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.2021.3102415, IEEE Access

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.2020.Doi Number

Power Transformer Fault Diagnosis based on DGA using a Convolutional Neural Network with Noise in Measurements

Ibrahim B. M. Taha[1], Saleh Ibrahim[1], Diaa-Eldin A. Mansour [2]
[1] Electrical Engineering Department, College of Engineering, Taif University, Taif 21974, Saudi Arabia
[2] Electrical Power and Machines Engineering Department, Faculty of Engineering, Tanta University, Tanta 31511, Egypt

Corresponding author: Diaa-Eldin A. Mansour (*mansour@f-eng.tanta.edu.eg*)

This work received financial support from Taif University Researchers Supporting Project Number (TURSP-2020/61), Taif University, Taif, Saudi Arabia.

ABSTRACT Fault type diagnosis is a very important tool to maintain the continuity of power transformer operation. Dissolved gas analysis (DGA) is one of the most effective and widely used techniques for predicting the power transformer fault types. In this paper, a convolutional neural network (CNN) model is proposed based on DGA approach to accurately predict transformer fault types under different noise levels in measurements. The proposed model is applied with three categories of input ratios: conventional ratios (Rogers'4 ratios, IEC 60599 ratios, Duval triangle ratios), new ratios (five gas percentage ratios and new form six ratios), and hybrid ratios (conventional and new ratios together). The proposed model is trained and tested based on 589 dataset samples collected from electrical utilities and literature with varying noise levels up to ±20%. The results indicate that the CNN model with hybrid input ratios has superior prediction accuracy. The high accuracy of the proposed model is validated in comparison with conventional and recently published AI approaches. The proposed model is implemented based on MATLAB/toolbox 2020b.

INDEX TERMS Power transformer, fault diagnosis, convolution neural network, noises in measurements.

I. INTRODUCTION

Power transformers are considered one of the vital equipment in the electric power system. Early detection of transformer faults avoids the discontinuity of the power network and reduces the loss of profits for the electric utilities. Various faults in power transformers are generated due to the deterioration of their insulation system. The insulation system consists of an insulation oil and an impregnated paper. The insulation deterioration results from exposure of the transformer to several stresses such as electrical, mechanical, and thermal stresses. These stresses lead to the formation of dissolved gases, some of them are combustible gases such as hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), and carbon mono-oxide (CO), and others are incombustible gases such as carbon dioxide (CO_2) [1]-[3].

These dissolved gases help determine possible failures inside the transformer using dissolved gas analysis (DGA). According to the values of the dissolved gases concentration and their ratios to each other, some theories and rules were established that link the proportions of these gases and the expected failure [4], [5]. In [4], Key gas method, Dornenburg method, and Rogers' method were presented as three DGA techniques used to interpret the transformer fault based on the ratio limits between the combustible gases or the percentage to their sum. In [5], the transformer faults were divided into five types based on a dataset for transformers in service (IEC TC 10 database). It classified the transformer faults into five types in a triangular form called Duval triangle. These faults include the following types: 1) partial discharge (PD) represents by small carbonized captures in the paper; 2) low energy discharge (D1) causes large captures in paper and carbon particles in oil; 3) high energy discharge (D2) characterized by extensive carbonizations and metal fusion; 4) low and medium thermal faults (T1/T2) with oil temperature less than 300 °C for T1 and oil temperature greater than 300 °C and less than 700 °C for T2; 5) high thermal fault (T3) with oil temperature greater than 700 °C. These abovementioned conventional methods failed in several cases to interpret the transformer faults due to some issues such as an outage of gas ratio combinations from predefined codes or dependence on only three combustible gases in Duval triangle. Therefore, the diagnostic accuracy of such conventional methods in some cases is very poor.

New pentagon-based graphical representations [6], [7] enhanced the diagnostic accuracy of the conventional methods considering the main five combustible gases in their diagnosis. These new graphical representations could increase the diagnostic accuracy rather than the Duval triangle method [8]. In [9], heptagon shape was developed as another graphical representation method that determines the transformer faults based on the main five gases with CO and CO_2.

For further enhancement of diagnostic accuracy, artificial intelligence (AI) techniques and optimization techniques were proposed and used to overcome the poor diagnostic accuracy of conventional and graphical methods. In this regard, different AI classification techniques were applied, such as artificial neural networks (ANNs) [10], [11], fuzzy logic [12]-[14], neuro-fuzzy system [15], support vector machine (SVM) [16]-[18]. In addition, various optimization techniques were utilized, such as particle swarm optimization (PSO) [2], genetic algorithm (GA) for optimizing SVM parameters [16], hybrid grey wolf optimizer [19].

Finally, the deep learning approach was implemented in [20]-[22] to predict the transformer fault types. In [20], a depth learning and Softmax classifier model was presented to predict the transformer fault types. The model presented in [21] was built based on a deep belief neural network (DBNN) to diagnose the transformer fault types. In [22], a long short-term memory (LSTM) with DBNN (LSTM-DBNN) model was implemented to detect the transformer fault types, whereas a deep parallel diagnostic (DPD) model was introduced in [23].

The noises in DGA measurements are considered one of the most critical issues that decrement the diagnostic accuracy of DGA methods. Accordingly, these noises should be considered during the evaluation process of the diagnostic method. Noises can be originated either during oil sampling, sample storage, or gas separation and measurement. The noises due to sampling and storage can reach about 14%, while measurement noises lie in the range of 5% [24]. Most of the previous DGA methods didn't deal effectively with noises in DGA measurements, which is the main aim of this paper.

In this paper, we develop a noise-resistant DGA method by using Convolution Neural Network (CNN). Our main contributions are

- the augmentation of CNN training dataset with noisy points to improve the DGA diagnosis accuracy, and
- solving the DGA problem using different combinations of gas ratios and identifying the ratios that achieve the best detection accuracy.

II. CONVOLUTIONAL NEURAL NETWORKS

In contrast to general neural networks, a convolutional neural network (CNN) contains one or more *convolution layers* that work as filters [25]. The filter function is applied to each neighborhood of nodes of the previous layer, producing a corresponding set of outputs each time. Fig. 1 illustrates the convolution process in one dimension. The grey nodes represent zero-valued *padding* at the edges of the input layer to simplify filter processing at the boundaries. By sliding the convolution kernel, the kernel is applied to each subset of neighboring nodes, called the input window. Two important configuration hyperparameters define a convolution layer, namely, the kernel size and the stride. The *kernel size* indicates the number of inputs processed by the convolution kernel in one application. The *stride*, on the other hand, represents the number of nodes by which the filter is displaced after each application. Multiple convolution filters (kernels) can be applied in a single convolution layer to increases the learning capacity to extract more features.

