IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 16 January 2024, accepted 11 February 2024, date of publication 19 February 2024, date of current version 26 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3366905

RESEARCH ARTICLE

Deep Machine Learning-Based Asset Management Approach for Oil-Immersed Power Transformers Using Dissolved Gas Analysis

LAN JIN, DOWON KIM, KIT YAN CHAN, (Member, IEEE), AND AHMED ABU-SIADA, (Senior Member, IEEE)
Electrical and Computer Engineering Discipline, Curtin University, Perth, WA 6102, Australia
Corresponding author: Ahmed Abu-Siada (a.abusiada@curtin.edu.au)

ABSTRACT Reliable operation of oil-immersed power transformers is crucial for electrical transmission and distribution networks. However, the aging of high voltage assets including power transformers along with the increasing of load demand have heightened the importance of adopting cost-effective asset management strategies. Dissolved gas analysis (DGA) has been recognized as a valuable diagnostic tool for detecting potential faults and monitoring the condition of oil-immersed power transformers. Traditional offline DGA method involves periodic sampling and laboratory analysis, which often results in delayed detection and response to emerging faults. To address these limitations, online DGA approach has been emerged to provide real-time monitoring and continuous data acquisition. This paper presents a new asset management approach for mineral oil-immersed power transformers by analysing the online DGA data using convolutional neural networks. The proposed approach provides real time solutions to classify emerging fault type and predict transformer health deterioration level with high accuracy. Results show that the accuracy of fault diagnostics of the proposed approach is approximately 87%.

INDEX TERMS Power transformers, dissolved gas analysis, condition monitoring, asset management, remnant life estimation.

I. INTRODUCTION

The reliable operation of a power system is largely dependent on the health condition and performance of its key equipment, particularly power transformers. Faults or malfunctions in a power transformer can directly impact the safety and reliability of the entire power grid. Therefore, it is essential to develop cost-effective asset management methods to assess its health condition and provide a timely decision to rectify emerging faults and avoid any potential catastrophic consequences. Over decades, many condition monitoring methods have been evolved to detect various faults in power transformers [1]. In industry practice, one of the most widely used methods for analysing power transformer oil to detect incipient faults is the Dissolved Gas Analysis (DGA) method [2]. This technique has been proven to be effective in identifying potential issues in power transformers and has become an essential asset management tool.

The fundamental principle of DGA is to measure the levels of various dissolved gases in the transformer oil. These gases are generated as a result of overheating, arcing and partial discharge events [3]. By analysing the type and concentration of dissolved gases in transformer oil samples, potential faults can be identified, and rate of insulation degradation can be assessed; allowing timely maintenance and repair plans to prevent potential severe damages to the transformer. The measured dissolved gases in transformer oil include Hydrogen (H_2), Methane (CH_4), Ethylene (C_2H_4), Ethane (C_2H_6), Acetylene (C_2H_2), Carbon Monoxide (CO) and Carbon Dioxide (CO_2). These gases have been considered as key

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

indicators of potential transformer faults, and their concentrations can provide valuable insights into the overall health condition of the transformer.

Several conventional DGA interpretation methods, derived from ANSI/IEEE standard and IEC publication 599, have been extensively utilized in the power industry [3], [4]. These methods include Key gas method, Rogers Ratios, Doernenburg Ratios, Duval Triangles and Pentagons graphical methods. However, each of these methods exhibits certain limitations such as out-of-code ratios, distinct boundaries, and the exclusion of gas evolution, which may lead to incorrect and inconsistent fault diagnoses [5]. As a result, the accuracy of diagnostic results relies more on the level of experience of the professionals conducting the test. To overcome such subjective interpretations, researchers have developed several artificial intelligence (AI)-based methods to improve the diagnostic accuracy. In [5], various AI-based methods for DGA interpretation have been presented. One of the methods is Fuzzy Logic, which replaces the precise values of input-output variables with a range of values to handle the uncertainties and imprecisions in the DGA data [6], [7]. Other AI methods such as Decision Tree [8], Random Forest [9], k-Nearest Neighbours (KNN) [10], [11], Logistic Regression [12], [13], Support Vector Machine (SVM) [14], [15], Bayesian Network [16], Artificial Neural Network (ANN) [17], [18], [19], Adaptive Neural Fuzzy Inference System (ANFIS) [20], [21] have also shown promising results in enhancing the analysis of DGA data. These methods are designed for various levels of data complexity, often requiring the use of multiple training models to analyse DGA data comprehensively. For example, ANFIS model encounters difficulties and may get stuck in the training process. On the other hand, fuzzy logic requires ample number of fuzzy rules, which complicate and reduce the accuracy of the DGA models of multiple input and output parameters.

In recent years, another powerful AI tool called Deep Machine Learning algorithm has emerged to solve complex problems that were previously challenging for traditional AI methods. Deep learning is modelled to mimic the hierarchical structure of the human brain and is designed to process data in a similar way, starting with lower-level features and gradually building up to higher-level concepts [22]. By doing so, deep learning models are able to handle complex problems with large data sets. This has made them highly effective for a wide range of applications, including image and sound recognition.

As the concept of future smart grids continues to evolve, the online monitoring of key assets including power transformers has become more prevalent. This will lead to a substantial rise in the amount of sophisticated data being collected and analysed. Some of these raw data are presented in numerical form, while others may be on the form of images or sound signals, such as thermal and vibration analyses [23], [24]. This requires adaptable diagnostic methods with enhanced learning and feature extraction capabilities to effectively reflect meaningful insights from the measured data. In [25] and [26], a specific type of deep machine learning called the probabilistic neural network with optimizer was developed to provide fault diagnosis in power transformers based on five DGA gas measurements. The output of the method identifies four potential fault conditions, including high/low temperature fault, partial discharge, and arc discharge.

This paper aims to provide a more comprehensive asset management solution for mineral oil-immersed power transformers. In addition to the fault diagnostic module, the developed asset management system also comprises a life management module to provide asset managers with the deterioration level of the solid insulation based on the amount of CO_2 and CO gases obtained from online DGA measurement sensors. The fault diagnostic module encompasses ''no fault'' condition, in addition to identifying thermal fault, arc discharge and partial discharge as will be elaborated below.

II. METHODOLOGIES UTILIZED IN DEVELOPING THE ASSET MANAGEMENT MODEL

A. DATA PRE-PROCESSING–NORMALIZATION

The aim of data normalization in the data pre-processing stage is to bring features into a comparable scale to enhance the model performance and improve the training stability [27]. In the training datasets, the gas concentration exhibits a wide range from 0 to 100,000s ppm. To prevent the dominance of large values on training weights, which could potentially distort the training results [28], a commonly utilized normalization method called Minimum-Maximum normalization as given by (1) is employed [29]. The Minimum-Maximum normalization technique scales the data into a range between 0 and 1, based on the minimum and maximum values in the datasets. Figure 1 (a) depicts the raw data distribution of H_2 gas measurements (in ppm) collected from an online DGA sensor, while Figure 1 (b) illustrates the data distribution after normalization. Comparison of the two figures shows that the normalization process does not alter the essential features of the collected data.

$$X_{nor} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where X, is the original value before normalization. X_{nor} is the X value after normalization. X_{max} and X_{min} represent the maximum and minimum of values of X value; respectively.

B. PRE-PROCESSING–BALANCE DATASETS

The training datasets often exhibit uneven distribution, where minority classes are vastly outnumbered by majority classes. When the model is trained using such imbalanced dataset, machine learning algorithms tend to favour the majority classes, which may lead to potential misclassification of the minority classes [30]. Based on the datasets presented in Table 1, it can be observed that the various transformer conditions are not evenly distributed. Specifically, the occurrence of partial discharge faults constitutes a relatively small percentage of all conditions (accounting for only 8.7%). On the other hand, energy discharge faults and thermal faults are

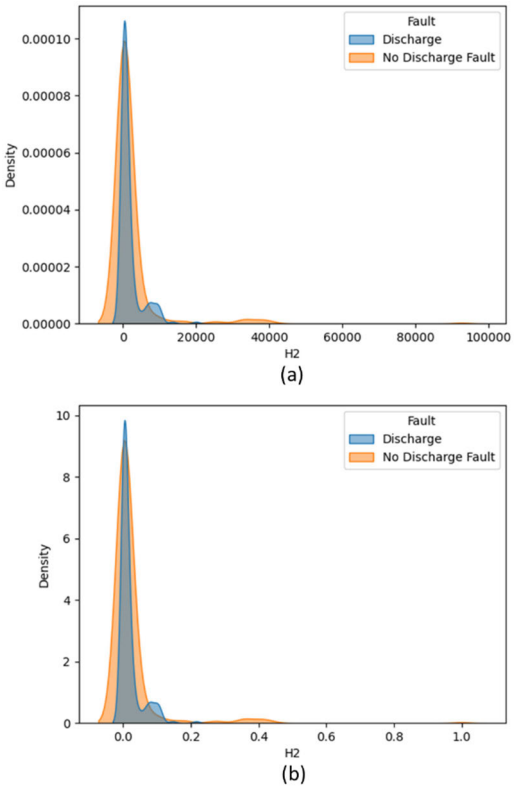


FIGURE 1. (a) Raw H2 data (in ppm) collected from DGA sensor. (b) Data processed using Minimum-Maximum normalization.

much more prevalent, comprising approximately 35% and 41.3% of the total conditions, respectively. Interestingly, the normal condition makes up 15.7% of the conditions, which is noteworthy given that in real-world scenarios, normal conditions tend to be the most commonly observed results [31]. The imbalance nature of the datasets could potentially have a significant impact on the outcomes of machine learning algorithms.

