



OPEN

## Discernment of transformer oil stray gassing anomalies using machine learning classification techniques

M. K. Ngwenyama<sup>1,2</sup>✉ & M. N. Gitau<sup>1,2</sup>

This work examines the application of machine learning (ML) algorithms to evaluate dissolved gas analysis (DGA) data to quickly identify incipient faults in oil-immersed transformers (OITs). Transformers are pivotal equipment in the transmission and distribution of electrical power. The failure of a particular unit during service may interrupt a massive number of consumers and disrupt commercial activities in that area. Therefore, several monitoring techniques are proposed to ensure that the unit maintains an adequate level of functionality in addition to an extended useful lifespan. DGA is a technique commonly employed for monitoring the state of OITs. The understanding of DGA samples is conversely unsatisfactory from the perspective of evaluating incipient faults and relies mainly on the proficiency of test engineers. In the current work, a multi-classification model that is centered on ML algorithms is demonstrated to have a logical, precise, and perfect understanding of DGA. The proposed model is used to analyze 138 transformer oil (TO) samples that exhibited different stray gassing characteristics in various South African substations. The proposed model combines the design of four ML classifiers and enhances diagnosis accuracy and trust between the transformer manufacturer and power utility. Furthermore, case reports on transformer failure analysis using the proposed model, IEC 60599:2022, and Eskom (Specification—Ref: 240-75661431) standards are presented. In addition, a comparison analysis is conducted in this work against the conventional DGA approaches to validate the proposed model. The proposed model demonstrates the highest degree of accuracy of 87.7%, which was produced by Bagged Trees, followed by Fine KNN with 86.2%, and the third in rank is Quadratic SVM with 84.1%.

With the radical growth in the power system capacity, the demands for power generation, transmission, and distribution, have become greater<sup>1</sup>. As a significant piece of equipment for power distribution in power systems, the power transformer (PT) is critical for the secure operation of the complete power system. The occurrence of a fault in a PT will result in damage to the unit. The most severe faults might even cause the failure of the entire power system, adversely affecting the functioning of the total national economy. Thus, it is beneficial to examine fault diagnosis technology relating to PTs<sup>2</sup>. PT faults usually emerge from electrical and thermal stresses, such faults vary merely in their energy, site, and time of occurrence. The oil temperature increases and several gases will be generated when the fault occurs. Generally, the combustible gasses found in the TO in service are hydrogen (H<sub>2</sub>), methane (CH<sub>4</sub>), ethane (C<sub>2</sub>H<sub>6</sub>), ethylene (C<sub>2</sub>H<sub>4</sub>), and acetylene (C<sub>2</sub>H<sub>2</sub>)<sup>3,4</sup>. The pollutants in oil are mostly the consequence of the degradation of insulating elements (oil or sheet) because of faults or chemical responses in the apparatus in question.

The quality and quantity of disintegrated gases have a prominent function in assessing the fault type in PTs<sup>5,6</sup>. Many conventional techniques have been developed to analyze transformer faults with gas chromatography; a procedure where a chemical combination transported through a gas or liquid is broken down into its constituent parts as a result of the substances flowing differently along or above a static solution. Such schemes for fault analysis are usually categorized into three types, specifically, the distinctive gas scheme<sup>7–10</sup>, the gas production rate scheme<sup>10</sup>, and the three-ratio scheme<sup>11–13</sup>. In China, over 50% of the PT faults in the energy system were evaluated by employing DGA-based analysis schemes which analyze transformer fault types and their severity following the content, proportion to one another, and the gas production rate of the DGs in the TO<sup>13</sup>. Adding

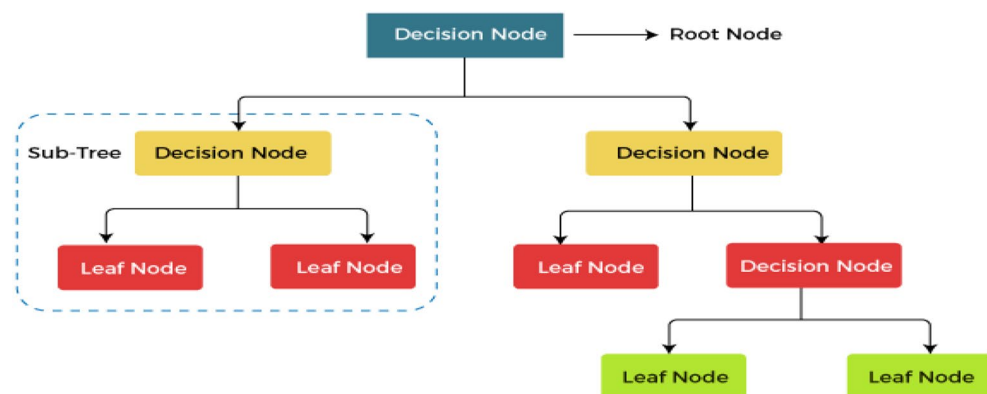
<sup>1</sup>Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa.

