

Received September 23, 2020, accepted September 30, 2020, date of publication October 5, 2020, date of current version October 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3028689

# Comparative Study of Full and Reduced Feature Scenarios for Health Index Computation of Power Transformers

SHERIF S. M. GHONEIM<sup>1,2</sup>, (Senior Member, IEEE), AND IBRAHIM B. M. TAHA<sup>1,3</sup>

<sup>1</sup>Department of Electrical Engineering, College of Engineering, Taif University, Taif 21944, Saudi Arabia

<sup>2</sup>Faculty of Technology and Education, Suez University, Suez 43527, Egypt

<sup>3</sup>Electrical Power and Machines Department, Faculty of Engineering, Tanta University, Tanta 31521, Egypt

Corresponding author: Ibrahim B. M. Taha (i.taha@tu.edu.sa)

This work was supported by Taif University Researchers Support Project Number (TURSP-2020/34).

**ABSTRACT** Power transformer health index (PTHI) computation is performed based on the results of different tests, such as dissolved gas analysis (DGA), oil quality (OQ) evaluation, and depolarization factor (DP) testing. In this study, PTHI computation is performed using 631 dataset samples from Malaysia and 730 samples from the Gulf Region. A new model is proposed to predict the PTHI state by adopting intelligent classification methods (e.g., decision tree, support vector machine, k-nearest neighbor, and ensemble methods). The model is built via two-stage data processing. The first stage separates the test results into three modules that represent DGA, OQ, and DP factor codes. In the second stage, the output of the three modules is processed to predict the PTHI state. The four classification methods are applied to the proposed model, and the prediction accuracy of the PTHI state is determined. Results indicate that the proposed model has superior classification accuracy for each AI method compared with recent work. Furthermore, feature reductions are applied to minimize the testing time, effort, and costs. The reduced-feature models reveal the effectiveness of the adopted feature reduction technique. A slight difference in accuracy is observed between the full- and reduced-feature scenarios. Thus, the reduced-feature scenario is considered to decrease the effort and time of the computation process and the experimental cost. The proposed model is validated against uncertain noise in features of up to  $\pm 20\%$ .

**INDEX TERMS** Power transformer health index, transformer tests, intelligent classification methods.

## I. INTRODUCTION

The transformer is the most expensive equipment in the electric power system. Monitoring the status of the insulation system in a power transformer is vital. Deterioration of the transformer insulating oil due to electrical and thermal stresses can lead to undesired transformer outage in electric power networks. Therefore, the continuous evaluation of the power transformer state has elicited much attention. The condition of a transformer can be evaluated using a health index [1]. Establishing the power transformer health index (PTHI) is a challenge. It involves the fusion of present data from several sensors and equally important historical data. PTHI is a practical method that uses transformer test results to indicate the state of a power transformer.

The associate editor coordinating the review of this manuscript and approving it for publication was Arpan Kumar Pradhan<sup>1b</sup>.

Common transformer tests include dissolved gas analysis (DGA), insulating oil quality (OQ) evaluation, and furan analysis (FA) [1]–[3].

Artificial intelligence (AI) methods have been used to construct a prediction model to identify PTHI directly from test results. A neuro-fuzzy (NF) network was constructed in [3], [4] to determine PTHI on the basis of the results of DGA, OQ, and FA. The results indicated that 62% of the 73 tested transformers had the same assessment results regardless of the type of data used. In [4], an artificial neural network (ANN) and adaptive NF inference system were used to construct a model for identifying PTHI. Technical and economical parameters were utilized as inputs to the constructed model. The technical parameters were DGA, OQ, and FA. The economical parameters included transformer aging variables and cost functions. A fuzzy logic (FL) system was utilized in [2], [5]–[7]. In [2], the effect of interfacial tension

List of abbreviations

Abbr.	Description	Abbr.	Description
PTHI	Power transformer health index	AI	Artificial intelligence
DGA	Dissolved gas analysis	ANN	Artificial neural network
OQ	Oil quality	DT	Decision tree
DP	Depolarization		
PF	Power factor	KNN	K-nearest neighbor
HI	Health index	EN	Ensemble methods
		FL	Fuzzy logic
OQF	Oil quality factor	$c$	Output classes
DPF	Depolarization factor	SVM	Support vector machine
$K_i$	Transformer condition weight factor	DGAF	Dissolved gas analysis factor
$W_i$	Gas weight factor	$S_i$	Gas score value
$Nc$	Number of corrected predicting state	$Nt$	Total number of states
$S$	Optimal set of features	$ S $	Number of features in $S$
$A_s$	Relevance	$B_s$	Redundancy
$x$ and $z$	Features belonging to set $S$	$I$	Mutual information
MRMR	Minimum redundancy maximum relevance	MIQ	Mutual information quotient
PCA	Principal component analysis		

and oil color was ignored, and the accuracy of the proposed FL model was not reported. In [5], an elaborate approach was used to construct separate FL systems for identifying individual transformer states. Then, a final FL system was built to compute PTHI on the basis of the results of the primary FL systems (thermal, electrical, mechanical, and dielectric integrity conditions). In [6], a Cauchy membership function for fust grade division and a fuzzy evidence fusion method were used to develop a PTHI prediction model. The model considers the bushing state and the state of other accessories. In [7], PTHI was estimated using the FL approach based on a distribution area of 69 KV or less. An analysis was conducted using the results of oil tests, such as water content, breakdown voltage, dissipation factor, furan,  $H_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_6$ , and  $C_2H_4$ . The ANN approach was applied to predict the PTHI state in [8]–[10]. In [8], an ANN was proposed for PTHI prediction by using a feature-based exhaustive method to eliminate the least significant tests. The constructed model was based on DGA, OQ, and FA. In [9], integrated transformer subsystems (insulation system, bushing, tap changer, core, and windings) were considered to evaluate PTHI. An intelligent multiple regression ANN model was constructed to build a quantitative PTHI. The model was applied on the basis of 345 transformer datasets with high-performance PTHI state prediction. In [10], an ANN model was built based on DGA, OQ, and FA factors to predict the PTHI state. The model exhibited an accuracy of 95% when it used the subset of input features, but its accuracy was 89% when it was tested with datasets of other utility networks.

A support vector machine (SVM) model was built in [11] to predict PTHI. The model was tested with 14 oil test results,

which were used to build the model on the basis of 724 test samples with high detection accuracy. In [12], the level of furan content in transformer oil was determined through the measurement of oil test factors, such as breakdown voltage, water content, and dissolved gas. A prediction model was built using the k-nearest neighbor (KNN) method with 90% detection accuracy. A wavelet model was developed in [13] on the basis of 19 oil test results to predict the PTHI state. The main objective of this work was to present a new method of assessing transformer conditions by adding a PTHI table that improves conventional approaches. The model was used to predict the PTHI state in 345 transformers, and it exhibited good detection performance.

A Bayesian information fusion approach was proposed in [14] to predict the PTHI state on the basis of data collected from transformer measurements, maintenance, and failure statistics. In [15], a decision-making model was built in consideration of reliability and economy to identify the best maintenance strategy for oil-filled transformers as a basis of PTHI prediction. Particle swarm optimization was used to construct the decision-making model on the basis of DGA, oil test, electrical test, reliability, and economic operation. In [16], a Markov model (MM) was utilized to determine the PTHI of 373 transformers. Dissolved gas, OQ, and furan compounds were used as model inputs. The MM model was built based on a probability decision process, and it was used for PTHI prediction. Meanwhile, a cluster-merging model was presented in [17] to predict the PTHI state. However, only DGA was considered in this work; all other tests for identifying the PTHI state were ignored. The method's accuracy was investigated by using one case only, which was not enough to verify the model's prediction capability. The proposed model in [18] was used to estimate the linear relation between PTHI and transformer tests. The researchers concluded that DGA, FA, and breakdown voltage test are sufficient and can provide an acceptable indication of PTHI compared with other techniques that involve many tests. However, the model was applied in 90 transformers only, which may not be enough to provide useful insights into the accuracy of the method. In [19], a feature reduction model was developed based on different reduced-feature approaches. The results showed that water content, breakdown voltage, furan, and acidity are the most important features in predicting the PTHI state.

The drawbacks of these previous studies are as follows. First, most of them predicted the PTHI state by using a one-stage process, which decreases the overall prediction accuracy. Second, they did not investigate the effect of uncertainty originating from various parameters, such as sampling temperature, sampling position, loading history, and maintenance history of the test transformer.

The current study presents a PTHI framework for transformers that is based on extensive research on ageing markers in transformers and how to classify them effectively for asset management. The framework is fed with training data and assesses classification accuracy by using a test dataset.

In addition, a new AI-based model for PTHI prediction was developed. The proposed model consists of two-stage data processing. The first stage uses four modules to process each type of test result, including DGA, OQ, and DP, and produces intermediate code factors. The codes are used as inputs to the second-stage module (module 4), which identifies the final PTHI state. All modules in the proposed model use AI classification methods, such as decision tree (DT), SVM, KNN, and ensemble method (EN). A total of 631 and 730 test dataset samples were collected from the test and maintenance laboratories of two electricity companies in two countries to evaluate the accuracy of the proposed models and alternative AI-optimized methods. The datasets were utilized for training and testing purposes. The results on PTHI prediction accuracy indicated that the proposed model and the four classification methods exhibited high prediction accuracy for PTHI states. The EN method demonstrated an advantage over the other AI methods. A feature reduction approach was used to reduce testing time, effort, and costs. The results of the feature reduction showed high PTHI detection performance. The uncertainty study revealed the robustness of the proposed method in predicting the PTHI state, and PTHI prediction accuracy exhibited only a slight change. The maximum change in PTHI accuracy was 2.8% and 5.3% for the KNN method at  $\pm 15\%$  and  $\pm 20\%$  uncertainty noise compared with that without uncertainty under full- and reduced-feature scenarios, respectively.