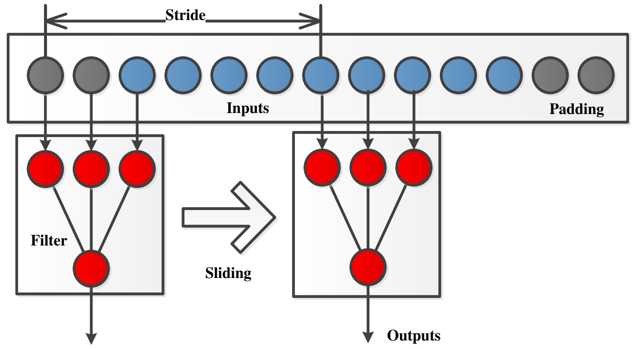

FIGURE 1. **Main CNN structure in one dimension.**

Other types of layers are commonly used within a CNN [25]. *Pooling layers* are special convolution layers with the main purpose of reducing dimensionality or subsampling. A *max-pooling* layer outputs the maximum input within its input window. Similarly, an *average-pooling* layer outputs the average input. *Fully connected layers* are important for mapping learned features into more comprehensive functions. *Threshold layers* such as rectified linear unit (ReLU) layers are often utilized to improve the nonlinear capacity of the model. *Batch normalization layers* are often applied to the output of convolution layers before the application of nonlinearities. Normalization improves the learning speed and dampens the effect of the random initial network weights.

CNNs are expected to be suitable for DGA because of their high noise resilience. Although CNNs are considered complex and expensive to train, this isn't a concern because training is performed offline, and the application of the resulting trained model sufficiently efficient. Therefore, we focus on obtaining a trained model that achieves the highest possible classification accuracy.

III. PROPOSED METHODOLOGY

A. Proposed neural network architecture

The proposed CNN architecture is illustrated in Fig. 2. To utilize CNNs in predicting transformer fault types based on DGA, the input points are treated as an image with dimension 9×1. Then, two convolution stages are performed. Each stage starts with a convolution layer followed by a batch normalization layer and a ReLU threshold layer. A max-pooling layer is inserted between the two convolution stages. A final classification stage includes a fully connected layer

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.2021.3102415, IEEE Access

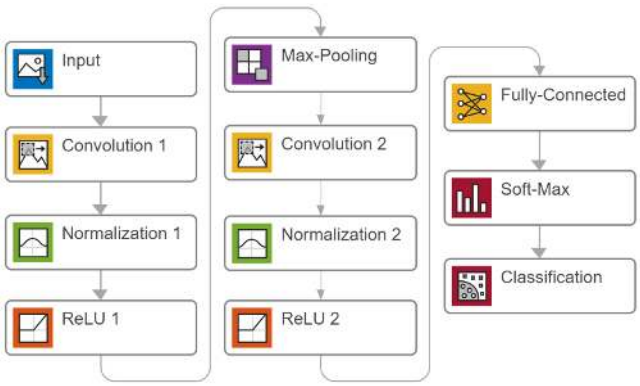

FIGURE 2. **Proposed CNN main configuration.**

followed by a soft-max layer and finally a classification layer. The proposed CNN was implemented and simulated using the 2020b MATLAB Deep Learning Toolbox [26].

The configuration parameters for each of the layers are summarized in Table 1. These parameters were fine-tuned through extensive trial-and-error simulations to maximize the CNN prediction accuracy. During training, the CNN weights are optimized using a stochastic gradient-based optimization algorithm known as Adaptive Moment Estimation (AdaM) [26], [27], which builds on the idea of adding a momentum term to the weight update formula to reduce oscillation along the steepest descent. AdaM also uses different learning rates for different weight vector elements based on a moving average of the first moment of the gradient and a moving average of its square. Finally, gradient clipping is employed to stabilize the training in the presence of gradient outliers. The algorithm is listed as Algorithm 1, and its hyperparameter settings are shown in Table 1.

B. Developing CNN model with noisy data

The CNN is developed using 589 dataset samples collected from literature [5, 16, 28-41], the Egypt electrical utility [40],

Algorithm 1. Adaptive Moment Estimation Optimization Algorithm

Inputs: α ▻ *step size*
$\beta_1, \beta_2 \lesssim 1$ ▻ *exponential decay rates*
$\omega_0 \in \mathbb{R}^d$ ▻ *initial weight vector*
$f(\omega): \mathbb{R}^d \to \mathbb{R}$ ▻ *stochastic objective function*
Output: $\omega \in \mathbb{R}^d$ ▻ *optimized weight vector*
Procedure:
1. $m_0 \leftarrow \langle 0 \rangle^d,\ v_0 \leftarrow \langle 0 \rangle^d,\ t \leftarrow 0$ ▻ *Initialize 1st and 2nd moment vectors, and time step.*
2. **do**
3. $t \leftarrow t+1$ ▻ *next step*
4. $g_t \leftarrow \nabla_\omega f(\omega_{t-1})$ ▻ *weight gradient*
5. $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1-\beta_1)\cdot g_t$ ▻ *update 1st moment*
6. $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1-\beta_2)\cdot g_t^2$ ▻ *update 2nd moment*
7. $\hat{m}_t \leftarrow m_t/(1-\beta_1^t)$ ▻ *bias-corrected 1st moment*
8. $\hat{v}_t \leftarrow v_t/(1-\beta_2^t)$ ▻ *bias-corrected 2nd moment*
9. $\omega_t \leftarrow \omega_{t-1} - \alpha\cdot \hat{m}_t/(\sqrt{\hat{v}_t}+\epsilon)$ ▻ *update weights*
10. **while** ω_t not converged

TABLE 1. CNN model selected parameters.

	Parameter	Value
Convolution Layer 1	Filter Size	3×1
	Number of filters	16
	Padding	1×0
Convolution Layer 2	Filter Size	3×1
	Number of filters	245
	Padding	1×0
Max-Pooling Layer	Stride	1
Fully Connected Layer	Outputs	6
Learning Algorithm	Step size, α	10^{-3}
	Gradient decay rate, β_1	0.9
	Gradient^2 decay rate, β_2	0.999
	Epsilon, ϵ	10^{-8}
	Gradient threshold	0.005

and the Indian utility in TIFAC laboratory [41]. The combined dataset is made available as part of DGALab's public code repository [42], and a summary of the dataset sources and fault type distribution is given in Table A1 in Appendix A. The complete dataset samples are divided into two subsets. The first set is used for training and represents 65% (383 samples), randomly selected from the complete dataset. The second set is used for the testing process and contains the remaining samples representing 35% (206 samples) of the complete dataset. The noise in measurement is introduced to each sample, $R = \langle r_i \rangle_{i=1}^{5} = \langle H2, CH4, C2H6, C2H4, C2H2 \rangle$, to generate a noisy sample, $R' = \langle r_i' \rangle_{i=1}^{5}$, using the equation adapted from [8].

$$r_i' = r_i \times \left\{1 + \frac{m(2\,n_i - 1)}{100}\right\} \quad (1)$$

where, m is the maximum noise level, and $N = \langle n_i \rangle_{i=1}^{5}$ is a 5×1 random vector with component values between 0 and 1. The maximum noise level is varied in the set $\{5, 10, 15, 20\}$ to generate four additional sets of noisy samples corresponding to noise levels up to ±5%, ±10%, ±15%, and ±20%. After augmenting the original dataset with the generated noisy samples, the total number of the training samples becomes (383×5=1915), while the total number of the testing samples becomes (206×5=1030).