<sup>2</sup>These authors contributed equally: M. K. Ngwenyama and M. N. Gitau. ✉email: u11265702@tuks.co.za

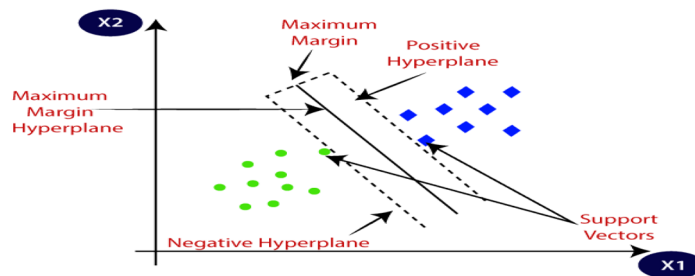
to the above three key conventional techniques, some enhanced schemes have emerged, like the Doernenburg scheme, the Rogers ratio scheme, the Duval triangle scheme, the International Electrotechnical Commission (IEC) ratio scheme, and the Key Gas (KG) scheme<sup>14–18</sup>. Such schemes usually employ numerous gas ratios or compare gas levels with the appointed criteria to analyze the state of a PT. However, most of these conventional analysis techniques provide a restricted impact to a transformer's fault analysis, which is unable to precisely identify its correct fault type. Particularly, it is extremely complex to precisely determine the fault state with several DGs, a great probability of misdiagnosis will occur when the calculated and analyzed gas ratio is near the critical value<sup>19</sup>. Furthermore, the more comprehensive the classifications of fault types are, the lesser the precision rate of fault analysis is, and vice versa. Moreover, rough classifications are not conducive to the fault analysis of a PT, and it is challenging to meet the demands of applications.

DGA is a technique for detecting and forecasting problems in OITs by (i) determining the levels of various gases contained in the insulation oil, as well as respective gas rates and gas proportions, (ii) fault detection utilizing diagnosis instruments such as KG<sup>20,21</sup>, IEC ratios<sup>22</sup>, Rogers ratios<sup>23</sup>, Doernenburg ratios<sup>24</sup> and Duval triangle<sup>23</sup>. Nevertheless, these instruments have certain flaws. In certain situations, the computed gas ratios deviate from the instruments' specified ratio codes. Faults that develop within the transformer might be undetectable<sup>25</sup>. Additionally, these instruments can produce various analytical outcomes for the equivalent dissolved gas (DG) file, making it challenging for experts to reach a definitive conclusion when confronted with such a wide range of data<sup>26</sup>. Due to these constraints, several scientists have developed systems that are integrated with ML approaches that use historical DGA information to forecast imminent or undiscovered faults for diagnosing faults. The complexity of identifying the appropriate fault situation and the analytic precisions for units under fault categories are defined by these aspects<sup>27,28</sup>. The KG ratios, as well as graphic depiction schemes, are all DGA schemes that are utilized as data inputs to ML classifiers for fault classification. In the current study, a multi-classification model that is centered on ML algorithms is shown to have an intelligible, precise, and clear understanding of DGA. This enthusiasm is supported by (i) efficient adaptation to fresh data in ML; (ii) for structural layout, ML needs minimal exertion (i.e. several control settings are involved.); and (iii) the capability of ML to categorize unpredictable issues<sup>29</sup>. Capitalizing on these benefits, the proposed model is used to analyze and evaluate the state and suitable gas name subscription of 138 TO samples that exhibited different stray gassing characteristics in various South African substations. The model uses four ML classifiers, namely: (i) Decision Tree (DT)<sup>30</sup>, (ii) Support Vector Machine (SVM)<sup>31</sup>; (iii) K-Nearest Neighbour (KNN)<sup>32</sup>; and (iv) Ensemble Classifier (EC)<sup>33</sup>. These classifiers are applied for oil sample classification and are selected based on their capacity to compare new data inputs to existing data to identify the class that closely resembles existing classes to place new data within. In MATLAB/Simulink, the proposed model serves as the framework underlying the various classifiers and is designed to aggregate ML algorithms for information-gathering activities. A detailed summary of the various ML classifiers utilized in this work is provided in the section that follows:

- DT: As shown in Fig. 1, the DT classifier<sup>34</sup> is an ML technique that makes predictions using a tree structure. It builds a flowchart-like tree structure where each internal node represents a feature test, each branch represents a test outcome, and each leaf node stores a class label. It is constructed by constantly splitting the training data into subsets depending on feature values until a stopping requirement is met, such as the maximum depth of the tree or the minimum number of samples needed to divide a node. The method replicates the operation for every split subgroup that is the offspring of a given node. Lastly, the tree is trimmed by deleting limbs that are not useful for classification.
- SVM: The working of the SVM classifier<sup>35</sup> can be understood by using Fig. 2. SVMs fall within the broad group of kernel schemes<sup>36</sup> that rely solely on data using mark pairings. To guarantee that the hyperplane is as broad as feasible across categories, the kernel function determines an estimation product for certain potentially large-scale feature regions. SVMs possess the benefits of becoming less mathematically intensive compared to different methods of classification, performing well in large-scale areas, as well as managing



**Figure 1.** Example of DT.



**Figure 2.** Example of SVM.

unpredictable classification effectively by utilizing the kernel trick, which subsequently converts the data area into a different large-scale feature area.

- **KNN:** The KNN classifier<sup>37</sup> is a monitored learning approach utilized for numerous machine learning scenarios. It arranges elements using the nearest trained samples in the characteristic domain. The goal underlying KNN is to locate a well-known amount of training data that is nearest in proximity to a particular querying case and estimate the querying case's category based on them. Regarding categorization, KNN is comparable to a DT method, except that rather than developing a tree, instead, it creates a route through the graph. KNNs are also quicker compared to DTs. The working of the KNN is shown in Fig. 3.
- **EC:** The ensemble classifier<sup>38</sup> produces classification forecasts using a set of classifiers, which achieves more accurate specialization than one classifier and results in an improved measurement grade. A dataset is used to train a list of classifiers, and the separate predictions made by each of the classifiers applied to the dataset form the basis of EC. The ensemble model then combines the outcomes of each classifier prediction to get the final result. This sort of classifier remains simple to simulate but is often appropriate for large samples. The working of the EC is shown in Fig. 4.

