IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Adaptive Dynamic Meta-heuristics for Feature Selection and Classification in Diagnostic Accuracy of Transformer Faults

SHERIF. S. M. GHONEIM[1] (Senior Member, IEEE), TAMER AHMED FARRAG[2], A. ALI RASHED[1] EL-SAYED M. EL-KENAWY[3] (Member, IEEE), ABDELHAMEED IBRAHIM[4] (Member, IEEE)

[1]Electrical Engineering Department, College of Engineering, Taif University, Taif 21944, Saudi Arabia
[2]Department of Computer Engineering, MISR Higher Institute for Engineering and Technology, Mansoura, 35511, Egypt
[3]Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology (DHIET), Mansoura, 35111, Egypt
[4]Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura, 35516, Egypt

Corresponding authors: Abdelhameed Ibrahim (e-mail: afai79@mans.edu.eg) and El-Sayed M. El-kenawy (e-mail: skenawy@ieee.org)

ABSTRACT Detection of transformer faults avoids the transformer's undesirable loss from service and ensures utility service continuity. Diagnosis of transformer faults is determined using dissolved gas analysis (DGA). Several traditional DGA techniques, such as IEC code 60599, Rogers' ratio method, Dornenburg method, Key gas method, and Duval triangle method, but these DGA techniques suffer from poor diagnosis transformer faults. Therefore, more research was used to diagnose transformer fault and diagnostic accuracy by combining traditional DGA techniques with artificial intelligence and optimization techniques. In this paper, a proposed Adaptive Dynamic Polar Rose Guided Whale Optimization algorithm (AD-PRS-Guided WOA) improves the classification techniques' parameters that were used to enhance the transformer diagnostic accuracy. The results showed that the proposed AD-PRS-Guided WOA provides high diagnostic accuracy of transformer faults as 97.1%, which is higher than other DGA techniques in the literature. The statistical analysis based on different tests, including ANOVA and Wilcoxon's rank-sum, confirms the algorithm's accuracy.

INDEX TERMS Diagnostic Accuracy, Dissolved Gas Analysis, Polar Rose, Artificial Intelligence, Data Classification

I. INTRODUCTION

THE power transformers are very crucial in the electrical power system, and the electricity utilities are keen to carry out inspections to monitor their status regularly. The malfunction in their operation will lead to disconnection of the system and consequently to revenue loss [1], [2]. The detection of the transformer faults is essential to avoid an unexpected and undesired outage from the system [3]. Dissolved gas analysis (DGA) is a method that interprets the cause of transformer faults and identifies the fault types [4]. Several DGA methods have been developed to address the rules used to diagnose the transformer faults, such as the Dornenburg, Rogers' ratios, IEC code 60599, and Duval triangle methods, as the traditional DGA methods [5]–[9]. Some graphical representations are designed to identify the transformer faults such as Duval Triangle [7], [8], pentagon [10], [11], and heptagon [12]. The poor accuracy of the traditional DGA methods is observed, and decreasing the errors between the estimated and actual diagnostic faults requires other tools to solve this shortcoming; then, the intelligent techniques are utilized.

Several attempts were carried out to increase the traditional DGA method's diagnostic accuracy to provide a correct diagnosis of transformer fault type using artificial intelligence techniques. Several intelligent techniques were addressed combining with traditional DGA methods to improve the accuracy of diagnosing the transformer faults such as artificial neural networks (ANN) [13]–[15], Fuzzy logic system [FLS] [10], [16]–[18], Neuro-fuzzy [19], [20], support vector machine (SVM) [21], [22], Dempster–Shafer Theory [23],

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

and k-NN [24]. Despite the combination of artificial intelligence methods with traditional DGA methods, the accuracy of diagnosing transformer faults is still less than researchers' aspiration. Therefore many of them resorted to trying to improve the accuracy of diagnosis through optimization techniques.

Some of the researchers' contributions using the optimization techniques were reported to improve the artificial intelligence methods' parameters to enhance the accuracy of the diagnosis of transformer faults. In [25], the genetic algorithm (GA) is used to adapt the wavelet networks (GAWNs) parameters such as the nodes' weighting values and wavelet nodes translation and dilation. The GAWNs, ANN, and conventional DGA accuracy results were compared. The bootstrap is used with genetic programming (GP) to enhance the extraction of classification features of each transformer fault type [26]. The features supplied the ANN, SVM, and K-nearest neighbor (KNN) classifiers as the inputs to identify each fault classification. The results revealed that the suggested algorithm improved the accuracy of the transformer faults diagnosis. A Genetic Neural Computing (GNC) is used in [27] to interpret and diagnose the transformer faults based on DGA data. According to IEEE C57.104, the faults are categorized into four subsets using GA. The clustered data are used as inputs to ANN to predict the transformer fault types. The suggested algorithm developed the required decision rules to identify the correct diagnoses of the transformer faults.

In [28], the Particle swarm optimizer (PSO) optimized the Parzan windows (PW) parameters to improve its ability for transformer fault classification. An intuitive interpretation of transformer faults and correct decision-making are the main advantages of this algorithm. Furthermore, the diagnostic accuracy is enhanced according to the other DGA methods. The ANN and PSO are merged to increase the accuracy of the transformer faults diagnosis based on DGA datasets [29]. It revealed that the accuracy is improved using the suggested algorithm. A hybrid modified evolutionary particle swarm optimization-time varying acceleration coefficient (MEPSO-TVAC) with artificial neural network (ANN) was suggested as a combined optimizer to identify the transformer faults based on DGA records [30].

The transformer faults can be diagnosed using an optimized regression ANN (GRNN), Cuckoo search algorithm (CSA), and rough set theory (RS) in [31]. The RS reduced and simplified the high dimensioned data to develop the better features of GRNN input. The CSA with Levy flight helps the GRNN to get a good global convergence. Real fault cases are used to validate the proposed algorithm, and the results explained that the algorithm could provide good accuracy for diagnosing the fault types based on the DGA dataset. The IEC 60599 code and Rogers' ratio method were utilized to identify the transformer faults, but they have poor diagnostic accuracy; then, the PSO-FS optimizer is used to determine the optimal limit of the two methods ratios to enhance the diagnostic accuracy [1]. The Fuzzy system (FS) identifies the transformer faults by the modified ratio limits developed by PSO. The proposed algorithm enhanced the diagnostic accuracy compared with the other DGA methods in the literature.

A hybrid grey wolf optimizer (GWO) with differential evaluation (DE) is presented for enhancing the diagnostic accuracy of DGA methods by avoiding the local optima, improving the diversity of the population, then fitting the relation between the exploration and exploitation [32]. A fault diagnosis model of GWO optimized least square support vector machine (HGWO-LSSVM) is suggested and applied to transformer fault diagnosis with the optimal hybrid DGA feature set selected as the model's input. The kernel principal component analysis (KPCA) is utilized to extract the features decreasing the model training time. A fuzzy system produced the rules to limit the gas ratio for transformer fault types considering three memberships for three regions of each gas percentage range [33]. The membership limits can be optimized using GWO, which developed the diagnostic code matrix. The proposed algorithm enhanced the diagnostic accuracy of the transformer faults (95.45 %) for the random testing data.

The traditional DGA techniques in IEC standard 60599 [5] and IEEE Standards C 57.104 [6] fail to interpret the transformer faults in most cases, then the diagnostic accuracy of these methods is lacking. Moreover, combining the DGA methods and artificial intelligence methods to enhance the diagnostic accuracy of traditional DGA techniques requires much work, and diagnostic accuracy is still low. Therefore, the researchers attempted to use the optimization techniques to optimize the DGA method parameters or to optimize the classification parameters for the classification techniques [1], [21], [22], [28]–[33]. The optimization techniques maximize the agreement of predicting and the actual faults to develop the highest diagnostic accuracy of the transformer faults. Table 1 illustrates the recent research diagnosing the transformer fault based on several artificial intelligence and optimization methods. The accuracy differs from the DGA method to another based on the used technique and the number of data samples. It is seen that only one approach develops high diagnostic accuracy, as in [34]. Still, the number of data samples is deficient, which can not indicate the robustness of the BA-PNN based DGA technique. The other recent DGA techniques develop diagnostic accuracy of about 90%, which is higher in most cases than traditional DGA methods.

In this work, 475 dataset samples were collected from the central chemical laboratory of the Electricity Authority of Egypt and literature. The proposed Adaptive Dynamic Polar Rose Guided Whale Optimization algorithm (AD-PRS-Guided WOA) enhances several classification methods' parameters, such as k-NN, ensemble classifier, and voting classifier, to improve their performance in classifying purposes. The proposed classification method developed an excellent diagnostic accuracy of the transformer faults. First, a binary version of the proposed (AD-PRS-Guided WOA) algorithm is used for feature selection from the tested dataset.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

TABLE 1: Recent research for classification in diagnostic accuracy of transformer faults

Reference	Methods	# of samples	Accuracy
S. S. M. Ghoneim et al. (2016), [4]	Multiple parallel ANN	184	87.8%
S. S. M. Ghoneim et al. (2016), [35]	Mathematical model based on gases ratios	386	84.71%
S. Koroglu et al. (2016), [36]	Multi-layer SVM	100	84-92%
I. B. M. Taha et al. (2017), [13]	Conditional probability	403	83.13–86.85%
T. Kari et al. (2018), [37]	Adaptive neuro-fuzzy inference system (ANFIS) and Dempster-Shafer Theory	697	84.4±3.7 %
Y. Kim et al. (2018), [38]	SVM and KNN	189	88%
X. Zhang et al. (2019), [34]	BA-PNN-based methods	139	98.46%
I. B. M. Taha et al. (2020), [1]	Particle Swarm Optimization Fuzzy-Logic Approach	481	88.65%
X. Wu et al. (2020), [39]	Deep learning frameworks based on CNN and LSTM	528	96.9%
S. S. M. Ghoneim et al. (2021), [40]	Teaching-learning based optimization (TLBO)	386	83.15%
I. B. M. Taha et al. (2021), [41]	Neural pattern-recognition techniques	335	92.8%

The binary AD-PRS-Guided WOA algorithm is evaluated in compared with the Grey Wolf Optimizer (GWO) [42], PSO [43], [44], Bat Algorithm (BA) [45], [46], WOA [47], [48], Bowerbird Optimizer (SBO) [49], Multiverse Optimization (MVO) [50], Biogeography-Based Optimizer (BBO) [51], Firefly Algorithm (FA) [52], and Genetic Algorithm (GA) [53]. Second, a voting classifier based on the proposed algorithm (voting AD-PRS-Guided WOA) is applied to the experiments' dataset. The output results are compared with voting WOA, voting GWO, voting GA, and Voting PSO. Finally, the diagnostic accuracy of the proposed classification algorithm of the randomly selected samples extracted by the optimizer from the total of 475 samples is investigated. The comparison of the proposed classification algorithm and the other DGA techniques in literature, such as Conditional probability [13] and NPR [41] and other DGA techniques, is illustrated in the experiments.

The main contributions of this work can be formed as follows.