Another crucial issue that must be considered when utilizing machine learning algorithms is their ability to generalize beyond the training datasets. In the construction of the training datasets, the transformer fault types are typically identified based on the expert judgements, which are often derived from conventional interpretation methods such as IEEE/IEC ratio and Duval Triangles/Pentagons. However, the heuristic nature of expert judgments, coupled with the fact that transformers may exhibit multiple faults simultaneously, can lead to varying degrees of inconsistency between datasets collected from different sources. This inconsistency in the training data can pose a significant challenge to the generalization capability of machine learning algorithms and may ultimately undermine their effectiveness in practical applications.

To overcome the above-mentioned issues of imbalanced data, Synthetic Minority Over-sampling Technique (SMOTE) has been implemented [53], [54]. The principle of SMOTE is to generate synthetic samples for minority classes. It begins by randomly selecting a data point from the minority class and identifying its k nearest neighbours. SMOTE then places a synthetic point along the line connecting the chosen data point and one of its nearest neighbours. These steps are repeated until the dataset is balanced; thereby ensuring a more even representation of all classes in the training data. Figure 2 presents a comparative histogram illustrating the effect of data balancing through SMOTE processing. Figure 2 (a) provides an overview of the distribution of multi-class targets within the initial training dataset before the application of SMOTE. Each bin within the histogram corresponds to a distinct label. Specifically, the x-axis denotes the labels associated with the dataset. In this context, label "1" pertains to Thermal fault, label "2" signifies partial discharge (PD) fault, label "4" indicates No Fault, label "8" represents Discharge fault, and label "9" corresponds to a combination of Discharge and Thermal faults. The y-axis represents the frequency (number of instances) that belongs to each label. On the other hand, Figure 2 (b) shows the distribution of labels in the training data after applying SMOTE. As can be observed, the frequency of labels "2", "4", "8", and "9" have been changed due to the introduction of synthetic samples.

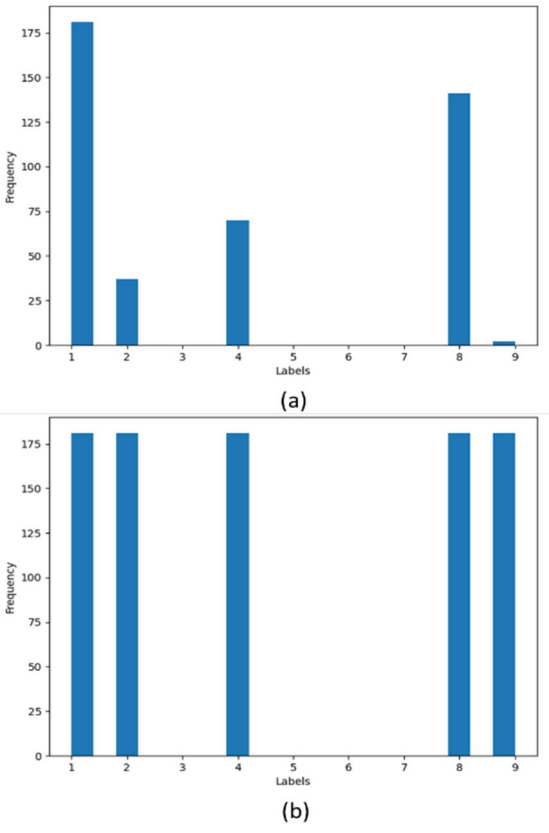


FIGURE 2. (a) Data distribution before SMOTE and (b) Data distribution after SMOTE. (for reference to label numbers, see Table 3).

C. CONVOLUTION NEURAL NETWORK

Convolution Neural Network (CNN) has emerged as a popular and effective deep learning technique, particularly for

TABLE 1. Datasets for fault diagnostic module training.

Reference	Energy Discharge	No fault	Partial Discharge	Thermal Fault	Total # of samples
[11]	6	0	5	9	20
[21]	3	0	1	3	6
[7]	3	0	2	6	11
[32]	4	0	0	6	10
[33]	4	0	2	4	10
[34]	6	12	4	12	34
[35]	2	0	1	3	6
[36]	2	0	1	3	6
[37]	8	0	4	12	24
[38]	6	13	3	11	33
[39]	10	5	3	7	25
[40]	28	9	0	38	75
[41]	8	1	1	12	22
[19]	15	8	0	8	31
[42]	3	0	1	4	7
[43]	4	2	0	10	16
[16]	1	2	1	2	6
[44]	10	0	3	23	36
[45]	4	0	7	9	20
[46]	3	1	1	5	10
[47]	4	0	1	6	8
[48]	2	2	0	0	4
[49]	0	4	1	5	10
[50]	4	0	3	2	9
[4]	74	34	9	34	151
[51]	8	6	1	15	30
[2]	5	2	0	13	20
[52]	6	3	3	12	24
Total	233	104	58	274	664
Percentage	35%	15.7%	8.7%	41.3%	

analysing visual data [23]. In recent years, there has been growing interest in leveraging CNN for condition monitoring applications. A study published in [23] has utilized CNN to identify six types of PD faults in power transformers by analysing Phase-Resolved Partial Discharge (PRPD) signals captured by Ultra-High Frequency (UFH) sensors. Another investigation in [24] focused on using CNN to assess transformer winding conditions through the analysis of vibration signatures. Furthermore, [55] explored the application of CNN in assessing oil quality based on oil aging images. Considering the future advancements and ongoing developments of online condition monitoring methods, CNN holds a great potential as a powerful tool that can provide more comprehensive condition assessment of power transformers in real-time. This will provide a more accurate and holistic understanding of transformer performance, facilitate timely maintenance interventions, and extend the operational lifespan of power transformers.

The structure of a CNN comprises two primary components: feature selection and conventional neural network (Figure 3) as shown in Figure 4. The feature selection layers encompass convolutional layers, pooling layers, batch normalization layers, and flatten layers [56]. The convolutional layers play a significant role in feature extraction, applying filters or kernels to the input data to capture patterns and spatial dependencies. The pooling layers reduce the spatial dimensions of the resulting feature maps; effectively summarizing the learned features. The batch normalization layers normalize the outputs of the previous layers to enhance training stability and accelerate convergence. Finally, the flatten layers transform the multidimensional feature maps into a one-dimensional vector. This process prepares the extracted features for further processing in neural network. The feature selection process in CNN provides the key advantage over traditional machine learning algorithms, such as Support Vector Machine (SVM). CNN can autonomously learn intricate features and patterns directly from the raw input data that eliminates the need for manual feature engineering. This capability significantly reduces the burden of feature extraction and enhances the overall efficiency of the model.

The neural network component comprises fully connected layers, which integrate the extracted features and make predictions based on the learned representations. In Figure 3, a neural network with 2 hidden layers is depicted. The inputs $x_1 \ldots x_i$, contain the features of the input data that are fed into the network. Within the hidden layers, each neuron takes

input from the previous layer, applies a weight (i.e., w_{ij} or w_{jk}) and a bias (i.e., b_j or B_k), and passes the results through an activation function as presented by (2). During the training process, the backpropagation algorithm determines the weights along with biases of the neural network to minimize the error of difference between the predicted output and the target or desired output. This adjustment is performed using an optimization algorithm, such as Adaptive Moment Estimation. The optimization algorithm updates the weights based on the calculated error and the network's learning rate, which controls the step size of the weights updates.

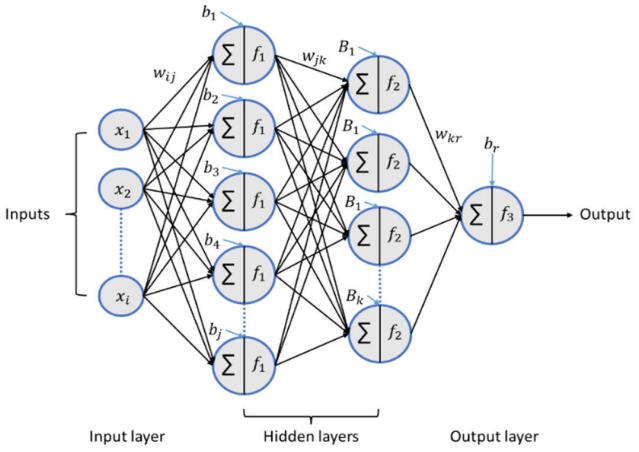


FIGURE 3. General structure of neural networks with 2 hidden layers.

Activation functions play a critical role in transforming the weighted sum of inputs and bias into an output value. For example, the Rectified Linear Unit (ReLU) activation function is commonly used in neural networks, which results in the input value if it is positive, and zero otherwise. Another example is Sigmoid function, which squashes the input value into a range between 0 and 1, in case output needs to be interpreted as probabilities.

$$Y_j = f_1\left(\sum_{j=1}^{i} (w_{ij} \times x_i + b_j)\right) \tag{2}$$

where, Y_j is the output of the first hidden layer .x_i is the input. f_1 is the activation function. w_{ij} is the weight and b_j is the bias.

III. PROPOSED ASSET MANAGEMENT MODEL

The asset management approach proposed, as depicted in Figure 5, utilizes gas measurements acquired from an online DGA sensor. This model consists of two distinct submodules: a Fault Diagnostic module and a Life Management module.

In this section, a detailed explanation of the training process for each module is presented, along with insights into the possible outputs derived from these modules.

A. FAULT DIAGNOSTIC MODULE

There are six primary types of faults that can be identified using DGA method, as outlined in Table 2 based on the IEC 60599 and IEEE57.104 [3]. Due to the limited information available in the datasets, fault types have been classified into 3 broader categories: Partial Discharge (PD), Energy Discharges and Thermal Faults.

Partial Discharge occurs when a localized area of solid or fluid insulation material, exposed to high voltage stress, undergoes a partial breakdown without fully bridging the gap between two conductive materials [2], [3], [57]. In this context, PD specifically refers to corona-type PD that transpires within gas bubbles or voids. During PD activities, air or nitrogen in the gas phase undergoes ionization, forming a plasma of ionized oxygen and nitrogen atoms.