In monitoring the insulation status in OITs, several chemical and electrical processes are employed, such as DGA and Furan Analysis (FA), which indicate the Degree of Polymerization (DP) of the cellulose paper<sup>6,39</sup>. DGA is one of the most common methods for detecting an incipient fault in PTs. DGA can be used to assess present-day transformer status, predict future failures, and identify inconvenient transformer operations to provide appropriate maintenance planning. Figure 5 illustrates the standard technique employed by the transformer manufacturing sector to collect transformer oil on-site for DGA at the testing facility.

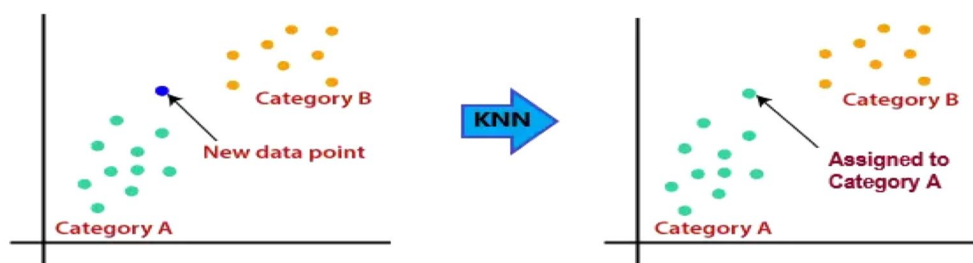
The presented DGA approaches do not contain any mathematical development, and the assessment depends on an experiential method that can vary depending on the expertise of the laboratory analyst, which results in unpredictable assessment<sup>40</sup>. To overcome this limitation, several computational models based on ML have been used in assessing incipient faults in PTs. In the proposed research work, recent related studies and their contributions to transformer fault diagnosis have been highlighted and a multi-classification model for transformer fault diagnosis is proposed. Table 1 presents a comparative study of the existing recent survey and the proposed model for transformer fault analysis.

## Contribution and novelty

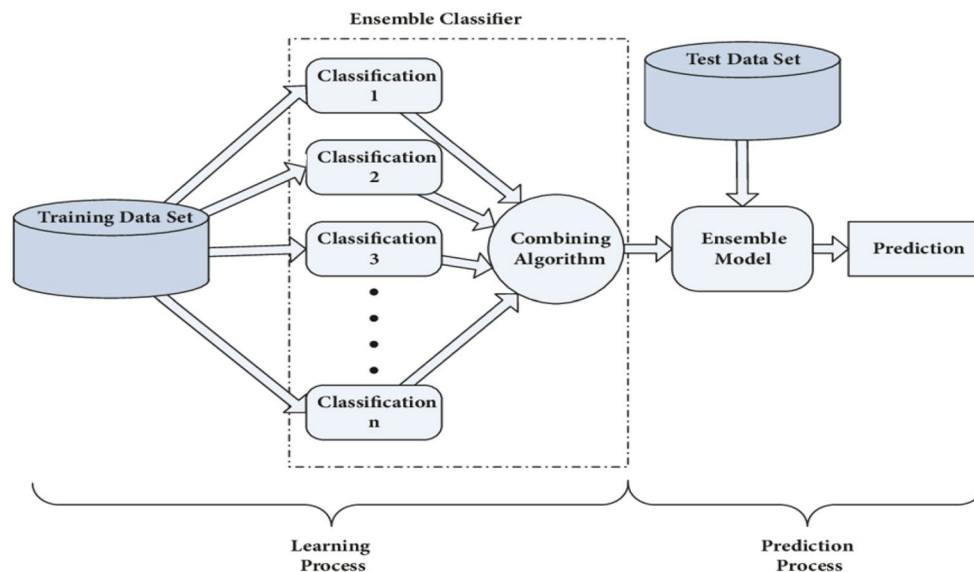
### Research contribution

This research work provided a summary of recent transformer fault analysis. Several ML-based techniques based on conventional DGA approaches have been discussed. The following are the contributions of the proposed research analysis:

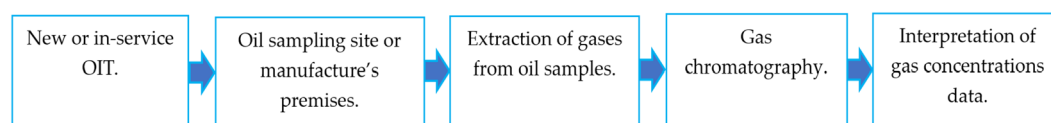
- A multi-classification model for fault diagnosis is proposed, that enhances diagnostic accuracy and optimism between transformer manufacturer and power utility.



**Figure 3.** Example of KNN.



**Figure 4.** Example of EC.



**Figure 5.** Extraction of transformer oil for DGA.

Ref	Year	Proposed technique	Contribution
41	2018	Doernenburg ratio approach, Roger's ratio approach, multi-layer ANN perceptron	A fault diagnostic analysis was performed by developing a hybrid Doernenburg and Rogers ratio technique to determine a gas ratio suitable to train a multi-layer ANN perceptron
42	2019	Mean Shift algorithm (MSA), ANN	An MSA-based ANN is proposed. The IEC 60,599:2007 standard consists of gases that are used to create parameters that will be trained using the proposed method. The MSA was used to effectively avoid the limitation of the number of training patterns (data size). The training and validation techniques both produced acceptable outcomes
43	2020	Fuzzy Logic, IEC ratio approach	A fuzzy logic-IEC ratio approach was proposed for transformer fault diagnosis. The outcomes demonstrate an improvement over the conventional IEC ratio technique
44	2022	ANFIS, Roger's ratio approach	A hybrid Rogers ratio technique-based ANFIS was proposed to detect transformer faults. The training was carried out by employing the gas ratios presented by the IEEE C57-104 and IEC 60,599 standards
Current study	2023	Multi-classification model	A multi-classification model for fault diagnosis is proposed, that enhances diagnostic accuracy and optimism between transformer manufacturer and power utility Case reports on transformer fault analysis utilizing the proposed multi-classification model, IEC 60599:2022, and Eskom (Specification—Ref: 240-75661431) standards for fault analysis have been presented

**Table 1.** Summary of recent related studies.

- Case reports on transformer fault analysis utilizing the proposed multi-classification model, IEC 60599:2022 and Eskom (Specification—Ref: 240-75661431) standards for fault analysis have been presented.