## II. POWER TRANSFORMER HEALTH INDEX

The determination of PTHI was performed based on the analysis in [1], [20], [21]. PTHI can be estimated as follows.

### A. DISSOLVED GAS ANALYSIS

The computation of the DGA factor (DGAF) was performed based on the following dissolved gases decomposed in transformer insulating oil:  $H_2$ ,  $CH_4$ ,  $C_2H_6$ ,  $C_2H_4$ ,  $C_2H_2$ ,  $CO$ , and  $CO_2$ . DGAF can be determined as

$$DGAF = \sum_{i=1}^7 \frac{S_i \times W_i}{W_i}, \quad (1)$$

where  $W_i$  and  $S_i$  refer to the weight and score of test parameter  $i$ , respectively, as shown in Table 12 in the appendix [20].

The transformer condition based on DGAF is represented by a range from “A” to “E”. For instance, when DGAF is less than 1.2, the transformer condition is good and referred to as code “A”. The transformer condition is considered poor when DGAF is greater than 3.0, which corresponds to code “E”. The codes from “A” to “E” are defined in Table 13 in the appendix [20].

### B. OIL QUALITY

The second important factor that influences PTHI is the oil quality factor (OQF), which depends on the dielectric strength, humidity, color, interfacial tension, acid amount of the insulating oil, and insulation dissipation factor (DF). The condition of the transformer can be monitored and evaluated

by measuring the insulation DF [20]. Insulation DF indicates the state of the insulation integrity of the winding and identifies the DF of the overall insulation, including the winding and bushing. The DF test is a routine test that measures the insulation capacitance and power factor of transformer insulation under 10 kV with 50 or 60 Hz [22]. Table 16 in the appendix (as in [20]) illustrates a ranking method of the transformer insulation DF. The score ( $S_i$ ) and weight factors ( $W_i$ ) of test parameter  $i$  in different transformer operating voltage ratings are listed in detail in Table 14 in the appendix, as in [20]. OQF can be calculated as follows:

$$OQF = \sum_{i=1}^6 \frac{S_i \times W_i}{W_i} \quad (2)$$

The transformer condition is identified based on OQF by using a code from “A” to “E” similar to DGAF.

### C. DEPOLARIZATION FACTOR

The degree of depolarization (DP) of the paper insulation can be measured via the furfural content in the insulating oil [23]. The furan test is recommended and should be performed periodically when the rate of change in  $CO$  and  $CO_2$  concentration increases or when the transformer age exceeds 25 years [22]. DP is an important factor to compute PTHI. However, when the oil has been replaced, the degradation of the cellulose paper cannot be identified based on the furan content. In this case, the age of the transformer can be used as a factor for health index computation [20], [23]. The range of furan content and the transformer age can be utilized to compute the DP factor (DPF) in accordance with Table 15 in the appendix, as in [20].

Table 1 shows the scoring system of the main representative factors for investigating the transformer health index factor (HIF). The HIF score varies from 4 to 0, which correspond to transformer criteria codes “A” to “E” respectively. On the basis of representative factors, the total PTHI can be computed as

$$PTHI = \frac{\sum_{i=1}^4 K_i \times HIF_i}{\sum_{i=1}^4 4K_i}, \quad (3)$$

where  $K_i$  is the weight factor for each transformer condition criterion [1], [20].  $K_i$  is identified in accordance with the importance of the test (DGA, OQ, and DP) in evaluating the PTHI state. The weighting factor of DGAF is 10 because its effect on evaluating the PTHI state is higher than that of OQF (8) and DPF (6). DGAF is the main factor that indicates faulty (poor) and non-faulty (good and fair) states. Utility engineers use the values of DGA features as indicators for faulty and non-faulty states of power transformers.

Table 2 defines three PTHI states and their corresponding threshold limits [24]. The “good” state represents the case where the PTHI value is greater than or equal to 85, and the “fair” state represents the case where PTHI ranges from 85 to 50. The threshold limits for the “poor” state are obtained when PTHI is lower than 50. The required actions corresponding to each state are defined in Table 2 [24].

TABLE 1. Health index scoring [20].

Transformer Condition Criteria	$K_i$	Transformer Condition	HIF
1 Dissolved gas analysis	10	A, B, C, D, E	4, 3, 2, 1, 0
2 Oil quality	8	A, B, C, D, E	4, 3, 2, 1, 0
3 Furan	6	A, B, C, D, E	4, 3, 2, 1, 0

TABLE 2. Threshold limit of transformer health state [24].

Health Index State	Threshold Limits	Required Maintenance
Good	$\geq 85$	No maintenance required
Fair	85 – 50	More diagnostic testing is required
Poor	$< 50$	The transformer must be out of service

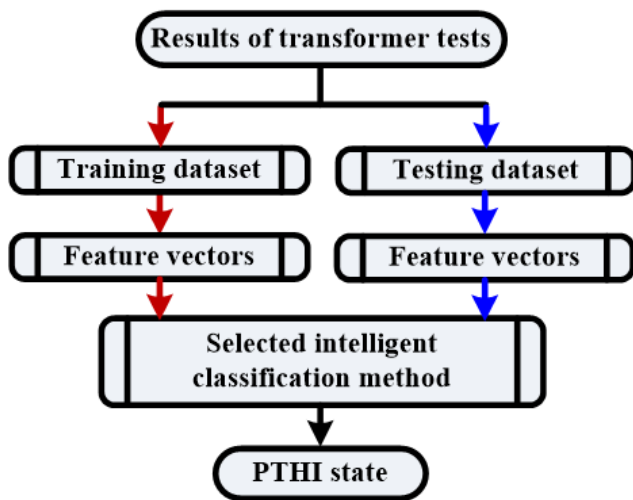


FIGURE 1. Strategy of the classification methodology.

III. METHODOLOGY AND SCENARIOS

A. FULL-FEATURE SCENARIO

PHTI is computed using AI classification methods (DT, SVM, KNN, and EN). Figure 1 presents the procedure used to compute the PHTI state in the proposed model depending on the full features of the three test results (DGA, OQ, and DP tests). The dataset containing test results collected from a chemical company in Malaysia and an electric company in the Gulf Region is divided into two subsets. The first subset is used for training, and the second one is used for testing. To calculate the target output for each feature vector of the training and testing datasets, PHTI is computed using Eq. (3) and mapped to the corresponding transformer health state in accordance with Table 2. Then, the selected AI classification method is applied. Finally, the PHTI state is predicted and reported against the selected AI classification method. Figure 2 shows the procedure of the proposed model.

In the full-feature scenario, the PHTI state is predicted in two stages, as illustrated in Figure 2. In the first stage, the 14 input parameters are divided into three groups corresponding to the three factors: DGAF, OQF, and DP. Each group and its corresponding factor are used for training an

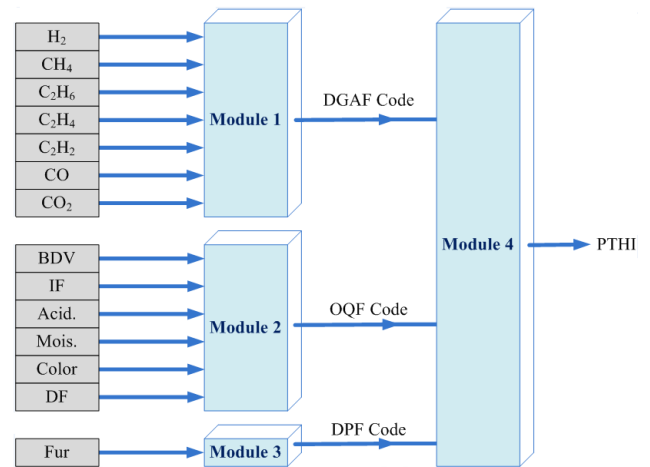


FIGURE 2. Proposed model scenari.

individual module. The first group consists of seven input parameters ( $H_2$ ,  $CH_4$ ,  $C_2H_6$ ,  $C_2H_4$ ,  $C_2H_2$ ,  $CO$ , and  $CO_2$ ) that are fed into the first classification module with the corresponding calculated DGAF. The second group consists of six inputs parameters (moisture, breakdown voltage, color, acidity, IF, and DF) that are supplied to the second classification module with the corresponding calculated OQF. The furan concentration is applied to the third classification module with the corresponding calculated DPF. The selected classification method is used to train the three classification modules individually. In the second stage, a fourth module is trained with a feature vector composed of DGAF, OQF, and DPF as the input and the PHTI state obtained from Table 2 as the output. After completing the training process, the testing dataset is used to measure the accuracy of the developed classification methods.

B. REDUCED-FEATURE SCENARIO

Feature selection (FS) is a process of considering relevant features and discarding irrelevant features that do not influence the output of a problem with minimum performance degradation. Given that PHTI evaluation considers all features, it requires considerable time for experiments, intensive labor, and high costs. FS is divided into three categories, namely, filters, wrappers, and embedded methods. In this work, the minimum redundancy maximum relevance (MRMR) approaches categorized as filter methods were used to determine the importance features. The feature importance scores were estimated using MRMR, as performed in [25]–[27].

