2. Additionally, it is notable that new created features including t6 to t17 are extracted from the EWDGA.

c. Filter feature selection approaches, including Relief F, Fisher Score, and Laplacian Score, are implemented to calculate and rank all features on the basis of importance. Afterwards, MVR is employed to integrate all rank lists and give a more informative and comprehensive rank list.

d. The obtained feature subsets are used as inputs for ELM, and WOA is adopted to optimize critical parameters, such as connection weights and thresholds.

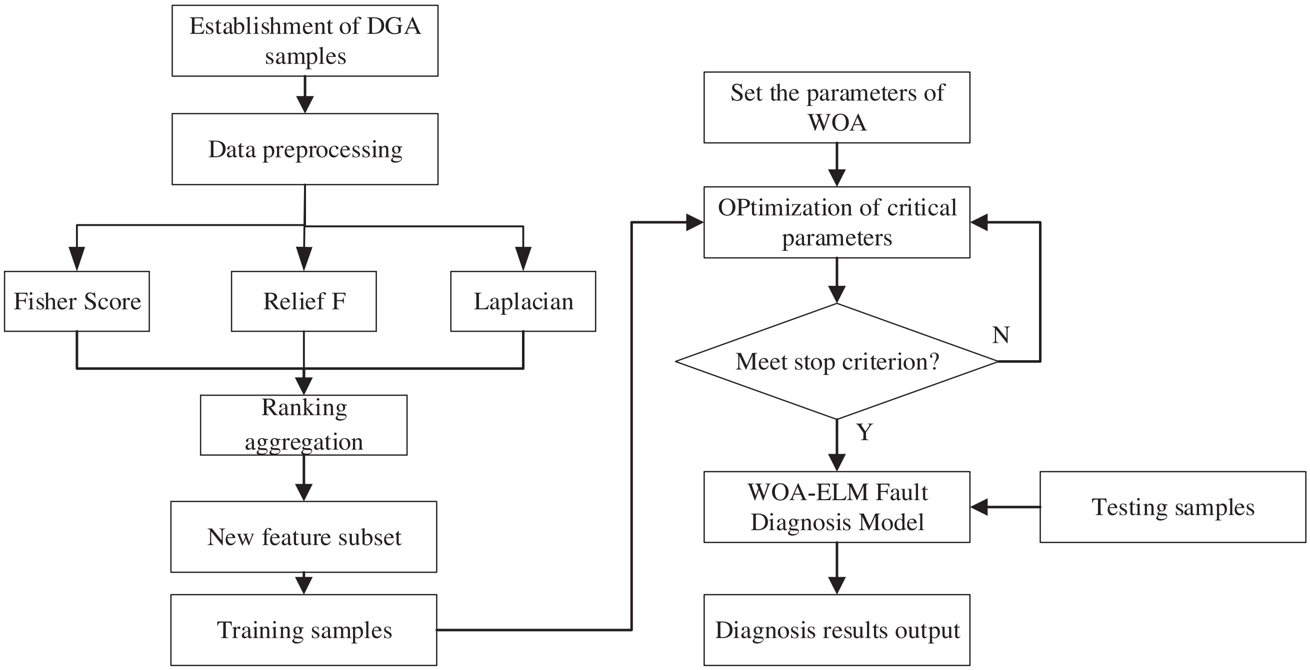

Figure 2. Flowchart for the proposed fault diagnosis method.

Table 2. Feature sets established for fault diagnosis.

Number	Feature	Number	Feature
t1	H_2	t14	H_2/EWTHH
t2	CH_4	t15	C_2H_4/EWTHH
t3	C_2H_6	t16	C_2H_2/EWTHH
t4	C_2H_4	t17	EWTCH
t5	C_2H_2	t18	CH_4/H_2
t6	H_2/EWTH	t19	C_2H_2/H_2
t7	CH_4/EWTH	t20	C_2H_4/H_2
t8	C_2H_6/EWTH	t21	C_2H_6/H_2
t9	C_2H_4/EWTH	t22	H_2/CH_4
t10	C_2H_2/EWTH	t23	C_2H_4/CH_4
t11	CH_4/EWTHD	t24	C_2H_6/CH_4
t12	C_2H_6/EWTHD	t25	C_2H_2/CH_4
t13	C_2H_2/EWTHD		

where, EWTHD=$CH_4+2.64C_2H_4+7.56C_2H_2$; EWTHH=$2.76H_2+2.64C_2H_4+7.56C_2H_2$;
EWTCH= $CH_4+1.63C_2H_6+2.64C_2H_4+7.56C_2H_2$;
EWTH=$2.76H_2+CH_4+1.63C_2H_6+2.64C_2H_4+7.56C_2H_2$

4. RESULTS AND DISCUSSION

4.1. Data Preprocessing

To create an accurate and reliable fault diagnosis model and assess its effectiveness, DGA samples from local power transformer plants and published literature are used. Dissolved gas features are classified into three types: gas concentration values, gas ratio, and energy weighted DGA. A thorough feature set is developed and displayed in Table 2 in accordance with accepted methods and published references.