- A novel Adaptive Dynamic Polar Rose Guided Whale Optimization algorithm (AD-PRS-Guided WOA) is proposed.
- A binary version of the proposed algorithm (binary AD-PRS-Guided WOA) is used for feature selection from the tested dataset.
- To test the statistical difference of the proposed binary AD-PRS-Guided WOA, a one-way analysis of variance (ANOVA) and a one-sample t-test tests are applied in this experiment.
- A voting classifier based on the proposed algorithm (voting AD-PRS-Guided WOA) is developed to improve the tested dataset classification accuracy.
- The ANOVA and Wilcoxon's rank-sum tests are investigated to test the statistical difference of the proposed voting (AD-PRS-Guided WOA) algorithm.
- The importance of the current work is applying a new optimization Polar Rose Guided Whale Optimization algorithm (ADPRS-GuidedWOA) to enhance several classification methods' parameters, such as k-NN, ensemble classifier, and voting classifier.
- The AD-PRS-Guided WOA algorithm is used to improve the classification method performance in classifying purposes and apply it in a new application of high voltage engineering to diagnose the transformer faults achieving high diagnostic accuracy of the transformer faults.
- The proposed binary and voting algorithms can be generalized and applied to different types of datasets.

II. MATERIALS AND METHODS

This section will discuss the dataset samples tested in this work and introduce the essential machine learning algorithms, including mathematical discussion of the traditional classifiers and ensemble methods. The original whale optimization algorithm will also be introduced in this section.

A. DATASET

In this study, 475 dataset samples were collected from the central chemical laboratory, Egyptian electricity holding company, and the literature. The distribution of the 475 samples is based on the transformer fault types, and their references are illustrated in Table 2. Table 2 shows that 242 data samples for real cases from the central chemical laboratory and Egyptian electricity holding company are included [54].

TABLE 2: Distribution of the data samples according to transformer fault types and literature

Ref.	PD	D1	D2	T1	T2	T3	Total
[7]	3	0	0	3	1	0	7
[8]	13	27	52	1	0	18	111
[21]	0	2	1	1	3	1	8
[54]	27	42	57	70	18	28	242
[55]	1	0	5	2	0	1	9
[56]	3	0	4	4	3	5	19
[57]	1	1	2	1	0	1	6
[58]	1	4	2	0	2	8	17
[59]	0	2	2	0	0	0	4
[60]	1	0	4	4	0	2	11
[61]	0	0	0	0	1	2	3
[62]	0	0	0	0	0	2	2
[63]	1	0	0	4	2	14	21
[64]	0	1	4	2	1	0	8
[65]	0	3	1	2	0	1	7
Total	51	82	134	94	31	83	475

B. MACHINE LEARNING ALGORITHMS

In this section, traditional classifiers and ensemble methods are considered as they are employed in the experiments for the dataset illustrated in Table 2. The traditional classifiers are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

Multilayer Perceptron (MLP) and k-Nearest Neighbors (k-NN). Ensemble methods cover two main types of ensemble methods: bagging classifier and random forest as a type of averaging technique, and Adaboost and voting as a type of boosting technique.

1) Traditional classifiers

Artificial neural networks (ANN) are excellent classification algorithms due to their ability to learn a non-linear decision boundary, like MLP, containing two or more layers. Thus, it is very flexible in solving real-world tasks. It consists of many processing elements (PEs) that are called artificial neurons and connections. These PEs try to emulate our human nervous system's operation using special training algorithms (i.e., ADAM, SGD, and L-BFGS-B) [66]. The MLP neural network can include input and output layers and one layer between them named hidden layer. For the node output value calculations, the weighted sum is computed as

$$S_j = \sum_{i=1}^{n} w_{ij} I_i + \beta_j \tag{1}$$

where I_i indicates input variable i and w_{ij} is connection weight between I_i and neuron j in the hidden layer. β_j is a bias value. The node j output can be defined using the sigmoid activation function as

$$f_j(S_j) = \frac{1}{1 + \exp^{-S_j}} \tag{2}$$

The network output is then defined using the value of $f_j(S_j)$ for all hidden layer neurons as

$$y_k = \sum_{j=1}^{m} w_{jk} f_j(S_j) + \beta_k \tag{3}$$

where w_{jk} is the weights between neuron j in the hidden layer and output node k and β_k is the bias value for the output layer.

The k-NN algorithm classifies samples or cases based on their similarity measure after storing all variable samples. In this algorithm, data is used directly for classification, and the classification process is done based on the nearest points or samples. The value of the nearest neighbors (k) is an adjustable parameter that can change to make the model more (i.e., small values of k) or less flexible (i.e., large values of k). Besides, the value of the nearest neighbors is one by default [67]. k-NN is employed to predict the output variables based on classification approaches. The dataset is divided into training and testing data in this approach. A similarity measure is used by the k-NN model to compare the given testing to training data. The Euclidean distance is commonly used between training data (x_{train}) and data testing (x_{test}), and it can be as follows.

$$D\left(x_{train,i}, x_{test,i}\right) = \sqrt{\sum_{i=1}^{k} (x_{train,i} - x_{test,i})^2} \tag{4}$$

In predicting output variables, the k-NN model chooses k training data that can be close to the testing data. To predict the output value of the unknown testing data, the output value of k training data is selected to be the nearest neighbors. To predict a value, the following formula is used by k-NN regression.

$$\hat{y} = \frac{1}{k} \sum_{j=1}^{k} y_j \tag{5}$$

where k represents number of nearest neighbors of y_j. In case of time series data, Eq. 5 is less efficient because the correlation between observations (time) is not considered. To predict the data testing, the following general formulation is applied.

$$\hat{y} = \sum_{j=1}^{k} w_j y_j \tag{6}$$

where w_j is weighted for the jth neighbor. This weighting is adjusted based on the observed data, as $w_j = j/n$, with n represents number of training data. This model is considered as a time series model of k-NN.

2) Averaging Ensemble Classifiers

The ensemble methods seek to merge ML classifiers' predictions to improve performance over a single classifier because of the variance reduction. Bagging or averaging methods are based on building several classifiers independently and then averaging their predictions (e.g., bagging classifier, randomized trees (random forest), and voting methods (soft and hard). These algorithms are less affected by the overfitting problem. Random decision forests (RF) is one of the most popular and successful ensemble algorithms used for classification, regression. RF has gained massive interest due to its accuracy and immunity to noise than single classifiers did. This means that small changes in training data do not make any reasonable change in the tree. This is because of the hierarchical architecture of the tree classifiers. High variance is a drawback for this algorithm. Generally, RF performance is lower than gradient boosted trees (GBT) and higher than decision trees (DT). This technique is a very effective technique for high-dimensional classification tasks [68].

For the RF training algorithm, set $X = x_1, \ldots, x_n$ with responses set as $Y = y_1, \ldots, y_n$. With B times, the bagging selects a sample randomly with the training set replacement and fits trees to these samples. Let $b = 1, ..., B$, a sample, with replacement, for n training examples from X and Y; named X_b and Y_b. Then, train the classification/regression tree f_b on X_b and Y_b. After the training process, the predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x' as follows.

$$\hat{f} = \frac{1}{B} \sum_{b=1}^{B} f_b(x') \tag{7}$$

In the classification trees case, the majority vote is applied.

3) Boosting Ensemble Classifiers

Adaboost is a type of ensemble boosting method that can improve accuracy by combining multiple classifiers. It builds a high-performance classifier by merging a set weak classifier (low accurate). This algorithm's philosophy is to set the weights of classifiers and to train the data case in each iteration to ensure the exact predictions of unusual observations. Any ML technique can be used as a base estimator if it accepts weights on the training set. Bagging classifier is an acronym from Bootstrap aggregating. It gets its name because it merges two techniques (i.e., Bootstrapping and Aggregation) into one model. It is a type of ensemble averaging method with simple implementation and improves the performance of ML algorithms [69]. It works by learning different classifiers on randomly generated training sets to get a final prediction. All classifiers in the ensemble are used to classify the test sample by merging all models' predictions using uniform averaging or voting methods over class labels. This technique can be used as a variance reduction method because of the randomization into its structure procedure, and then we can create an ensemble out of it. The concept behind this method is to merge a set of ML estimators and use a majority vote (i.e., the output of each estimator) called hard voting. On the opposite side, soft voting or average predicted probabilities returns the class label as argmax of the sum of predicted probabilities. The expected class probabilities are collected for each classifier. It is then multiplied by a specific weight for each classifier and is then averaged. The final class label (i.e., the highest average probability) is then derived from the class label. This technique is suitable for equally well-performing ML classifiers to balance their weaknesses [68].

C. WHALE OPTIMIZATION ALGORITHM

WOA has shown its advantages in the optimization area, and it is considered one of the most effective algorithms in the literature. However, it suffers from a low exploration capability of the search space in some applications [68]. The inspiration in the WOA algorithm is from the behaviour of whales in finding food [47], [70]. There is an n-dimensional search space in which whales swim. n represents the number of variables. The global solution can be found if each solution's position in the space search is updated. The main mechanism of this algorithm uses this equation to update solution' positions.

$$X(t+1) = X^*(t) - A.D, D = |C.X^*(t) - X(t)| \tag{8}$$

where $X(t)$ indicates a solution at iteration t and $X^*(t)$ indicates the prey' position (best solution). The "." is pairwise multiplication and $X(t+1)$ indicates the updated solution's position [71], [72]. The A and C vectors are updated in each iteration by $A = 2a.r_1 - a$ and $C = 2.r_2$. a is changing from 2 to 0 linearly. r_1 and r_2 are random values in $[0, 1]$.

III. PROPOSED AD-PRS-GUIDED WOA ALGORITHM

The proposed Adaptive Dynamic Polar Rose Guided Whale Optimization Algorithm (AD-PRS-Guided WOA) has the following main parts that are different from the original WOA algorithm. These changes are used to explore the search space while being affected by the leader's position to enhance exploration performance. The AD-PRS-Guided WOA algorithm is shown in Algorithm (1).

- The algorithm follows three random solutions instead of one solution.
- It uses exponentially change instead of linearly one to change between exploration and exploitation processes smoothly.
- It calculates a list of generated walks in a diffusion process, according to the best solution, as a polar rose function.

Each of these changes, the Guided WOA Algorithm, Adaptive Dynamic Technique, and Polar Rose Function, then the proposed algorithm, will be explained in detail in the following subsections.

A. GUIDED WOA ALGORITHM

In the AD-PRS-Guided WOA algorithm, the updating positions mechanism of the WOA algorithm is modified to follow three random solutions instead of one solution. The three random solutions are named X_{o1}, X_{o2}, and X_{o3}. These solutions are updated each iteration to enhance the algorithm performance and reach the best solution at minimum time.

$$\begin{aligned} X(t+1) = {} & w_1 * X_{o1} \\ & + z * w_2 * (X_{o2} - X_{o3}) \\ & + (1-z) * w_3 * (Q - X(t)) \end{aligned} \tag{9}$$

where $X(t)$ represents the solution at iteration t and $X(t+1)$ represents the updated solution position. Q indicates the best solution (prey' position). w_1, w_2, and w_3 are random values in $[0, 0.5]$, $[0, 1]$, and $[0, 1]$, respectively. z is updated between exploitation and exploration smoothly and is calculated as follows

$$z = 1 - \left(\frac{t}{iters_{max}}\right)^2 \tag{10}$$

where $iters_{max}$ is maximum iterations and t is an iteration.

B. ADAPTIVE DYNAMIC TECHNIQUE

A fitness value is determined for each solution in the population after initialization. The algorithm then finds the best solution with the best fitness value. After that, the algorithm splits individuals from the population into two groups as shown in Fig. 1: the exploration and exploitation groups. Some individuals in the exploitation group are going toward the best "leaders" option, and other individuals search in the area around the leaders. Most individuals in any of the subgroups of the population dynamically change. To guarantee a balance between exploration and exploitation, the algorithm starts with a (50/50) population.