C. Proposed input ratios

The CNN is trained using the training samples but with different input ratios. The CNN inputs used are (i) the old conventional ratios, (ii) new ratios, and (iii) hybrid ratios (conventional and new ratios together), while the output of the CNN is the transformer fault types. Table 2 presents the different ratios used for training the CNN model. The transformer fault types diagnosed by the CNN are (i) partial discharge (F1), (ii) low energy discharge (F2), (iii) high energy discharge (F3), (iv) low thermal (F4, oil temperature less than 300 °C), (v) medium thermal (F5, oil temperature greater than 300 °C and less than 700 °C), and (vi) high thermal (F6, oil temperature greater than 700 °C).

Fig. 3 presents the proposed methodology diagram. Firstly, the five input dataset gasses (H_2, CH_4, C_2H_6, C_2H_4, C_2H_2) in ppm are used to generate the noise data with different noise levels up to ±20%. Then, the dataset transformation process,

TABLE 2. Ratios used for training the CNN.

Conventional Ratios	Rogers' 4 ratios {1}	$C2H6/CH4$, $C2H4/C2H6$, $C2H2/C2H4$, $CH4/H2$
	IEC 60599 ratios {2}	$C2H4/C2H6$, $C2H2/C2H4$, $CH4/H2$
	Duval Ratios {3}	$C2H2/(TD) \times 100$ [a] $C2H4/(TD) \times 100$ $CH4/(TD) \times 100$
New Ratios	Percentage ratios {4}	$H2/(TCDG) \times 100$ [b] $CH4/(TCDG) \times 100$ $C2H6/(TCDG) \times 100$ $C2H4/(TCDG) \times 100$ $C2H2/(TCDG) \times 100$
	New form {5}	$H2/max(H2, CH4, C2H6, C2H4, C2H2) \times 100$ $CH4/max(H2, CH4, C2H6, C2H4, C2H2) \times 100$ $C2H6/max(H2, CH4, C2H6, C2H4, C2H2) \times 100$ $C2H4/max(H2, CH4, C2H6, C2H4, C2H2) \times 100$ $C2H2/max(H2, CH4, C2H6, C2H4, C2H2) \times 100$ $\ln\{\frac{max(H2, CH4, C2H6, C2H4, C2H2)}{TCDG} \times 100\}$
Hybrid Ratios	Percentage ratios + Rogers' 4 ratios {6}	
	Percentage ratios+ ln (Rogers' 4 ratios) {7}	
	Percentage ratios + ln (IEC 60599) {8}	
	New form + ln (Rogers' 4 ratios) {9}	
	New form + ln (IEC 60599) {10}	

[a] $TD = C2H4 + CH4 + C2H2$,
[b] $TCDG = H2 + CH4 + C2H6 + C2H4 + C2H2$

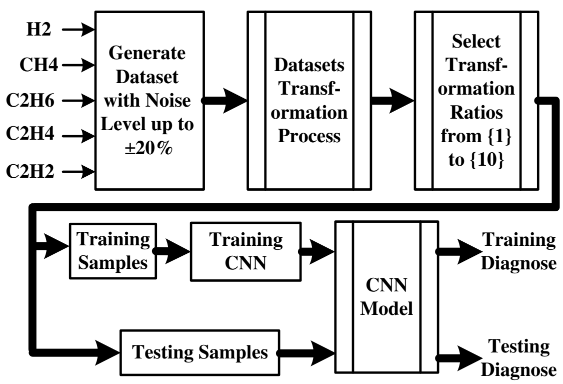

FIGURE 3. Proposed methodology diagram.

according to that presented in Table 2, is implemented. The transformation ratios of {1} to {10} are selected, and then the dataset is randomly divided into training and testing sets. Thus, the training dataset is used for training the CNN model. Finally, the training and testing datasets are applied to the generated CNN model to obtain the output diagnosis for both.

IV. RESULTS AND DISCUSSION

The convolution neural network (CNN) is implemented and carried out using MATLAB/toolbox 2020b. The CNN model is developed based on the training samples and the prediction evaluated. Like other optimization methods, a CNN depends on random initialization, which means that different results are obtained each time the CNN is trained using the same dataset. Therefore, we train the CNN ten times for each of the ratios introduced in Table 2. Finally, the CNN model accuracy for each ratio is evaluated based on the mean value of the ten training results.

The CNN prediction accuracy can be estimated as follows:

$$\%\eta = \frac{PT+NT}{PT+NT+PF+NF} \times 100 \qquad (2)$$

where, PT and NT are the positive and negative class true rate, respectively, and PF and NF are the positive and negative class false rate, respectively.

The CNN loss can be expressed as follows:

$$Loss = \sum_{i=1}^{n} \sum_{j=1}^{C} k_{ij} O_{ij} \qquad (3)$$

where, n is the number of dataset samples, C is the number of classes, k_{ij} is the probability that the i^{th} sample belongs to the j^{th} class, O_{ij} is the output of the dataset sample i in the class j, which is the output of the SoftMax layer.

Table 3 presents the statistical analysis of the prediction accuracy obtained through ten training attempts of the CNN model when its inputs are the new form ratios {5}. After each attempt, we observed the prediction accuracy for three cases: (i) using the training samples with all noise levels, (ii) using the testing samples with all noise levels, and (iii) using the complete dataset (both training and testing samples) for each maximum noise level, 0%, ±5%, ±10%, ±15%, and ±20%, separately. The training accuracy varies from 96.8% to 98.1% with mean and STD values of 97.7% and 0.37%, respectively, while the testing accuracy varies from 92.8% to 94.6% with a mean and STD values of 93.7% and 0.54%, respectively. The CNN model exhibits good accuracy of detecting the fault types from noisy samples with maximum noise levels up to ±20%. The mean values of the prediction accuracy over the 10 training attempts are 97.4%, 97.2%, 96.5%, 96.1% and 94.3% for 0%, ±5%, ±10%, ±15% and ±20% maximum noise levels, respectively.