Transformer faults are categorized into seven categories in this paper: partial discharge (PD), low-energy discharge (LED), high-energy discharge (HED), thermal fault of low temperature (LT), thermal fault of medium temperature (MT), thermal fault of high temperature (HT) and normal condition (NC). The distribution of all DGA samples used in this paper is shown in Table 3.

Table 3. Sample distributions for each fault type.

Fault type	NC	LT	MT	HT	PD	LED	HED	Total
Samples Form Literatures	21	-	-	20	15	14	18	88 [30]
	3	8	14	15	8	16	46	110 [31]
	-	4	4	4	4	4	-	20 [32]
	-	10	10	10	-	-	1	31 [33]
	-	4	4	4	4	4	-	20 [34]
	9	-	-	-	2	3	-	14 [35]
	6	4	5	2	-	1	2	20 [36]
	-	2	-	-	12	-	-	14 [37]
DGA case	61	68	63	45	55	58	33	383
Total	100	100	100	100	100	100	100	700

4.2. Hybrid Feature Selection Analysis

To obtain desirable fault diagnosis performance with an informative, distinguishable and lower dimensional feature subset, multi-criteria filter feature selection techniques are adopted in this paper. Features derived from dissolved gases and weighted energy DGA, which is displayed in Table 2, are firstly normalized in order to avoid data singularity and improve fault diagnosis performance with Equation (14)

$$x_{nik} = \frac{x_{ik} - x_{k\ \min}}{x_{ik\ \max} - x_{k\ \min}} \tag{14}$$

where x_{ik} and x_{nik} are the value of the i_{th} sample for the k_{th} feature before and after normalization; $x_{k\max}$ and $x_{k\min}$ are the maximum and minimum values of the k_{th} feature.

Relief F, Fisher Score, and Laplacian techniques are used to sort all features in ascending order according to corresponding importance. The obtained rank lists are presented in Figure 3 and Appendix Table A1. It can be seen that different feature selection approaches supply different rank orders. The discrepant results are led by different evaluation criteria for each feature selection algorithm. To be specific, the first five features and the last few features represent the gas concentration values and conventional gas ratios, which have relatively lower ranked order than other features. Features with numbers between 6 and 16, which stand for weighted energy DGA, have relatively better sorted orders than that of other features. The obtained results suggest that it is necessary and important to employ multi-criteria feature selection and need to take EWDGA into account.

Afterwards, a hybrid feature selection method (HFS), which aggregates all sorted order supplied by Relief F, Fisher Score, and Laplacian method with MVR approach, is proposed to give a more distinguishable and informative feature set. The process of MVR approach to integrate all rank lists is described as follows:

Firstly, the consistency matrix $C_{25\times25}$ and decision matrix $X_{25\times25}$ are obtained from Equations (2) and (3). Then, the column sums of X are calculated and sorted in ascending order to obtain the ranking, and the details of this process can be found in Appendix A. Next, the final aggregated results are presented in Figure 4. The larger the summation is, the more important the feature will be. On the contrary, the smaller the rank order is, the more significant the feature will be. So, it can

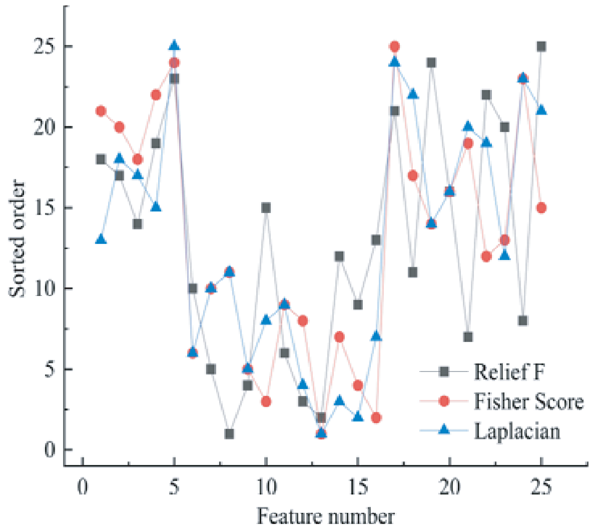

Figure 3. Sorted order of all features under different filter feature selection methods.