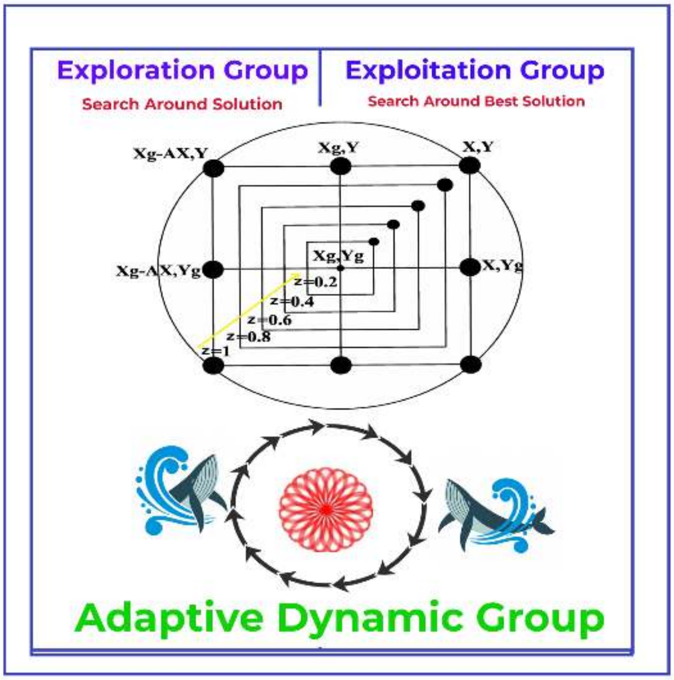

FIGURE 1: Groups balance in the proposed AD-PRS-Guided WOA algorithm.

C. POLAR ROSE FUNCTION

The polar rose function is employed in the proposed algorithm to search around the best solution to find another good solution. It can be applied as follows.

$$X(t+1) = k \sin\left(\frac{a}{b}\theta\right) \tag{11}$$

where a and b are within $[-10, 10]$ and $0 \leq \theta \leq 12\pi$. The k value decreases exponentially and is calculated as $2 - \frac{2\times t^2}{(iters_{max})^2}$ for iteration t and maximum iterations $iters_{max}$.

To show how this function helps search about the best solution and how it can cover an enormous range of areas around the selected solution, Fig. 2 shows the output of the polar rose function based on different values of a and b.

D. COMPUTATIONAL COMPLEXITY

The proposed AD-PRS-Guided WOA algorithm' computational complexity, as shown in Algorithm 1, can be expressed as follow for a number of population $n = n_1+n_2$ and number of iterations $iters_{max}$.

- Population settings: O (1).
- Parameters settings w_1, w_2, w_3: O (1).
- Collection of configuration parameters: O (1).
- Calculate objective function for n solutions: O (n).
- Finding best solution Q: O (n).
- Selecting three random solutions X_{o1}, X_{o2}, and X_{o3}: O ($iters_{max} \times n$).
- setting z value: O ($iters_{max} \times n$).
- Positions' updating for each solution: O ($iters_{max} \times n$).
- Positions' updating for each solution in exploration group: O ($iters_{max} \times n_1$).
- Positions' updating for each solution in exploitation group: O ($iters_{max} \times n_2$).
- Increasing solutions in exploration group: O ($iters_{max} \times n_1$).
- Decreasing solutions in exploitation group: O ($iters_{max} \times n_2$).
- Updating random solutions in exploration group X_{o1}, X_{o2}, X_{o3}, and Q: O ($iters_{max} \times n_1$).
- Mutate solutions in exploration group: O ($iters_{max} \times n_1$).
- Positions' updating for each solution: O ($iters_{max} \times n_1$).
- Updating random solutions in exploitation group X_{o1}, X_{o2}, X_{o3}, and Q: O ($iters_{max} \times n_2$).
- Move solutions in exploitation group: O ($iters_{max} \times n_2$).
- Searching around best solution in exploitation group: O ($iters_{max} \times n_2$).
- Amend solutions: O (n).
- Update fitness: O (n).
- Returning the best solution Q: O (1).

This analysis shows that the proposed AD-PRS-Guided WOA algorithm' complexity of computations is O ($iters_{max} \times n$), since $n \geq n_1$ and $n \geq n_2$. In case of a d dimension' problem, the algorithm' complexity will be O ($iters_{max} \times n \times d$).

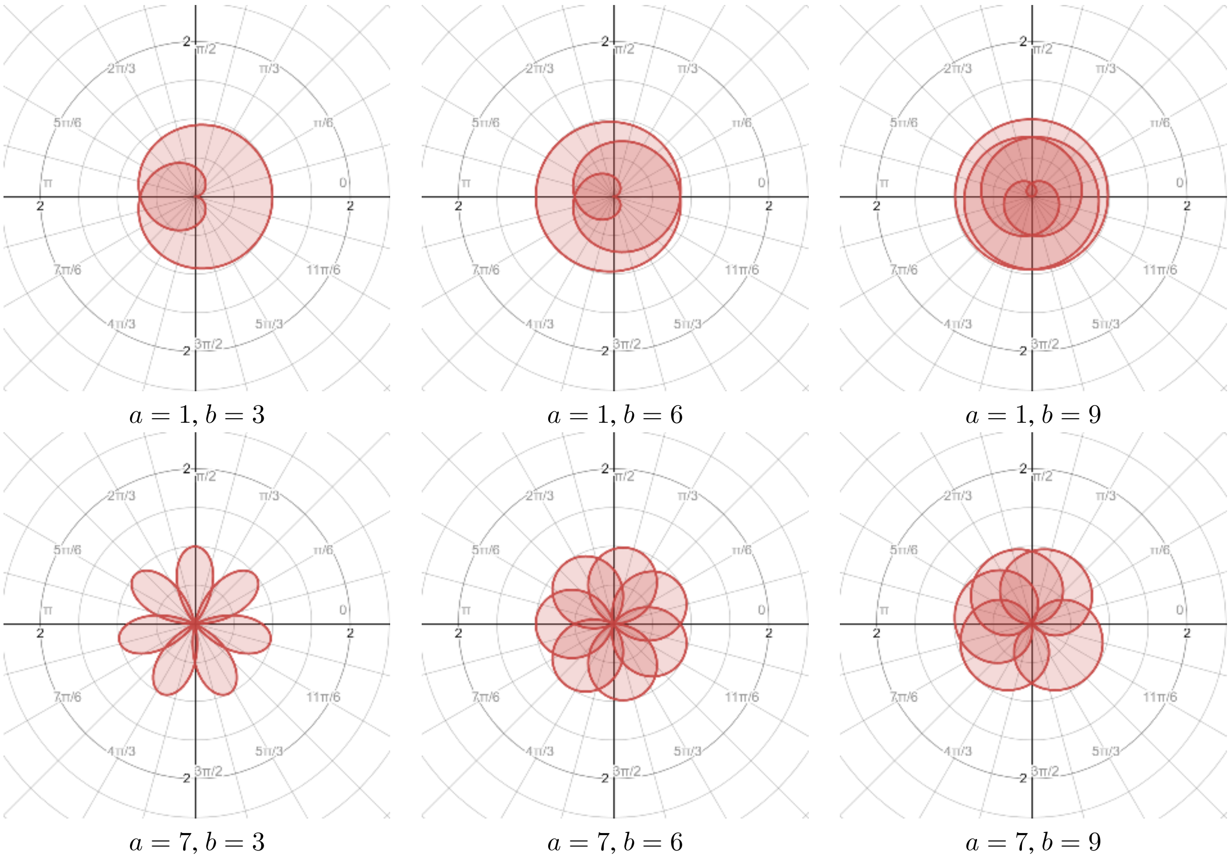

FIGURE 2: Polar rose function based on different values of a and b.

E. BINARY OPTIMIZER

For the problems of feature selection, the solutions are only binary with values of 0 or 1. Thus, the proposed AD-PRS-Guided WOA algorithm's continuous values can be converted into binary [0,1] to achieve the feature selection process. The following equation based on the $Sigmoid$ function is applied in this work.

$$\begin{aligned} X(t+1) &= \begin{cases} 1 & \text{if } Sigmoid(x) \geq 0.5 \\ 0 & otherwise \end{cases}, \\ Sigmoid(x) &= \frac{1}{1+\exp^{-10(x-0.5)}}, \end{aligned} \tag{12}$$

where $X(t+1)$ indicates the binary solution at iteration t. $Sigmoid$ can scale the output values to be binary. $Sigmoid(x) \geq 0.5$ converts the value to be 1. x represents the best solution of the proposed algorithm. The binary AD-PRS-Guided WOA Algorithm in explained step by step in Algorithm 2.

F. OBJECTIVE FUNCTION

The objective function is applied to get the optimizer solutions' quality. To evaluate the quality of a solution, the following equation is used.

$$F_n = \alpha ER(D) + \beta \frac{|s|}{|f|} \tag{13}$$

where $ER(D)$ gives classier' error rate, s indicates number of selected features, f represents the hole number of features. $\alpha \in [0,1], \beta = 1 - \alpha$ shows the number of the selected feature importance for population. The solution is a good solution if it can get a subset of features that can give low classification error rate with lower number of selected features.

IV. DIAGNOSTIC ACCURACY OF TRANSFORMER FAULTS FRAMEWORK

The proposed step-by-step framework for the diagnostic accuracy of transformer faults is shown in Fig. 3. The framework consists of three main steps: First step of data processing, Second step of training base model, and Third step of training voting ensemble model. In the first step of the framework, the processes of removing null values, feature scaling, and correlation analysis are applied to the input data samples as a data processing stage. The binary AD-PRS-Guided WOA algorithm, shown in Algorithm 2, is applied here for the feature selection from the processed data in the input dataset. The data is then divided randomly into 80% for training purposes, and the remaining samples

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

Algorithm 1 : Proposed AD-PRS-Guided WOA algorithm

Set population $X_i (i = 1, 2, ..., n)$, objective function F_n, size n, maximum iterations $iters_{max}$.
Set parameters w_1, w_2, w_3
Collection AD-PRS-Guided WOA configuration parameters
Calculate objective function F_n for all solutions X_i
Set Q = best agent position
while $t \leq iters_{max}$ **do**
  for $(i = 1 : i \leq n)$ **do**
    Select three random solutions X_{o1}, X_{o2}, and X_{o3}
    Set $z = 1 - \left(\frac{t}{iters_{max}}\right)^2$
    Update position of current search agent as
    $X(t + 1) = w_1 * X_{o1} + z * w_2 * (X_{o2} - X_{o3}) + (1 - z) * w_3 * (Q - X(t))$
  end for
  Update Solutions in exploration group (n_1) and exploitation group (n_2)
  if (Best F_n is same for three iterations) **then**
    Increase solutions of exploration group (n_1)
    Decrease solutions of exploitation group (n_2)
  end if
  for $(i = 1 : i \leq n_1)$ **do**
    (exploration group update)
    update three random solutions X_{o1}, X_{o2}, X_{o3}, and Q (The best solutions were elitism)
    if ($Q <$ Any of the best solutions) **then**
      Mutate the solution by
      $X(t + 1) = k + \left(\frac{\sum X_{o1} + X_{o2} + X_{o3}}{e^{zk}}\right)$, $k = 2 - \frac{2 \times t^2}{(iters_{max})^2}$
    else
      Update agent position by
      $X(t + 1) = w_1 * X_{o1} + z * w_2 * (X_{o2} - X_{o3}) + (1 - z) * w_3 * (Q - X(t))$
    end if
  end for
  for $(i = 1 : i \leq n_2)$ **do**
    (exploitation group update)
    update three random solutions X_{o1}, X_{o2}, X_{o3}, and Q (The best solutions were elitism)
    if ($Q <$ Any of the best solutions) **then**
      Move towards the best solution by
      $X(t + 1) = w_1 * X_{o1} + z * w_2 * (X_{o2} - X_{o3}) + (1 - z) * w_3 * (Q - X(t))$
    else
      Search around the best solution
      $X(t + 1) = k \sin\left(\frac{a}{b}\theta\right)$
    end if
  end for
  Amend solutions
  Update fitness
end while
Return best agent Q

are used for testing the models. The base models of NN, k-NN, and Random forest are trained on the second step of the framework using the processed data in the first step. The NN model is based on Eq. 1, Eq. 2, and Eq. 3. The k-NN model is based on Eq. 4, Eq. 5, and Eq. 6, while the Random Forest model uses Eq. 7.