Fig. 4 illustrates the prediction accuracy and the loss against the iteration number during the second training attempt in Table 2. The results indicate that the training accuracy is near one hundred percent, while the loss is low near zero, which means a good training accuracy of the CNN model.

Fig. 5 compares the actual fault types (F1 to F6) against the predicted fault type to illustrate the prediction accuracy of the CNN model at 0%, 5%, 10%, and 20% noise levels. The

TABLE 3. The statistical analysis of the prediction accuracy of ten training attempts of the proposed CNN model, indicating the training accuracy, the testing accuracy, and the overall accuracy for each noise level.

No.	Training accuracy	Testing accuracy	Overall accuracy for each noise level				
			0%	±5%	±10%	±15%	±20%
Max.	98.1	94.6	98.0	97.8	97.3	97.5	94.9
Min.	96.8	92.8	96.6	95.8	95.6	95.4	93.4
Mean	97.7	93.7	97.4	97.2	96.5	96.1	94.3
STD	0.37	0.54	0.48	0.61	0.54	0.58	0.48

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.2021.3102415, IEEE Access

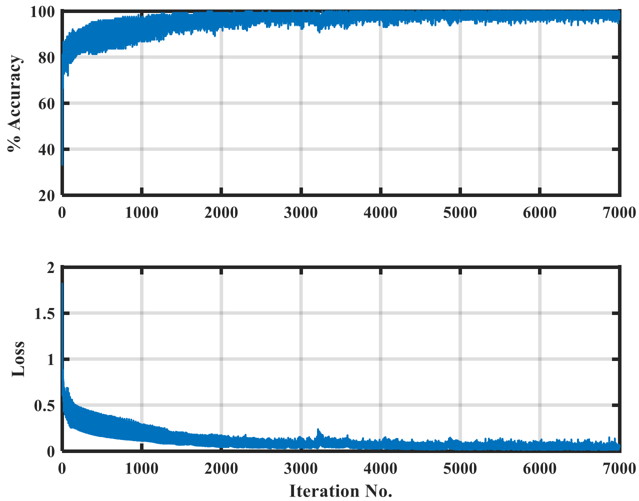

FIGURE 4. Prediction accuracy during training process with new ratio form input to the CNN model for training No. 2.

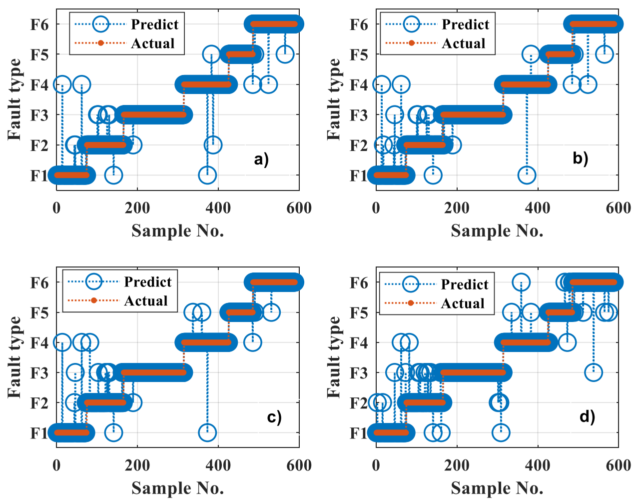

FIGURE 5. Comparison between the actual and predict samples with input {5} to the CNN model for training No. 2.

results show that the CNN model has a good detecting accuracy at 0%, 5%, 10%, and 20% noise levels with high prediction accuracy of 96.9%, 96.6%, 96.3%, and 94.7%, respectively.

The CNN model was trained with different input ratios. To increase confidence in results, the mean prediction accuracy for ten attempts of training and testing episodes at each noise level was calculated and presented in Table 4. The results illustrate that the prediction accuracy with the new ratios is better than that with the conventional ratios as an input of the CNN model. Furthermore, the results illustrate that the accuracy with the hybrid ratios {7} (Percentage ratios + ln{Percentage ratios}) has the highest accuracy of 95% for overall testing samples and the highest prediction accuracy of 95.8% with up to ±20% noise level.

Fig. 6 presents the prediction accuracy and loss against the iteration number during one of the training attempts with input ratio {7} to the CNN model. The results illustrate that the training accuracy is the nearest to one hundred percent, while

TABLE 4. The CNN prediction accuracy with different input ratios

Ratios	Training	Testing	Noise Level				
			0%	±5%	±10%	±15%	±20%
{1}	94.6	90.1	94.1	93.9	93.3	92.4	91.5
{2}	89.2	85.0	89.1	89.1	87.9	86.8	85.7
{3}	73.2	68.0	73.0	72.7	72.1	70.4	68.6
{4}	97.2	94.0	97.1	96.8	96.3	95.6	94.4
{5}	97.7	93.8	97.4	97.2	96.5	96.1	94.4
{6}	98.0	94.5	97.8	97.7	97.2	96.3	95.8
{7}	98.1	95.0	97.9	97.7	97.2	96.5	95.8
{8}	98.4	94.7	98.1	97.9	97.3	96.6	95.7
{9}	97.9	94.7	97.9	97.8	96.8	96.4	95.1
{10}	98.2	94.7	98.0	97.9	97.2	96.5	95.3

the loss is low near zero, indicating a good training accuracy of the CNN model.

Fig. 7 illustrates a comparison between the fault types (F1

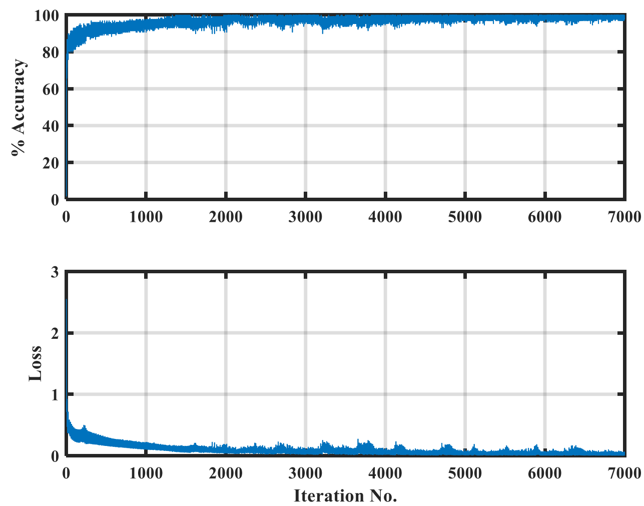

FIGURE 6. Prediction accuracy during training process with input {7} to the CNN model for training No. 10.