TABLE 2. Six fundamental types of faults diagnosed using DGA.

Code	Primary Faults
T1	Thermal fault; T < 300°C
T2	Thermal fault; 300°C < T < 700°C
T3	Thermal fault; T > 700°C
D1	Low energy discharge
D2	High energy discharge
PD	Corona Partial Discharge

This plasma interacts with the surrounding oil or cellulose, leading to the generation of hydrogen as the primary by-product.

Energy Discharges in oil-immersed power transformers occur when there is an energy discharge that creates a localized conducting path or short circuit between conductive materials [2], [3], [57]. This leads to sparking around loose connections within the transformer. When low energy arcs, denoted as D1 in Table 2, occur in transformer oil, only a thin layer of the oil makes contact with the path of the arc. The high temperature of the arc, exceeding 3000°C, causes decomposition of this small oil layer. The decomposition primarily yields acetylene, with traceable amount of ethylene being produced. In contrast, high energy arcs, referred to as D2 in Table 2, involve a greater current flowing through the arc path, resulting in a longer duration. The extended duration allows for a larger volume of oil to be heated by the arc. The convective flow of cooler surrounding oil contributes to this process. Consequently, a significant temperature gradient is established in the oil surrounding the arc path, ranging from around 3000°C to 500°C. Interestingly, despite the higher energy content of D2 arcs, their average oil temperature is lower compared to D1 arcs. However, D2 arcs generate a substantial amount of C_2H_4 in addition to C_2H_2 due to the temperature gradient and longer duration of the arc.

Thermal Faults in oil-immersed power transformers arise from the circulation of electric current within the insulating paper due to excessive dielectric losses [2], [3], [57]. Thermal faults can be classified into three categories: T1, T2 and T3. T1 fault occurs when there is an increase in the average winding temperature, typically caused by increased load or

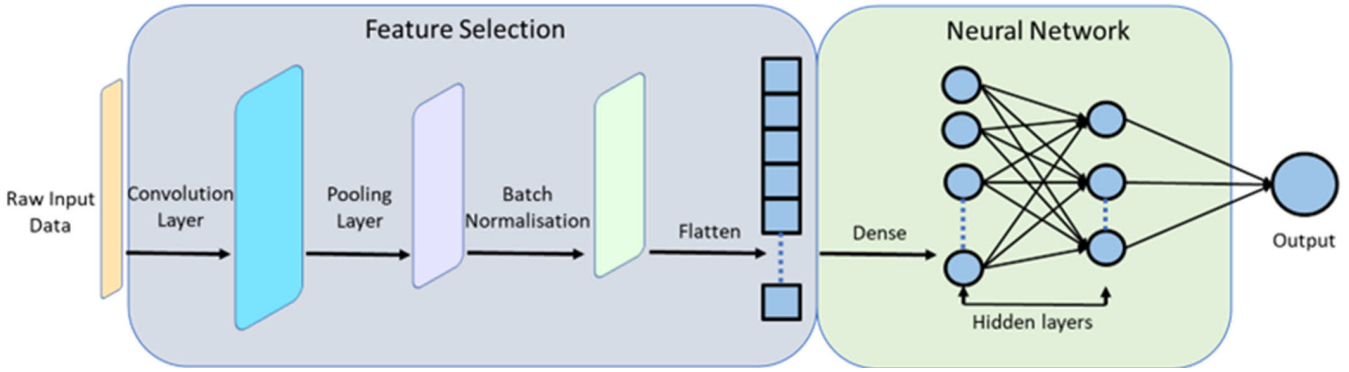


FIGURE 4. The structure of convolution neural network.

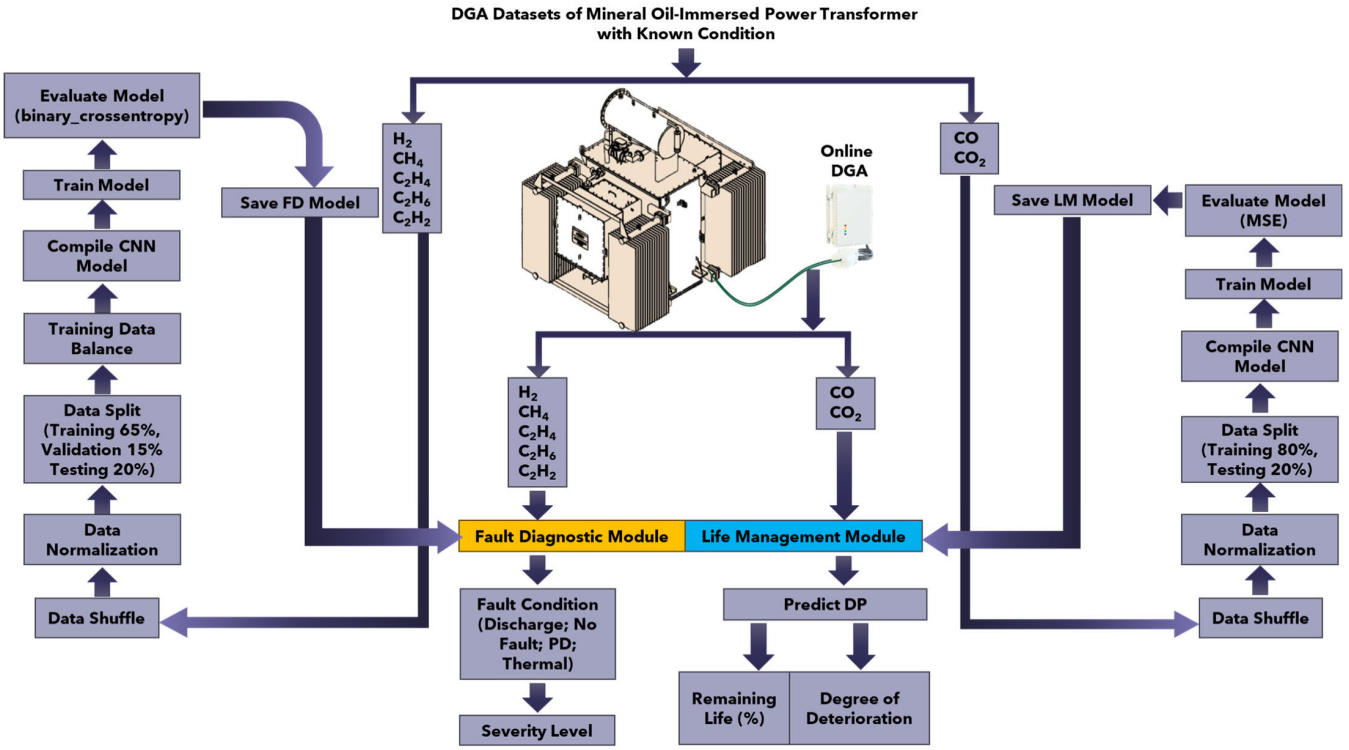


FIGURE 5. Proposed workflow for the oil-immersed power transformer asset management model.

ambient temperature. T2 fault, on the other hand, involves localized hotspots within the winding, resulting from electrical contact or excessive current density. Localized hotspots generate higher temperatures, which accelerate the insulation aging and compromise the transformer's overall performance. Lastly, T3 fault encompasses more severe conditions, such as arcing or short circuits, which cause significant temperature rise and pose an immediate risk to the transformer's operation and safety.

The proposed Fault Diagnostic module shown in Figure 5 employs CNN to analyse the measurements of five gases: H_2, CH_4, C_2H_4, C_2H_6, and C_2H_2. As introduced in the previous section, CNN has proven to be highly effective in processing complex data, making it ideal candidate for fault diagnosis in power transformers. Based on the concentrations of these gases, the module classifies the transformer's condition into four categories: ''Discharge'', ''no Fault'', ''PD'', and ''Thermal''. Moreover, the module is capable of diagnosing combined fault conditions, such as ''Discharge and Thermal'', adding further versatility to improve diagnostic accuracy.

In the context of enabling the module for multi-label classification tasks, a binary representation, as shown in Table 3 has been adopted. The proposed approach utilizes a four-digit binary representation based on 2^n, where 'n' is set to four in this application, accommodating the representation of up to sixteen possible cases.

Based on the output of the Fault Diagnostic module, an asset management decision will be provided to users with the severity level as listed in Table 4.

TABLE 3. Possible conditions represented using binary numbers.

Case	Discharge	no Fault	PD	Thermal	Diagnosis
0	0	0	0	0	No Fault
1	0	0	0	1	Thermal Fault
2	0	0	1	0	PD Fault
3	0	0	1	1	PD, Thermal Fault
4	0	1	0	0	No Fault
5	0	1	0	1	Thermal Fault
6	0	1	1	0	PD Fault
7	0	1	1	1	PD, Thermal Fault
8	1	0	0	0	Discharge Fault
9	1	0	0	1	Discharge, Thermal Fault
10	1	0	1	0	Discharge, PD Fault
11	1	0	1	1	Discharge, PD &Thermal Fault
12	1	1	0	0	Discharge Fault
13	1	1	0	1	Discharge, Thermal Fault
14	1	1	1	0	Discharge, PD Fault
15	1	1	1	1	Discharge, PD &Thermal Fault

TABLE 4. Fault diagnostic module outputs and corresponding asset management statement.

Module Output	Asset Management Statement – severity level
No Fault	"No Fault detected, very low likelihood of failure."
PD	"Sign of Partial Discharge fault, low risk of failure."
Thermal Fault	"Sign of Thermal fault, moderate risk of failure"
PD & Thermal Fault	"Sign of Thermal fault, moderate risk of failure"
Discharge Fault	"Sign of Discharge fault, very high risk of failure."
Discharge & Thermal	"Sign of Discharge fault, very high risk of failure."
Discharge & PD	"Sign of Discharge fault, very high risk of failure."
Discharge, PD & Thermal	"Sign of Discharge fault, very high risk of failure."