The last step of the proposed framework as shown in Fig. 3 is the training voting ensemble model. In this step, a voting classifier is presented using the proposed AD-PRS-Guided WOA algorithm illustrated in Algorithm 1 based on the Guided WOA Algorithm, Adaptive Dynamic Technique, and Polar Rose Function. The voting algorithm improves the ensemble's accuracy by aggregating the NN, k-NN, and Random Forest classifiers. Voting is based on merging a set of ML algorithms and returns the class label as argmax of the sum of predicted probabilities. The predicted class probabilities are collected for each classifier. Then, it is multiplied by a specific weight for each classifier and is then averaged.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

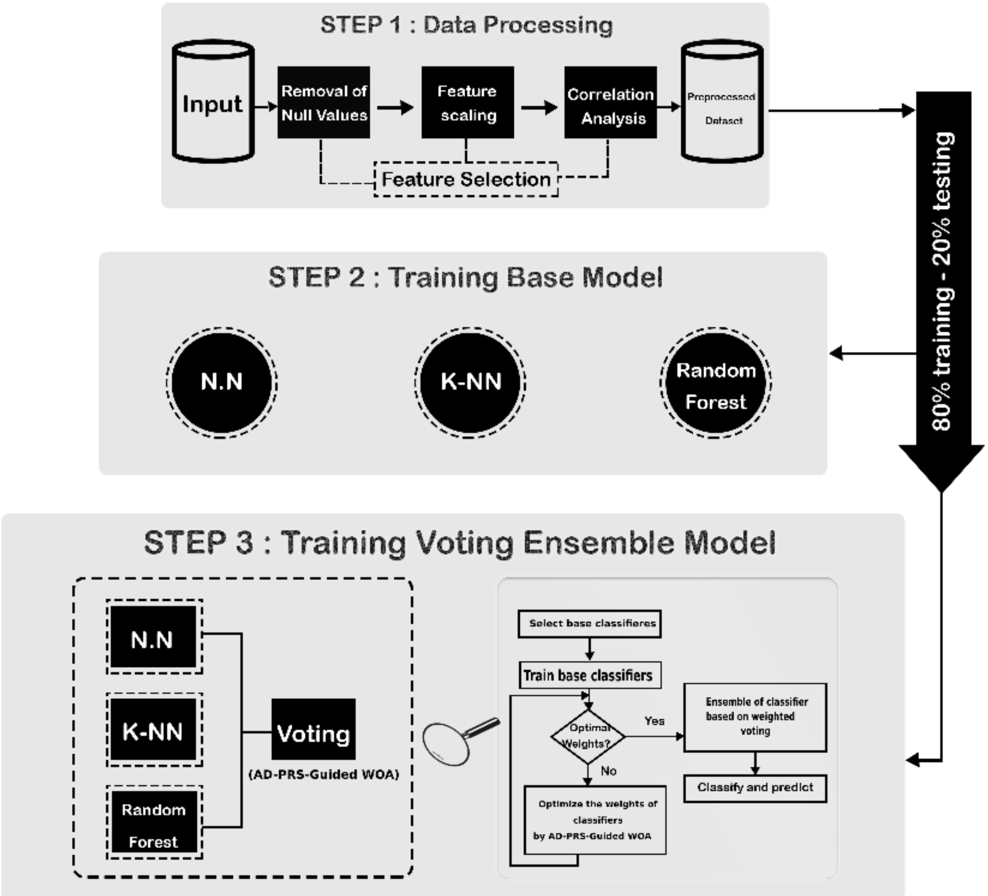

FIGURE 3: Proposed Framework based on AD-PRS-Guided WOA algorithm.

Algorithm 2 : Proposed binary AD-PRS-Guided WOA Algorithm

Set AD-PRS-Guided WOA population, parameters, configuration.
Convert solutions to binary [0,1]
Calculate objective function and select best solutions
Train k-NN and calculate error
while $t \leq iters_{max}$ **do**
  Apply AD-PRS-Guided WOA algorithm
  Convert updated solution to binary by Eq. 12
  Calculate fitness
  Update parameters
end while
Return best solution

The final class label (i.e., the highest average probability) is then derived from the class label. In addition, weights are optimized using the AD-PRS-GUIDE WOA algorithm. This process will guarantee the best model performance. The ensemble of the classifier is based on weighted voting. The data is then classified, and the output is predicted.

V. EXPERIMENTAL RESULTS

There are two main parts of the experiments. The first part considers the feature selection ability of the proposed binary AD-PRS-Guided WOA algorithm. The second part is designed for the classification problem of the tested dataset based on the algorithm. The 475 dataset samples are divided into 80% training and 20% testing. The training samples are used to train the proposed optimizer, and the testing samples are used for the evaluation.

A. FEATURE SELECTION SCENARIO

The proposed binary algorithm is used in this scenario for feature selection from the tested dataset. The binary AD-PRS-Guided WOA algorithm is evaluated in compared with the Grey Wolf Optimizer (GWO), PSO [43], Bat Algorithm (BA) [45], [46], WOA [47], Bowerbird Optimizer (SBO) [49], Multiverse Optimization (MVO) [50], Biogeography-Based Optimizer (BBO) [51], Firefly Algorithm (FA) [52], and Genetic Algorithm (GA) [53]. Table 3 shows the perfor-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

TABLE 3: Performance metrics for feature selection

Metric	Value
Average Error	$1-\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\sum_{i=1}^{N}Match(C_i,L_i)$
Average Select Size	$\frac{1}{M}\sum_{j=1}^{M}\frac{size(g_j^*)}{D}$
Average Fitness	$\frac{1}{M}\sum_{j=1}^{M}g_j^*$
Best Fitness	$Min_{j=1}^{M}g_j^*$
Worst Fitness	$Max_{j=1}^{M}g_j^*$
Standard Deviation	$\sqrt{\frac{1}{M-1}\sum(g_j^*-Mean)^2}$

TABLE 4: AD-PRS-Guided WOA algorithm configuration

Parameter	Value
Agents	10
Iterations	80
Repetitions	20
Dimension	Number of features
a	$[-10, 10]$
b	$[-10, 10]$
θ	$[0, 12\pi]$
α of F_n	0.99
β of F_n	0.01

TABLE 5: Compared algorithms configuration

Algorithm	Parameter (s)	Value (s)
GWO	a	2 to 0
PSO	Inertia W_{max}, W_{min}	[0.9,0.6]
	Acceleration constants C_1, C_2	[2,2]
BA	Pluse rate	0.5
	Loudness	0.5
	Frequency	[0,1]
WOA	a	2 to 0
	r	[0,1]
BBO	Immigration Probability	[0,1]
	Mutation Probability	0.05
	Habitat modification Probability	1.0
	Step size	1.0
	Migration rate	1.0
	Max immigration	1.0
MVO	Wormhole existence probability	[0.2,1]
SBO	Step size	0.94
	Probability of Mutation	0.05
	Upper and lower limit difference	0.02
FA	Fireflies	10
GA	Mutation ratio	0.1
	Crossover	0.9
	Selection mechanism	Roulette wheel

mance metrics for feature selection tested in this experiment. The configuration of the proposed algorithm is presented in Table 4. The parameters of the objective function α and β are set to 0.99 and 0.01. The configuration of the compared algorithms is shown in Table 5. The average error of (0.4515) achieved by the proposed binary AD-PRS-Guided WOA, shown in Table 6 is the minimum error among the compared algorithms. Other metrics, including the standard deviation of (0.0337), approve the algorithm's superiority in this kind of problem. Figure 4 shows the fast convergence of the proposed AD-PRS-Guided WOA algorithm to find the optimal solution compared to other techniques.

To test the statistical difference of the proposed (AD-PRS-Guided WOA), a one-way analysis of variance (ANOVA) test is applied in this experiment. A null hypothesis is set as (H_0: $\mu_{AD-PRS-Guided\,WOA} = \mu_{GWO} = \mu_{PSO} = \mu_{BA} = \mu_{WOA} = \mu_{BBO} = \mu_{MVO} = \mu_{SBO} = \mu_{FA} = \mu_{GA}$) and an alternate hypothesis is confirmed as (H_1: No equal means). Table 7 shows the ANOVA test results. Figure 5 presents the proposed and compared algorithms ANOVA test results considering the function F_n. It is noted that the alternate hypothesis H_1 can be accepted for this test.

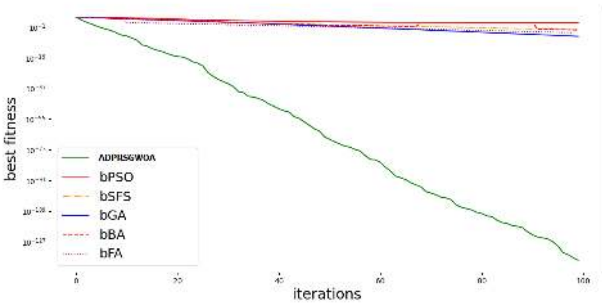

FIGURE 4: Convergence curves of the proposed and compared algorithms.

Another test, named one sample t-test, is conducted for the evaluation at a significance level of 0.05. In this test, a null hypothesis is set as (H_0: $\mu_A = \mu_{GWO}$, $\mu_A = \mu_{PSO}$, $\mu_A = \mu_{BA}$, $\mu_A = \mu_{WOA}$, $\mu_A = \mu_{BBO}$, $\mu_A = \mu_{MVO}$, $\mu_A = \mu_{SBO}$, $\mu_A = \mu_{FA}$, $\mu_A = \mu_{GA}$), for $A = AD-PRS-Guided\,WOA$, and an alternate hypothesis is formed as (H_1: No equal means). Table 8 show the test results of 20 runs (Repetitions) as indicated in Table 4. This confirm that the p-values are less than 0.05 which shows the statistical significant difference between groups. Thus, the hypothesis H_1 can be accepted accepted.

The residual values and plots can observe the possible problems better than the plot of the original dataset. Some datasets cannot be good candidates for the feature selection process. Figure 5 shows the residual, heteroscedasticity, quantile-quantile (QQ) plots, and the heatmap. These plots provide a visual view between the prediction errors and the predicted dependent variable scores. Any violation can be quickly determined to improve the accuracy of the research findings. Note that the QQ plot points' distributions are approximately near to the line. These plots indicate that the actual and the predicted residuals are linearly related, confirming the proposed algorithm performance.