MRMR ranks the importance features of the classification problem. The main objective of MRMR is to minimize the redundancy of a feature set and maximize the relevance of a feature set to classification output  $c$  by using mutual information  $I$ , as follows [18], [19]:

$$A_s = \frac{1}{|S|} \sum_{x \in S} I(x, c), \tag{4}$$

$$B_s = \frac{1}{|S|^2} \sum_{x, z \in S} I(x, z), \tag{5}$$

The algorithms of the four modules are introduced as follows:

---

#### Module 1 Algorithm

---

1. Insert the DGA features to the classification method ( $H_2$ ,  $CH_4$ ,  $C_2H_6$ ,  $C_2H_4$ ,  $C_2H_2$ ,  $CO$ , and  $CO_2$ ).
2. Calculate DGAF by using Eq. (1) on the basis of the DGA feature scores and weights in Table 12.
3. Evaluate the DGAF code according to Table 13.
4. Insert the DGA features (for the training dataset) as the input and the corresponding DGAF code as the output to each AI classification method for the training process to build classification module 1.
5. Use the constructed classification module 1 to predict the DGAF code of the testing dataset samples.

---

#### Module 2 Algorithm

---

1. Insert the OQ features to the classification method (BVD, IF, acidity, moisture, color, and DF).
2. Calculate OQF by using Eq. (2) on the basis of the OQ feature scores and weights in Table 12.
3. Evaluate the OQF code according to Table 14.
4. Insert the OQ features (for the training dataset) as the input and the corresponding OQF code as the output to each AI classification method for the training process to build classification module 2.
5. Use the constructed classification module 2 to predict the OQF code of the testing dataset samples.

---

#### Module 3 Algorithm

---

1. Insert the furan feature to the classification method.
2. Evaluate the DPF code according to Table 15.
3. Insert the furan features (for the training dataset) as the input and the corresponding DPF code as the output to each AI classification method for the training process to build classification module 3.
4. Use the constructed classification module 3 to predict the DPF code of the testing dataset samples.

---

#### Module 4 Algorithm

---

1. Insert the DGAF, OQF, and DPF codes.
  2. Evaluate the PTHI state according to Tables 1 and 2.
  3. Insert the three output codes of Modules 1, 2, and 3 (for the training dataset) as the input and the corresponding PTHI code as the output to each AI classification method for the training process to build classification module 4.
  4. Use the constructed classification module 4 to predict the PTHI state of the testing dataset samples.
- 

where  $S$  is the optimal set of features that maximize relevance  $A_s$  and minimize redundancy  $B_s$ ,  $|S|$  is the number of features in  $S$ ,  $x$  and  $z$  are two features that belong to set  $S$ ,  $c$  denotes the output classes, and  $I$  is the mutual information

that can be expressed as follows [18], [20]:

$$I = \sum_{b \in B} \sum_{a \in A} p(a, b) \times \log \frac{p(a, b)}{p(a) \times p(b)}, \quad (6)$$

where  $A$  and  $B$  are two sets;  $a$  and  $b$  are features in subsets  $A$  and  $B$ , respectively;  $p(a, b)$  is the probability of  $a$  and  $b$ ; and  $p(a)$  and  $p(b)$  are the probability of  $a$  and  $b$ , respectively.

Calculating an optimal set  $S$  requires considering all  $2^w$  combinations, where  $w$  is the entire feature site. The mutual information quotient (MIQ) is used to facilitate the estimation of MRMR, as follows:

$$MIQ_x = \frac{A_x}{B_x}, \quad (7)$$

where  $A_x$  and  $B_x$  are the relevance and redundancy of a feature, respectively, and they can be evaluated as follows:

$$A_x = I(x, c), \quad (8)$$

$$B_x = \frac{1}{|S|} \sum_{z \in S} I(x, z) \quad (9)$$

## IV. RESULTS AND DISCUSSIONS

### A. FULL-FEATURE RESULTS

Each of the four selected classification methods (DT, SVM, KNN, and EN) was implemented using 2019b MATLAB/software. The methods were optimized and implemented in accordance with the MATLAB classification learner application toolbox. A brief description of the four intelligent classification methods is presented in the appendix.

The cross-fold validation approach with five folds was used to train the four AI-optimized classification methods. The training and testing processes were carried out on the 1,361 dataset samples. The dataset samples were collected from two countries. The first set (631 samples) was collected from the transmission regions of Malaysia Electricity Company. The dataset includes samples of power transformers of 220 kV at the transmission stage. The second set (730 samples) was collected from the Gulf Region's medium-voltage region (66 kV). Each set was divided randomly into 65% and 35% training and testing data samples, respectively. The training samples (885) of the two sets were collected for training, and the testing samples (476) of the two sets were collected for testing. Table 3 shows the distribution of the samples.

The accuracy of each AI classification method for training, testing, and overall stages was calculated as:

$$\%Accuracy = \frac{Nc}{Nt} \times 100, \quad (10)$$

where  $Nc$  is the number of correctly predicted state samples from each AI-optimized classification method. The correctly predicted state sample was identified by comparing the result of the AI classification method with the result of the PTHI of Eq. (3) for each sample.  $Nt$  is the total number of state samples.

The classification learner toolbox application in MATLAB 2019b was used to build the four classification methods

**TABLE 3.** Data distribution of the samples used in training and testing stages.

State	Good	Fair	Poor
Training	662	204	19
Testing	341	120	15
All	1003	324	34

**TABLE 4.** Main optimization parameters of training DT, SVM, KNN, and EN methods.

Optimizer	Bayesian optimization
Acquisition function	Probability of improvement
Maximum number of iterations	30

(DT, SVM, KNN, and EN). For example, the DT method types (fine tree, medium tree, and coarse tree) were used, and the issue was determining which one develops the best predictor. Hence, the hyperparameter optimization in the classification learner toolbox was used to select the suitable classification method and the corresponding parameters of the two suggested scenarios.

The minimum estimated error was determined as follows:

$$\text{minimum error} = \min\left(1 - \frac{N_{ci}}{N_t}\right), \quad (11)$$

where  $N_{ci}$  is the number of correctly predicted states of classification type  $i$ .

AI classification algorithms (DT, SVM, KNN, and EN) require an optimization technique to determine the best optimization parameters for each AI classification method. The optimal parameters of DT methods are the maximum number of splits and split criterion, and that of SVM are multiclass method, box constraint level, kernel function used, and standardized data. Several optimization approaches, such as Bayesian optimization (BO), grid search, and random search, are frequently used with classification methods. The BO approach is effective for optimization problems and can be used for most machine learning techniques for hyperparameter optimization [28]. BO iteratively explores the hyperparameter space, where a probabilistic model of estimation is built based on prior calculation. Then, the probabilistic model is utilized to evaluate the optimal parameters by using the probability values of its position and selecting the parameters related to the highest probability [29]. The detailed model of BO used for estimating the optimal hyperparameters of machine learning techniques was introduced in [29], [30].

The training process was conducted through the following steps. First, the main optimization parameters were selected before the training process of each classification method (DT, SVM, KNN, and EN). Second, the optimization training process was carried out for each classification method. Third, an optimized model was generated for each classification method. Lastly, Steps 2 and 3 were repeated for each module of the proposed model shown in Figure 2. The main optimization parameters of training DT, SVM, KNN, and EN methods are presented in Table 4.

**TABLE 5.** Classification method optimization and range parameters for the full-feature scenario.

	DT	SVM	KNN	EN
Module 1	Maximum number of splits: 81	Multiclass method: one vs. one	No. of neighbors: 1	EN Method: bag
	Split criterion: max. deviance reduction	Box constraint level: 75.0521	Distance metric: city block	No. of learners: 492
	Minimum error: 0.07	Kernel function: linear	Distance weight: squared inverse	Maximum number of splits: 439
	Accuracy: 93%	Standardized data: true	Standardized data: true	No. of predictors to samples: 7
Module 2	Maximum number of splits: 59	Multiclass method: one vs. one	No. of neighbors: 5	EN Method: AdaBoost
	Split criterion: towing rule	Box constraint level: 4.2248	Distance metric: city Block	No. of Learners: 494
	Minimum error: 0.078	Kernel function: cubic	Distance weight: squared inverse	Learning rate: 0.43465
	Accuracy: 92.2%	Standardized data: true	Standardized data: true	Maximum number of splits: 42
Module 3	Maximum number of splits: 5	Multiclass method: one vs. one	No. of neighbors: 7	EN Method: bag
	Split criterion: max. deviance reduction	Box constraint level: 998.31	Distance metric: Chebyshev	No. of Learners: 65
	Minimum error: 0.002	Kernel function: quadratic	Distance weight: squared inverse	Maximum number of splits: 665
	Accuracy: 99.8%	Standardized data: false	Standardized data: false	No. of predictor to samples: 1
Module 4	Maximum number of splits: 48	Multiclass method: one vs. all	No. of neighbors: 1	EN Method: AdaBoost
	Split criterion: towing rule	Box constraint level: 194.323	Distance metric: Chebyshev	No. of Learners: 254
	Minimum error: 0.0033	Kernel function: quadratic	Distance weight: squared inverse	Learning Rate: 0.04795
	Accuracy: 99.7%	Standardized data: false	Standardized data: true	Max. No. of Splits: 1
	Minimum error: 0.00212	Minimum error: 0.002	Minimum error: 0.0754	Minimum error: 0.04978
	Accuracy: 99.8%	Accuracy: 99.8%	Accuracy: 92.5%	Accuracy: 95%
	Minimum error: 0.00212	Minimum error: 0.002	Minimum error: 0.0754	Minimum error: 0.04978
	Accuracy: 99.8%	Accuracy: 99.8%	Accuracy: 92.5%	Accuracy: 95%
	Minimum error: 0.00212	Minimum error: 0.002	Minimum error: 0.0754	Minimum error: 0.04978
	Accuracy: 99.8%	Accuracy: 99.8%	Accuracy: 92.5%	Accuracy: 95%

The DT, SVM, KNN, and EN methods were optimized for the four modules of the full-feature approach on the basis of the 885-sample training dataset. The optimal parameters of the four classification methods for the four modules with the full-feature scenario are given as in Table 5.