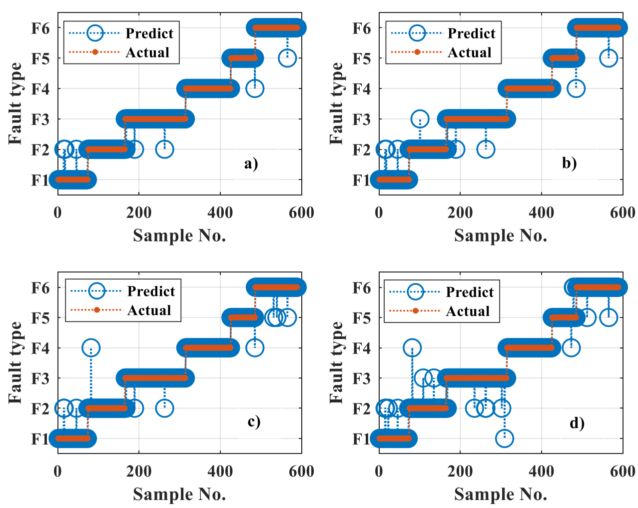

FIGURE 7. Comparison between the actual and predict samples with input {7} to the CNN model for training No.10 (highest training accuracy).

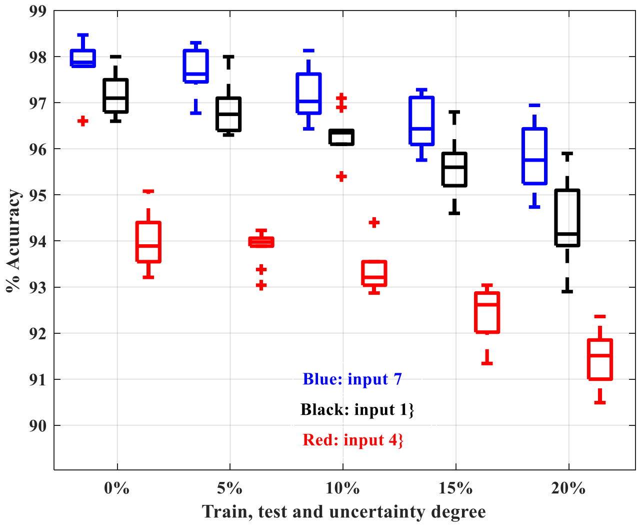

FIGURE 8. Boxplot prediction accuracy of the CNN model with different input {1}, {4} and {7} at different noise levels.

to F6) predicted by the CNN model against the actual fault types. The results show that the CNN model has a good performance at 0%, 5%, 10%, and 20% noise levels with high prediction accuracy of 98.5%, 98.3%, 98%, and 96.6%, respectively.

Fig. 8 presents a boxplot comparison of the CNN prediction accuracy with inputs {1}, {4} and {7} at noise levels of 0%, ±5%, ±10%, ±15% and ±20%, respectively. It illustrates that the prediction accuracy of the CNN with input {7} (hybrid input ratios, five percentage ratios + ln [Rogers' 4 ratios]) is a superior one compared to that of inputs {1} (conventional input ratios, Rogers' 4 ratios) and {4} (new input ratios, five gas percentage ratios) respectively.

V. MODEL VALIDATION

A. CNN model against ANN with noisy data

The performance of the proposed CNN model is compared with the artificial neural network method (ANN). The ANN method is built using MATLAB ANN toolbox version 2020b. The trained dataset (1915 samples was applied to the 9-variable input ratio {7}) is divided into three sets (70% for training, 15% for validations, and 15% for testing). The ANN model is trained under different hidden layer numbers. One hundred twenty-five neurons are used for hidden layers that give the best training predicting accuracy. The training performance for both training, testing, and validation dataset is introduced in Fig. 9. The training and testing results of the CNN model with 9-ratio {7} are introduced in Table 5.

Six dataset samples with different noise levels (0%, 5%, 10%, 15%, and 20) are used in Table 6 as case studies for comparing the CNN and ANN models. The original samples are shaded in grey, followed by the noisy samples derived from the original sample using (1). The ACT column indicates the actual fault type. The CNN and ANN columns indicate the corresponding diagnosis generated by each of the two

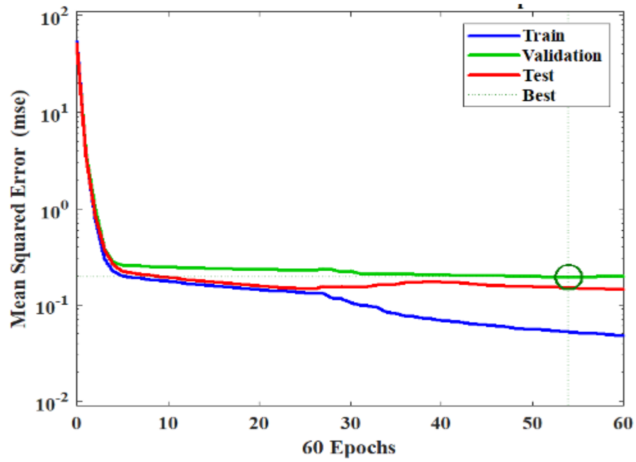

FIGURE 9. Mean squared error against number of epochs during training process of ANN.

TABLE 5. CNN AND ANN COMPARISONS FOR BOTH TRAINING AND TESTING STAGES.

FT	Training		Testing	
	CNN {7}	ANN	CNN {7}	ANN
PD	**95**	85.7	**92.7**	79.3
D1	**97.6**	91.5	**99.4**	89.1
D2	**96.4**	90.5	**99.1**	89.8
T1	**100**	93.4	**100**	91.9
T2	**100**	96.7	**93**	90.4
T3	**99.1**	96.6	**97.5**	96
All	**97.9**	92.3	**97.4**	89.8

methods. The results illustrate the effectiveness of the CNN model with different transformer fault types and all different noise levels up to 20%, while the ANN model fails to diagnose the transformer fault types with high noise levels (Highlighted as bold).

B. CNN model against machine learning models with noisy data

Three machine learning approaches are used decision tree method (DT), support vector machine method (SVM), and ensemble method (EN) are built based on the MATLAB/classification learner toolbox (2020b). The CNN, DT, SVM, and EN methods were applied to the 9-variable input ratio {7}. Then, the generated models were compared using the testing samples. Fig. 10 presents the minimum error against iteration number during the training stage of the three machine learning methods (DT, SVM, and EN methods). The cross-fold validation with ten folds during training of the three machine learning methods. The optimization technique used to determine the optimal parameters of each method are Bayesian optimization method, while the acquisition function used is expected improvement per second plus. The optimal parameters of the DT, SVM, and EN methods are introduced in Table 7.

TABLE 6. CNN and ANN comparisons for several case studies with various noise levels.