B. LIFE MANAGEMENT MODULE

The life expectancy of power transformers predominantly relies on their paper insulation [5]. The composition of transformer paper is primarily comprised of 90% cellulose by weight. Cellulose is an organic compound characterized by lengthy chains of glucose rings, typically ranging between 1000 to 1200 per chain for new paper [5]. The Degree of Polymerization (DP) refers to the average number of glucose rings present within each chain. Based on experience, it is commonly considered that transformers reach their end-of-life when the DP of paper declines to 200, which means the tensile strength decreases to approximately 40% of its initial value [58], [59]. However, measuring DP requires a paper sample from the transformer, which is not feasible for in-service transformers.

In the present industry practice, the estimation of DP value is often conducted through the analysis of furan compounds. Furans are generated as by-products during the degradation of paper insulation impregnated with oil. Among the five furan compounds, 2-furfural (2-FAL) is commonly utilized as a predictor of DP due to its higher stability compared to other compounds [58]. However, the development of online furan measuring methods is still an ongoing research area.

Previous research has revealed a correlation coefficient of 0.87 between 2-FAL and DP [58]. Notably, the ratio of carbon dioxide to carbon monoxide (CO_2/CO) exhibits the highest correlation coefficient of 0.97 [58], emphasizing its superior stability as an indicator for assessing the condition of insulation paper. It is important to highlight that CO_2 and CO are also generated through the oxidation of oil [58]. The utilization of this ratio is justified by the fact that in situations of high thermal and arcing faults, CO shows a much more rapid increase compared to CO_2 [60]. Conversely, during significant heat generation in normal operation conditions, CO_2 increases at a faster rate than CO.

Unlike conventional machine learning approaches, deep learning has the capability to directly incorporate measurements of CO_2 and CO as input features, thereby eliminating the requirement for extra features such as CO_2/CO ratio.

The proposed Life Management module, also depicted in Figure 5, utilizes a CNN regression approach, incorporating simply two-gas measurements CO_2 and CO to forecast the DP value.

Based on the DP value, the Life Management module delivers an evaluation of the paper insulation condition and categorize it into four levels: ''Healthy insulation'', ''Moderate deterioration'', ''Extensive deterioration'', and ''End of insulation life''.

TABLE 5. DP values and their associated significance [61].

DP Value	Significance
1200-700	Healthy Insulation
700-450	Moderate Deterioration
450-250	Extensive Deterioration
< 250	End of Life

The inclusion of the estimated percentage of remaining life within the asset management framework does not only enhance the precision of the assessments but also provide a dynamic understanding of the insulation condition. An estimation for the percentage of remaining life based on the DP value can be conducted using (3) [62]. Upon review of Table 5, ''Healthy Insulation'' category corresponds to DP

values ranging from 700 to 1200. In order to align the model with real-world applications more effectively, a DP value of 800 is considered to be corresponding to 100% transformer remaining life as per (3).

$$\%\ \text{of remaining life} = 166.1 \times \log_{10}(DP) - 382.2 \quad (3)$$

IV. RESULTS AND DISCUSSION

A. FAULT DIAGNOSTIC MODULE

The database comprises a collection of 1083 DGA samples incorporated from the 29 literatures listed in Table 1. 65% of the samples were randomly selected for the training phase, with an additional 15% allocated for the validation phase while the remaining 20% were designated for the testing phase.

The Fault Diagnostic module employs the capabilities of a one-dimensional CNN, which has an architecture adept at processing sequential data. Within this framework, several adjustable parameters play crucial roles in shaping the model's performance. These parameters include the filter size, which determines the width of the convolutional filters employed to extract features from the input data. Additionally, the kernel size dictates the scope of each convolutional operation, influencing the receptive field of the network. The choice of padding, whether 'valid' or 'same', modifies the dimensions of the output feature maps. Lastly, the activation function governs the non-linearity introduced within the network, contributing to its ability to capture complex patterns and relationships within the data.

The module training process involves an exhaustive exploration of various configuration settings. Different options were tested at filter sizes: 32, 64, and 128, along with varying numbers of neural network layers. Throughout these experiments, layers have been systematically added to assess their impact on the model's performance. Despite the array of layer configurations tested, the final set of layers and parameters that yielded optimal results are as shown in Table 6.

TABLE 6. Optimal parameters of the developed fault diagnostic module.

Layer	Parameter	Setting
Convolutional layer	Filter size	64
	Kernel size	3
	Padding	'same'
	Activation function	'ReLU'
Dense layer	Number of neurons	16
	Activation function	'ReLU'
Dense_1 layer	Activation function	'sigmoid'

The ReLU activation function transforms negative input values to zero while leaving positive values unchanged. The ReLU function finds extensive application in neural networks across diverse domains owing to its efficient computation and improved gradient propagation, thus facilitating the extraction of significant features from the input data.

The sigmoid activation function possesses the ability to condense input values within a range between 0 and 1, as depicted by (4).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

The sigmoid function exhibits an S-shaped curve is capable of transforming both positive and negative input values into probabilities. Thus, binary classification will be provided. For example, if both "Discharge" and "Thermal" faults are present, the predicted probabilities might be presented as [0.8, 0.2, 0.4, 0.7] (["Discharge", "no Fault", "PD", "Thermal"]); with more probability assigned to the present faults. It's noteworthy that each output probability is determined independently, meaning the prediction for one condition does not influence the prediction of another.

During the model compilation phase, the 'Nadam' optimizer, which is a combination of the Nesterov Accelerated Gradient (NAG) and Adam optimizers has been applied. The selection of 'binary_crossentropy' as the loss function, as given by (5), is a common choice for binary classification problems. It measures the dissimilarity between predicted probabilities and true labels (0 or 1), thus optimizing the model to achieve accurate binary predictions.

$$\text{Binary Cross} - \text{Entropy Loss} = -[y_i \times \log(\hat{y}_i) + (1 - y_i) \times \log(1 - \hat{y}_i)] \quad (5)$$

where, y_i is the actual target value (0 or 1) of the i-th data point. $\hat{y}_i$ is the predicted value of the i-th data point generated by the model.

Lastly, 'accuracy', a standard evaluation metric, is used for classification tasks. It calculates the ratio of correctly predicted instances to the total number of instances to provide insight into the model's overall performance.

During the last phase of model training, the model's weights are updated based on the provided training data. The training process involves passing the training data through the network, computing predictions, comparing them with the actual targets, and then backpropagating the error to update the model's weights. 'Epochs' defines the number of times the model will iterate over the entire training dataset, which is 1000 times in this case. The batch size determines the number of training examples the model processes in each update of the gradient. Smaller batch sizes may lead to more frequent updates, while larger batch sizes can speed up the training process. A batch size of 16 has been chosen for the developed model based on running through many simulations with different batch sizes.

The training process of the model randomly runs due to the random initial weights, leading to varying results in accuracy and loss. Following parameters adjustments, the training model has been executed several times, and the run producing the highest accuracy and lowest loss is selected. The generated plots shown in Figure 6 provide valuable insights into the training process and the performance of the developed CNN model. The alignment or divergence of the curves reveals the overfitting or underfitting phenomena and guides potential adjustments in the model architecture

or hyperparameters for optimal performance. In Figure 6(a), the 'Training loss' curve, depicted in yellow, shows how the model's loss decreases as it learns to better fit the training data. The 'Validation loss' curve, depicted in red, demonstrates the model's performance on unseen validation data. A decreasing validation loss over epochs indicates successful generalization of the model. In some runs, it was noticed the validation loss started to rise after a certain number of epochs while the training loss curve was stable. This overfitting phenomenon indicates that the model has started to memorize the training data instead of capturing underlying patterns. In Figure 6(b), the 'Training acc' curve in yellow illustrates the model's accuracy using the training data, whereas the 'Validation acc' curve in red showcases the model's performance using validation data. As epochs progress, observed increasing validation accuracy that aligns with the training accuracy reflects the model's ability to generalize and predict unseen data accurately.

The satisfied model performance was achieved with an accuracy of 0.8479 and a corresponding loss of 0.2989 using the test dataset. This model was subsequently saved for the Fault Diagnostic module. To test the Fault Diagnostic module, new gas measurements were fed into the module, enabling the prediction of outputs through its learned capabilities.

In order to evaluate the performance of the Fault Diagnostic module, a dataset comprising 151 samples from the IEC TC10 database was used [4]. Subsequently, a thorough assessment of misdiagnoses within each class was carried out. Notably, the analysis revealed that misdiagnosis predominantly occurred within the "No Fault" condition as revealed by the confusion matrix of the Fault Diagnostic module shown in Figure 7. This may be attributed to the fact that the majority of the collected datasets were predominantly geared towards the identification of fault conditions.

The samples presented in Table 7 originate from the IEC TC 10 database (samples 1-21) and Korea Electric Power Corporation (KEPCO) historical data [4], [31]. The second last column in the table shows the actual condition of the transformers as determined through physical inspection while the last column lists the diagnostic results generated from the proposed Fault Diagnostic module. Certain discrepancies have been identified in the samples #4, #13, #16, #17, #20, and KEPCO's samples. In Table 8, traditional IEEE and IEC DGA interpretation methods are used to analyse all samples and compare the results with those obtained from the module. The following observations can be drawn out of these comparisons:

1) In the case of sample #4, the actual condition is described as "Tracking to the ground in glue of central beam", categorized as low energy discharge. Traditional methods such as Duval Triangle 1, IEEE and IEC ratios can identify this discharge fault. However, the developed module did not capture this specific fault.
2) In the case of sample #17, where the inspection outcome indicated no fault, the Roger ratio method indicated a discharge fault. However, both the Duval Triangle 1 and the developed module reached the same conclusion, identifying a thermal fault.