Figure 3 shows the minimum errors during the optimization process of the training stage for DT, SVM, KNN, and EN method corresponding to the four modules introduced in Figure 2 in the full-feature scenario. SVM had a minimum error for module 1, EN had a minimum error for module 2, and the four methods had an equal minimum error for modules 3 and 4.

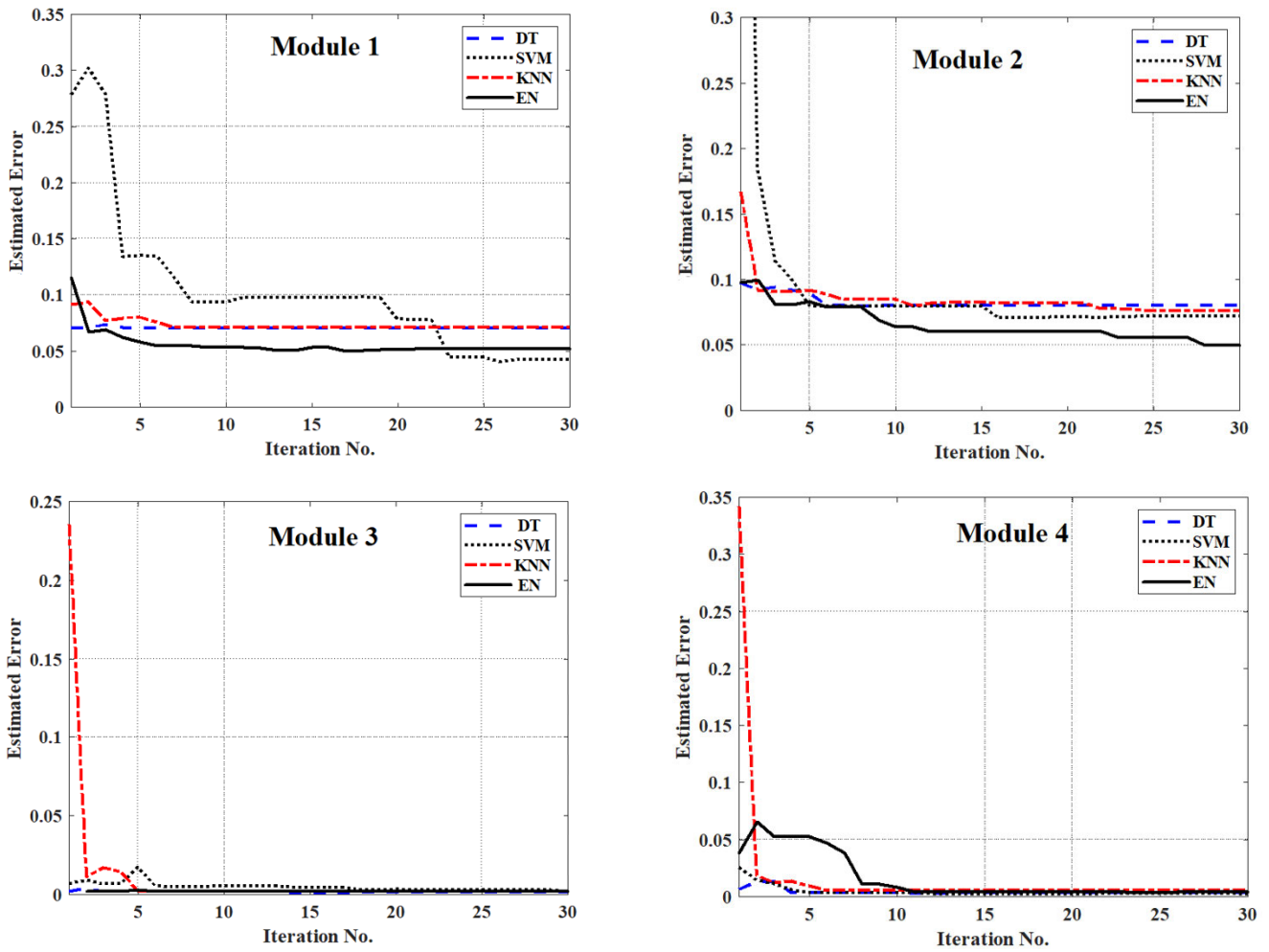


FIGURE 3. Comparisons of the minimum errors of DT, SVM, KNN, and EN methods for different modules of the two-stage model classificatio.

True class				Efficiency
	1	2	3	
1	653	9	0	98.6%
2	11	190	3	93.1%
3	0	11	8	42.1%
	1	2	3	96.2%

FIGURE 4. Confusion matrix for the training process of the SVM method.

Figure 4 presents the confusion matrix for the training process of SVM with the full-feature scenario. The state classes of “good,” “fair,” and “poor” are denoted by 1, 2, and 3, respectively. The results showed that the classification process for the “good” state had the highest classification accuracy ( $653/662 = 98.6\%$ ), whereas the “poor” state had the lowest classification accuracy ( $8/19 = 42.1\%$ ). The overall accuracy of predicting the PTHI state was ( $851/885 = 96.2\%$ ).

Figure 5 illustrates the overall accuracy for training and testing stages in the full-feature scenario. The overall

accuracy of predicting the PTHI states for DT, SVM, KNN, and EN methods corresponding to the training and testing process was (95.3%, 96.2%, 95.4%, and 96%) and (93.9%, 95.2%, 95.2%, and 95.6%), respectively. These results indicate good prediction accuracy during the training and testing stages.

Table 6 presents the number of total samples correctly predicted by each of the four methods against each PTHI state. The detection accuracy of SVM and EN was better than that of the other methods for the “good” state. KNN had the highest accuracy for the “fair” state, and SVM method had the highest accuracy for the “poor” state. The EN method achieved superior classification accuracy (95.9%) in the full-feature scenario.

**B. REDUCED-FEATURE RESULTS**

This section introduces the procedure of evaluating the rank of important redundancy features for DGAF and OQF by using the MRMR approach presented in Section III (B). The sufficient features for each factor were determined using the

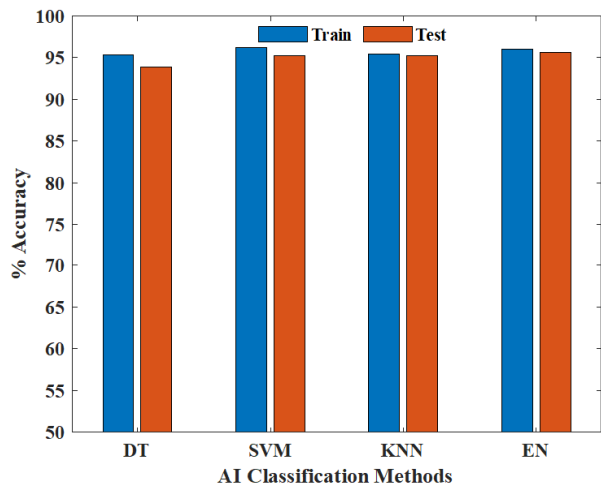


FIGURE 5. Training and testing accuracy of the four methods for the full-feature scenario.

TABLE 6. Overall accuracy of the classification methods for the full-feature model.

HI	Good	Fair	Poor	% Accuracy
No. of samples	1003	324	34	
DT	987	280	23	94.8
SVM	989	282	26	95.3
KNN	984	298	22	95.8
EN	989	295	21	95.9

principal component analysis (PCA) facility in the MATLAB classification learner toolbox after the training process. The sufficient features for each factor were those that had a variance greater than 95%. The training process was carried out with PCA for DGAF and OQF. The minimum feature for DGAF was only one feature with explained variance ratios in the order of 98.9%, 1%, and other features <0.1%. Meanwhile, the number of features required for OQF was at least three, with explained variance ratios in the order of 72.4%, 15.7%, 9.3%, 2.5%, 0.2%, and 0%. The importance features for DGAF and OQF were selected based on the highest feature scores obtained from the MRMR method. The importance feature scores of DGAF and OQF are presented in Figures (6) and (7), respectively. The results showed that CO<sub>2</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>6</sub>, and C<sub>2</sub>H<sub>4</sub> (Figure 6) were the most important factors of DGAF, whereas color, BDV, IF, and moisture (Figure 7) were the most important features of OQF. The training process was repeated with the selected features for DGAF and OQF. Although the variance of DGAF was only one (CO<sub>2</sub>), it was not enough to represent DGAF because CO<sub>2</sub> and CO illustrate the insulation paper state, and the other features illustrate the faulty state of transformer oil. Hence, one of them was selected with CO<sub>2</sub> as an input of module 1 to predict the DGAF state.