B. CLASSIFICATION SCENARIO

The second part of the experiment is based on the proposed voting classifier (voting AD-PRS-Guided WOA) algorithm. The algorithm results are compared with voting WOA, voting GWO, voting GA, and Voting PSO against the Area Under The Curve (AUC) and the Mean Square Error (MSE). The AUC or balanced accuracy is calculated using an average of sensitivity and specificity in the following equation.

$$AUC = (Sensitivity + Specificity)/2 \tag{14}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

TABLE 6: Feature selection results of the proposed and compared algorithms

	AD-PRS-Guided WOA	bGWO	bPSO	bBA	bWOA	bBBO	bMVO	bSBO	bFA	bGA
Average error	0.4515	0.4687	0.5025	0.5121	0.5023	0.4707	0.4792	0.5108	0.5009	0.4823
Average Select size	0.4043	0.6043	0.6043	0.7437	0.7677	0.7681	0.7008	0.7746	0.6388	0.5467
Average Fitness	0.5147	0.5309	0.5293	0.5522	0.5371	0.535	0.559	0.569	0.5812	0.5423
Best Fitness	0.4165	0.4512	0.5096	0.4419	0.5012	0.5247	0.4842	0.5121	0.4999	0.4456
Worst Fitness	0.515	0.5181	0.5773	0.5435	0.5773	0.6112	0.6022	0.5918	0.5975	0.5607
Standard deviation Fitness	0.337	0.3417	0.3411	0.351	0.3433	0.386	0.3918	0.402	0.3779	0.3433

TABLE 7: ANOVA test results of the proposed algorithm

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	0.07437	9	0.008263	F (9, 190) = 77.51	P < 0.0001
Residual (within columns)	0.02026	190	0.0001066		
Total	0.09462	199			

TABLE 8: One sample t-test results of the proposed and compared algorithms

	AD-PRS-Guided WOA	GWO	PSO	BA	WAO	BBO	MVO	SBO	FA	GA
Theoretical mean	0	0	0	0	0	0	0	0	0	0
Actual mean	0.4515	0.4687	0.502	0.5111	0.5028	0.4712	0.4787	0.5103	0.5009	0.4823
Number of values	20	20	20	20	20	20	20	20	20	20
One sample t-test										
t, df			t=195.9 df=19	t=171.4 df=19	t=440.5 df=19	t=99.66 df=19	t=186.8 df=19	t=276.4 df=19	t=345.2 df=19	t=332.4 df=19
P value (two tailed)			0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
P value summary			****	****	****	****	****	****	****	****
Significant (alpha=0.05)?			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

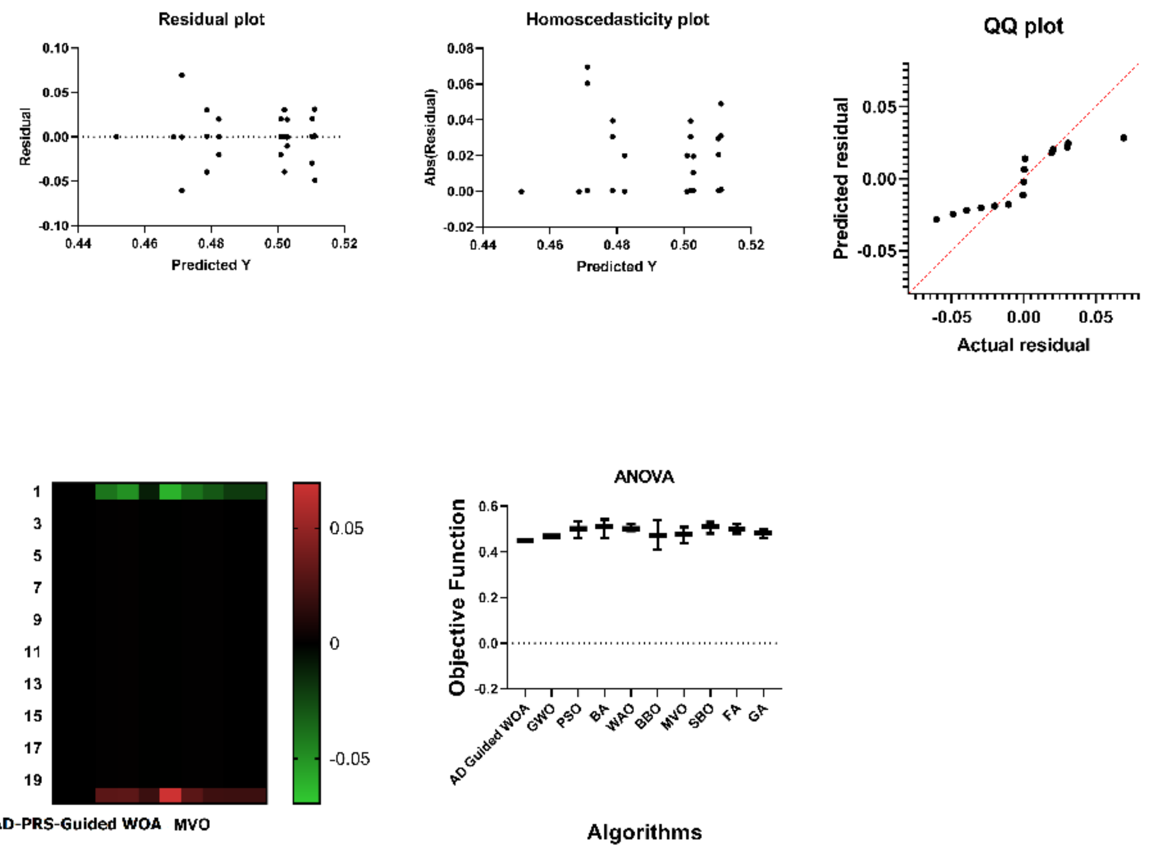

FIGURE 5: Different curves for the feature selection techniques.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

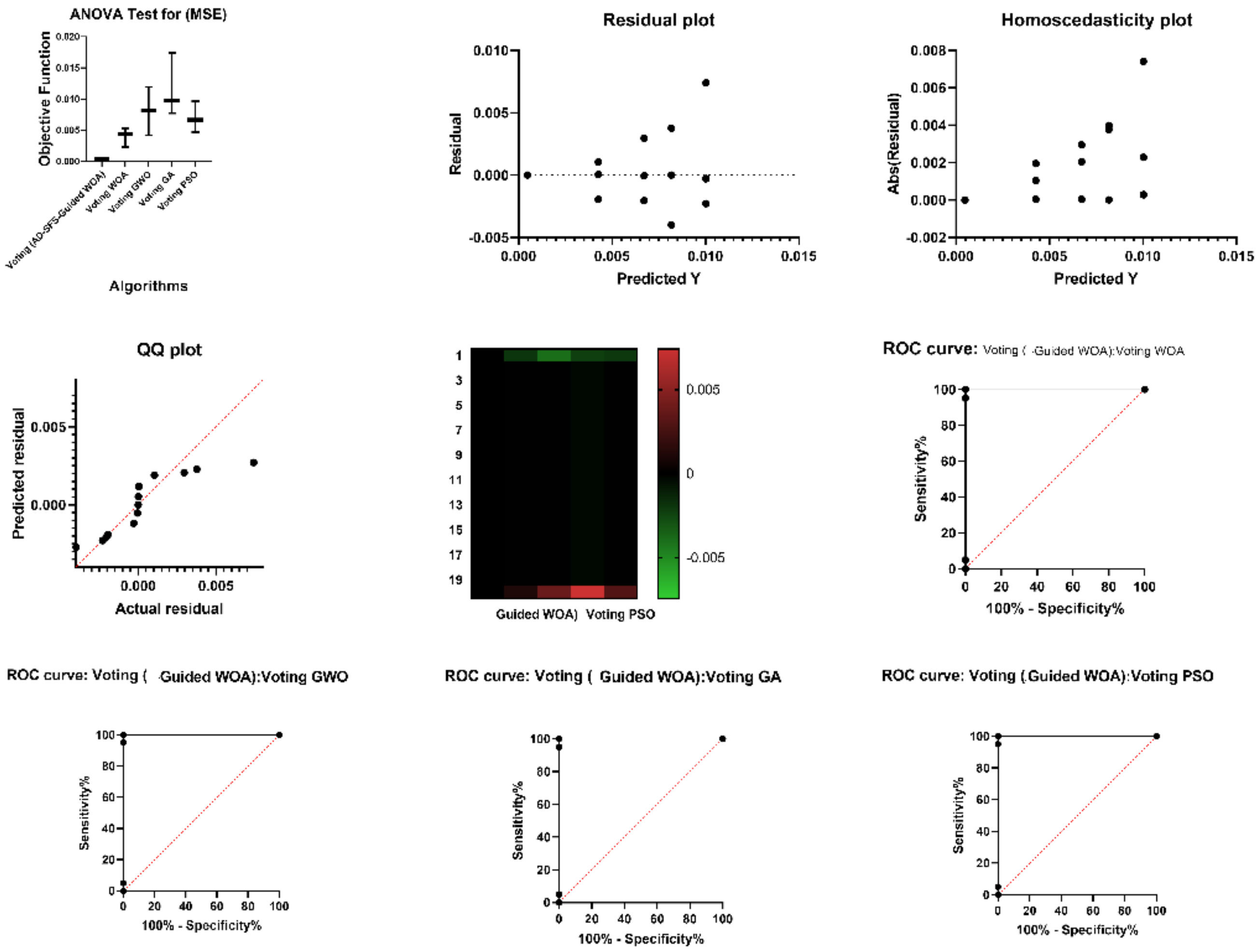

FIGURE 6: Different curves for the classification techniques.

TABLE 9: Classification Models Parameters

Classifier	Parameter (s)	Value (s)
NN	hidden_layer_sizes	20
	learning_rate_init	0.007
	validation_fraction	0.1
	beta_1	0.6
	beta_2	0.899
	epsilon	1e-06
k-NN	n_neighbors	3
	leaf_size	20
	p	2
Random Forest	n_estimators	40
	min_samples_split	2
	min_samples_leaf	1
	min_weight_fraction_leaf	0.0
Voting	estimators	*
	voting	hard
	weights	'optimal' #AD-PRS-Guided WOA
	flatten_transform	True

TABLE 10: Single classification models' results

	NN	k-NN	Random Forest
AUC	0.744	0.7387	0.797
MSE	0.080226	0.098999	0.04887

The MSE value is evaluated by calculating the difference between the required and the actual output of the classifiers according to this equation:

$$MSE = \sum_{x=1}^{n} (o_x^h d_x^h)^2 \tag{15}$$

where for n number of outputs, d_x^h is the xth input neuron optimal value when applying hth training instance. When the hth training instance appears in the input, o_x^h is the optimal output actual value of the xth input neuron.

Classification models parameters settings are presented in Table 9. The table includes the values of the parameters of NN, k-NN, Random Forest, and Voting classifiers.