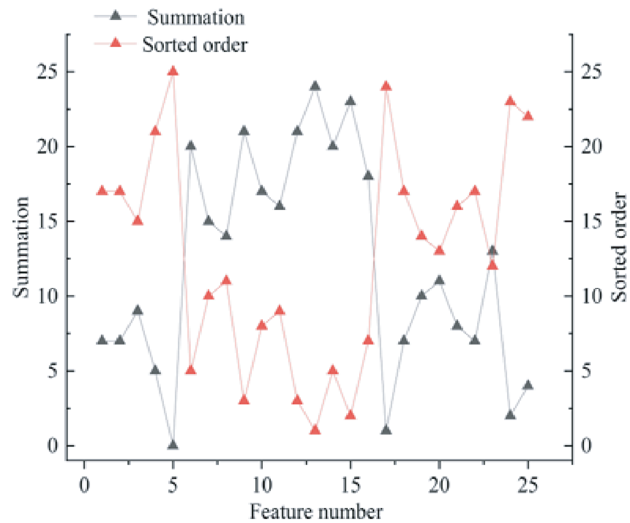

Figure 4. Summation result and sorted order by MVR method.

be inferred that the first four significant features obtained by MVR approach are C_2H_2/EWTHD, C_2H_4/EWTHH, C_2H_4/EWTH, and C_2H_6/EWTHD. In other words, these selected features are related to energy-weighted theory. Besides, feature numbers between t6 and t16, which represents the weighted energy DGA, have better sorted order than that of other features. In summary, the obtained results verify the validation and effectiveness of features based on energy weighted theory.

To demonstrate the superiority of MVR to aggregate feature lists, two popular feature aggregating approaches, majority voting strategy (MVS) [38] and Dowdall method (DM) [39], are employed to fuse and provide the sorted orders. When the MVS method is employed for combining rank orders, if the ranking of the feature is in the top 50%, it will be treated as 1 vote; otherwise, it is recorded as 0 votes. In this work, features with a voting frequency greater than one are retained and used to form the optimal subset; otherwise, they are discarded. When rank orders are combined using the DM method, each feature has a score in each ranking list equal to the reciprocal of the ranking. The scores of each feature in each list are added together to get a total score, which is used as the basis for ranking. The results obtained by the MVS and DM are shown in the Appendix Table A2.

It can be seen from Table 4 that the three aggregation approaches mentioned above supply different optimal feature subset and diagnosis accuracy. The diagnosis model based on MVS offers the lowest diagnosis accuracy with the largest feature size, while the employed MVR methods give the best diagnosis accuracy up to 98.57% with smallest feature size. However, the DM approach needs 10 features, which is inferior to models based on MVR. All in all, fault diagnosis performance based on MVR approach is superior to models based on other aggregation approaches, which can provide better fault diagnosis accuracy with smaller feature size.

Table 4. Comparison of diagnosis results under different ranking aggregation methods.

Aggregation method	Feature size	Accuracy
MVS	11	97.85%
DM	10	98.57%
Proposed	5	98.57%

4.3. Analysis of Fault Diagnosis Results Based on WOA-ELM

To confirm the accuracy and superiority of the proposed approach, fault diagnosis models based on different inputs are established in this paper. WOA algorithm is used to optimize crucial parameters of ELM. The feature selection and fault diagnosis are carried out in a MATLAB environment on a DELL server. Parameters of ELM and WOA for all experiments are set as follows: the neuro number of hidden layers is 40. The population size and iteration times are 20 and 200. The diagnosis performance under different feature selection techniques is shown as Figure 5.

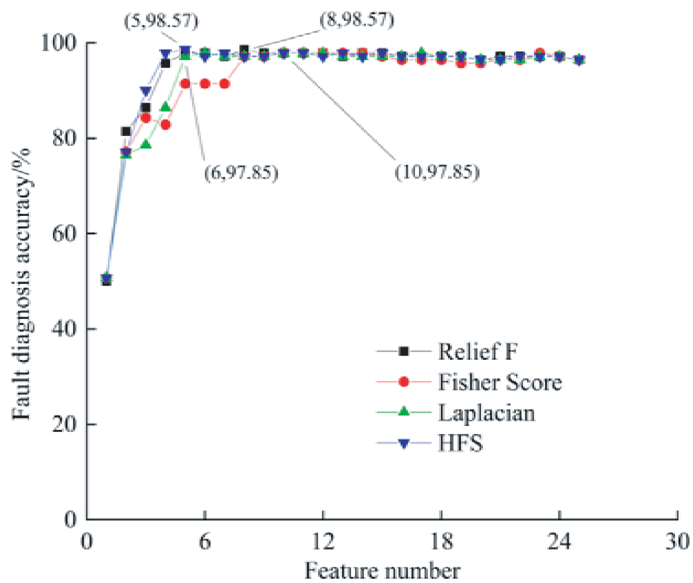

Figure 5. Fault diagnosis accuracy with different feature selection techniques.

According to the obtained results, with the increase of feature size, fault diagnosis accuracy gets improved. When only one feature is used to build a fault diagnosis model, a poor diagnosis accuracy around 50% is obtained. When the first five features are supplied by HFS approach, the WOA-ELM gets the best accuracy 98.57%. However, the feature size for Relief F, Fisher Score, and Laplacian to obtain the best diagnosis accuracy is 8, 10, and 6, respectively (except that the best accuracy of Relief F is 98.57%, and others are 97.85%). The results imply that fault diagnosis model based on HFS has simpler and better diagnosis performance than that of other approaches, which verify the effectiveness and validity of the proposed methods. In addition, the optimal feature subset for the best diagnosis accuracy is listed in Table 5 (It is worth noting that although both t6 and t14 are ranked 5th, the accuracy obtained after experimental validation is higher when t6 is used.). It is shown that the optimal feature subset is based on energy weighted theory, which suggests again that it is necessary and reasonable to take those features into account. Fitness development curve of WOA typical curves

Table 5. The obtained optimal feature subset.