After selecting the required features for each module (modules 1 and 2), the same procedure for the full-feature scenario was carried out for the reduced-feature scenario. Different features were selected for DGAF and OQF and applied to

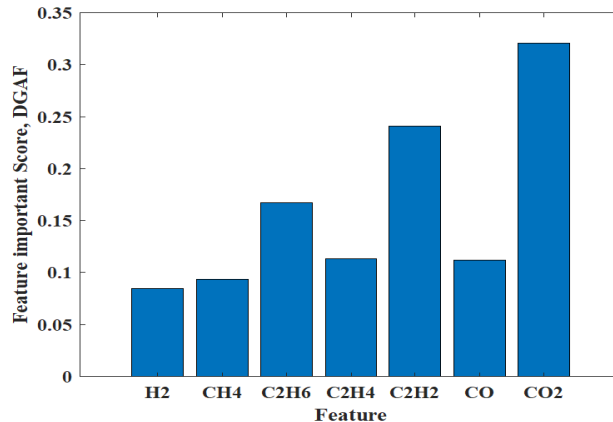


FIGURE 6. Feature importance of DGAF.

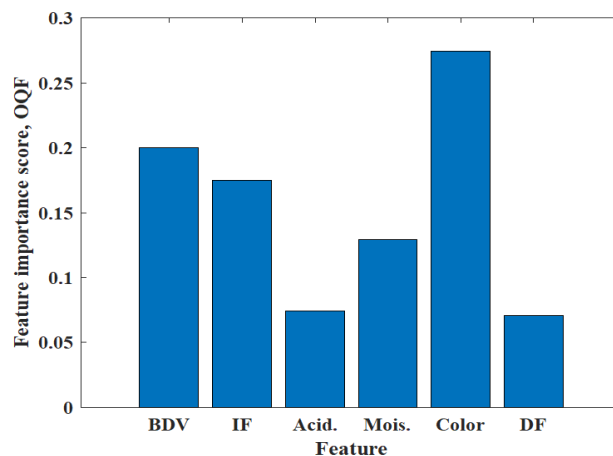


FIGURE 7. Feature importance scores of OQF.

modules 1 and 2, respectively, whereas furan was inserted into module 3. The output codes of modules 1, 2, and 3 were applied to module 4 to obtain the final PTHI state. The obtained model for each case was used at the testing stage to predict the PTHI by using the testing data samples. Different numbers of features were considered for DGAF and OQF. The overall accuracy in each case is reported in Table 7. In Table 7, the numbers 7, 5, 3, and 4 denote CO<sub>2</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>6</sub>, and C<sub>2</sub>H<sub>4</sub> on the DGAF column, respectively, and the numbers 12, 8, 9, and 11 denote color, BDV, IF, and moisture on the OQF column, respectively. The highlighted case has the best prediction accuracy of 93.5%, 94.3%, 94.9, and 95.8% for DT, SVM, KNN, and EN methods, respectively. The selected features for DGAF that satisfy the minimum requirements of PCA and achieve the highest prediction accuracy for the PTHI state were CO<sub>2</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>6</sub>, and C<sub>2</sub>H<sub>4</sub>; the corresponding features for OQF were color, BDV, IF, and moisture.

Figure 8 shows the confusion matrix for the training process of the EN method in the reduced-feature scenario. The classification process for the “good” state had the highest classification accuracy (655/662 = 98.9%), whereas the “poor” state had the lowest classification accuracy



**TABLE 7.** Accuracy of the classification methods for the selected features of DGAF and OQF.

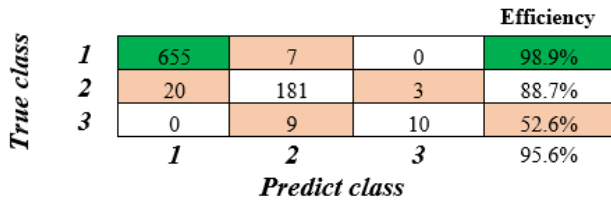
Selected Features		% Accuracy			
DGAF	OQF	DT	SVM	KNN	EN
7, 5	12, 8, 9	90.1	89.7	88.7	90.2
7, 5, 3	12, 8, 9	92.9	93.2	91.8	93.3
7, 5, 3, 4	12, 8, 9	93	93.4	92.4	93.8
7, 5	12, 8, 9, 11	90.6	90.7	91.4	92.4
7, 5, 3	12, 8, 9, 11	93.5	93.9	94.6	95.3
7, 5, 3, 4	12, 8, 9, 11	93.5	94.3	94.9	95.8

**TABLE 8.** Overall accuracy of the classification methods for the reduced-feature model.

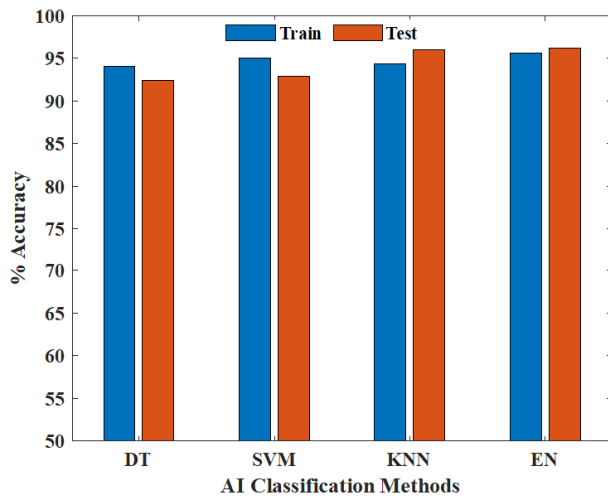
HI	Good	Fair	Poor	% Accuracy
No. of samples	1003	324	34	
DT	975	281	17	93.5
SVM	984	279	20	94.3
KNN	986	285	21	94.9
EN	990	295	19	95.8

**TABLE 9.** Comparison of full-feature and reduced-feature scenarios.

	Full-feature Scenario				Reduced-feature Scenario			
	DT	SVM	KNN	EN	DT	SVM	KNN	EN
Good	98.4	98.6	98.1	98.6	97.2	98.1	98.3	98.7
Fair	86.4	87.0	92.0	91.0	86.7	86.1	88.0	91.0
Poor	67.6	76.5	64.7	61.8	50	58.8	61.8	55.9
ALL	94.8	95.3	95.8	95.9	93.5	94.3	94.9	95.8



**FIGURE 8.** Confusion matrix for the training process of the EN method.



**FIGURE 9.** Training and testing accuracy of the four methods for the reduced-feature scenario.

(10/19 = 52.6%). The overall accuracy of detecting the PTHI state was (846/885 = 95.6%).

Figure 9 presents the overall accuracy for training and testing stages in the reduced-feature scenario. The overall accuracy of predicting the PTHI states for DT, SVM, KNN, and EN methods during the training and testing processes was (94.1%, 95%, 94.4% and 95.6%) and (92.4%, 92.9%, 96% and 96.2%), respectively. The results reveal a good prediction accuracy during the training and testing stages, especially with the EN method.

Table 8 presents the correct number of samples for the classification methods in comparison with each PTHI state. The results indicated that the detection accuracy of the EN method was better than that of the other methods for the “good” and “fair” states. KNN exhibited the highest accuracy for the “poor” state. The EN method achieved the highest classification accuracy (95.8%) in the reduced-feature scenario. Only a slight difference in accuracy was observed between the

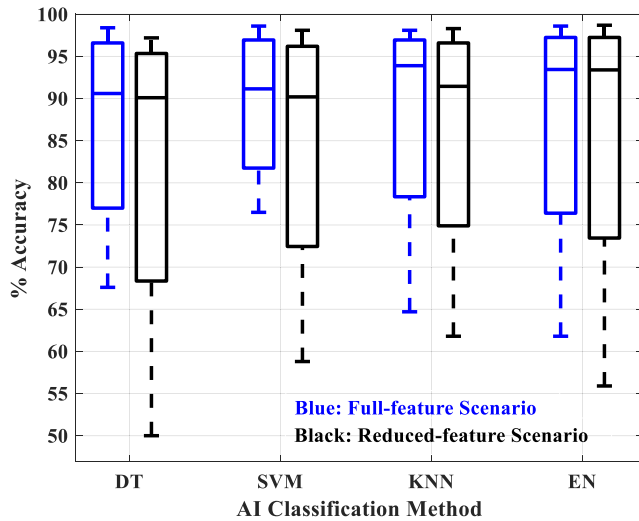
full- and reduced-feature scenarios. Thus, the reduced-feature scenario was considered to decrease the effort and time of the computation process and the experimental cost.

**C. COMPARISON OF FULL AND REDUCED FEATURES**

Comparisons of the full- and reduced-feature scenarios were implemented to investigate the difference in their prediction accuracy. All features were used for training and testing the full-feature scenario, but only nine features (CO<sub>2</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub>, color, BDV, IF, moisture, and furan) were adopted for the reduced-feature scenario. Applying the reduced features resulted in a short training time (for training modules 1 and 2), less time and effort for measuring the features in the laboratory, and fewer oil samples required for measuring the features.

Table 9 presents a detailed comparison of the full- and reduced-feature scenarios for “good,” “fair,” “poor,” and overall PTHI states. The prediction accuracies of each method in the two scenarios showed good agreement. The prediction accuracy of “good,” “fair,” “poor,” and overall PTHI states for the EN method was (98.6%, 91%, 61.8%, 95.9%) and (98.7%, 91%, 55.9%, 95.8%) for the full- and reduced-feature scenarios, respectively. The results show a slight difference between the full- and reduced-feature scenarios, especially when the EN method was involved.