*Research novelty*

The main objective of the current research is to contribute to the practice of TO analysis. Although various current research studies have concentrated on TO analysis, minimal and occasional research has been published on the adoption of a multi-classification model, IEC 60599:2022, and Eskom (Specification—Ref: 240-75661431) standards for OIT analysis. The proposed model is a significant technique for overcoming the inadequacies of

the IEC gas ratio technique to create an effective oil analysis tool. The seven fault categories utilized in the IEC 60599:2022 standard were considered and concluded that the degree of accuracy for fault detection is not ideal as a consequence of the limits defined by the gas ratio codes, and leads to "not detectable" in certain cases scenarios. However, after introducing the proposed model, the analysis is on an equal footing with the actual fault analysis. Furthermore, this research work addressed the optimal ratios of fault analysis. It is crucial to train the proposed model. Consequently, the DGA data utilized to train the proposed model is made up of samples that cover all known types of faults as defined by the IEC 60599:2022 standard. According to the findings of this research, the forecasting of transformer faults employing the proposed model as well as the IEC 60599:2022 gas ratio technique is comparable to actual fault analysis and offers an improvement over the IEC 60599:2022 gas ratio technique.

### Paper organization

The rest of the work is structured as follows: Section "Review of existing DGA approaches" provides an overview of current DGA approaches. Section "Applicable works" provides an overview of techniques employed by researchers to explore DGA. Section "Proposed approach" discusses the research approach and model. Section "Materials and protocols" presents the materials and protocols of the study. Section "Results" presents the results and discussions that validate the proposed model. and finally, Section "Conclusions" presents the conclusions of this work.

### Review of existing DGA approaches

There are several procedures for diagnosing deformities in transformer insulation. DGA analysis strategies are dependent on scientific hypotheses and practical knowledge gained by specialists across the world<sup>45,46</sup>. However, if these analysis strategies are not implemented with caution, they might detect abnormalities erroneously since they simply signal potential faults<sup>47</sup>. DGA strategies can vary regarding diagnosed faults in several instances, which is undesirable for an accurate fault analytic technology<sup>48</sup>. Flammable and non-flammable gases can be discovered within the gases contained in the oil, as shown in Table 2. These gases can also be categorized based on the type of fault that induced them, as shown in Table 3. Gas levels, KGs, KG ratios, and graphic interpretations are mutual ideologies adopted in analysis strategies<sup>49</sup>. The DGA can recognize different faults such as partial discharge, excessive heat, as well as arcing in a wide range of PTs. A single dataset is required for the efficiency analysis and analytical comparison of conventional DGA approaches<sup>50</sup>. Figure 6 illustrates a dataset of potential transformer faults. Furthermore, DGA can give the timely detection required to maximize the probability of establishing a suitable remedy<sup>51,52</sup>. Numerous informative techniques based on DGA to identify the emerging fault type have been stated. In this work, seven of the DGA approaches were explored: (i) the CIGRE approach, (ii) the Doernenburg ratio approach, (iii) the KG approach, (iv) the Nomograph approach, (v) the IEC ratio approach, (vi) Duval triangle approach, and (viii) Rogers ratio approach.

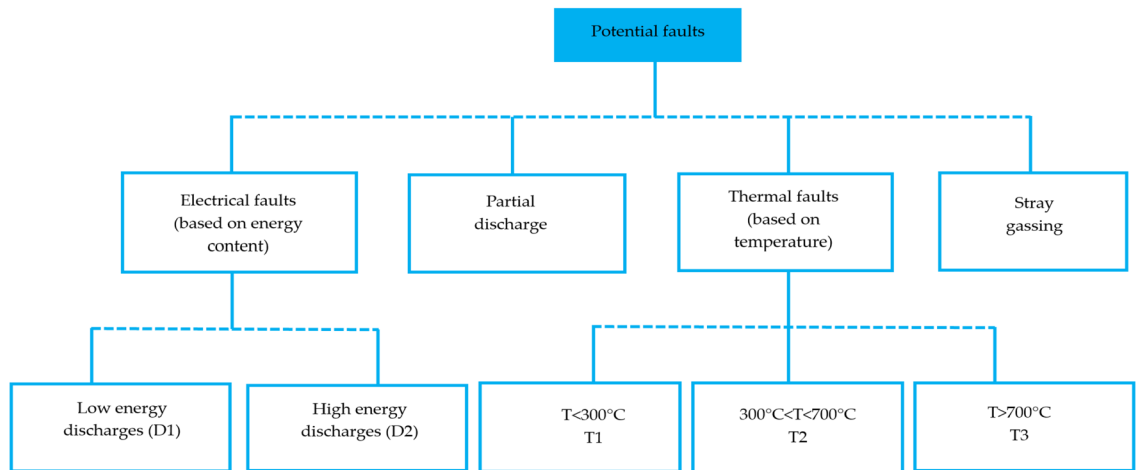
These approaches are experimental, with assumptions established on relationships between gases discovered through gas analysis<sup>54</sup>. For instance, The Rogers Ratio technique considers the ratios of H<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub>, and C<sub>2</sub>H<sub>2</sub> to create code acknowledging fault analysis. The ratio range, related codes, and related diagnostics for the