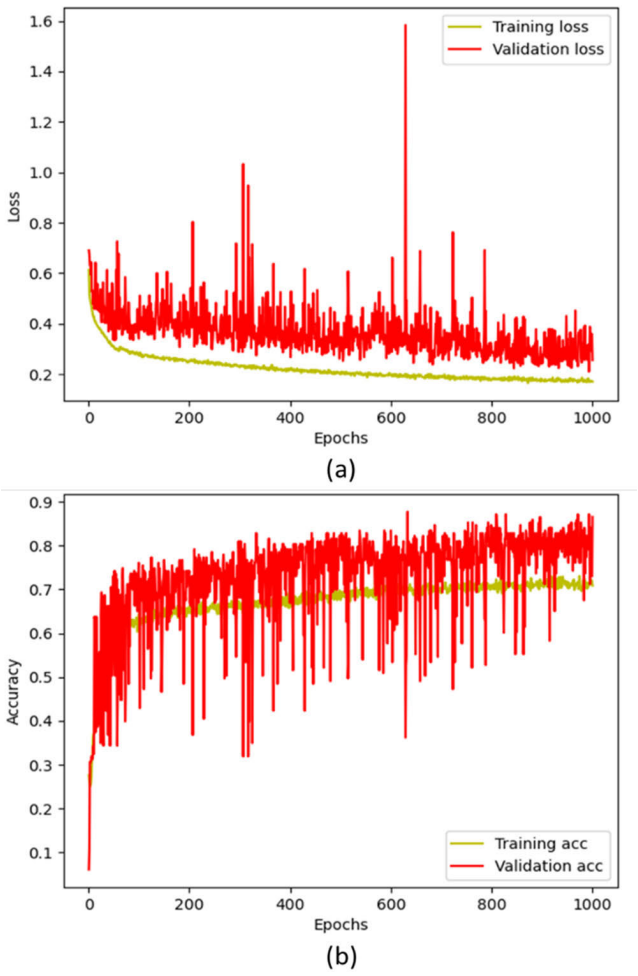


FIGURE 6. Fault diagnostic module: (a) Training and validation loss plot. (b) Training and validation accuracy plot.

Predicted condition \ Actual condition	Discharge	No Fault	PD	Thermal	
Discharge	70 46.36%	8 5.30%	0 0%	2 1.32%	87.50% 12.50%
No Fault	4 2.65%	23 15.23%	0 0%	3 1.99%	76.70% 23.30%
PD	0 0%	0 0%	9 5.96%	0 0%	100% 0%
Thermal	0 0%	3 1.99%	0 0%	29 19.20%	90.60% 9.40%
	94.60% 5.40%	67.60% 32.40%	100% 0%	85.30% 14.70%	86.75% 13.25%

FIGURE 7. Confusion matrix of the proposed fault diagnostic module.

TABLE 7. Comparison between the diagnostic result from the fault diagnostic module and actual condition from inspection. (all gases are measured in ppm).

IEC TC10 Samples		H_2	CH_4	C_2H_4	C_2H_6	C_2H_2	Actual condition from inspection	Module's Result
1		543	120	411	41	1880	Discharge	Discharge
2		1230	163	233	27	692	Discharge	Discharge
3		645	86	110	13	317	Discharge	Discharge
4		60	10	4	4	4	Discharge	No Fault
5		95	10	11	0	39	Discharge	Discharge
6		32930	2397	0	157	0	PD	PD
7		37800	1740	8	249	8	PD	PD
8		92600	10200	0	0	0	PD	PD
9		8266	1061	0	22	0	PD	PD
10		9340	995	6	60	7	PD	PD
11		1270	3450	1390	520	8	Thermal	Thermal
12		3420	7870	6990	1500	33	Thermal	Thermal
13		360	610	260	259	9	Thermal	Discharge; Thermal
14		0	18900	540	410	330	Thermal	Thermal
15		960	4000	1560	1290	6	Thermal	Thermal
16		2031	149	3	20	0	Thermal	PD; Thermal
17		125	100	150	100	20	No Fault	Thermal
18		175	0	375	100	3	No Fault	No Fault
19		50	30	0	0	5	No Fault	No Fault
20		250	150	250	150	150	No Fault	Discharge; Thermal
21		134	224	224	550	154	No Fault	No Fault
KEPCO Samples		**H_2**	**CH_4**	**C_2H_4**	**C_2H_6**	**C_2H_2**	**KEPCO Health State**	**Module's result**
Transformer 1	1999	0	6	2	2	0	No Fault	No Fault
	2000	0	25	13	9	0	No Fault	No Fault
	2001	0	35	37	31	0	No Fault	Thermal
	2002	0	44	28	85	0	No Fault	Thermal
	2003	251	139	256	123	1064	Electrical Fault	Discharge
Transformer 2	2011	10	7	2	5	0	No Fault	No Fault
	2012	13	11	3	26	0	No Fault	No Fault
	2013	48	24	12	63	14	No Fault	Discharge
	2015	335	246	1324	150	1123	Electrical Fault	Discharge
Transformer 3	2000	0	1	5	0	0	No Fault	No Fault
	2002	0	7	11	14	0	No Fault	No Fault
	2003	0	64	150	99	0	No Fault	Thermal
	2004	218	744	1743	264	7	Thermal Fault	Thermal
Transformer 4	2000	5	44	4	9	0	No Fault	No Fault
	2001	6	42	10	9	0	No Fault	No Fault
	2002	6	44	12	10	0	No Fault	No Fault
	2003	7	56	12	10	0	No Fault	No Fault
	2004	628	1381	1873	351	2.8	Thermal Fault	Discharge; Thermal

3) In the case of sample #20, the inspection result denoted no fault. In contrast, the module and the Duval Triangle 1 method yielded results indicating combined discharge and thermal faults. The Roger ratio method, on the other hand, identified a discharge fault.
4) Observations reveal that the module occasionally indicated the presence of multiple faults, as seen in samples #13 and #16. In the case of sample #13, the module detected a discharge fault in addition to a thermal fault. Notably, if the traditional methods were applied to analyse sample #13, the results would align with the inspection result. For sample #16, the module detected a partial discharge fault alongside a thermal fault. Interestingly, when the IEC ratio and Duval Triangle methods were applied to sample #16, both identified a partial discharge fault, in contrast to the inspection result, which revealed a thermal fault.

TABLE 8. Comparison of diagnostic results: proposed module vs. the traditional methods. (all gases are measured in ppm).

IEC TC10 Samples	Actual condition from inspection	Roger Ratio Method	Doernenburg Ratio Method	IEC Ratio Method	Duval Triangle	Module's result
1	Discharge	N/A	Discharge	Discharge	Discharge	Discharge
2	Discharge	Discharge	Discharge	Discharge	Discharge	Discharge
3	Discharge	Discharge	Discharge	Discharge	Discharge	Discharge
4	Discharge	Discharge	N/A	Discharge	Discharge	No Fault
5	Discharge	N/A	Discharge	N/A	Discharge	Discharge
6	PD	N/A	N/A	N/A	PD	PD
7	PD	N/A	PD	PD	PD	PD
8	PD	N/A	N/A	N/A	PD	PD
9	PD	N/A	N/A	N/A	PD	PD
10	PD	N/A	N/A	N/A	PD	PD
11	Thermal	Thermal	Thermal	Thermal	Thermal	Thermal
12	Thermal	Thermal	Thermal	Thermal	Thermal	Thermal
13	Thermal	Thermal	Thermal	Thermal	Thermal	Discharge; Thermal
14	Thermal	N/A	N/A	N/A	Thermal	Thermal
15	Thermal	Thermal	Thermal	Thermal	Thermal	Thermal
16	Thermal	N/A	N/A	PD	PD	PD; Thermal
17	No Fault	Discharge	N/A	N/A	Thermal	Thermal
18	No Fault	N/A	N/A	N/A	Thermal	No Fault
19	No Fault	N/A	N/A	N/A	Discharge	No Fault
20	No Fault	Discharge	N/A	N/A	Discharge; Thermal	Discharge; Thermal
21	No Fault	N/A	N/A	Thermal	Discharge	No Fault

KEPCO Samples		KEPCO Health State	Roger Ratio Method	Doernenburg Ratio Method	IEC Ratio Method	Duval Triangle	Module's result
Transformer 1	1999	No Fault	N/A	No Fault	N/A	Thermal	No Fault
	2000	No Fault	N/A	No Fault	N/A	Thermal	No Fault
	2001	No Fault	N/A	No Fault	N/A	Thermal	Thermal
	2002	No Fault	N/A	N/A	N/A	Thermal	Thermal
	2003	Electrical Fault	N/A	Discharge	N/A	Discharge	Discharge
Transformer 2	2011	No Fault	Thermal	No Fault	N/A	Thermal	No Fault
	2012	No Fault	N/A	No Fault	N/A	Thermal	No Fault
	2013	No Fault	N/A	N/A	N/A	Discharge	Discharge
	2015	Electrical Fault	Discharge	Discharge	Discharge	Discharge	Discharge
Transformer 3	2000	No Fault	N/A	No Fault	N/A	Thermal	No Fault
	2002	No Fault	N/A	No Fault	N/A	Thermal	No Fault
	2003	No Fault	N/A	N/A	N/A	Thermal	Thermal
	2004	Thermal fault	Thermal	Thermal	Thermal	Thermal	Thermal
Transformer 4	2000	No Fault	Thermal	No Fault	N/A	Thermal	No Fault
	2001	No Fault	Thermal	No Fault	Thermal	Thermal	No Fault
	2002	No Fault	Thermal	No Fault	Thermal	Thermal	No Fault
	2003	No Fault	Thermal	No Fault	Thermal	Thermal	No Fault
	2004	Thermal Fault	Thermal	Thermal	Thermal	Thermal	Discharge; Thermal

TABLE 9. Optimal parameters of the developed life management module.