Figure 10 presents a comparison boxplot of the classification methods (DT, SVM, KNN, and EN) for the full- and reduced-feature scenarios. It presents the minimum, maximum, and standard deviation of each method in the two scenarios. A slight difference in the maximum and mean values was observed between the two scenarios for the four classification methods, especially the EN method. The KNN and EN methods had small differences in their minimum values under the full- and reduced-feature scenarios, whereas DT and SVM showed a large difference. The EN method is thus the preferred method for the reduced-feature scenario; it requires only nine features, and its accuracy is close to that of the full-feature accuracy.



**FIGURE 10.** Accuracy of the full- and reduced-feature scenarios with different states and overall PTHI states.

**TABLE 10.** Comparison of the proposed methods and the methods in [24] by using two dataset samples.

	Proposed models			Models in [24]				
	DT	SVM	KNN	EN	ANN	MLR	J48	RF
*	96	96.4	96.7	96.4	94.9	95.5	95.6	96.6
**	94.7	95.2	95.5	95.6	95.1	95.3	95.3	96.6

#### D. PROPOSED MODEL VALIDATION

##### 1) PROPOSED MODEL AGAINST THE RESULTS IN [24]

The effectiveness of the proposed model was checked by using two approaches. The first one involved comparisons with the results published in [24], which used dataset samples of System 2 (Gulf Region). The second approach was an uncertainty check.

Table 10 presents the comparison between the results obtained by the proposed methods and those presented in [24] for the full- and reduced-feature scenarios. The results in [24] are based on the Gulf Region dataset and on the methods NN, MLR, J48, and RF (the nine methods in [24] exhibited high accuracy). The proposed methods demonstrated higher accuracy compared with the methods presented in [24] for the full-feature (\*) scenario and acceptable accuracy for the reduced-feature (\*\*) scenario. For the full-feature scenario, the highest accuracy achieved by the proposed model was 96.7% for the KNN classification method, whereas the highest accuracy achieved by the method in [24] was 96.6% for RF classification. In the reduced-feature scenario, the results of the proposed methods were better than those in [24], except for the RF method that has 96.6%. The best among the proposed methods (the EN method) achieved 95.6%.

##### 2) PROPOSED MODEL AGAINST UNCERTAINTY

The process of preparing the datasets of the power transformers was carried out offline in three main steps. The first step was extracting samples from the power transformers. The second step was extracting the gases from the transformer

oil, and the third step was predicting the PTHI state. The oil samples were extracted using special syringes. The extracted samples were stored and transferred to laboratories. Many factors, such as storage time and temperature, affect gas concentration. The extraction process of gases is often conducted using several techniques. Air bubble is the most critical factor that affects gas concentrations [31]. Air bubbles reduce dissolved gases because of the diffusion of gases into the air bubbles, thus leaving the oil [32]. Hence, uncertainty during measurements affects PTHI state prediction. The uncertainty during measurements must therefore be considered by the AI classification techniques used for PTHI state prediction. An uncertainty level of  $\pm 14\%$  is caused by the sample's storage and temperature effects, and an uncertainty level of up to  $\pm 5\%$  is caused by measurement accuracy [33]. In this work, an uncertainty level of up to  $\pm 20\%$  was considered.

An uncertainty evaluation was carried out by applying percentage noise on the input data. Percentage noise can be calculated as follows [21]:

$$N_l = \{100 - m + (2m \times R_L) / 100, \quad (12)$$

where  $N_l$  is the noise vector,  $l$  represents the number of applied samples in the uncertainty evaluation (476 testing samples),  $m$  is the required noise percentage level (5%, 10%, 15%, or 20%), and  $R_L$  is a random vector with a length of 14 and varies from 0 to 1.

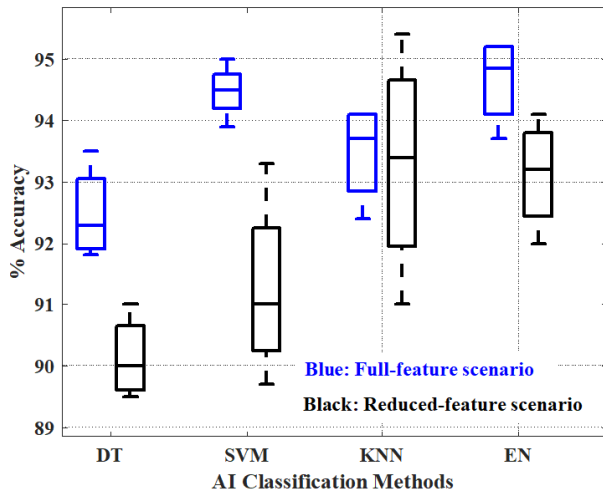
Noise vector  $N_l$  is a vector that has numbers varying from 0.95 to 1.05, 0.9 to 1.1, 0.85 to 1.15, or 0.8 to 1.2 when  $m$  has values of 5%, 10%, 15%, or 20%, respectively. The original input feature vector and  $N_l$  were produced element by element to obtain new data (data with uncertain noise). These new data were inserted into the proposed model by using different AI classification methods for full- and reduced-feature scenarios. The test dataset was used to measure the performance of the four AI methods during the uncertainty evaluation.

Table 11 presents the results of full- and reduced-feature scenarios against uncertainty data (476 testing dataset samples) for the four AI classification methods. Tables 6 and 10 show the results of the four AI methods with the testing data for the full- and reduced-feature scenarios, respectively. The four AI methods exhibited good deprecation performance against uncertainty, and the EN method showed the best performance in the full-feature scenario. Its accuracy decreased from 95.6% to 95.2%, 95.2%, 93.7%, and 94.5% with uncertainty noise of  $\pm 5\%$ ,  $\pm 10\%$ ,  $\pm 15\%$ , and  $\pm 20\%$ , respectively. The maximum percentage error was obtained with the KNN method ( $(95.2 - 92.4) / 95.2 \times 100 = 2.8\%$ ) at 15% uncertainty noise.

Furthermore, the four AI classification methods exhibited good performance against uncertainty in the reduced-feature scenario. The EN method had the best performance. Its accuracy decreased from 96.2% to 94.1%, 93.5%, 92.9%, and 92.0% with uncertainty noise of  $\pm 5\%$ ,  $\pm 10\%$ ,  $\pm 15\%$ , and  $\pm 20\%$ , respectively. The maximum percentage error was obtained with the KNN method ( $(96.2 - 91) / 96.2 \times 100 = 5.4\%$ )

**TABLE 11.** Accuracy of the full- and reduced-feature scenarios against uncertainty data.

	Full-feature Scenario				Reduced-feature Scenario			
	% Uncertainty Noise				% Uncertainty Noise			
	5%	10%	15%	20%	5%	10%	15%	20%
DT	93.5	92.0	92.6	91.8	91.0	90.3	89.5	89.7
SVM	95.0	94.5	93.9	94.5	93.3	91.2	90.8	89.7
KNN	94.1	94.1	92.4	93.3	95.4	93.9	92.9	91.0
EN	95.2	95.2	93.7	94.5	94.1	93.5	92.9	92.0



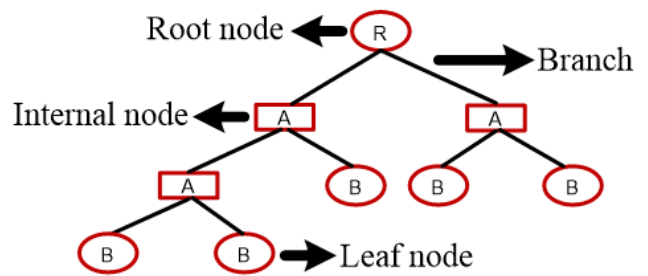
**FIGURE 11.** Accuracy of the full- and reduced-feature scenarios with uncertainty noise in the four AI methods.

at  $\pm 20\%$  uncertainty noise. The performance of the four AI classification methods confirmed that the uncertainty noise effect was limited in the proposed scenarios.

Figure 11 presents the change in the accuracy of the four AI classification methods against uncertainty noise of 0% to  $\pm 20\%$  under the two scenarios. The degradation in the accuracy of the four methods was limited in the two scenarios against an uncertainty level of up to  $\pm 20\%$ . EN had the best performance against uncertainty in the full- and reduced-feature scenarios.

**V. CONCLUSION**

In this work, full- and reduced-feature scenarios were proposed to predict the PTHI state by using four intelligent classification methods. The full-feature scenario included four classification modules for predicting the PTHI state. The four proposed AI classification methods were DT, SVM, KNN, and EN. The overall prediction accuracy in the full-feature scenario was 94.8%, 95.3%, 95.8%, and 95.9% for DT, SVM, KNN, and EN, respectively. These results indicate that the recommended method for PTHI state prediction is the EN method for the full-feature scenario. Furthermore, the results of the proposed model are superior to those presented in [24]. The MRMR feature reduction method was also used to reduce the input features of the full-feature scenario from 7 to 4 and from 6 to 4 for DGAF and OQF, respectively. CO<sub>2</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub>, color, BDV, IF, moisture, and furan were the final selected features for the reduced-feature scenario.