%N	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	ACT	CNN	ANN
0%	111.7	19.4	104.1	6.4	3.8	PD	PD	PD
±5%	112.3	19.1	101.2	6.3	3.8	PD	PD	PD
±10%	103.6	20.3	105.4	6.9	3.7	PD	PD	**D1**
±15%	115.8	19.4	116.8	6.2	4.1	PD	PD	PD
±20%	115.3	22.9	113.2	6.6	3.5	PD	PD	**D1**
0%	169.0	38.0	48.5	6.5	5.8	D1	D1	D1
±5%	173.5	39.3	49.3	6.3	5.9	D1	D1	D1
±10%	171.5	35.8	50.4	6.5	6.2	D1	D1	D1
±15%	188.5	34.9	48.6	6.7	5.9	D1	D1	**PD**
±20%	148.3	36.9	48.2	7.4	5.0	D1	D1	D1
0%	235.5	333.6	177.5	1202	148.9	D2	D2	D2
±5%	232.6	324.8	182.9	1157	153.5	D2	D2	D2
±10%	232.9	346.5	165.6	1142	155.6	D2	D2	D2
±15%	270.7	381.0	158.9	1113	148.0	D2	D2	**T1**
±20%	198.4	339.7	145.8	1357	138.9	D2	D2	D2
0%	42.0	124.0	1.0	8.0	0.0	T1	T1	T1
±5%	42.4	124.5	1.0	8.2	0.0	T1	T1	T1
±10%	38.5	133.7	0.9	7.6	0.0	T1	T1	T1
±15%	44.6	138.1	1.1	9.0	0.0	T1	T1	T1
±20%	49.6	124.3	1.1	8.6	0.0	T1	T1	**T2**
0%	27.0	90.0	42.0	63.0	0.2	T2	T2	T2
±5%	28.0	86.9	41.5	60.5	0.2	T2	T2	T2
±10%	26.4	88.0	41.5	58.2	0.2	T2	T2	T2
±15%	29.4	78.1	48.3	65.1	0.2	T2	T2	T2
±20%	31.5	101.4	34.1	53.9	0.2	T2	T2	**T3**
0%	3420	7870	1500	6990	33.0	T3	T3	T3
±5%	3366	7530	1435	6796	32.6	T3	T3	T3
±10%	3517	8486	1493	7534	30.8	T3	T3	**T2**
±15%	3917	7490	1403	7427	36.8	T3	T3	T3
±20%	2930	9198	1218	8160	29.7	T3	T3	**T2**

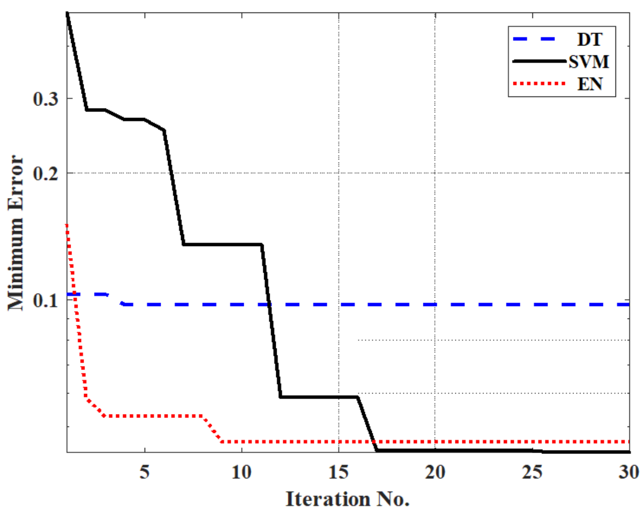

FIGURE 10. Minimum Error against iteration number during training process of DT, SVM and EN methods.

Table 8 illustrates the results of the proposed CNN model and the results of DT, SVM, and EN methods during training and testing stages. The results illustrate the superiority of the proposed CNN model compared to other methods. The results of the proposed CNN model are compared with the conventional and recently published AI methods with different noise levels.

TABLE 7. Optimal parameters and optimizer options of DT, SVM and EN methods.

Method	DT	SVM	EN
Optimal parameters	Maximum number of splits: 277 Split criterion: Maximum deviance reduction	Multi class method: One-vs-one Box constraint level: 481.864 Kernel scale: 8.7605 Kernel function: Gaussian Standardize data: false	Ensemble method: Bag Number of learners: 400 Maximum number of splits: 1080 Number of predictors to sample: 1
No. of iterations	30		
Optimizer	Gaussian optimization		
Acquisition function	Expected improvement per second plus		

TABLE 8. Comparisons of the proposed CNN {7} prediction accuracy with DT, SVM and EN with training and testing datasets.

FT	Training				Testing			
	CNN {7}	DT	SVM	EN	CNN {7}	DT	SVM	EN
PD	**95**	89.2	93.2	92.8	**92.7**	82	91	91
D1	**97.6**	78.6	91.8	90.1	**99.4**	89.4	92.1	92.7
D2	**96.4**	89.9	95.3	96.0	**99.1**	89.7	97.1	96.3
T1	**100**	96.2	99.2	97.8	**100**	95.8	100.0	100
T2	**100**	90.7	95.1	94	**93**	91.5	97.4	93.2
T3	**99.1**	95.2	97.7	97.7	**97.5**	94	99.4	98.8
All	**97.9**	90.3	95.7	95.1	**97.4**	90.7	96.5	95.8

C. CNN model against recently published research

Table 9 presents the results of the best training attempt of the proposed CNN model with input ratio {7} side by side with the results of Rogers'4 ratios, IEC 60599, Duval triangle, conditional probability [8], modified-Rogers'4, modified-IEC 60599 [2] and code-tree [19] methods. The results indicate the superiority of the proposed CNN model compared to the other methods.

TABLE 9. Comparisons between the proposed CNN prediction accuracy with other methods with different noise levels

Method	Noise Level				
	0%	±5%	±10%	±15%	±20%
Duval	64.3	63.8	62.6	62.5	62.3
Rogers'4	47.9	47.9	47.2	46.9	47.0
IEC 60599	56.2	55.9	56.0	53.3	55.5
Probability	80.0	79.5	78.9	77.9	78.4
Mod-Rog.	81.2	80.8	79.5	77.8	76.9
Mod-IEC	82.2	82.3	80.1	79.1	77.8
Code-tree	80.8	80.8	79.5	78.8	78.6
CNN {7}	**98.5**	**98.3**	**98.0**	**97.3**	**96.6**

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.2021.3102415, IEEE Access

VI. CONCLUSIONS

This research proposed the CNN model to detect the transformer fault types under different uncertain noise levels in measurements. The DGA data were collected from various resources, including literature and electrical utilities. The noise in DGA data was introduced to all samples with various levels ranging up to 20%. The dataset samples were randomly divided into two subsets, a training set with 65% of data samples and a testing set with the remaining data samples. Moreover, the data samples were presented with different input ratios to the CNN. The superiority of CNN in detecting various fault types with different noise levels was evaluated through several indicators as follows:

1- It was found that the predicting accuracy of the CNN with the input of five percentage ratios plus ln [Rogers' 4 ratios] is a superior one compared to other inputs.
2- The CNN model had high predicting accuracy of 98.5%, 98.3%, 98%, 97.3% and 96.6% for noise levels of 0%, ±5%, ±10%, ±15% and ±20%, respectively. This predicting accuracy validates the strong immunity of the proposed CNN against noises in measurements.
3- The comparisons between the predicting accuracy of the proposed CNN model and other methods indicated significantly superior performance for the CNN model.