Gases	
Flammable	Non-flammable
Carbon monoxide (CO)	Oxygen (O <sub>2</sub> )
Hydrogen (H <sub>2</sub> )	–
Methane (CH <sub>4</sub> )	Nitrogen (N <sub>2</sub> )
Ethane (C <sub>2</sub> H <sub>6</sub> )	–
Ethylene (C <sub>2</sub> H <sub>4</sub> )	Carbon dioxide (CO <sub>2</sub> )
Acetylene (C <sub>2</sub> H <sub>2</sub> )	

**Table 2.** DGs in TO.

Fault type	Flammable	Non-flammable
Overheating of windings	CO	–
Oil overheating	CH <sub>4</sub> , C <sub>2</sub> H <sub>6</sub> , C <sub>2</sub> H <sub>4</sub>	–
Partial discharge	H <sub>2</sub>	–
Arcing	C <sub>2</sub> H <sub>2</sub>	–
Minimal temperatures	CH <sub>4</sub> , C <sub>2</sub> H <sub>6</sub>	–
Excessive temperatures	C <sub>2</sub> H <sub>4</sub> , H <sub>2</sub> , CH <sub>4</sub> , C <sub>2</sub> H <sub>6</sub>	–
All faults	–	O <sub>2</sub> , N <sub>2</sub> , CO <sub>2</sub>

**Table 3.** Types of DGs according to the fault type and material concerned.



**Figure 6.** Dataset of potential faults<sup>53</sup>.

several code combinations are provided in Table 13. The ratios of the approaches are illustrated as: **R1**:(CH<sub>4</sub>/H<sub>2</sub>), **R2**:(C<sub>2</sub>H<sub>2</sub>/C<sub>2</sub>H<sub>4</sub>); **R3**:(C<sub>2</sub>H<sub>2</sub>/CH<sub>4</sub>); **R4**:(C<sub>2</sub>H<sub>6</sub>/C<sub>2</sub>H<sub>2</sub>); and **R5**:(C<sub>2</sub>H<sub>4</sub>/C<sub>2</sub>H<sub>6</sub>).

**CIGRE approach**

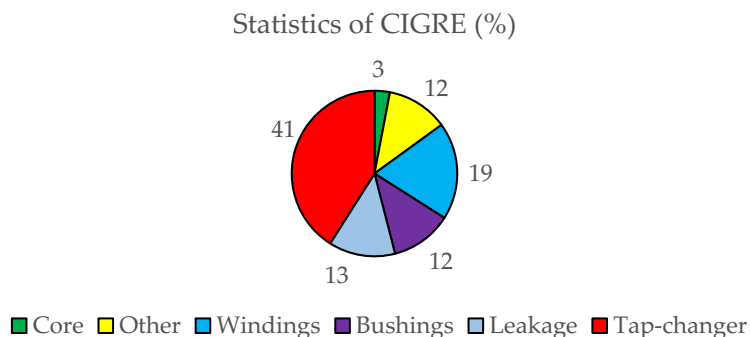
This approach<sup>55</sup> explores KG ratios and gas levels. The 5 KG ratios evaluated using this approach are C<sub>2</sub>H<sub>2</sub>/C<sub>2</sub>H<sub>6</sub>, H<sub>2</sub>/CH<sub>4</sub>, C<sub>2</sub>H<sub>4</sub>/C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>2</sub>/H<sub>2</sub>, and CO/CO<sub>2</sub>. A transformer is declared efficient if consecutive deployments of these approaches produce gas ratios and levels that are within permissible thresholds. The incidence of catastrophes in PTs discussed in Ref.<sup>56</sup> is graphically shown in Fig. 7. The catastrophe statistics of CIGRE consist of approximately 800 catastrophes<sup>57</sup>.

**Doernenburg ratio approach**

Reserve integrated safety sensors (RIS2) are required in this approach, and a significant volume of gas is required to demonstrate its application. RIS2 is an essential accessory in the protection of oil transformers. It allows continuous control of the pressure, temperature, oil level, and gas states. The work presented in Ref.<sup>58</sup> states that gas ratio approaches utilize encoding algorithms that designate certain pairings of codes to specific fault categories for efficient fault identification. The codes are formed by estimating gas level ratios as well as comparing them to predefined values acquired from experience and constantly modified. The author discovered that when a gas composition matches the code for a certain fault, a fault state is identified. In Ref.<sup>59</sup>, the approach was applied to diagnose faults by monitoring gas levels of CH<sub>4</sub>/H<sub>2</sub>, C<sub>2</sub>H<sub>2</sub>/CH<sub>4</sub>, C<sub>2</sub>H<sub>4</sub>/C<sub>2</sub>H<sub>6</sub> and C<sub>2</sub>H<sub>2</sub>/C<sub>2</sub>H<sub>4</sub>. To determine whether there is an actual fault with the unit, the concentration of the gases must initially exceed the acceptable limits, and is adequate formation of each gas for the ratio analysis must be present to be valid<sup>60</sup>. Table 4 demonstrates the KGs as well as their concentration limits.

**KG approach**

This approach<sup>61,62</sup> monitors the gases emitted from TO upon a failure, which causes the temperature in the transformer to rise. It must be noted that the utility of oil in the transformer provides insulation, and cooling, and helps quench arc. This approach is the most critical and commonly utilized since it presents the earliest signal of an incident. Table 5 demonstrates the diagnostic explanations by using different KG concentrations. The ppm concentration standard value limit detected in PTs according to IEC 60599:2022 is specified in Table 6.