**FIGURE 12.** Decision tree construction.

A slight difference in accuracy was observed between the full- and reduced-feature scenarios. The highest overall accuracy in the full- and reduced-feature scenarios was 95.9% and 95.8%, respectively; these values were obtained with the EN method. Thus, using the reduced-feature scenario is more reasonable than using the full-feature scenario. An uncertainty approach was applied to investigate the robustness of the proposed model. The results indicated that the maximum errors of classification accuracy were 2.8% at  $\pm 15\%$  uncertainty noise and 5.4% at  $\pm 20\%$  uncertainty noise for the full- and reduced-feature scenarios, respectively; these values were obtained with the KNN classification method.

**APPENDIX A  
DECISION TREE (DT)**

The DT method is a classification approach. The first step of the decision tree is the root node [34]. In this step, data are classified into different branches of the tree to obtain the specified classes. The second step of the decision tree is to identify the class condition, and the classification forms a sub-tree. The same sorting process is repeated until all data in a branch are of the same type.

Figure 12 shows the construction of a decision tree that consists of the root node. Two or more paths branch from the root node to represent an outcome of the test on the training dataset. These paths end with an internal node, which denotes a test on an attribute. At the end, the leaf node holds a numeric prediction that considers numeric class. The DT method has different types, such as fine, medium, and coarse.

**APPENDIX B  
ENSEMBLE METHOD (EN)**

This classifier method is used to reduce variance in prediction models. The EN method has different types, such as bag, AdaBoost, RUS boost, logit boost, and gentle boost. The most important EN method is the bag trees method. It accomplishes the task by extracting new generation datasets using combinations with repetitions from the original dataset for training [35], [36]. It fits the base classifiers of the random generation subsets from the original dataset. Hence, it aggregates the individual predictions of these subsets to develop the final decision. This type of classification is based on a tree structure consisting of a master node called the “root” and a group of internal and final nodes called “terminals.” The models generated by this classification are characterized

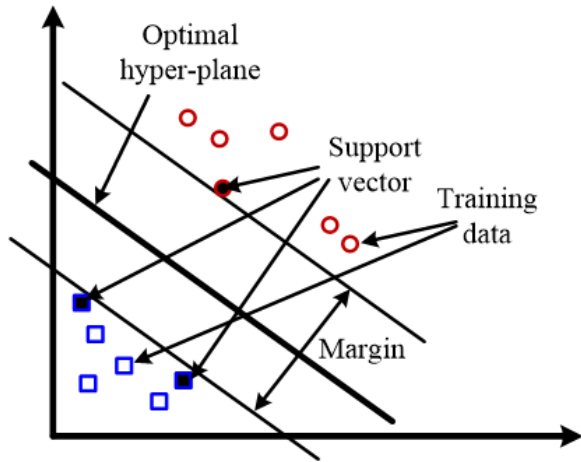


FIGURE 13. Separation between two classes by SVM.

by high accuracy and speed in model construction. It can also be applied to multiclass data and can be interpreted and understood by decision tree analysis. The proposed model was validated against uncertainty noise of up to ±20%.

APPENDIX C

SUPPORT VECTOR MACHINE (SVM)

SVM is one of the most common machine learning methods, and it is used as a classifier for data analysis [37], [38]. SVM finds the optimal separating hyperplane to maximize the margin between data samples, as indicated in Figure 13. The filled circles denote the support vectors, and the unfilled circles refer to the training data.

SVM regression evaluates a function on the basis of input and output data, as follows [38]:

$$f(x) = w \cdot x + b = \sum_{k=1}^m w_k \cdot x_k + b = 0, \quad (A1)$$

where  $w$  refers to the weight factor and  $b$  is the bias term.

They are used to identify the location of the hyperplane that satisfies certain constraints, as follows [38]:

$$\begin{cases} y_k (w \cdot x_k + b) \geq 1, k = 1, 2, \dots, m \\ \min(\frac{1}{2} \|w\|^2) \end{cases} \quad (A2)$$

Therefore, SVM depends on the features of each collected sample. SVM constructs a hyperplane to segregate the samples of different classes. The hyperplane is built according to the training datasets, and it is used as a classifier for a new sample to obtain the actual class of each tested sample. A popular function used to identify the hyperplane is the kernel function. The kernel function has different types, such as Gaussian, linear, quadratic, and cubic [39].

APPENDIX D

K-NEAREST NEIGHBOR (KNN) CLASSIFIER

KNN is a machine learning classification method. KNN stores all available data on certain states and classifies the new sample on the basis of a similarity measure by measuring the

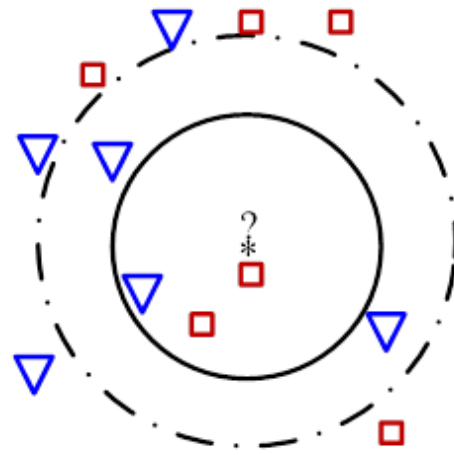


FIGURE 14. Influence of a neighbor's number on the classification [40].

TABLE 12. Scoring and weight factors for gas levels [ppm].

Gas	$S_i$						$W_i$
	1	2	3	4	5	6	
$H_2$	$\leq 100$	100–200	200–300	300–500	500–700	$>700$	2
$CH_4$	$\leq 75$	75–125	125–200	200–400	400–600	$>600$	3
$C_2H_6$	$\leq 65$	65–80	80–100	100–120	120–150	$>150$	3
$C_2H_4$	$\leq 50$	50–80	80–100	100–150	150–200	$>200$	3
$C_2H_2$	$\leq 3$	3–7	7–35	35–50	50–80	$>80$	5
$CO$	$\leq 350$	350–700	700–900	900–1,100	1,100–1,400	$>1400$	1
$CO_2$	$\leq 2500$	$\leq 3000$	$\leq 4000$	$\leq 5000$	$\leq 7000$	$>7000$	1

TABLE 13. Transformer rating based on DGAF.

Rating Code	Condition	Description
A	Good	$DGAF < 1.2$
B	Acceptable	$1.2 \leq DGAF < 1.5$
C	Needs caution	$1.5 \leq DGAF < 2$
D	Poor	$2 \leq DGAF < 3$
E	Very poor	$DGAF \geq 3$

distance functions. KNN operation theory determines the distances between a query and all samples of the datasets. Then, by selecting the specified number of samples ( $k$ ) close to the query, KNN votes on the most frequent label or averages the labels. Figure 14 shows the operation of KNN.

Several classes can be represented by different shapes, such as squares and triangles. A test sample is expressed using a star to identify the class that it belongs to (class 1 [square] or class 2 [triangle]). The label of  $k$  determines the class. For example,  $k = 3$  expresses class 1 because only two squares and only one triangle exist. By contrast,  $k = 5$  refers to class 2 because the number of samples is five, which consists of three triangles and two squares [40]. The different metric distance methods used for KN are Euclidean, city block, Chebyshev, Minkowski (cubic), Mahalanobis, cosine, Spearman correlation, Hamming, and Jaccard.

TABLE 14. Grading method for oil quality test parameters.

	U<69 kV	69<U<230 kV	U>230 kV	$S_i$	$W_i$
Dielectric strength kV (2 mm gap)	≥45	≥52	≥60	1	3
	35–45	45–52	50–60	2	
	30–35	35–45	40–50	3	
IFT dyne/cm	≤30	≤35	≤40	4	2
	≥25	≥30	≥32	1	
	20–25	23–30	25–32	2	
	15–20	18–23	20–25	3	
Acidity	≤15	≤18	≤20	4	1
	≤0.05	≤0.04	≤0.03	1	
	0.05–0.1	0.04–1	0.03–0.07	2	
	0.1–0.2	0.1–0.15	0.07–0.1	3	
Moisture (ppm)	≥0.2	≥0.15	≥0.1	4	4
	≤20			1	
	20–30			2	
	30–40			3	
Color	≥40			4	2
	≤1.5			1	
	1.5–2			2	
	2–2.5			3	
	≥2.5			4	

TABLE 15. Furan test rating code or age rating when testing is not available.

Rating code	Furan (ppm)	Age year
A	0–0.1	<20
B	0.1–0.5	20–40
C	0.5–1	40–60
D	1–5	>60
E	>5	-

TABLE 16. Rating code of the maximum insulation dissipation factor.

Rating Code	Maximum insulation power factor
A	$DF_{max} < 0.5$
B	$0.5 \leq DF_{max} < 1$
C	$1 \leq DF_{max} < 1.5$
D	$1.5 \leq DF_{max} < 2$
E	$DF_{max} \geq 2$

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support received from Taif University Researchers Supporting Project Number (TURSP-2020/34), Taif University, Taif, Saudi Arabia. The authors thank Prof. Ayman Al-Hag and Prof. Norazhar Abu Bakar for their valuable discussion in this work and for helping us to obtain the datasets.

REFERENCES

[1] A. Jahromi, R. Piercy, S. Cress, J. Service, and W. Fan, "An approach to power transformer asset management using health index," *IEEE Elect. Insul. Mag.*, vol. 25, no. 2, pp. 20–34, Mar. 2009.

[2] C. Ranga, A. K. Chandel, and R. Chandel, "Condition assessment of power transformers based on multi-attributes using fuzzy logic," *IET Sci., Meas. Technol.*, vol. 11, no. 8, pp. 983–990, Nov. 2017.