Number	Feature	Number	Feature
1	C_2H_2/EWTHD	4	C_2H_6/EWTHD
2	C_2H_4/EWTHH	5	H_2/EWTH
3	C_2H_4/EWTH		

where, EWTHD=CH_4+2.64C_2H_4+7.56C_2H_2; EWTHH=2.76H_2+2.64C_2H_4+7.56C_2H_2;
EWTH=2.76H_2+CH_4+1.63C_2H_6+2.64C_2H_4+7.56C_2H_2

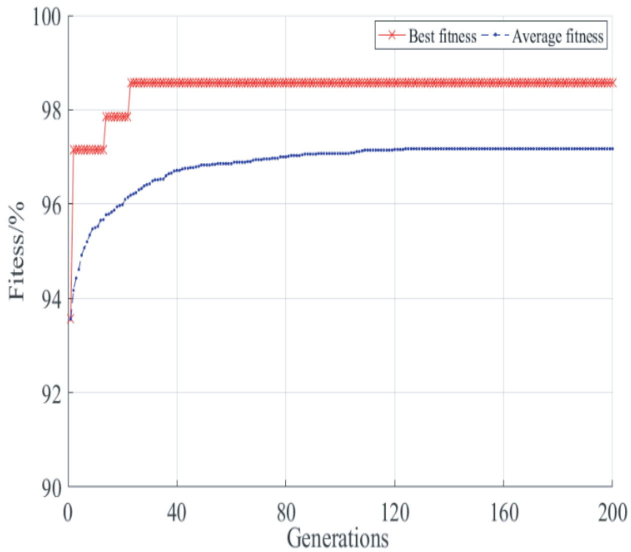

Figure 6. Fitness development curve of WOA (accuracy = 98.57%).

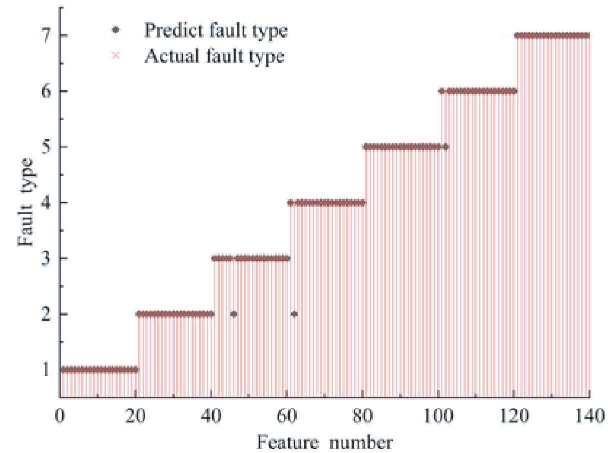

Figure 7. Fault diagnosis results for testing samples based on OFS-WOA-ELM.

of average and best fitness of WOA is described in Figure 6.

In addition, when more features larger than the optimal size are used to established models, the fault diagnosis performance will remain unchanged or get worse. The main reason for decreased accuracy is that the extra features are highly irrelative or redundant to fault diagnosis. Moreover, the unfavorable effects of these extra features include high-dimension inputs, complicated model structure, and low computational efficiency. Therefore, it is necessary to select the optimal feature subset to diagnose the fault of power transformer.

4.4. Comparisons of Fault Diagnosis Results

To further study and confirm fault diagnosis performance of the proposed approach, four popular conventional methods, including IEC method, improved three ratio method, BP and ELM are used to diagnose faults. Moreover, a similar feature subset (SFS) not taking energy weighted theory into account, shown as Table 6, is proposed to demonstrate the superiority and necessity of introducing energy weighted theory.

Table 6. Similar optimal feature subset for fault diagnosis.

Number	Feature	Number	Feature
1	C_2H_2/THD	4	C_2H_6/THD
2	C_2H_4/THH	5	H_2/TH
3	C_2H_4/TH		

where, THD=CH_4+C_2H_4+C_2H_2; THH=H_2+C_2H_4+C_2H_2; TH=H_2+CH_4+C_2H_6+C_2H_4+C_2H_2

The optimal feature subset and similar feature subset are used as input to establish the fault diagnosis model. For all models, 20% of collected samples are randomly selected and used to test all models' fault diagnosis performance. In this comparison, default parameters are applied in BP, while WOA is applied to tune the critical parameters of ELM. It is worth noting that all experiments are implemented 50 times, and average accuracy is adopted to evaluate diagnosis performance. A specific testing result is shown in Figure 7. The obtained results are described in Table 7.

Table 7. Comparison of diagnosis performance with various methods.