Layer	Parameter	Setting
Convolutional layer	Filter size	128
	Kernel size	1
	Activation function	'ReLU'
Dense layer	Number of neurons	64
	Activation function	'ReLU'
Dense_1 layer	Number of neurons	1

TABLE 10. CO_2 and CO concentration ranges and the corresponding deterioration level [64].

CO_2	CO	Significance
0 - 2500	0 - 350	Health Insulation
2500 - 4000	350 - 570	Moderate Deterioration
4000 - 10000	570 - 1400	Extensive Deterioration
≥ 10000	≥ 1400	End of Life

5) Another assessment was conducted on four KEPCO transformers based on their annual DGA measurements. The results exhibited the capability of the developed module in early fault detection for transformers 1, 2 and 3, which enables proactive maintenance measures to be taken. In the case of transformer 4, the module not only detected the thermal fault but also identified a potential discharge fault.
6) In the overall assessment of diagnostic accuracy using IEC TC10 data, the developed module consistently demonstrates a better accuracy when compared to traditional methods. It's worth noting that the Doernenburg ratio method appears to have higher accuracy when applied to KEPCO data. This discrepancy can be attributed to a specific requirement outlined in IEEE C57.104 [3], which mandates a minimum concentration limit for at least one key gas used in the ratios before the Doernenburg method can be applied. In the KEPCO dataset, the ''no Fault'' cases appear to have very low gas centration levels, therefore, identified as ''no Fault''. On the other hand, in the IEC TC10 data, from samples #17 to #21, the gas measurements exceed the concentration limit, resulting in inconclusive outcomes when utilizing the Doernenburg ratio method. This issue also manifests in the KEPCO data, producing inconclusive results for 12 or 24 months before the actual failures may really happen.

From the above analysis, it can be concluded that the developed CNN-based fault diagnostic module comprises the following unique features:

- *Ability to identify multi-label classification*: Conventional DGA interpretation techniques like Duval Triangle 1, IEEE and IEC ratios methods predominantly pinpointed singular faults. Among them, Duval Triangle 1 method can only identify a combined discharge and thermal fault. In contrast, the CNN model exhibits the ability to discern distinct features across all four different conditions.
- *Using a single training model*: To distinguish three pivotal faults; Discharge, Partial Discharge, and Thermal faults, the conventional machine learning methods require three distinct training models to collectively assess the transformer's overall condition. However, CNN method streamlines this process by utilizing a single training model, which minimizes the training duration substantially. This simplified approach not only expedites the training phase but also produces a remarkable level of accuracy.
- *Avoiding data manipulation*: The CNN model possesses inherent feature selection capabilities, which facilitates the use of raw data directly and independently. This mechanism circumvents the need for manually engineered features like gas ratios or gas percentage. Additionally, this intrinsic capability contributes to reduce execution time, making it particularly suitable for real-time condition monitoring.

B. LIFE MANAGEMENT MODULE

The training of the Life Management module employed 47 datasets sourced from three distinct literatures. These datasets were divided into a training set comprising 80% of the data and a testing set constituting the remaining 20%.

The training model in this context also incorporated a one-dimensional CNN architecture. The specific parameters utilized in configuring the CNN architecture are detailed in Table 9.

In the phase of compiling the model, the Adaptive Moment Estimation 'adam' optimizer and 'mse' loss function are adopted. As stated above, 'adam' enhances optimization by independently adjusting learning rates for each parameter based on the historical gradients. The Mean Squared Error 'mse' loss function as given by (6), quantifies the average squared difference between the predicted values generated by a model and the target values in the dataset. It is often used as a loss function for regression models to guide the optimization process.

$$\text{MSE} = \frac{1}{n} \times \sum (y_i - \hat{y}_i)^2 \tag{6}$$

where, n is the total number of samples in the dataset used to evaluate the model performance. y_i is the actual observed value while $\hat{y}_i$ is the corresponding predicted value by the model of the i-th data point.

By setting the number of epochs to 5000 and the batch size to 16, an optimal model performance has been observed. The progression of the training process is visually depicted

TABLE 11. Comparison between the predicted result from the Life Management module and actual condition from dataset.

Sample #	CO_2	CO	Targeted DP	Condition based on targeted DP	Predicted DP	Condition based on predicted DP (using CO, CO_2)
1	812	62	1304	Healthy	1020	Healthy
2	2628	370	490	Moderate	540	Moderate
3	1498	132	963	Healthy	984	Healthy
4	1878	164	1189	Healthy	727	Healthy
5	2298	214	1149	Healthy	596	Moderate
6	4400	594	266	Extensive	338	Extensive
7	2562	146	846	Healthy	565	Moderate
8	2587	387	500	Moderate	544	Moderate
9	2502	353	654	Moderate	556	Moderate
10	4348	576	362	Extensive	339	Extensive
11	4206	586	268	Extensive	341	Extensive
12	2984	503	600	Moderate	489	Moderate
13	1815	211	785	Healthy	748	Healthy
14	4217	566	276	Extensive	342	Extensive
15	2421	372	652	Moderate	564	Moderate

in Figure 8. A substantial reduction in both the training and validation losses over the course of training can be observed from the figure.

To assess the accuracy of the developed module, a database that was not included in the training process was analysed as outlined in Table 11 [63]. As can be seen in the Table, only two samples (#5 and #7) out of the total 15 results deviate from the expected target values. This reflects an accuracy level of 86.7% for the developed model.

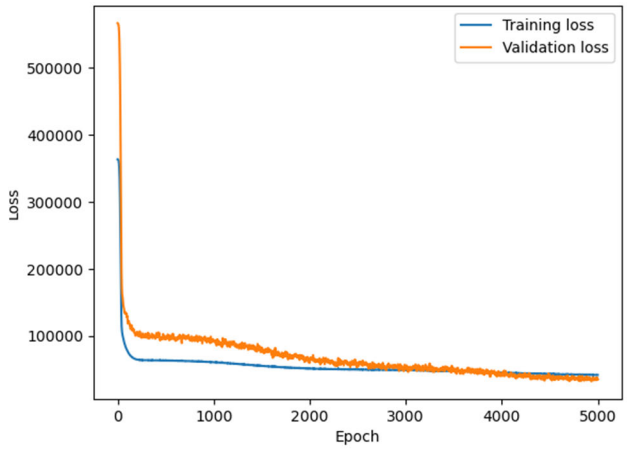


FIGURE 8. The training and validation loss plot for the life management module.

In the past, there was lack of records connecting CO_2 and CO measurements with the degree of deterioration of insulating paper. To pursue further module testing, an additional 131 CO_2 and CO measurements have been collected from diverse sources of the literature. The CO_2 and CO measurements are fed into the life management module for processing based on the ranges and corresponding conditions given in Table 10. Upon comparing the module's outcomes with the estimated results, a variance was observed in 20 out of 131 cases, demonstrating an approximate 85% concurrence rate.

C. CONTRIBUTION AND SIGNIFICANCE

Results show that the developed asset management module offers a generalized approach to evaluate power transformer condition. Constructed using diverse datasets from reputable literature sources, this model emerges as a comprehensive tool for users at all stages. The model features a continuous learning capacity, progressively enhancing its performance as it encounters new data, thus ensuring its adaptability to specific requirements, including power transformers operating within distinct conditions and environments.

The model's foundation lies in the employment of CNN deep machine learning algorithms, empowering it to assimilate fresh information and fine-tune its predictions accordingly. By harnessing this innovative model, users can access invaluable insights into the well-being of their power transformers, facilitating well-informed decisions regarding maintenance and potential replacements.

While the model was developed mainly for mineral oil immersed-power transformers due to the availability of required data, same concepts can be used to modify the model to other transformer types once sufficient data are available to train the model.

Overall, key advancements and contributions highlighted in this study include:

- Developed transformer asset management solely relying on measurements obtained from online DGA sensors.
- Empowerment of the model to interpret online DGA measurements accurately and diagnose multiple faults to provide more insights into transformer health condition.

- Estimating the DP value based on CO and CO_2 measurements and hence eliminating the need to measure furan compounds offline. This feature facilitates the online implementation of the developed asset management model.
- Employing CNN, which utilizes a single training model and requires minimal input features.

V. CONCLUSION

This paper presents a comprehensive approach for transformer asset management through the integration of two modules: Fault Diagnostics and Life Management. Both modules have undergone training utilizing deep CNN machine learning technique. This technique empowers the model to harness the potential of online DGA measurements, providing asset managers with a streamlined means to obtain highly accurate insights into the health condition of power transformers. Furthermore, the model offers indication of paper insulation deterioration in real time using the measurements of CO and CO2 that can be obtained using online DGA sensors. This feature is crucial information for effective real time asset management schemes. The adoption of CNN not only simplifies the process, but it utilizes a single training model and requires minimal input features. The proposed approach enhances the precision of predictions, hence facilitating informed decision-making for asset managers. This comprehensive strategy, encompassing fault diagnostics and life assessment, demonstrates the integration of cutting-edge technology into asset management practices, and contributes to the enhanced reliability and longevity of power transformers.

ACKNOWLEDGMENT

The authors sincerely acknowledge the support from BHP for providing a scholarship to the author Lan Jin to conduct her research on high voltage assets condition monitoring methodologies. They also acknowledge the support from Curtin University, Australia, for facilitating research resources.

REFERENCES

[1] L. Jin, D. Kim, A. Abu-Siada, and S. Kumar, "Oil-immersed power transformer condition monitoring methodologies: A review," *Energies*, vol. 15, no. 9, p. 3379, May 2022, doi: 10.3390/en15093379.