Single classification models' results for the tested dataset based on the NN, k-NN, Random forest techniques are shown in Table 10. The Random forest classifier achieved an AUC of (0.797) and an MSE of (0.04887), which are better than the NN and k-NN classifiers for the current problem. However, the results can be improved. The results of the proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

TABLE 11: Proposed optimization ensemble (AD-PRS-Guided WOA) compared to other algorithms

	Bagging	AdaBoost	Majority voting	Voting (AD-PRS-Guided WOA)
AUC	0.802	0.827	0.887	0.971
MSE	0.022941	0.015132	0.005813	4.77E-04

TABLE 12: Proposed optimization ensemble voting (AD-PRS-Guided WOA) compared to other voting based algorithms

	Voting (AD-PRS-Guided WOA)	Voting WOA	Voting GWO	Voting GA	Voting PSO
AUC	0.971	0.946	0.926	0.917	0.938
MSE	4.77E-04	0.004324	0.008195	0.0097441	0.00667

TABLE 13: Descriptive statistics of the proposed optimization ensemble (AD-PRS-Guided WOA) compared to other algorithms

	Voting (AD-PRS-Guided WOA)	Voting WOA	Voting GWO	Voting GA	Voting PSO
Number of values	20	20	20	20	20
Minimum	0.000477	0.002324	0.004195	0.007744	0.00467
25% Percentile	0.000477	0.004324	0.008195	0.009744	0.00667
Median	0.000477	0.004324	0.008195	0.009744	0.00667
75% Percentile	0.000477	0.004324	0.008195	0.009744	0.00667
Maximum	0.000477	0.005324	0.01195	0.01744	0.00967
Range	0	0.003	0.007755	0.009697	0.005
Mean	0.000477	0.004274	0.008183	0.01003	0.00672
Std. Deviation	0	0.0005104	0.001259	0.001801	0.000826
Std. Error of Mean	0	0.0001141	0.0002814	0.0004027	0.000185
Coefficient of variation	0.000%	11.94%	15.38%	17.96%	12.29%
Geometric mean	0.000477	0.004236	0.008076	0.009917	0.006675
Geometric SD factor	1	1.16	1.192	1.153	1.125
Sum	0.00954	0.08548	0.1637	0.2006	0.1344

TABLE 14: ANOVA test results of the proposed optimization ensemble (AD-PRS-Guided WOA) algorithm

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	0.0011	4	0.000275	F (4, 95) = 238.2	P < 0.0001
Residual (within columns)	0.0001096	95	1.15E-06		
Total	0.001209	99			

TABLE 15: Wilcoxon Signed Rank Test results of the proposed and compared voting-based algorithms

	Voting (AD-PRS-Guided WOA)	Voting WOA	Voting GWO	Voting GA	Voting PSO
Theoretical median	0	0	0	0	0
Actual median	0.000477	0.004324	0.008195	0.009744	0.00667
Number of values	20	20	20	20	20
Wilcoxon Signed Rank Test					
Sum of signed ranks (W)	210	210	210	210	210
Sum of positive ranks	210	210	210	210	210
Sum of negative ranks	0	0	0	0	0
P value (two tailed)	0.0001	0.0001	0.0001	0.0001	0.0001
Exact or estimate?	Exact	Exact	Exact	Exact	Exact
P value summary	****	****	****	****	****
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes	Yes

(voting AD-PRS-Guided WOA) algorithm compared to Bagging, AdaBoost, and Majority, voting ensemble techniques, are presented in Table 11.

The voting AD-PRS-Guided WOA algorithm reached an AUC value of (0.971) and an MSE value of (4.77E-04). These results show the algorithm's performance compared to single classifier models, including Random forest and ensemble-based techniques. The algorithm is compared to voting WOA, voting GWO, voting GA, and voting PSO algorithms to confirm the proposed voting algorithm's classification accuracy. Table 12 presents the voting-based algorithms results. This table confirms that the AUC value of (0.971) and the MSE value of (4.77E-04) are the optimal results that can be achieved based on the tested dataset. The descriptive statistics of voting AD-PRS-Guided WOA, voting WOA, voting GWO, voting GA, and voting PSO algorithms are shown in Table 13 which confirms the superiority of the proposed voting algorithm.

The ANOVA test is also applied in this experiment to test the statistical difference of the proposed voting (AD-PRS-Guided WOA). A null hypothesis is set as (H_0: $\mu_{A1} = \mu_{B1} = \mu_{C1} = \mu_{D1} = \mu_{E1}$), where $A1$: Voting (AD-PRS-Guided WOA), $B1$: Voting WOA, $C1$: Voting GWO, $D1$: Voting GA, and $E1$: Voting PSO, and an alternate hypothesis is formed as (H_1: No equal means). Table 14 shows the ANOVA test results. Figure 6 presents the proposed and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

compared algorithms ANOVA test results considering the function F_n. It is noted that the alternate hypothesis H_1 can be accepted for this test.

One more test, named Wilcoxon's rank-sum test, is employed in this part. This test can discover whether the proposed algorithm results have a significant difference compared to other algorithms. If the p-value < 0.05, it will indicate that algorithm has significant superiority. A null hypothesis is set as (H_0: $\mu_{A1} = \mu_{B1}, \mu_{A1} = \mu_{C1}, \mu_{A1} = \mu_{D1}, \mu_{A1} = \mu_{E1}$), and an alternate hypothesis is formed as (H_1: No equal means). Table 15 show the test results of 20 runs (Repetitions) as indicated in Table 4. This confirms that the p-values are less than 0.05, which shows the statistically significant difference between groups. Thus, the hypothesis H_1 can be accepted.

The residual values and plots can observe the possible problems better than the plot of the original dataset. Some datasets can be not also good candidates for the classification process. Figure 6 shows the residual, heteroscedasticity, and quantile-quantile (QQ) plots and the heatmap for the tested dataset classification process. These plots indicate that the actual and the predicted residuals are linearly related, confirming the proposed voting (AD-PRS-Guided WOA) algorithm performance. Figure 6 also includes the ROC curves of the proposed voting (AD-PRS-Guided WOA) algorithm mapped to the compared voting algorithms. The ROC curves indicate that the proposed voting algorithm can distinguish different cases with a high AUC value near 1.0.

VI. VALIDATION AND DISCUSSION

The proposed classification algorithm is validated by comparing its results with the other DGA techniques in the literature. A total of 74 samples were extracted from the 475 data samples as testing samples. A total of the 74 samples was randomly selected by the optimization method. The distribution of the testing samples is illustrated in Table 16. In this table, the samples were categorized as 6 labels for PD, 13 for D1, 24 for D2, 16 for T1, 4 for T2, and 10 for T3 fault. It also shows the number of samples that were collected from the practical cases (39 real samples from [54]) and the credited published articles as 17 samples from [8]. The results of the diagnostic accuracy of the proposed classification algorithm are illustrated in Table 17 comparing with the diagnostic accuracy of other DGA techniques in the literature. It showed from Table 17 that the overall diagnostic accuracy of the proposed classification algorithm is 94.6%, which is greater than that of the other DGA techniques where the highest diagnostic accuracy that close to the proposed algorithm is Conditional probability [13] and NPR [41] providing 90.54%. On the other hand, the other traditional DGA techniques have poor overall diagnostic accuracies such as IEC 60599 (50%), Rogers Ratio method (45.95%), and Duval triangle method (66.27%). The proposed classification algorithm's diagnostic accuracy results revealed that the proposed classification algorithm's a high ability for correct diagnoses of the transformer faults.

TABLE 16: Distribution of the 74 testing samples according to the fault type and references

Ref.	PD	D1	D2	T1	T2	T3	Total
[8]	1	6	8	1		1	17
[54]	6	6	11	11	3	2	39
[56]			2	1			3
[57]						1	1
[58]						2	2
[59]			1				1
[60]		1	1	1			3
[61]						1	1
[63]				2	1	3	6
[64]			1				1
Total	7	13	24	16	4	10	74

The proposed binary AD-PRS-Guided WOA algorithm in the feature selection process achieved an average error of (0.4515) which is the minimum error among the compared algorithms, and other metrics, including the standard deviation of (0.0337), approve the algorithm's superiority in this kind of problem. The algorithm also shows a fast convergence in finding find the optimal solution compared to other techniques. The proposed voting AD-PRS-Guided WOA algorithm achieved an AUC (balanced accuracy) value of (0.971) and a MSE value of (4.77E-04) which are the best results that can be achieved based on the tested dataset compared to other techniques.

For the feature selection scenario, the ANOVA test was applied to test the statistical difference of the proposed (AD-PRS-Guided WOA) algorithm. Another test, named the one-sample t-test, was also conducted for the evaluation at a significance level of 0.05. For the classification scenario, the ANOVA test was applied in the experiment to test the statistical difference of the proposed voting (AD-PRS-Guided WOA) algorithm. One more test, named Wilcoxon's rank-sum test, was also employed in this part. This test can discover whether the proposed algorithm results have a significant difference compared to other algorithms. The statistical analysis based on different tests confirmed that the proposed algorithm is a statistically significant difference.

VII. CONCLUSION

Several traditional dissolved gas analysis (DGA) techniques, such as IEC code 60599 and the Duval triangle method, were used to diagnose the transformer faults. This paper proposed a novel Adaptive Dynamic Polar Rose Guided Whale Optimization algorithm (AD-PRS-Guided WOA) to improve classification techniques' parameters of several classification techniques. A binary version of the proposed algorithm (binary AD-PRS-Guided WOA) was used for feature selection from the tested dataset. ANOVA and one-sample t-test tests were applied in this experiment to test the statistical difference of the proposed binary AD-PRS-Guided WOA. A voting classifier based on the proposed algorithm (voting AD-PRS-Guided WOA) was developed to improve the tested dataset classification accuracy. The ANOVA and Wilcoxon's rank-sum tests were investigated to show the proposed voting's statistical difference (AD-PRS-Guided WOA) algo-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

TABLE 17: Diagnostic accuracy of the 74 testing samples of the suggested DGA algorithm and the other DGA techniques in literature

Fault Type	Samples	Adaptive	IEC-60599 [5]	Rog. 4 Ratios [6]	IEC 60599 Modified [1]	Rog. Modified [1]	Duval [7], [8]
PD	7	100	28.57	14.29	100	100	42.86
D1	13	92.31	30.77	0	61.54	61.54	69.23
D2	24	91.67	41.67	50	87.5	79.17	75
T1	16	93.95	68.75	100	100	100	56.25
T2	4	100	75	25	100	100	0
T3	10	100	70	40	100	100	100
Overall	74	94.6	50	45.95	89.19	86.49	66.27
Fault Type	Samples	Clustering [73]	Cond. Prob. [13]	CSUS-ANN [4]	NPR [41]	SVM [41]	
PD	7	100	100	85.71	100	85.71	
D1	13	69.23	61.54	53.85	76.92	76.92	
D2	24	91.67	100	87.5	87.5	91.67	
T1	16	93.75	87.5	93.75	100	100	
T2	4	75	100	75	75	25	
T3	10	100	100	80	100	100	
Overall	74	89.19	90.54	81.08	90.54	87.84	

rithm. The proposed AD-PRS-Guided WOA algorithm provided high diagnostic accuracy of transformer faults, higher than other DGA techniques in the literature. The proposed algorithm's diagnostic accuracy results have a high ability for correct diagnoses of the transformer faults. The proposed algorithm's diagnostic accuracy based on randomly selected samples from the tested dataset approved the algorithm's performance compared to other DGA techniques. The proposed binary and voting algorithms can be generalized and applied to different datasets in the future.