VII. ACKNOWLEDGEMENT

The authors would like to acknowledge the financial support received from Taif University Researchers Supporting Project Number (TURSP-2020/61), Taif University, Taif, Saudi Arabia.

VIII. REFERENCES

[1] J. Faiz and M. Soleimani, "Dissolved gas analysis evaluation in electric power transformers using conventional methods a review," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 2, pp. 1239-1248, Apr. 2017.

[2] 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 Trans. Dielectr. Electr. Insul.*, vol. 27, no. 1, pp. 222-230, Feb. 2020

[3] S. A. Ward, A. El-Faraskoury, M. Badawi, S. A. Ibrahim, K. Mahmoud, M. Lehtonen and M. M. F. Darwish, "Towards precise interpretation of oil transformers via novel combined techniques based on DGA and partial discharge sensors," *Sensors*, vol. 21, no. 6, p. 2223, Mar. 2021.

[4] *IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers*, IEEE Std C57.104-2019, Nov. 2019.

[5] M. Duval and A. dePabla, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases," *IEEE Electr. Insul. Mag.*, vol. 17, no. 2, pp. 31-41, Mar. 2001.

[6] D. A. Mansour, "A new graphical technique for the interpretation of dissolved gas analysis in power transformers," in *Proc. Annu. Rep. Conf. Electr. Insul. Dielectr. Phenomena*, Montreal, QC, Canada, pp. 195-198, Oct. 2012.

[7] M. Duval and L. Lamarre, "The Duval pentagon – a new complementary tool for the interpretation of dissolved gas analysis in transformers," *IEEE Electr. Insul. Mag.*, vol. 30, no. 6, pp. 9-12, Nov. 2014.

[8] I. B. M. Taha, D. A. Mansour, S. S. M. Ghoneim and N. I. Elkalashy "Conditional probability-based interpretation of dissolved gas analysis for transformer incipient faults," *IET Gener. Transm. Distrib.*, vol. 11, no. 4, pp. 943-951, Mar. 2017.

[9] O. E. Gouda, S. H. El-Hoshy and H. H. El-Tamaly, "Proposed heptagon graph for DGA interpretation of oil transformers," *IET Gener. Transm. Distrib.*, vol. 12, no. 2, pp. 490-498, Jan. 2018.

[10] J. L. Guardado, J. L. Naredo, P. Moreno and C. R. Fuerte, "A comparative study of neural network efficiency in power transformers diagnosis using dissolved gas analysis," *IEEE Trans. Power Del.*, vol. 16, no. 4, pp. 643-647, Oct. 2001.

[11] 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 Trans. Dielectr. Electr. Insul.*, vol. 23, no. 3, pp. 1838-1845, June 2016.

[12] A. Abu-Siada and S. Hmood, "A new fuzzy logic approach to identify power transformer criticality using dissolved gas-in-oil analysis," *Int. J. Electr. Power Energy Syst.*, vol. 67, pp. 401-408, May 2015.

[13] I. B. M. Taha, S. S. M. Ghoneim and H. G. Zaini, "A fuzzy diagnostic system for incipient transformer faults based on DGA of the insulating transformer oils," *Int. Rev. Electr. Eng.*, vol. 11, no. 3, pp. 305-313, July 2016.

[14] .M Noori, R. Effatnejad, P. Hajihosseini, "Using dissolved gas analysis results to detect and isolate the internal faults of power transformers by applying a fuzzy logic method," *IET Gener. Transm. Distrib.*, 2017, vol. 11, no. 10, pp. 2721-2729 .July ٢٠١٧ ,

[15] S. A. Khan, M. D. Equbal and T. Islam, "A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 1, pp. 590-596, Feb. 2015.

[16] 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 Trans. Dielectr. Electr. Insul.*, vol. 23, no. 2, pp. 1198-1206, Apr. 2016.

[17] Y. Benmahamed, M. Teguar and A. Boubakeur, "Application of SVM and KNN to Duval Pentagon 1 for transformer oil diagnosis," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 6, pp. 3443-3451, Dec. 2017.

[18] Y. Zhang, X. Li, H. Zheng, H. Yao, J. Liu, C. Zhang, H. Peng and J. Jiao, "A fault diagnosis model of power transformers based on dissolved gas analysis features selection and improved krill herd algorithm optimized support vector machine," *IEEE Access*, vol. 7, pp. 102803-102811, 2019.

[19] A. Hoballah, D. 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.

[20] X. Ji, Y. Zhang, H. Sun, J. Liu, Y. Zhuang and Q. Lei, "Fault diagnosis for power transformer using deep learning and softmax regression," *2017 Chinese Automation Congress (CAC)*, Jinan, China, pp. 2662-2667, 2017.

[21] J. Dai, H. Song, G. Sheng and X. Jiang, "Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 5, pp. 2828-2835, Oct. 2017.

[22] J. Lin, L. Su, Y. Yan, G. Sheng, D. Xie and X. Jiang, "Prediction method for power transformer running state based on LSTM_DBN network," *Energies*, vol. 11, no. 7, p. 1880, Jul. 2018.

[23] X. Wu, Y. He and J. Duan, "A deep parallel diagnostic method for transformer dissolved gas analysis," *Appl. Sci.*, vol. 10, no. 4, p. 1329, Feb. 2020.

[24] 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*, D1-204, 2008.

[25] T. Singh, "*Deep Learning with Applications Using Python*," Navin Kumar Manaswi, Bangalore, Karnataka, India, 2018.

[26] https://ch.mathworks.com/discovery/convolutional-neural-network-matlab.html.

[27] D. P. Kingma and J. L. Ba. "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[28] 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.

[29] M. Duval, "A review of faults detectable by gas-in-oil analysis in transformers," *IEEE Electr. Insul. Mag.*, vol. 18, no. 3, pp. 8-17, May 2002.