Method	Accuracy
IEC	$51.43\% \pm 0\%$
Improved Three Ratio	$83.57\% \pm 0\%$
BPNN	$83.91\% \pm 4.78\%$
ELM	$92.34\% \pm 1.36\%$
SFS-WOA-ELM	$95.57\% \pm 0.43\%$
OFS-WOA-ELM	$97.27\% \pm 0.57\%$

It can be shown from Table 7 that the traditional methods, such as IEC, improved three ratio method, BPNN, and ELM, have mediocre accuracy which is lower than that of ELM-based models. Fault diagnosis models based on ELM have higher accuracy than all conventional approaches, while models established with a similar feature subset, the SFS-WOA-ELM model provides impressive diagnosis average accuracy up to $95.57 \pm 0.43\%$. However, fault diagnosis accuracy is still lower than that of models based on optimal feature subset. The proposed approach comes up with the best average accuracy up to $97.27 \pm 0.57\%$, which indicates that the optimal feature subset (OFS) based on EWDGA and MVR could supply more stable and accurate diagnostic performance. The experimental results indicate the validation and effectiveness of the proposed approach.

4.5. Validation of Fault Diagnosis Performance

To verify the generalization and robustness of the suggested fault detection approach based on selected optimized feature subset and WOA-ELM, a public dataset IEEE Dataport [40], containing 185 DGA samples, is employed. The obtained result is shown in Table 8, and the rest is in accordance with Section 4.4.

Table 8. Comparison of diagnosis performance with various methods.

Method	Accuracy
IEC	$51.1\% \pm 0\%$
Improved Three Ratio	$56.59\% \pm 0\%$
BP	$62.9\% \pm 4.5\%$
ELM	$67.9\% \pm 1.47\%$
SFS-WOA-ELM	$76.4\% \pm 1.45\%$
OFS-WOA-ELM	$77.2 \pm 1.52\%$

It can be seen from Table 8 that the models based on optimal feature subset fault diagnosis accuracy comes up with the best accuracy up to $77.2 \pm 1.52\%$, which is better than that of other five approaches, and the results confirm again the feasibility and generalization ability of the optimal feature subset based on EWDGA and MVR.

5. CONCLUSION

In order to choose the best input feature subset and improve the parameters of the fault diagnostic model, a new hybrid feature selection approach that combines the whale optimization algorithm and extreme learning machine is implemented in this research. The MVR technique creates a more thorough and instructive feature set by aggregating Relief F, Fisher Score, and Laplacian Score feature selection methods. The comparison findings show that feature subsets obtained by MVR can provide results that are more accurate than those generated by MVS or DM aggregating approaches. The fused feature selection method in this paper can take into account the rationality of each feature selection algorithm and the differences in importance of each fault feature, providing a more reasonable ranking for the selection subsystem and creating a basis for subsequent screening of the preferred fault features

The final selected optimal feature subsets C_2H_2/EWTHD, C_2H_4/EWTHH, C_2H_4/EWTH C_2H_6/EWTHD, and H_2/EWTH stand for weighted energy DGA. In light of the comparison of feature subsets and diagnosis outcomes, the optimum fault diagnosis performance, with an accuracy of up to $97.27 \pm 0.57\%$, can be provided by the chosen optimal feature subsets. Testing findings demonstrate the universality and robustness of the chosen optimal feature subsets, confirming the value and superiority of the optimal feature subset and the suggested methodology. The energy-weighted processing of the dissolved gas components in the oil effectively characterizes the information contained in the internal fault state of the transformer, significantly improving the sensitivity and accuracy of the diagnosis of transformer faults. Finally, through the validation of the IEEE Dataport database, it is demonstrated that the new fault characteristics in this paper can achieve higher transformer fault diagnosis accuracy than the traditional numerical signs and ratio signs, and have certain generalization applicability and extension adaptability.

In the future work, more features derived from dissolved gas analysis should be investigated. Besides, other new optimization algorithms, wrapper and embed feature selection techniques, and classifiers need to be applied to study the diagnosis performance from various aspects. In addition, most of the current work on transformer fault feature selection is focused on the data-driven perspective. Some attempts should be made to use the micro-molecular perspective to explain the superiority of the preferred fault features and thus gain a deeper understanding of the nature of transformer faults.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of national natural science foundation of China (52067021), natural science foundation of Xinjiang (2022D01C35), and the excellent youth scientific and technological talents plan of Xinjiang (2019Q012).

APPENDIX A.

Table A1 displays the supplement feature ranks that three filter algorithms produced. It is clear that feature ranks are influenced by feature selection algorithms and data processing techniques.

Table A1. Feature ranks obtained by different filter algorithms.