REFERENCES

[1] I. B. M. Taha, A. Hoballah, and S. S. M. 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, pp. 222–230, Feb. 2020. [Online]. Available: https://doi.org/10.1109/tdei.2019.008395

[2] S. S. M. Ghoneim and I. B. M. Taha, "Comparative study of full and reduced feature scenarios for health index computation of power transformers," IEEE Access, vol. 8, pp. 181 326–181 339, 2020. [Online]. Available: https://doi.org/10.1109/access.2020.3028689

[3] O. E. Gouda, S. H. El-Hoshy, and H. H. E.L.-Tamaly, "Proposed three ratios technique for the interpretation of mineral oil transformers based dissolved gas analysis," IET Generation, Transmission & Distribution, vol. 12, no. 11, pp. 2650–2661, Apr. 2018. [Online]. Available: https://doi.org/10.1049/iet-gtd.2017.1927

[4] S. S. M. Ghoneim, I. B. M. Taha, and N. I. Elkalashy, "Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 23, no. 3, pp. 1838–1845, Jun. 2016. [Online]. Available: https://doi.org/10.1109/tdei.2016.005301

[5] IEC 60599: Mineral oil-filled electrical equipment in service - Guidance on the interpretation of dissolved and free gases analysis. IEC, Geneva, Switzerland, Edition 2.1, 2007.

[6] "Ieee guide for the interpretation of gases generated in oil-immersed transformers," IEEE Std C57.104-2008 (Revision of IEEE Std C57.104-1991), pp. 1–36, 2009.

[7] M. Duval, "A review of faults detectable by gas-in-oil analysis in transformers," IEEE Electrical Insulation Magazine, vol. 18, no. 3, pp. 8–17, May 2002. [Online]. Available: https://doi.org/10.1109/mei.2002.1014963

[8] M. Duval and A. dePabla, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases," IEEE Electrical Insulation Magazine, vol. 17, no. 2, pp. 31–41, Mar. 2001. [Online]. Available: https://doi.org/10.1109/57.917529

[9] N. A. Bakar and A. Abu-Siada, "Fuzzy logic approach for transformer remnant life prediction and asset management decision," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 23, no. 5, pp. 3199–3208, Oct. 2016. [Online]. Available: https://doi.org/10.1109/tdei.2016.7736886

[10] D.-E. A. Mansour, "Development of a new graphical technique for dissolved gas analysis in power transformers based on the five combustible gases," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 22, no. 5, pp. 2507–2512, Oct. 2015. [Online]. Available: https://doi.org/10.1109/tdei.2015.004999

[11] "The duval pentagon-a new complementary tool for the interpretation of dissolved gas analysis in transformers," IEEE Electrical Insulation Magazine, vol. 30, no. 6, pp. 9–12, Nov. 2014. [Online]. Available: https://doi.org/10.1109/mei.2014.6943428

[12] O. E. Gouda, S. H. El-Hoshy, and H. H. El-Tamaly, "Proposed heptagon graph for DGA interpretation of oil transformers," IET Generation, Transmission & Distribution, vol. 12, no. 2, pp. 490–498, Jan. 2018. [Online]. Available: https://doi.org/10.1049/iet-gtd.2017.0826

[13] I. B. Taha, D.-E. A. Mansour, S. S. Ghoneim, and N. I. Elkalashy, "Conditional probability-based interpretation of dissolved gas analysis for transformer incipient faults," IET Generation, Transmission & Distribution, vol. 11, no. 4, pp. 943–951, Mar. 2017. [Online]. Available: https://doi.org/10.1049/iet-gtd.2016.0886

[14] M. D. Equbal, 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 and Computer Science, vol. 26, pp. 77 – 88, 2018.

[15] M. Ou, H. Wei, Y. Zhang, and J. Tan, "A dynamic adam based deep neural network for fault diagnosis of oil-immersed power transformers," Energies, vol. 12, no. 6, p. 995, Mar. 2019. [Online]. Available: https://doi.org/10.3390/en12060995

[16] E. T. Mharakurwa, G. N. Nyakoe, and A. O. Akumu, "Power transformer fault severity estimation based on dissolved gas analysis and energy of fault formation technique," Journal of Electrical and Computer Engineering, vol. 2019, pp. 1–10, Feb. 2019. [Online]. Available: https://doi.org/10.1155/2019/9674054

[17] M. Noori, R. Effatnejad, and P. Hajihosseini, "Using dissolved gas analysis results to detect and isolate the internal faults of power transformers by applying a fuzzy logic method," IET Generation, Transmission & Distribution, vol. 11, no. 10, pp. 2721–2729, Jul. 2017. [Online]. Available: https://doi.org/10.1049/iet-gtd.2017.0028

[18] J. Aghaei, A. Gholami, H. A. Shayanfar, and A. Dezhamkhooy, "Dissolved gas analysis of transformers using fuzzy logic approach," European Transactions on Electrical Power, pp. 630–638, 2009. [Online]. Available: https://doi.org/10.1002/etep.343

[19] K. U. Mulyodinoto, Suwarno, R. Prasojo, and A. Abu-Siada, "Applications of anfis to estimate the degree of polymerization using transformer dissolve gas analysis and oil characteristics," Polymer sciences, vol. 14, no. 2, pp. 1–9, 2018.

[20] M. Allahbakhshi and A. Akbari, "Novel fusion approaches for the dissolved gas analysis of insulating oil," IJST, Transactions of Electrical Engineering, vol. 35, no. E, pp. 13–24, 2011.

[21] K. Bacha, S. Souahlia, and M. Gossa, "Power transformer fault diagnosis based on dissolved gas analysis by support vector machine," Electric Power Systems Research, vol. 83, no. 1, pp. 73–79, Feb. 2012. [Online]. Available: https://doi.org/10.1016/j.epsr.2011.09.012

[22] Y. Benmahamed, M. Teguar, and A. Boubakeur, "Application

of SVM and KNN to duval pentagon 1 for transformer oil diagnosis," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 24, no. 6, pp. 3443–3451, Dec. 2017. [Online]. Available: https://doi.org/10.1109/tdei.2017.006841

[23] D. Bhalla, R. K. Bansal, and H. O. Gupta, "Integrating AI based DGA fault diagnosis using dempster–shafer theory," International Journal of Electrical Power & Energy Systems, vol. 48, pp. 31–38, Jun. 2013. [Online]. Available: https://doi.org/10.1016/j.ijepes.2012.11.018

[24] Y. Benmahamed, Y. Kemari, M. Teguar, and A. Boubakeur, "Diagnosis of power transformer oil using KNN and nave bayes classifiers," in 2018 IEEE 2nd International Conference on Dielectrics (ICD). IEEE, Jul. 2018. [Online]. Available: https://doi.org/10.1109/icd.2018.8514789

[25] Y.-C. Huang, "A new data mining approach to dissolved gas analysis of oil-insulated power apparatus," IEEE Transactions on Power Delivery, vol. 18, no. 4, pp. 1257–1261, Oct. 2003. [Online]. Available: https://doi.org/10.1109/tpwrd.2003.817736

[26] A. Shintemirov, W. Tang, and Q. Wu, "Power transformer fault classification based on dissolved gas analysis by implementing bootstrap and genetic programming," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 39, no. 1, pp. 69–79, Jan. 2009. [Online]. Available: https://doi.org/10.1109/tsmcc.2008.2007253

[27] S. Al-Janabi, S. Rawat, A. Patel, and I. Al-Shourbaji, "Design and evaluation of a hybrid system for detection and prediction of faults in electrical transformers," International Journal of Electrical Power & Energy Systems, vol. 67, pp. 324–335, May 2015. [Online]. Available: https://doi.org/10.1016/j.ijepes.2014.12.005

[28] Z. Richardson, J. Fitch, W. Tang, J. Goulermas, and Q. Wu, "A probabilistic classifier for transformer dissolved gas analysis with a particle swarm optimizer," IEEE Transactions on Power Delivery, vol. 23, no. 2, pp. 751–759, Apr. 2008. [Online]. Available: https://doi.org/10.1109/tpwrd.2008.915812

[29] H. A. Illias, X. R. Chai, A. H. A. Bakar, and H. Mokhlis, "Transformer incipient fault prediction using combined artificial neural network and various particle swarm optimisation techniques," PLOS ONE, vol. 10, no. 6, p. e0129363, Jun. 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0129363

[30] H. A. Illias, X. R. Chai, and A. H. A. Bakar, "Hybrid modified evolutionary particle swarm optimisation-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis," Measurement, vol. 90, pp. 94–102, Aug. 2016. [Online]. Available: https://doi.org/10.1016/j.measurement.2016.04.052

[31] A. Li, X. Yang, Z. Xie, and C. Yang, "An optimized GRNN-enabled approach for power transformer fault diagnosis," IEEJ Transactions on Electrical and Electronic Engineering, vol. 14, no. 8, pp. 1181–1188, Jun. 2019. [Online]. Available: https://doi.org/10.1002/tee.22916

[32] B. Zeng, J. Guo, W. Zhu, Z. Xiao, F. Yuan, and S. Huang, "A transformer fault diagnosis model based on hybrid grey wolf optimizer and LS-SVM," Energies, vol. 12, no. 21, p. 4170, Nov. 2019. [Online]. Available: https://doi.org/10.3390/en12214170

[33] A. Hoballah, D.-E. A. Mansour, and I. B. M. Taha, "Hybrid grey wolf optimizer for transformer fault diagnosis using dissolved gases considering uncertainty in measurements," IEEE Access, vol. 8, pp. 139 176–139 187, 2020. [Online]. Available: https://doi.org/10.1109/access.2020.3012633

[34] X. Yang, W. Chen, A. Li, C. Yang, Z. Xie, and H. Dong, "BA-PNN-based methods for power transformer fault diagnosis," Advanced Engineering Informatics, vol. 39, pp. 178–185, Jan. 2019. [Online]. Available: https://doi.org/10.1016/j.aei.2019.01.001

[35] S. S. Ghoneim and I. B. Taha, "A new approach of DGA interpretation technique for transformer fault diagnosis," International Journal of Electrical Power & Energy Systems, vol. 81, pp. 265–274, Oct. 2016. [Online]. Available: https://doi.org/10.1016/j.ijepes.2016.02.018

[36] S. Koroglu and A. Demircali, "Diagnosis of power transformer faults based on multi-layer support vector machine hybridized with optimization methods," Electric Power Components and Systems, vol. 44, no. 19, pp. 2172–2184, Nov. 2016. [Online]. Available: https://doi.org/10.1080/15325008.2016.1219427

[37] T. Kari, W. Gao, D. Zhao, Z. Zhang, W. 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, pp. 360–371, Feb. 2018. [Online]. Available: https://doi.org/10.1109/tdei.2018.006746

[38] Y. Kim, T. Park, S. Kim, N. Kwak, and D. Kweon, "Artificial intelligent fault diagnostic method for power transformers using a new classification system of faults," Journal of Electrical Engineering & Technology, vol. 14, no. 2, pp. 825–831, Feb. 2019. [Online]. Available: https://doi.org/10.1007/s42835-019-00105-0

[39] X. Wu, Y. He, and J. Duan, "A deep parallel diagnostic method for transformer dissolved gas analysis," Applied Sciences, vol. 10, no. 4, p. 1329, Feb. 2020. [Online]. Available: https://doi.org/10.3390/app10041329

[40] S. S. M. Ghoneim, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, "Enhancing diagnostic accuracy of transformer faults using teaching-learning-based optimization," IEEE Access, vol. 9, pp. 30 817–30 832, 2021. [Online]. Available: https://doi.org/10.1109/access.2021.3060288

[41] I. B. Taha, S. S. Dessouky, and S. S. Ghoneim, "Transformer fault types and severity class prediction based on neural pattern-recognition techniques," Electric Power Systems Research, vol. 191, p. 106899, Feb. 2021. [Online]. Available: https://doi.org/10.1016/j.epsr.2020.106899

[42] E.-S. El-Kenawy and M. Eid, "Hybrid gray wolf and particle swarm optimization for feature selection," International Journal of Innovative Computing, Information and Control, vol. 16, no. 3, pp. 831–844, 2020.