[2] G. Odongo, R. Musabe, and D. Hanyurwimfura, "A multinomial DGA classifier for incipient fault detection in oil-impregnated power transformers," *Algorithms*, vol. 14, no. 4, p. 128, Apr. 2021, doi: 10.3390/a14040128.

[3] *IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers*, IEEE Standard C57.104-2019, Revision of IEEE Std C57.104-2008, 2019, pp. 1–98, doi: 10.1109/IEEESTD.2019.8890040.

[4] M. Duval and A. dePabla, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases," *IEEE Elect. Insul. Mag.*, vol. 17, no. 2, pp. 31–41, Mar. 2001, doi: 10.1109/57.917529.

[5] L. Jin, D. Kim, and A. Abu-Siada, "State-of-the-art review on asset management methodologies for oil-immersed power transformers," *Electr. Power Syst. Res.*, vol. 218, May 2023, Art. no. 109194, doi: 10.1016/j.epsr.2023.109194.

[6] A. Abu-Siada, S. Hmood, and S. Islam, "A new fuzzy logic approach for consistent interpretation of dissolved gas-in-oil analysis," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 20, no. 6, pp. 2343–2349, Dec. 2013, doi: 10.1109/TDEI.2013.6678888.

[7] M. Žarković and Z. Stojković, "Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics," *Electr. Power Syst. Res.*, vol. 149, pp. 125–136, Aug. 2017, doi: 10.1016/j.epsr.2017.04.025.

[8] A. G. C. Menezes, M. M. Araujo, O. M. Almeida, F. R. Barbosa, and A. P. S. Braga, "Induction of decision trees to diagnose incipient faults in power transformers," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 29, no. 1, pp. 279–286, Feb. 2022, doi: 10.1109/TDEI.2022.3148453.

[9] N. Haque, A. Jamshed, K. Chatterjee, and S. Chatterjee, "Accurate sensing of power transformer faults from dissolved gas data using random forest classifier aided by data clustering method," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5902–5910, Mar. 2022, doi: 10.1109/JSEN.2022.3149409.

[10] Y. Kim, T. Park, S. Kim, N. Kwak, and D. Kweon, "Artificial intelligent fault diagnostic method for power transformers using a new classification system of faults," *J. Electr. Eng. Technol.*, vol. 14, no. 2, pp. 825–831, Mar. 2019, doi: 10.1007/s42835-019-00105-0.

[11] O. Kherif, Y. Benmahamed, M. Teguar, A. Boubakeur, and S. S. M. Ghoneim, "Accuracy improvement of power transformer faults diagnostic using KNN classifier with decision tree principle," *IEEE Access*, vol. 9, pp. 81693–81701, 2021, doi: 10.1109/ACCESS.2021.3086135.

[12] Y. D. Almoallem, I. B. M. Taha, M. I. Mosaad, L. Nahma, and A. Abu-Siada, "Application of logistic regression algorithm in the interpretation of dissolved gas analysis for power transformers," *Electronics*, vol. 10, no. 10, p. 1206, May 2021, doi: 10.3390/electronics10101206.

[13] M Božić, M Stojanović, Z Stajić, and D. Vukić, "Power transformer fault diagnosis based on dissolved gas analysis with logistic regression," *Przeglad Elektrotechniczny*, vol. 6, pp. 83–87, Jan. 2013.

[14] H. A. Illias and W. Z. Liang, "Identification of transformer fault based on dissolved gas analysis using hybrid support vector machine-modified evolutionary particle swarm optimisation," *PLoS One*, vol. 13, no. 1, Jan. 2018, Art. no. e0191366, doi: 10.1371/journal.pone.0191366.

[15] 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, doi: 10.1016/j.epsr.2011.09.012.

[16] J. I. Aizpurua, V. M. Catterson, B. G. Stewart, S. D. J. McArthur, B. Lambert, B. Ampofo, G. Pereira, and J. G. Cross, "Power transformer dissolved gas analysis through Bayesian networks and hypothesis testing," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 2, pp. 494–506, Apr. 2018, doi: 10.1109/TDEI.2018.006766.

[17] L. Cheng and T. Yu, "Dissolved gas analysis principle-based intelligent approaches to fault diagnosis and decision making for large oil-immersed power transformers: A survey," *Energies*, vol. 11, no. 4, p. 913, Apr. 2018.

[18] J. Faiz and M. Soleimani, "Assessment of computational intelligence and conventional dissolved gas analysis methods for transformer fault diagnosis," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 5, pp. 1798–1806, Oct. 2018, doi: 10.1109/TDEI.2018.007191.

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

[20] T. Kari, W. Gao, D. Zhao, Z. Zhang, W. Mo, Y. Wang, and L. Luan, "An integrated method of ANFIS and Dempster–Shafer theory for fault diagnosis of power transformer," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 1, pp. 360–371, Feb. 2018, doi: 10.1109/TDEI.2018.006746.

[21] R. A. Hooshmand, M. Parastegari, and Z. Forghani, "Adaptive neuro-fuzzy inference system approach for simultaneous diagnosis of the type and location of faults in power transformers," *IEEE Elect. Insul. Mag.*, vol. 28, no. 5, pp. 32–42, Sep. 2012, doi: 10.1109/MEI.2012.6268440.

[22] Y. Wang and L. Zhang, "A combined fault diagnosis method for power transformer in big data environment," *Math. Problems Eng.*, vol. 2017, May 2017, Art. no. 9670290, doi: 10.1155/2017/9670290.

[23] T.-D. Do, V.-N. Tuyet-Doan, Y.-S. Cho, J.-H. Sun, and Y.-H. Kim, "Convolutional-neural-network-based partial discharge diagnosis for power transformer using UHF sensor," *IEEE Access*, vol. 8, pp. 207377–207388, 2020, doi: 10.1109/ACCESS.2020.3038386.

[24] K. Hong, M. Jin, and H. Huang, "Transformer winding fault diagnosis using vibration image and deep learning," *IEEE Trans. Power Del.*, vol. 36, no. 2, pp. 676–685, Apr. 2021, doi: 10.1109/TPWRD.2020.2988820.

[25] L. Tao, X. Yang, Y. Zhou, and L. Yang, "A novel transformers fault diagnosis method based on probabilistic neural network and bio-inspired optimizer," *Sensors*, vol. 21, no. 11, p. 3623, May 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/11/3623

[26] Y. Zhou, L. Tao, X. Yang, and L. Yang, "Novel probabilistic neural network models combined with dissolved gas analysis for fault diagnosis of oil-immersed power transformers," *ACS Omega*, vol. 6, no. 28, pp. 18084–18098, Jul. 2021, doi: 10.1021/acsomega.1c01878.

[27] L. Zhang, G. Sheng, H. Hou, and X. Jiang, "A fault diagnosis method of power transformer based on cost sensitive one-dimensional convolution neural network," in *Proc. 5th Asia Conf. Power Electr. Eng. (ACPEE)*, Jun. 2020, pp. 1824–1828, doi: 10.1109/ACPEE48638.2020.9136223.

[28] C. Sun, Y. Chen, and N. Tang, "Fault diagnosis of power transformer based on DGA and information fusion," in *Proc. IEEE/IAS Ind. Commercial Power Syst. Asia (I&CPS Asia)*, Jul. 2022, pp. 247–251, doi: 10.1109/ICPSAsia55496.2022.9949927.

[29] J. Li, C. Hai, Z. Feng, and G. Li, "A transformer fault diagnosis method based on parameters optimization of hybrid kernel extreme learning machine," *IEEE Access*, vol. 9, pp. 126891–126902, 2021, doi: 10.1109/ACCESS.2021.3112478.

[30] Y. Cui, H. Ma, and T. Saha, "Improvement of power transformer insulation diagnosis using oil characteristics data preprocessed by SMOTE-Boost technique," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 21, no. 5, pp. 2363–2373, Oct. 2014, doi: 10.1109/TDEI.2014.004547.

[31] S. Kim, S.-H. Jo, W. Kim, J. Park, J. Jeong, Y. Han, D. Kim, and B. D. Youn, "A semi-supervised autoencoder with an auxiliary task (SAAT) for power transformer fault diagnosis using dissolved gas analysis," *IEEE Access*, vol. 8, pp. 178295–178310, 2020, doi: 10.1109/ACCESS.2020.3027830.

[32] S. A. Khan, M. D. Equbal, and T. Islam, "ANFIS based identification and location of paper insulation faults of an oil immersed transformer," in *Proc. 6th IEEE Power India Int. Conf. (PIICON)*, Dec. 2014, pp. 1–6, doi: 10.1109/POWERI.2014.7117715.

[33] 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, doi: 10.1109/TDEI.2017.006841.

[34] M. S. Katooli and A. Koochaki, "Detection and classification of incipient faults in three-phase power transformer using DGA information and rule-based machine learning method," *J. Control, Autom. Electr. Syst.*, vol. 31, no. 5, pp. 1251–1266, Oct. 2020, doi: 10.1007/s40313-020-00625-5.

[35] 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, doi: 10.1109/TDEI.2017.005959.

[36] M. Elsisi, M. Tran, K. Mahmoud, D.-E.-A. Mansour, M. Lehtonen, and M. M. F. Darwish, "Effective IoT-based deep learning platform for online fault diagnosis of power transformers against cyberattacks and data uncertainties," *Measurement*, vol. 190, Feb. 2022, Art. no. 110686, doi: 10.1016/j.measurement.2021.110686.

[37] I. B. M. Taha, S. S. M. Ghoneim, and H. G. Zaini, "Improvement of Rogers four ratios and IEC code methods for transformer fault diagnosis based on dissolved gas analysis," in *Proc. North Amer. Power Symp. (NAPS)*, Oct. 2015, pp. 1–5, doi: 10.1109/NAPS.2015.7335098.