Feature	Obtained orders			Feature	Obtained orders		
	Relief F	Fisher Score	Laplacian		Relief F	Fisher Score	Laplacian
t1	18	21	13	t14	12	7	3
t2	17	20	18	t15	9	4	2
t3	14	18	17	t16	13	2	7
t4	19	22	15	t17	21	25	24
t5	23	24	25	t18	11	17	22
t6	10	6	6	t19	24	14	14
t7	5	10	10	t20	16	16	16
t8	1	11	11	t21	7	19	20
t9	4	5	5	t22	22	12	19
t10	15	3	8	t23	20	13	12
t11	6	9	9	t24	8	23	23
t12	3	8	4	t25	25	15	21
t13	2	1	1				

The results obtained by the MVS and DM are shown in Table A2.

Table A2. The aggregated results obtained by the MVS and DM.

Feature	**MVS**	**DM**	**Feature**	**MVS**	**DM**
t1	0	0.180	t14	3	0.560
t2	0	0.164	t15	3	0.861
t3	0	0.186	t16	2	0.720
t4	0	0.165	t17	0	0.129
t5	0	0.125	t18	1	0.195
t6	3	0.433	t19	0	0.185
t7	3	0.400	t20	0	0.188
t8	3	1.182	t21	1	0.245
t9	3	0.650	t22	1	0.181
t10	2	0.525	t23	1	0.210
t11	3	0.389	t24	1	0.212
t12	3	0.708	t25	0	0.154
t13	3	2.500			

The consistency matrix $C_{25\times25}$ and decision matrix $X_{25\times25}$ are obtained from Equations (2) and (3), and the results are shown at the end of Appendix A.

$$C = \left[\begin{array}{ccccccccccccccc}
0 & 1 & 1 & -3 & -3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
-1 & 0 & 3 & -1 & -3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
-1 & -3 & 0 & -1 & -3 & 3 & 3 & 3 & 3 & 1 & 3 & 3 & 3 & 3 & 3 \\
3 & 1 & 1 & 0 & -3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
3 & 3 & 3 & 3 & 0 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
-3 & -3 & -3 & -3 & -3 & 0 & -1 & -1 & 3 & -1 & -1 & 1 & 3 & -1 & 3 \\
-3 & -3 & -3 & -3 & -3 & 1 & 0 & -1 & 3 & 1 & 1 & 3 & 3 & 1 & 1 \\
-3 & -3 & -3 & -3 & -3 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-3 & -3 & -3 & -3 & -3 & -3 & -3 & -1 & 0 & -1 & -3 & 1 & 3 & -1 & 1 \\
-3 & -3 & -1 & -3 & -3 & 1 & -1 & -1 & 1 & 0 & -1 & 1 & 3 & 1 & 1 \\
-3 & -3 & -3 & -3 & -3 & 1 & -1 & -1 & 3 & 1 & 0 & 3 & 3 & 1 & 1 \\
-3 & -3 & -3 & -3 & -3 & -1 & -3 & -1 & -1 & -1 & -3 & 0 & 3 & 1 & 1 \\
-3 & -3 & -3 & -3 & -3 & -3 & -3 & -1 & -3 & -3 & -3 & -3 & 0 & -3 & -3 \\
-3 & -3 & -3 & -3 & -3 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & 3 & 0 & 3 \\
-3 & -3 & -3 & -3 & -3 & -3 & -1 & -1 & -1 & -1 & -1 & -1 & 3 & -3 & 0 \\
-3 & -3 & -3 & -3 & -3 & 1 & -1 & -1 & 1 & -3 & -1 & 1 & 3 & 1 & 1 \\
3 & 3 & 3 & 3 & -1 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
-1 & -1 & -1 & -1 & -3 & 3 & 3 & 3 & 3 & 1 & 3 & 3 & 3 & 1 & 3 \\
1 & -1 & -1 & -1 & -1 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
-1 & -3 & -1 & -1 & -3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
-1 & -1 & 1 & -1 & -3 & 1 & 3 & 3 & 3 & 1 & 3 & 3 & 3 & 1 & 1 \\
1 & 1 & 1 & 1 & -3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
-1 & -1 & -1 & -1 & -3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
1 & 1 & 1 & 1 & -3 & 1 & 3 & 3 & 3 & 1 & 3 & 3 & 3 & 1 & 1 \\
1 & 1 & 1 & 1 & -1 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3
\end{array}\right.$$