**Figure 7.** Catastrophe statistics of distinct transformer parts reported by CIGRE<sup>57</sup>.

Doernenburg ratio scheme	
KG	Concentration (ppm)
C <sub>2</sub> H <sub>2</sub> (acetylene)	35
C <sub>2</sub> H <sub>6</sub> (ethane)	65
C <sub>2</sub> H <sub>4</sub> (ethylene)	50
CH <sub>4</sub> (methane)	120
CO(carbon monoxide)	350
H <sub>2</sub> (hydrogen)	100

**Table 4.** Concentration for Doernenburg ratio scheme.

KG scheme	
Gas detected	Interpretation
C <sub>2</sub> H <sub>2</sub> (acetylene)	Electric fault (arc, spark)
C <sub>2</sub> H <sub>6</sub> (ethane)	Secondary indicator of thermal fault
C <sub>2</sub> H <sub>4</sub> (ethylene)	Thermal fault (overheating local)
CH <sub>4</sub> (methane)	Secondary indication of arc or severe excessive heat
CO(carbon monoxide)	Paper degradation
CO <sub>2</sub> (carbon dioxide)	Paper degradation
H <sub>2</sub> (hydrogen)	Electromagnetic disposal
O <sub>2</sub> (oxygen)	Transformer seal fault

**Table 5.** Gas dissolved in oil for analysis.

Gas	H <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	CO	CO <sub>2</sub>
Concentration (ppm)	100	120	1	50	65	350	2500

**Table 6.** Limit concentrations of DGs for values observed in the transformer.

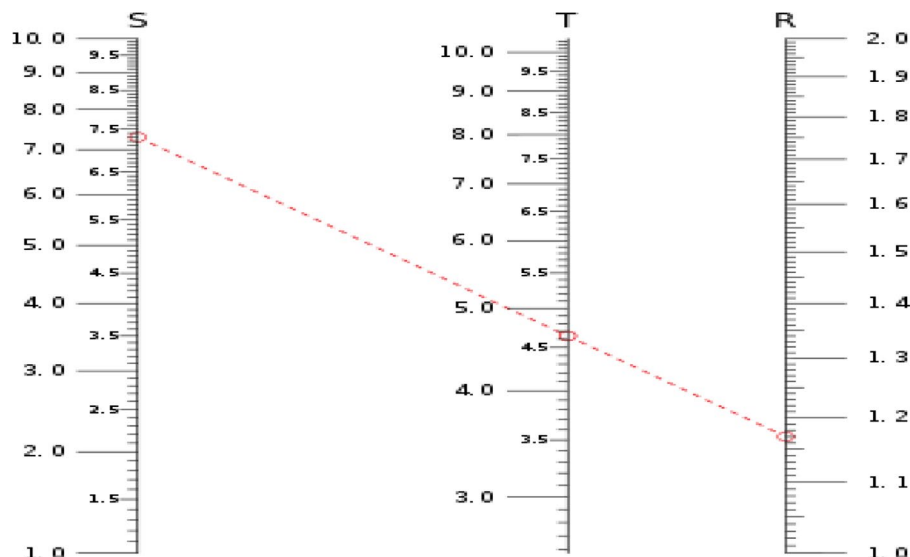
The faults in this approach are compared with the gas concentration profile. According to the IEEE standard, KGs are gases produced in OITs that alert to observational fault-type diagnostics, depending on which gases are common or prominent at certain temperature levels<sup>63</sup>. If there is no previous DG data is provided for analysis, hazards in the apparatus can be detected and evaluated using the guidelines indicated in Table 7. Healthy operation is represented by State 1. State 2 signifies that the instrument is possibly malfunctioning, with overall gases exceeding normal concentrations. State 3 implies a high amount of degradation. State 4 indicates that excessive deterioration and continued operation may end in failure or breakdown<sup>64,65</sup>.

### Nomograph approach

The author<sup>66</sup> proposed the Nomograph approach to enhance fault analysis precision through the combination of fault gas ratios<sup>67</sup>. It was proposed to supply both a visual demonstration of fault-gas data as well as the resources to understand its importance. The Nomograph is built from upright exponential measurements that depict the amounts of various gases. Using this methodology, straight lines are created between neighboring measurements to link the dots reflecting different gas level values. The gradients of these lines serve as a guide for diagnosing the type of fault. The key (T) across the two axes specifies the fault type for the two axes. The positioning of the line about the intensity scales allows you to determine the degree of the fault. The example to calculate the value of T is shown in Fig. 8.

States	H <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	CO	CO <sub>2</sub>	Total gas
State 1	100	120	1	50	65	350	2500	720
State 2	101–700	121–400	2–9	51–100	66–100	351–570	2500–4000	721–1920
State 3	701–1800	401–1000	10–35	101–200	101–150	571–1400	4001–10,000	1921–4630
State 4	> 1800	> 1000	> 35	> 200	> 150	> 1400	> 10,000	> 4630

**Table 7.** Risk assessment in transformers using DG levels (ppm).



**Figure 8.** The logarithmic nomograph.

The formula to calculate the value of  $T$  is given in (1).

$$T = (1.84S + 4.66)^{0.37} + (1.21R)^{1.333}, \tag{1}$$

where  $S$  is the pre-test gas,  $R$  is the post-test gas and  $T$  is the likelihood fault type ratio.

### IEC ratio approach

This approach<sup>68</sup> analyzes and predicts transformer faults using the first five gases provided in Table 8. The gases are employed to harvest three gas ratios, namely:  $C_2H_2/C_2H_4$ ,  $CH_4/H_2$  as well as  $C_2H_2/C_2H_6$ . There are two critical considerations to emphasize regarding the approach. Different countries utilize differing ratios, as well as (the  $C_2H_2/C_2H_6$ ) ratio usually employed to substitute the ( $CH_4/H_2$ ) ratio. The ratios should be identified mainly when one of the DGs contains a significant concentration and/or a fast-rising rate<sup>69</sup>. Table 9 depicts the IEC standard for describing fault types and provides estimates for the 3 KG ratios using prescribed fault analysis. Whenever key-gas ratios exceed certain limitations, incipient faults in the transformer are to be foreseen<sup>70</sup>.