[30] S. Agrawal and A. K. Chandel, "Transformer incipient fault diagnosis based on probabilistic neural network," *2012 Students Conference on Engineering and Systems (SCES)*, Allahabad, Uttar Pradesh, India, pp. 1-5, 2012.

[31] M.-H. Wang, "A novel extension method for transformer fault diagnosis," *IEEE Trans. Power Del.*, vol. 18, no. 1, pp. 164-169, Jan. 2003.

[32] Z. Yong-Li, W. Fang and G. Lan-Qin, "Transformer fault diagnosis based on naive bayesian classifier and SVR," *2006 IEEE Region 10 Conf. (TENCON)*, Hong Kong, China, pp. 1-4, Nov. 2006.

[33] D. V. S. S. Siva Sarma and G. N. S. Kalyani, "ANN approach for condition monitoring of power transformers using DGA," *2004 IEEE Reg. 10 Conf. (TENCON)*, Chiang Mai, Thailand, pp.444-447, Nov. 2004.

[34] G. Zhang, K. Yasuoka, S. Ishii, Li Yang and Zhang Yan, "Application of fuzzy equivalent matrix for fault diagnosis of oil-immersed insulation," in *Proc. IEEE 13th Int. Conf. Dielectr. Liquids (ICDL)*, Nara, Japan, pp. 400-403, Jul. 1999.

[35] E. Li, L. Wang and B. Song, "Fault diagnosis of power transformers with membership degree," *IEEE Access*, vol. 7, pp. 28791-28798, 2019.

[36] J. Hu, L. Zhou and M. Song, "Transformer fault diagnosis method of gas hromatographic analysis using computer image analysis," *Second Int. Conf. Intelligent System Design and Engineering Application*, Sanya, China pp. 1169-1172, Jan. 2012.

[37] O. E. Gouda, S. H. El-Hoshy and H. H. El-Tamaly, "Proposed three ratios technique for the interpretation of mineral oil transformers based dissolved gas analysis," *IET Gener. Transm. Distrib.*, vol. 12, no. 11, pp. 2650-2661, Jun. 2018.

[38] S. Seifeddine, B. Khmais and C. Abdelkader, "Power transformer fault diagnosis based on dissolved gas analysis by artificial neural network," *First Int. Conf. Renewable Energies and Vehicular Technology*, Nabeul, Tunisia, pp. 230-236, Mar. 2012.

[39] M. Rajabimendi and E. P. Dadios, "A hybrid algorithm based on neural-fuzzy system for interpretation of dissolved gas analysis in power transformers," *2012 IEEE Reg. 10 Conf. (TENCON)*, Cebu, Philippines, pp. 1-6, Nov. 2012.

[40] Egyptian Electricity Holding Company (EEHC) Reports, 1991-2016.

[41] Technology Information Forecasting and Assessment Council (TIFAC) laboratory, Department of science and technology, Govt of India, Delhi.

[42] S. I. Ibrahim, S. S. M. Ghoneim and I. B. M. Taha, "DGALab: an extensible software implementation for DGA," *IET Gener. Transm. Distrib.*, vol. 12, no. 18, pp. 4117-4124, Oct. 2018.

Ibrahim B. M. Taha Received B.Sc. degree from the Faculty of Engineering at Tanta, Tanta University, Egypt, in 1995. He received M.Sc. degree from the Faculty of Engineering at Mansoura, Mansoura University, Egypt, in 1999. Starting from 1996 he was a teaching staff at the Faculty of Engineering, Tanta University, Egypt. In 2007, he earned a Ph.D. Degree in Electrical power and machines, Faculty of Engineering-Tanta University (2007). He is currently an Assistant Professor with the Electrical Engineering Department, Taif University, Saudi Arabia. His research focuses in steady state and transient stability of HVDC systems, FACTS, load forecasting, Multi-Level Inverters, Dissolved gas analysis, and artificial intelligent technique applications, PV system fault detection and distance adaptive protective relays.

Saleh Ibrahim received the B.Sc. and M.Sc. degrees in computer engineering from Cairo University, Egypt, in 2000 and 2004, respectively, and the Ph.D. degree in computer science and engineering from the University of Connecticut, USA, in 2010. He is currently an Assistant Professor with the Electrical Engineering Department, Taif University, Saudi Arabia. He has been an Assistant Professor with the Computer Engineering Department, Cairo University, since 2011. He has published several research articles in high impact journals and international conferences. His current research interests include information security and computer networks.

Diaa-Eldin A. Mansour (S'06-M'10-SM'17) was born in Tanta, Egypt on 31 December 1978. He received the B.Sc. and M.Sc. degrees in electrical engineering from Tanta University, Tanta, Egypt, in 2000 and 2004, respectively, and the Ph.D. degree in electrical engineering from Nagoya University, Nagoya, Japan, in 2010. Since 2000, he has been with the Department of Electrical Power and Machines Engineering, Faculty of Engineering, Tanta University, Egypt. Currently, he is working as a full professor and director of the High Voltage and Superconductivity Laboratory in the same department. From 2010, he was a foreign researcher for three months at EcoTopia Science Institute, Nagoya University, Nagoya, Japan. His research interests are high voltage engineering, nanodielectrics, applied superconductivity, renewable energy, and smart grids. He received the best presentation award two times from IEE of Japan in 2008 and 2009, Prof. Khalifa's Prize from the Egyptian Academy of Scientific Research and Technology in 2013, Tanta University Encouragement Award in 2016, and Egypt-State Encouragement Award in the field of Engineering Sciences in 2018. Also, he has been listed among world's top 2% scientists by Stanford University, USA in 2020.

Appendix A

The 589-sample dataset used for the development of the CNN is collected from various sources indicated in Table A1. The table presents fault type statistics for each source and the overall statistics of the composite dataset.

TABLE A1. Datasets source references and fault distribution of the 589 dataset samples.

Fault type / Ref.	PD	D1	D2	T1	T2	T3	ALL
[5]	10	22	49	1	4	20	106
[16]	0	0	0	2	0	0	2
[28]	0	1	1	1	1	0	4
[29]	3	0	0	2	0	0	5
[30]	1	0	0	2	0	1	4
[31]	2		4	3	3	5	17
[32]	0	1	1	1	0	1	4
[33]	1	3	2		1	8	15
[34]	1	0	0	4	15	15	35
[35]	5	8	9	5	10	10	47
[36]	0	3	2			0	5
[37]	0	0	3	2	1	0	6
[38]	0	0	4	1	0	2	7
[39]	0	0	0	0	1	2	3
[40]	12	29	30	69	17	29	186
[41]	39	24	44	18	7	11	143
Total	74	91	149	111	60	104	589