[43] R. Bello, Y. Gomez, A. Nowe, and M. M. Garcia, "Two-step particle swarm optimization to solve the feature selection problem," in Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007), Oct 2007, pp. 691–696.

[44] A. Ibrahim, M. Noshy, H. A. Ali, and M. Badawy, "PAPSO: A power-aware VM placement technique based on particle swarm optimization," IEEE Access, vol. 8, no. 1, pp. 81 747–81 764, 2020. [Online]. Available: https://doi.org/10.1109/access.2020.2990828

[45] I. Karakonstantis and A. Vlachos, "Bat algorithm applied to continuous constrained optimization problems," Journal of Information and Optimization Sciences, pp. 1–19, Mar. 2020. [Online]. Available: https://doi.org/10.1080/02522667.2019.1694740

[46] Q. Al-Tashi, S. J. Abdul Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary optimization using hybrid grey wolf optimization for feature selection," IEEE Access, vol. 7, pp. 39 496–39 508, 2019.

[47] S. Mirjalili and A. Lewis, "The whale optimization algorithm," Advances in Engineering Software, vol. 95, pp. 51 – 67, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0965997816300163

[48] E. M. Hassib, A. I. El-Desouky, L. M. Labib, and E.-S. M. El-kenawy, "Woa+brnn: An imbalanced big data classification framework using whale optimization and deep neural network," Soft Computing, vol. 24, no. 8, pp. 5573–5592, Mar. 2019. [Online]. Available: https://doi.org/10.1007/s00500-019-03901-y

[49] S. H. Samareh Moosavi and V. Khatibi Bardsiri, "Satin bowerbird optimizer," Eng. Appl. Artif. Intell., vol. 60, no. C, pp. 1–15, Apr. 2017. [Online]. Available: https://doi.org/10.1016/j.engappai.2017.01.006

[50] S. Mirjalili, S. M. Mirjalili, and A. Hatamlou, "Multi-verse optimizer: A nature-inspired algorithm for global optimization," Neural Comput. Appl., vol. 27, no. 2, pp. 495–513, Feb. 2016. [Online]. Available: https://doi.org/10.1007/s00521-015-1870-7

[51] D. Simon, "Biogeography-based optimization," IEEE Transactions on Evolutionary Computation, vol. 12, no. 6, pp. 702–713, Dec. 2008. [Online]. Available: https://doi.org/10.1109/tevc.2008.919004

[52] I. Fister, X.-S. Yang, I. Fister, and J. Brest, "Memetic firefly algorithm for combinatorial optimization," Tech. Rep. arXiv:1204.5165, Apr 2012, comments: 14 pages; Bioinspired Optimization Methods and their Applications (BIOMA 2012). [Online]. Available: https://cds.cern.ch/record/1443422

[53] M. M. Kabir, M. Shahjahan, and K. Murase, "A new local search based hybrid genetic algorithm for feature selection," Neurocomputing, vol. 74, no. 17, pp. 2914 – 2928, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231211002748

[54] "Egyptian electricity holding company (eehc) reports," http://www.moee.gov.eg/english_new/report.aspx, accessed: 2021-03-06.

[55] S. Agrawal and A. K. Chandel, "Transformer incipient fault diagnosis based on probabilistic neural network," in 2012 Students Conference on Engineering and Systems. IEEE, Mar. 2012. [Online]. Available: https://doi.org/10.1109/sces.2012.6199110

[56] M.-H. Wang, "A novel extension method for transformer fault diagnosis," IEEE Transactions on Power Delivery, vol. 18, no. 1, pp. 164–169, Jan. 2003. [Online]. Available: https://doi.org/10.1109/tpwrd.2002.803838

[57] Z. Yong-li, W. fang, and G. Lan-qin, "Transformer fault diagnosis based on naive bayesian classifier and SVR," in TENCON 2006 - 2006 IEEE Region 10 Conference. IEEE, 2006. [Online]. Available: https://doi.org/10.1109/tencon.2006.343895

[58] D. V. S. S. Siva Sarma and G. N. S. Kalyani, "Ann approach for condition monitoring of power transformers using dga," in 2004 IEEE Region 10 Conference TENCON 2004., vol. C, 2004, pp. 444–447 Vol. 3.

[59] J. Hu, L. Zhou, and M. Song, "Transformer fault diagnosis method of gas hromatographic analysis using computer image analysis," in 2012 Second International Conference on Intelligent System Design and Engineering Application, 2012, pp. 1169–1172.

[60] S. Seifeddine, B. Khmais, and C. Abdelkader, "Power transformer fault diagnosis based on dissolved gas analysis by artificial neural network," in 2012 First International Conference on Renewable Energies and Vehicular Technology, 2012, pp. 230–236.

[61] M. Rajabimendi and E. P. Dadios, "A hybrid algorithm based on neural-fuzzy system for interpretation of dissolved gas analysis in power transformers," in TENCON 2012 IEEE Region 10 Conference, 2012, pp. 1–6.

[62] R. Soni and K. Chaudhari, "An approach to diagnose incipient faults of power transformer using dissolved gas analysis of mineral oil by ratio methods using fuzzy logic," in 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), 2016, pp. 1894–1899.

[63] G. Zhang, K. Yasuoka, S. Ishii, Li Yang, and Zhang Yan, "Application of fuzzy equivalent matrix for fault diagnosis of oil-immersed insulation," in Proceedings of 1999 IEEE 13th International Conference on Dielectric Liquids (ICDL'99) (Cat. No.99CH36213), 1999, pp. 400–403.

[64] O. E. Gouda, S. M. Saleh, and S. H. El-hoshy, "Power transformer incipient faults diagnosis based on dissolved gas analysis," Indonesian Journal of Electrical Engineering and Computer Science, vol. 1, no. 1, pp. 10–16, Jan. 2016. [Online]. Available: https://doi.org/10.11591/ijeecs.v1.i1.pp10-16

[65] J. Li, Q. Zhang, K. Wang, J. Wang, T. Zhou, and Y. Zhang, "Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 23, no. 2, pp. 1198–1206, Apr. 2016. [Online]. Available: https://doi.org/10.1109/tdei.2015.005277

[66] E.-S. M. El-kenawy, S. Mirjalili, A. Ibrahim, M. Alrahmawy, M. El-Said, R. M. Zaki, and M. M. Eid, "Advanced meta-heuristics, convolutional neural networks, and feature selectors for efficient COVID-19 x-ray chest image classification," IEEE Access, pp. 1–1, 2021. [Online]. Available: https://doi.org/10.1109/access.2021.3061058

[67] E.-S. M. El-Kenawy, M. M. Eid, M. Saber, and A. Ibrahim, "MbGWO-SFS: Modified binary grey wolf optimizer based on stochastic fractal search for feature selection," IEEE Access, vol. 8, no. 1, pp. 107 635–107 649, 2020. [Online]. Available: https://doi.org/10.1109/access.2020.3001151

[68] E.-S. M. El-kenawy, A. Ibrahim, S. Mirjalili, M. M. Eid, and S. E. Hussein, "Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images," IEEE Access, vol. 8, pp. 179 317 – 179 335, 2020. [Online]. Available: https://doi.org/10.1109/access.2020.3028012

[69] A. Ibrahim, A. Tharwat, T. Gaber, and A. E. Hassanien, "Optimized superpixel and adaboost classifier for human thermal face recognition," Signal, Image and Video Processing, vol. 12, pp. 711 – 719, 2018. [Online]. Available: https://doi.org/10.1007/s11760-017-1212-6

[70] S. Mirjalili, S. M. Mirjalili, S. Saremi, and S. Mirjalili, Whale Optimization Algorithm: Theory, Literature Review, and Application in Designing Photonic Crystal Filters. Cham: Springer International Publishing, 2020, pp. 219–238. [Online]. Available: https://doi.org/10.1007/978-3-030-12127-3_13

[71] E. Cuevas, F. Fausto, and A. González, Metaheuristics and Swarm Methods: A Discussion on Their Performance and Applications. Cham: Springer International Publishing, 2020, pp. 43–67. [Online]. Available: https://doi.org/10.1007/978-3-030-16339-6_2

[72] F. Fausto, A. Reyna-Orta, E. Cuevas, Á. G. Andrade, and M. A. Ṕerez-Cisneros, "From ants to whales: metaheuristics for all tastes," Artificial Intelligence Review, vol. 53, pp. 753–810, 2019.

[73] S. I. Ibrahim, S. S. Ghoneim, and I. B. Taha, "DGALab: an extensible software implementation for DGA," IET Generation, Transmission & Distribution, vol. 12, no. 18, pp. 4117–4124, Sep. 2018. [Online]. Available: https://doi.org/10.1049/iet-gtd.2018.5564

SHERIF S. M. GHONEIM (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Faculty of Engineering at Shoubra, Zagazig University, Egypt, in 1994 and 2000, respectively, and the Ph.D. degree in electrical power and machines from the Faculty of Engineering, Cairo University, in 2008. Since 1996, he has been teaching at the Faculty of Industrial Education, Suez Canal University, Egypt. From the end of 2005 to the end of 2007, he was a Guest Researcher with the Institute of Energy Transport and Storage (ETS), University of Duisburg–Essen, Germany. He joined Taif University as an Associate Professor with the Electrical Engineering Department, Faculty of Engineering. His research interests include grounding systems, dissolved gas analysis, breakdown in SF6 gas, and AI technique applications.

TAMER AHMED FARRAG was born in 1981. He received the B.Sc., M.Sc., and Ph.D. degrees from the Department of Computers and Systems Engineering, Mansoura University, Egypt, in 2002, 2006, and 2012, respectively. In Egypt, he worked as an Assistant Professor with the Department of Computer Engineering, MISR Higher Institution of Engineering and Technology. His research interests include artificial intelligence, optimization, programming languages, and computing systems.

PLACE PHOTO HERE

A. ALI RASHED is a staff member at Taif University, Taif, KSA. His field of experience includes electrical power systems, machine learning. He has authored or coauthor in more than 6 technical articles.

EL-SAYED M. EL-KENAWY (Member, IEEE) is an assistant professor at Delta Higher Institute for Engineering & Technology (DHIET), Mansoura, Egypt. Inspiring and motivating students by providing a thorough understanding of a variety of computer concepts. He has published over 30 publications with over 650 citations and an H-index of 18. He has pioneered and launched independent research programs. Adept at sometimes explaining complex concepts in an easy-to-understand manner. He serves as a Reviewer for the Journal IEEE Access. His research interests include optimization, artificial intelligence, machine learning, deep learning, data science, and digital marketing.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3083593, IEEE Access

ABDELHAMEED IBRAHIM (Member, IEEE) received the bachelor's and master's degrees in engineering from the Computer Engineering and Systems Department, in 2001 and 2005, respectively, and the Ph.D. degree in engineering from the Faculty of Engineering, Chiba University, Japan, in 2011. He was with the Faculty of Engineering, Mansoura University, Egypt, from 2001 to 2007, where he is currently an Associate Professor of computer engineering with the Faculty of Engineering. He has published over 50 publications with over 850 citations and an H-index of 17. He serves as a Reviewer for the Journal of Electronic Imaging, Optical Engineering, the IEEE Journal of Biomedical and Health Informatics, IEEE Access, Computer Standards and Interfaces, Journal of Healthcare Engineering, IET Image Processing, Multimedia Tools and Applications, and other respected journals. His research interests include pattern recognition, optimization, machine learning, virtualization, and live virtual machine migration.