$$
\left.\begin{matrix}
3 & -3 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 \\
3 & -3 & 1 & 1 & 3 & 1 & -1 & 1 & -1 & -1 \\
3 & -3 & 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 \\
3 & -3 & 1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 \\
3 & 1 & 3 & 1 & 3 & 3 & 3 & 3 & 3 & 1 \\
-1 & -3 & -3 & -3 & -3 & -1 & -3 & -3 & -1 & -3 \\
1 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 \\
1 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 \\
-1 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 \\
3 & -3 & -1 & -3 & -3 & -1 & -3 & -3 & -1 & -3 \\
1 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 \\
-1 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 \\
-3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 & -3 \\
-1 & -3 & -1 & -3 & -3 & -1 & -3 & -3 & -1 & -3 \\
-1 & -3 & -3 & -3 & -3 & -1 & -3 & -3 & -1 & -3 \\
0 & -3 & -1 & -3 & -3 & -1 & -3 & -3 & -1 & -3 \\
3 & 0 & 3 & 1 & 3 & 3 & 1 & 3 & 3 & 1 \\
1 & -3 & 0 & 1 & 1 & 1 & 1 & 1 & -1 & 1 \\
3 & -1 & -1 & 0 & -1 & -1 & 1 & 3 & -1 & -3 \\
3 & -3 & -1 & 1 & 0 & -1 & -1 & 1 & -1 & -1 \\
1 & -3 & -1 & 1 & 1 & 0 & 1 & 1 & -3 & -1 \\
3 & -1 & -1 & -1 & 1 & -1 & 0 & 1 & -1 & -3 \\
3 & -3 & -1 & -3 & -1 & -1 & -1 & 0 & -1 & -3 \\
1 & -3 & 1 & 1 & 1 & 3 & 1 & 1 & 0 & 1 \\
3 & -1 & -1 & 3 & 1 & 1 & 3 & 3 & -1 & 0
\end{matrix}\right]
$$

$$
X = \begin{bmatrix}
0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0
\end{bmatrix}
$$

REFERENCES

1. Kong, R. and L. Du, "A series-series compensated contactless power transfer based on the rotary transformer for the drilling system," *Progress In Electromagnetics Research M*, Vol. 113, 129–138, 2022.
2. Zhang, C., Y. He, B. Du, L. Yuan, and B. Li, "Transformer fault diagnosis method using IoT based monitoring system and ensemble machine learning," *Future Generation Computer Systems*, Vol. 108, 533–545, 2020.
3. IEC 60597: Guide for the sampling of gases and of oil-filled electrical equipment and for the analysis of free and dissolved gas, 2005.
4. IEC 599: Interpretation of the analysis of gases in transformers and other oil-filled electrical equipment in service, 1991.
5. Li, Z., G. Pu, Y. Lu, H. Chen, and F. Wan, "Research on power transformer fault diagnosis using fuzzy decision making," *Electrical Technology*, Vol. 21, No. 10, 77–82, 2020.
6. Taha, B. M., 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, 111162–111170, 2021.
7. Hao, H., D.-X. Wang, and Z. Wang, "A permittivity measurement method based on back propagation neural network by microwave resonator," *Progress In Electromagnetics Research C*, Vol. 110, 27–38, 2021.
8. Zhang, Z. and H. Zhu, "Decoupling control on outer rotor coreless bearingless permanent magnet synchronous motor using LS-SVM generalized inverse," *Progress In Electromagnetics Research M*, Vol. 111, 65–76, 2022.
9. Wang, J., S. Lu, S. Wang, and Y. Zhang, "A review on extreme learning machine," *Multimedia Tools and Applications*, 1–50, 2020.
10. Kari, T., W. Gao, D. Zhao, W. X. Mo, Y. Wang, and L. Luan, "An integrated method of ANFIS and Dempster-Shafer theory for fault diagnosis of power transformer," *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 25, No. 1, 360–371, 2018.
11. Yu, S., W. Tan, C. Zhang, Y. Fang, and C. Tang, "Research on hybrid feature selection method of power transformer based on fuzzy information entropy," *Advanced Engineering Informatics*, Vol. 50, 101433, 2021.
12. Mo, W., T. Kari, H. Wang, L. Luan, and W. S. Gao, "Power transformer fault diagnosis using support vector machine and particle swarm optimization," *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, 511–515, 2017.
13. Li, J., C. Hai, Z. Feng, and G. Li, "A transformer fault diagnosis method based on parameters optimization of hybrid kernel extreme learning machine," *IEEE Access*, Vol. 9, 126891–126902, 2021.
14. Equbal, M. D., S. A. Khan, and T. Islam, "Transformer incipient fault diagnosis on the basis of energy-weighted DGA using an artificial neural network," *Turkish Journal of Electrical Engineering & Computer Sciences*, Vol. 26, No. 1, 77–88, 2018.
15. Dhal, P. and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," *Applied Intelligence*, 1–39, 2021.
16. Agrawal, P., H. F. Abutarboush, T. Ganesh, and A. W. Mohamed, "Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019)," *IEEE Access*, Vol. 9, 26766–26791, 2021.
17. Ghosh, M. and G. Sanyal, "Performance assessment of multiple classifiers based on ensemble feature selection scheme for sentiment analysis," *Applied Computational Intelligence and Soft Computing*, Vol. 2018, 8909357.1–8909357.12, 2018.
18. Kari, T., W. Gao, D. Zhao, K. Abiderexiti, W. X. Mo, Y. Wang, and L. Luan, "Hybrid feature selection approach for power transformer fault diagnosis based on support vector machine and genetic algorithm," *IET Generation Transmission & Distribution*, Vol. 12, No. 21, 5672–5680, 2018.