[38] A. Peimankar, S. J. Weddell, T. Jalal, and A. C. Lapthorn, "Evolutionary multi-objective fault diagnosis of power transformers," *Swarm Evol. Comput.*, vol. 36, pp. 62–75, Oct. 2017, doi: 10.1016/j.swevo.2017.03.005.

[39] A. Kirkbas, A. Demircali, S. Koroglu, and A. Kizilkaya, "Fault diagnosis of oil-immersed power transformers using common vector approach," *Electric Power Syst. Res.*, vol. 184, Jul. 2020, Art. no. 106346, doi: 10.1016/j.epsr.2020.106346.

[40] L. Ganyun, C. Haozhong, Z. Haibao, and D. Lixin, "Fault diagnosis of power transformer based on multi-layer SVM classifier," *Electr. Power Syst. Res.*, vol. 74, no. 1, pp. 1–7, Apr. 2005, doi: 10.1016/j.epsr.2004.07.008.

[41] M.-H. Wang, "A novel extension method for transformer fault diagnosis," *IEEE Trans. Power Del.*, vol. 18, no. 1, pp. 164–169, Jan. 2003, doi: 10.1109/TPWRD.2002.803838.

[42] J. Fan, F. Wang, Q. Sun, F. Bin, F. Liang, and X. Xiao, "Hybrid RVM-ANFIS algorithm for transformer fault diagnosis," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 14, pp. 3637–3643, Sep. 2017, doi: 10.1049/iet-gtd.2017.0547.

[43] H. Hu, S. Qian, and J. Cao, "Monitoring and fault diagnosing system design for power transformer based on temperature field model and DGA feature extraction," in *Proc. 7th World Congr. Intell. Control Autom.*, Jun. 2008, pp. 1800–1805, doi: 10.1109/WCICA.2008.4593195.

[44] M. Badawi, S. A. Ibrahim, D. A. Mansour, A. A. El-Faraskoury, S. A. Ward, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, "Reliable estimation for health index of transformer oil based on novel combined predictive maintenance techniques," *IEEE Access*, vol. 10, pp. 25954–25972, 2022, doi: 10.1109/ACCESS.2022.3156102.

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

[46] Y. Zhang, X. Ding, Y. Liu, and P. J. Griffin, "An artificial neural network approach to transformer fault diagnosis," *IEEE Trans. Power Del.*, vol. 11, no. 4, pp. 1836–1841, Oct. 1996, doi: 10.1109/61.544265.

[47] X. Hao and S. Cai-xin, "Artificial immune network classification algorithm for fault diagnosis of power transformer," *IEEE Trans. Power Del.*, vol. 22, no. 2, pp. 930–935, Apr. 2007, doi: 10.1109/TPWRD.2007.893182.

[48] V. Duraisamy, N. Devarajan, D. Somasundareswari, A. A. M. Vasanth, and S. N. Sivanandam, "Neuro fuzzy schemes for fault detection in power transformer," *Appl. Soft Comput.*, vol. 7, no. 2, pp. 534–539, Mar. 2007, doi: 10.1016/j.asoc.2006.10.001.

[49] N. Yadaiah and N. Ravi, "Internal fault detection techniques for power transformers," *Appl. Soft Comput.*, vol. 11, no. 8, pp. 5259–5269, Dec. 2011, doi: 10.1016/j.asoc.2011.05.034.

[50] A. R. G. Castro and V. Miranda, "Knowledge discovery in neural networks with application to transformer failure diagnosis," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 717–724, May 2005, doi: 10.1109/TPWRS.2005.846074.

[51] M. M. Ibrahim, M. M. Sayed, and E. E. A. El-Zahab, "Diagnosis of power transformer incipient faults using fuzzy logic-IEC based approach," in *Proc. IEEE Int. Energy Conf. (ENERGYCON)*, May 2014, pp. 242–245, doi: 10.1109/ENERGYCON.2014.6850435.

[52] H. Malik and S. Mishra, "Selection of most relevant input parameters using principle component analysis for extreme learning machine based power transformer fault diagnosis model," *Electr. Power Compon. Syst.*, vol. 45, no. 12, pp. 1339–1352, Jul. 2017, doi: 10.1080/15325008.2017.1338794.

[53] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[54] K. N. V. P. S. Rajesh, U. M. Rao, I. Fofana, P. Rozga, and A. Paramane, "Influence of data balancing on transformer DGA fault classification with machine learning algorithms," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 30, no. 1, pp. 385–392, Feb. 2023, doi: 10.1109/TDEI.2022.3230377.

[55] M. M. Alam, G. Karmakar, S. Islam, J. Kamruzzaman, M. Chetty, S. Lim, G. Appuhamillage, G. Chattopadhyay, S. Wilcox, and V. Verheyen, "Assessing transformer oil quality using deep convolutional networks," in *Proc. 29th Australas. Universities Power Eng. Conf. (AUPEC)*, Nov. 2019, pp. 1–6, doi: 10.1109/AUPEC48547.2019.211896.

[56] J. Lin, J. Ma, J. G. Zhu, and Y. Cui, "A transfer ensemble learning method for evaluating power transformer health conditions with limited measurement data," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022, doi: 10.1109/TIM.2022.3175268.

[57] *D1/A2 Technical Brochure—Advances in DGA Interpretation*, CIGRE, Paris, France, Jul. 2019.

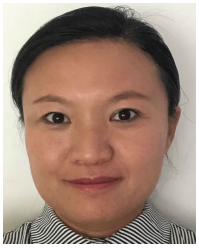
[58] A. Teymouri and B. Vahidi, "CO_2/CO concentration ratio: A complementary method for determining the degree of polymerization of power transformer paper insulation," *IEEE Elect. Insul. Mag.*, vol. 33, no. 1, pp. 24–30, Jan. 2017, doi: 10.1109/MEI.2017.7804313.

[59] M. Duval, A. De Pablo, I. Atanasova-Hoehlein, and M. Grisaru, "Significance and detection of very low degree of polymerization of paper in transformers," *IEEE Elect. Insul. Mag.*, vol. 33, no. 1, pp. 31–38, Jan. 2017, doi: 10.1109/MEI.2017.7804314.

[60] F. R. Barbosa, O. M. Almeida, A. P. S. Braga, M. A. B. Amora, and S. J. M. Cartaxo, "Application of an artificial neural network in the use of physicochemical properties as a low cost proxy of power transformers DGA data," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 19, no. 1, pp. 239–246, Feb. 2012, doi: 10.1109/TDEI.2012.6148524.

[61] S. Forouhari and A. Abu-Siada, "Application of adaptive neuro fuzzy inference system to support power transformer life estimation and asset management decision," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 3, pp. 845–852, Jun. 2018, doi: 10.1109/TDEI.2018.006392.

[62] S. S. M. Ghoneim, "The degree of polymerization in a prediction model of insulating paper and the remaining life of power transformers," *Energies*, vol. 14, no. 3, p. 670, Jan. 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/3/670
[63] M. M. Nezami, M. D. Equbal, S. A. Khan, S. Sohail, and S. S. M. Ghoneim, "Classification of cellulosic insulation state based on smart life prediction approach (SLPA)," *Processes*, vol. 9, no. 6, p. 981, Jun. 2021. [Online]. Available: https://www.mdpi.com/2227-9717/9/6/981
[64] M. M. Nezami, M. D. Equbal, S. A. Khan, and S. Sohail, "An ANFIS based comprehensive correlation between diagnostic and destructive parameters of transformer's paper insulation," *Arabian J. Sci. Eng.*, vol. 46, no. 2, pp. 1541–1547, Feb. 2021.



LAN JIN received the B.Eng. degree in electrical and electronic engineering from The University of Melbourne, Australia, in 2007. She is currently pursuing the master's degree with the School of Electrical and Computer Engineering, Computing and Mathematical Science, Curtin University, Perth, Australia. Her primary research interests include application of artificial intelligence algorithms in power transformer condition monitoring and fault diagnosis.

DOWON KIM received the Ph.D. degree in electrical engineering from Curtin University, Perth, WA, Australia, in 2021. From 1998 to 2011, he was a Transmission and Substation Engineer with Korea Electric Power Corporation, and has been working in the fields of power system study, and condition monitoring and protection in Australia, since 2012. Since August 2020, he has been a Lecturer with the School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University. His research interests include power system protection, high voltage condition monitoring and diagnosis, and wireless power transfer systems.

KIT YAN CHAN (Member, IEEE) received the Ph.D. degree in computing from London South Bank University, London, U.K., in 2006. He was a full-time Researcher with The Hong Kong Polytechnic University, from 2004 to 2009, and Curtin University, Perth, WA, Australia, from 2009 to 2013. He is currently a Senior Lecturer with the School of Electrical Engineering, Computing and Mathematical Science, Curtin University. He has published more than 100 journal articles and several books. His research interests include artificial intelligence, machine learning, and their applications to power flow optimization, underwater acoustic communications, and load consumption forecasting. He serves as an associate editor for several reputed journals.

AHMED ABU-SIADA (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Ain Shams University, Egypt, and the Ph.D. degree in electrical engineering from Curtin University, Australia. He is currently a Full Professor and the Head of the Electrical and Computer Engineering Discipline, Curtin University. He has published more than 300 journal articles and conference papers and holds several patents on condition monitoring, fault diagnosis and remnant life estimation of high voltage assets. He has delivered several keynote speeches in various IEEE and international conferences and has been the General Chair of the IEEE iSPEC 2022 and CMD 2018 conferences. He is the Founding Editor-in-Chief of two international journals and an Editor of several journals, such as *IET Generation, Transmission & Distribution*.