[3] E. Kadim, N. Azis, J. Jasni, S. Ahmad, and M. Talib, "Transformers health index assessment based on neural-fuzzy network," *Energies*, vol. 11, no. 4, p. 710, Mar. 2018.

[4] H. Zeinoddini-Meymand and B. Vahidi, "Health index calculation for power transformers using technical and economical Parameters," *IET Sci., Meas. Technol.*, vol. 10, no. 7, pp. 823–830, 2016.

[5] A. Chantola, M. Sharma, and A. Saini, "Integrated fuzzy logic approach for calculation of health index of power transformer," in *Proc. 2nd Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Coimbatore, India, Apr. 2018, pp. 1045–1050.

[6] F. Tian, Z. Jing, H. Zhao, E. Zhang, and J. Liu, "A synthetic condition assessment model for power transformers using the fuzzy evidence fusion method," *Energies*, vol. 12, no. 5, p. 857, 2019.

[7] A. E. B. Abu-Elanien, M. M. A. Salama, and M. Ibrahim, "Calculation of a health index for oil-immersed transformers rated under 69 kV using fuzzy logic," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 2029–2036, Oct. 2012.

[8] A. Y. Alqudsi and A. H. ElHag, "A cost effective artificial intelligence based transformer insulation health index," in *Proc. 3rd Int. Conf. Condition Assessment Techn. Electr. Syst. (CATCON)*, Rupnagar, India, Nov. 2017, pp. 108–111.

[9] M. Islam, G. Lee, S. N. Hettiwatte, and K. Williams, "Calculating a health index for power transformers using a subsystem-based GRNN approach," *IEEE Trans. Power Del.*, vol. 33, no. 4, pp. 1903–1912, Aug. 2018.

[10] A. Alqudsi and A. El-Hag, "Assessing the power transformer insulation health condition using a feature-reduced predictor mode," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 3, pp. 853–862, Jun. 2018.

[11] K. Ibrahim, R. M. Sharkawy, H. K. Temraz, and M. M. A. Salama, "Selection criteria for oil transformer measurements to calculate the health index," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, no. 6, pp. 3397–3404, Dec. 2016.

[12] K. B. Shaban, A. H. El-Hag, and K. Benhmed, "Prediction of transformer furan levels," *IEEE Trans. Power Del.*, vol. 31, no. 4, pp. 1778–1779, Aug. 2016.

[13] R. M. A. Velasquez and J. V. M. Lara, "Health index for transformer condition assessment," *IEEE Latin Amer. Trans.*, vol. 16, no. 12, pp. 2843–2849, Dec. 2018.

[14] S. Li, H. Ma, T. Saha, and G. Wu, "Bayesian information fusion for probabilistic health index of power transformer," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 2, pp. 279–287, Jan. 2018.

[15] M. Dong, H. Zheng, Y. Zhang, K. Shi, S. Yao, X. Kou, G. Ding, and L. Guo, "A novel maintenance decision making model of power transformers based on reliability and economy assessment," *IEEE Access*, vol. 7, pp. 28778–28790, 2019.

[16] M. Yahaya, N. Azis, M. Z. A. A. Kadir, J. Jasni, M. H. Hairi, and M. A. Talib, "Estimation of transformers health index based on the Markov chain," *Energies*, vol. 10, no. 11, p. 1824, 2017.

[17] B. Qi, P. Zhang, Z. Rong, J. Wang, C. Li, and J. Chen, "Rapid transformer health state recognition through canopy cluster-merging of dissolved gas data in high-dimensional space," *IEEE Access*, vol. 7, pp. 94520–94532, 2019.

[18] A. E. B. Abu-Elanien and M. M. A. Salama, "Evaluation of transformer health condition using reduced number of tests," *Electr. Eng.*, vol. 101, no. 2, pp. 357–368, Jun. 2019.

[19] K. Benhmed, A. Mooman, A. Younes, K. Shaban, and A. El-Hag, "Feature selection for effective health index diagnoses of power transformers," *IEEE Trans. Power Del.*, vol. 33, no. 6, pp. 3223–3226, Dec. 2018.

[20] A. Naderian, S. Cress, R. Piercy, F. Wang, and J. Service, "An approach to determine the health index of power transformers," in *Proc. Conf. Rec. IEEE Int. Symp. Electr. Insul.*, Vancouver, BC, Canada, Jun. 2008, pp. 192–196.

[21] A. Azmi, J. Jasni, N. Azis, and M. Z. A. A. Kadir, "Evolution of transformer health index in the form of mathematical equation," *Renew. Sustain. Energy Rev.*, vol. 76, pp. 687–700, Sep. 2017.

[22] M. Wang, A. J. Vandermaar, and K. D. Srivastava, "Review of condition assessment of power transformers in service," *IEEE Elect. Insul. Mag.*, vol. 18, no. 6, pp. 12–25, Nov. 2002.

[23] B. Gorgan, P. V. Notingham, L. V. Badicu, and G. Tanasescu, "Calculation of power transformers health indexes," *Ann. Univ. Craiova, Electr. Eng. Ser.*, no. 34, pp. 13–18, 2010.

[24] A. Alqudsi and A. El-Hag, "Application of machine learning in transformer health index prediction," *Energies*, vol. 12, no. 14, p. 2694, 2019.

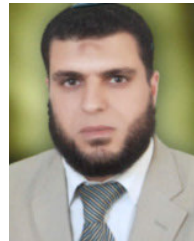
[25] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinf. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, Apr. 2005.

[26] G. A. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1315–1321, May 1999.

- [27] S. Ramírez-Gallego, I. Lastra, D. Martínez-Rego, V. Bolón-Canedo, J. M. Benítez, F. Herrera, and A. Alonso-Betanzos, "Fast-mRMR: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data," *Int. J. Intell. Syst.*, vol. 32, no. 2, pp. 134–152, Feb. 2017.
- [28] S. Putatunda and K. Rama, "A modified Bayesian optimization based hyper-parameter tuning approach for extreme gradient boosting," in *Proc. 15th Int. Conf. Inf. Process. (ICINPRO)*, Dec. 2019, pp. 1–6.
- [29] W. William, B. Burank, and P. Efratios, "Hyperparameter optimization of machine learning models through parametric programming," *Comput. Chem. Eng.*, vol. 139, pp. 1–12, Apr. 2020.
- [30] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, "Hyperparameter optimization for machine learning models based on Bayesian optimization," *J. Electron. Sci. Technol.*, vol. 17, pp. 26–40, Mar. 2019.
- [31] 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. 139176–139187, 2020.
- [32] S. M. Korobeynikov, A. V. Ridel, A. Y. Korobenkova, and D. V. Vagin, "Dissolution of bubbles with diagnostic gases in insulating liquids," in *Proc. IEEE 19th Int. Conf. Dielectr. Liquids (ICDL)*, Manchester, U.K., Jun. 2017, pp. 1–4.
- [33] S. Tenbohlen, J. Aragon-Patil, M. Fischer, M. Schäfer, Z. D. Wang, and I. H. Atanasova, "Investigation on sampling, measurement and interpretation of gas-in-oil analysis for power transformers," in *Proc. CIGRE*, 2008, pp. D1–D204.
- [34] D.-Y. Yeh, C.-H. Cheng, and S.-C. Hsiao, "Classification knowledge discovery in mold tooling test using decision tree algorithm," *J. Intell. Manuf.*, vol. 22, no. 4, pp. 585–595, Aug. 2011.
- [35] T. Muhlbacher, L. Linhardt, T. Moller, and H. Piringer, "TreePOD: Sensitivity-aware selection of Pareto-optimal decision trees," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 174–183, Jan. 2018.
- [36] P. K. Mishra, A. Yadav, and M. Pazoki, "A novel fault classification scheme for series capacitor compensated transmission line based on bagged tree ensemble classifier," *IEEE Access*, vol. 6, pp. 27373–27382, 2018.
- [37] K. Bacha, S. Souahlia, and M. Gossa, "Power transformer fault diagnosis based on dissolved gas analysis by support vector machine," *Electr. Power Syst. Res.*, vol. 83, no. 1, pp. 73–79, Feb. 2012.
- [38] A. D. Ashkezari, H. Ma, T. K. Saha, and C. Ekanayake, "Application of fuzzy support vector machine for determining the health index of the insulation system of in-service power transformers," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 20, no. 3, pp. 965–973, Jun. 2013.
- [39] *Hyperparameter Optimization in Classification Learner App*. [Online]. Available: <https://ch.mathworks.com/help/stats/hyperparameter-optimization-in-classification-learner-app.html>
- [40] Y. Benmahamed, Y. Kemari, M. Teguair, and A. Boubakeur, "Diagnosis of power transformer oil using KNN and Naïve Bayes classifiers," in *Proc. IEEE 2nd Int. Conf. Dielectr. (ICD)*, Jul. 2018, pp. 1–4.



**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 SF<sub>6</sub> gas, and AI technique applications.



**IBRAHIM B. M. TAHA** received the B.Sc. degree from the Faculty of Engineering, Tanta University, Egypt, in 1995, the M.Sc. degree from the Faculty of Engineering, Mansoura University, Egypt, in 1999, and the Ph.D. degree in electrical power and machines from the Faculty of Engineering, Tanta University, in 2007. Since 1996, he has been teaching at the Faculty of Engineering, Tanta University. He joined Taif University as an Associate Professor with the Electrical Engineering Department, Faculty of Engineering. His research interests include steady state and transient stability of HVDC systems, FACTS, multilevel inverters, dissolved gas analysis, and AI technique applications.

...