19. Cui, X., Y. Li, J. Fan, and T. Wang, "A novel filter feature selection algorithm based on relief," *Applied Intelligence*, 1–19, 2021.
20. Li, C. and J. Xu, "Feature selection with the Fisher score followed by the Maximal Clique Centrality algorithm can accurately identify the hub genes of hepatocellular carcinoma," *Scientific Reports*, Vol. 9, No. 1, 1–11, 2019.
21. Huang, R., W. Jiang, and G. Sun, "Manifold-based constraint Laplacian score for multi-label feature selection," *Pattern Recognition Letters*, Vol. 112, 346–352, 2018.
22. Jakob, F., P. Noble, and J. J. Dukarm, "A thermodynamic approach to evaluation of the severity of transformer faults," *IEEE Transactions on Power Delivery*, Vol. 27, No. 2, 554–559, 2011.
23. Guo, C., B. Wang, Z. Wu, M. Ren, Y. He, R. Albarracn, and M. Dong, "Transformer failure diagnosis using fuzzy association rule mining combined with case-based reasoning," *IET Generation, Transmission & Distribution*, Vol. 14, No. 11, 2202–2208, 2020.
24. Ghoneim, S. M., "Intelligent prediction of transformer faults and severities based on dissolved gas analysis integrated with thermodynamics theory," *IET Science, Measurement & Technology*, Vol. 12, No. 3, 388–394, 2018.
25. Yamamoto, Y., "A minimum violations ranking method," *Optimization and Engineering*, Vol. 13, No. 2, 349–370, 2012.
26. Li, D., S. Li, S. Zhang, J. Sun, L. Wang, and K. Wang, "Aging state prediction for super capacitors based on heuristic kalman filter optimization extreme learning machine," *Energy*, Vol. 250, 123773, 2022.
27. Zhang, H.-M., S. Zhou, C. Xu, and J. J. Zhang, "A robust approach for three-dimensional real-time target localization under ambiguous wall parameters," *Progress In Electromagnetics Research M*, Vol. 97, 145–156, 2020.
28. Yuan, P., C.-J. Guo, and Q. Zheng, "Synthesis of MIMO system with scattering using binary whale optimization algorithm with crossover operator," *Progress In Electromagnetics Research Letters*, Vol. 87, 21–28, 2019.
29. Qiu, Y., J. Zhou, M. Khandelwal, H. Yang, P. Yang, and C. Li, "Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration," *Engineering with Computers*, 1–18, 2021.
30. Kirkbas, A., A. Demircali, S. Koroglu, and A. Kizilkaya, "Fault diagnosis of oil-immersed power transformers using common vector approach," *Electric Power Systems Research*, Vol. 184, 106346, 2020.
31. Gouda, O. E., E. H. Salah, and H. E. L. Hassan, "Proposed three ratios technique for the interpretation of mineral oil transformers based dissolved gas analysis," *IET Generation, Transmission & Distribution*, Vol. 12, No. 11, 2650–2661, 2018.
32. Taha, I. B. M., A. Hoballah, and S. S. Ghoneim, "Optimal ratio limits of rogers' four-ratios and IEC 60599 code methods using particle swarm optimization fuzzy-logic approach," *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 27, No. 1, 222–230, 2020.
33. Yin, J., "Research on fault diagnosis method of oil-immersed power transformer based on correlation vector machine," North China Electric Power University, 2013.
34. Taha, I. B. M., S. S. Ghoneim, and H. G. Zaini, "Improvement of Rogers four ratios and IEC Code methods for transformer fault diagnosis based on Dissolved Gas Analysis," *2015 North American Power Symposium (NAPS)*, IEEE, 2015.
35. Taha, I. B. M., H. G. Zaini, and S. S. Ghoneim, "Comparative study between dorneneburg and roger's methods for transformer fault diagnosis based on dissolved gas analysis using Matlab Simulink Tools," *2015 IEEE Conference on Energy Conversion (CENCON)*, IEEE, 2015.
36. Seifeddine, S., B. Khmais, and C. Abdelkader, "Power transformer fault diagnosis based on dissolved gas analysis by artificial neural network," *2012 First International Conference on Renewable Energies and Vehicular Technology*, IEEE, 2012.

37. Liu, Z.-X., B. Song, E. W. Li, Y. Mao, and G. L. Wang, "Study of 'code absence' in the IEC three-ratio method of dissolved gas analysis," *IEEE Electrical Insulation Magazine*, Vol. 31, No. 6, 6–12, 2015.
38. Wahid, M. F., R. Tafreshi, and R. Langari, "A multi-window majority voting strategy to improve hand gesture recognition accuracies using electromyography signal," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 28, No. 2, 427–436, 2019.
39. Obata, T. and H. Ishii, "A method for discriminating efficient candidates with ranked voting data," *European Journal of Operational Research*, Vol. 151, No. 1, 233–237, 2003.
40. Li, E., L. Wang, and B. Song, "Fault diagnosis of power transformers with membership degree," *IEEE Access*, Vol. 7, 28791–28798, 2019.