Fault type	Report	$C_2H_2/C_2H_4$	$CH_4/H_2$	$C_2H_2/C_2H_6$
PD	Partial discharges	Insignificant value	< 0.1	< 0.2
D1	Low energy discharges	< 1.0	0.1–0.5	> 1.0
D2	High energy discharges	0.6–2.5	0.1–1.0	> 2.0
T1	Thermal faults, $T < 300\text{ }^\circ\text{C}$	Insignificant value	Insignificant value	< 1.0
T2	Thermal faults, $300\text{ }^\circ\text{C} < T < 700\text{ }^\circ\text{C}$	< 1.0	> 1.0	1.0–4.0
T3	Thermal faults, $T < 700\text{ }^\circ\text{C}$	< 0.2	> 1.0	> 4.0

**Table 8.** Evaluation of DGA data and categorization of fault categories by IEC 60,599.

Fault type	Fault definition	$CH_4$	$C_2H_4$	$C_2H_2$
PD	Partial discharges	98%	–	–
D1	Low energy discharges	–	23%	13%
D2	High energy discharges	–	23–40%	13–29%
T1	Thermal faults	–	20%	4%
T2	Thermal faults	–	20–50%	4%
T3	Thermal faults	–	50%	15%
D+T	Thermal and electrical faults	–	40–50%	4–29%

**Table 9.** Zone limitations are represented graphically.

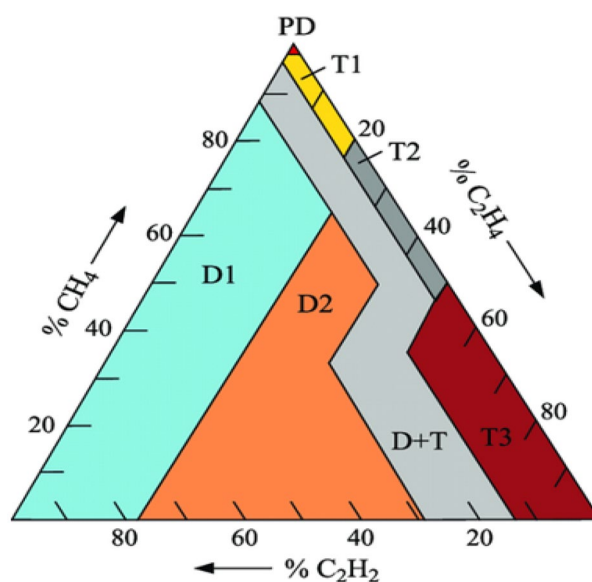


### Duval triangle approach

The work in Ref.<sup>71</sup> analyzes DG data by utilizing a triangle of comparative percentages of CH<sub>4</sub>, C<sub>2</sub>H<sub>2</sub>, and C<sub>2</sub>H<sub>4</sub>. These gases are converted into triangular data to represent the triangle. Even though this approach is regarded as simple to implement, incorrect classifications might occur because no section of the triangle is identified as a sample of typical aging<sup>72,73</sup>. Therefore, before adopting this approach to examine transformers that have been in operation for several years, the acceptable level of DG must be identified. A fault is detected by summing the quantities of the three Duval Triangle gases CH<sub>4</sub>, C<sub>2</sub>H<sub>2</sub>, and C<sub>2</sub>H<sub>4</sub> as well as extrication the sum by the volume of each gas to get the proportion of each gas overall. The intensities of CH<sub>4</sub>, C<sub>2</sub>H<sub>4</sub>, as well as C<sub>2</sub>H<sub>2</sub> are indicated as a proportion of the overall (CH<sub>4</sub> + C<sub>2</sub>H<sub>4</sub> + C<sub>2</sub>H<sub>2</sub>) and specify a point (%CH<sub>4</sub>, %C<sub>2</sub>H<sub>4</sub>, and %C<sub>2</sub>H<sub>2</sub>) in an organized structure denoted in a triangle, which has been sub-divided in separate zones<sup>74</sup>. Each zone is correlated to a specific class of fault. The Duval triangle solely contains areas correlated to fault events; there is no area for normal states. As a result, this approach can only be utilized to classify the kind of fault in the situation of a defective transformer<sup>75,76</sup>. Figure 9 shows the fault type identified in each zone. Internal transformer failures are classified into five kinds based on DL/T 722–2000 and IEC 60599–2022 standards: partial discharge (PD), low energy discharge (D1) and high energy discharge (D2), thermal faults; <300 °C (T1), thermal fault; 300 °C–700 °C (T2) and a combination of thermal and electrical faults (DT)<sup>77,78</sup>. Figure 9 may be interpreted in a table that shows the fault limitations, which are represented in Table 8.

### Rogers ratio approach

This approach adopts the exact protocol as the Doernenburg Approach, but just three proportions are utilized, and the approach's validity is not dependent on the gas concentration threshold<sup>71,80</sup>. This approach is proficient since it compares the outcomes of several failure diagnoses with the gas testing of each trial. Nevertheless, certain ratios in this approach are inconsistent with the analytic codes allocated for specific faults<sup>81</sup>. Faults are diagnosed using a basic coding technique based on ratio intervals, as illustrated in Tables 10 and 11<sup>82</sup>. Table 12 shows the proportions for DG in the oil as well as free gases, including the indicated failure detection based on the Rogers Approach. The coding results in 12 distinct forms of transformer malfunctions. Table 13 shows the different types of faults depending on the code.



**Figure 9.** Coordinates and zones of Duval triangle transformer fault diagnosis<sup>79</sup>.

Gas ratios	Ratio codes
CH <sub>4</sub> /H <sub>2</sub>	i
C <sub>2</sub> H <sub>6</sub> /CH <sub>4</sub>	j
C <sub>2</sub> H <sub>4</sub> /C <sub>2</sub> H <sub>6</sub>	k
C <sub>2</sub> H <sub>2</sub> /C <sub>2</sub> H <sub>4</sub>	l

**Table 10.** Gas ratio code.

Ratio code	Interval	Code
i	$\leq 0.1$	5
	$> 0.1, < 1.0$	0
	$\geq 1.0, < 3.0$	1
	$\geq 3.0$	2
j	$< 1.0$	0
	$\geq 1.0, < 3.0$	1
k	$< 1.0$	0
	$\geq 1.0, < 3.0$	1
	$\geq 3.0$	2
l	$< 0.5$	0
	$\geq 0.5, < 3.0$	1
	$\geq 3.0$	2

**Table 11.** Coding of Roger's ratios.

i	j	k	l	Analysis
0	0	0	0	Common degradation
5	0	0	0	Partial discharge
1–2	0	0	0	Slight overheating $< 150$ °C
1–2	1	0	0	High-temperature $150$ °C– $200$ °C
0	1	0	0	High-temperature $200$ °C– $300$ °C
0	0	1	0	General conductor overheating
1	0	1	0	Winding flowing currents
1	0	2	0	Core and tank overheating links
0	0	0	1	Flashover without power follow-through
0	0	1–2	1–2	Arc with energy follow-through
0	0	2	2	Constant flashing to free potential
5	0	0	1–2	Partial discharge with staking (note CO)

**Table 12.** Categorization based on Roger's ratio code.

Event	R2,C <sub>2</sub> H <sub>2</sub> /C <sub>2</sub> H <sub>4</sub>	R1,CH <sub>4</sub> /H <sub>2</sub>	R5,C <sub>2</sub> H <sub>4</sub> /C <sub>2</sub> H <sub>6</sub>	Catastrophe analysis recommended
0	$< 0.1$	$> 1.0, < 0.1$	$< 0.1$	Healthy unit
1	$< 0.1$	$> 1.0$	$< 0.1$	Partial discharge
2	1.0–3.0	0.1–1.0	$> 3.0$	High energy discharge
3	$< 0.1$	$> 1.0, < 0.1$	1.0–3.0	Low-temperature thermal failure
4	$< 0.1$	$> 1.0$	1.0–3.0	Thermal collapse $< 700$ °C
5	$< 0.1$	$> 1.0$	$> 3.0$	Thermal collapse $> 700$ °C

**Table 13.** Correlations for DG in oil, free gas, and catastrophe analysis are recommended by the approach of Rogers.

### Applicable works

The seven (7) DGA approaches provided in Section “[Review of existing DGA approaches](#)” are performed to monitor as well as synthesize the importance of gases existing in OITs. Concerning the faults of the above-mentioned conventional schemes, artificial intelligence (AI) schemes of PT fault analysis have attracted substantial consideration due to their superior flexibility and influential fault analysis presentation (e.g. expert system (EPS)<sup>83</sup>, fuzzy theory<sup>84</sup>, SVM<sup>85</sup>, extreme learning machine (ELM)<sup>86</sup>, as well as ANN<sup>87</sup>). EPS remains a clever AI setup scheme linked with skilled knowledge, which can analyze faults more thoroughly, precisely, and instantly.

For instance, in Refs.<sup>49,88</sup>, the author built an EPS for PT insulation fault analysis, which undertook DGA as the normal factor. The analysis results demonstrated that the suggested EPS can thoroughly examine the insulation state of a unit and detect the type of fault accurately. In Ref.<sup>89</sup>, the authors reported an instinctive fuzzy EPS to analyze PT faults, in such a manner that the approximation of KG ratio in the TO can be easier. The fuzzy concept mostly analyzes the interrelations amongst fuzzy matters, so it can handle these matters appropriately with fuzziness and ambiguity.

Also, the authors in Ref.<sup>90</sup> applied fuzzy logic linked with evaporated gas of crystal oil for PT fault analysis. Experimental outcomes proved that the extremely useful fault analysis scheme was to syndicate outputs from several DGA approaches as well as to combine them into a complete assessment.

The authors in Ref.<sup>91</sup> found the smart analysis logic centered on principal component analysis (PCA) as well as an adaptable evaluation scheme under fuzzy logic facilitates to forecasting initial fault analysis of PTs. SVM is an AI system based on the numerical learning hypothesis which holds impressive benefits in non-linear complications. The author<sup>92</sup> examined a new extension technique in which an SVM was utilized to examine the PT's faults and to elect the extremely applicable gas signature among the DGA conventional approaches and a new extension technique. The examination outcomes showed that the new extension technique as well as the SVM scheme can notably enhance the analysis precisions for PT fault categorization.

The authors in Ref.<sup>93</sup> suggested an improved prototype merging SVM with a genetic algorithm (SVMG) to analyze PT faults. The trial outcomes revealed that the SVMG technique can accomplish better ranking analytical precision compared to the IEC three ratios, typical SVM classifier, as well as ANN. ELM is a developing learning procedure that has been initiated for transformer fault analysis in current years. In Ref.<sup>3</sup>, the author employed ELM mixed with PCA to categorize the initial faults of PTs and assessed its execution with fuzzy logic as well as ANN. The evaluated outcomes demonstrated that ELM could supply decent analysis findings. Again in Ref.<sup>51</sup>, the author proposed an integrated particle swarm optimization (PSO) as well as an ELM technique to analyze PT faults.

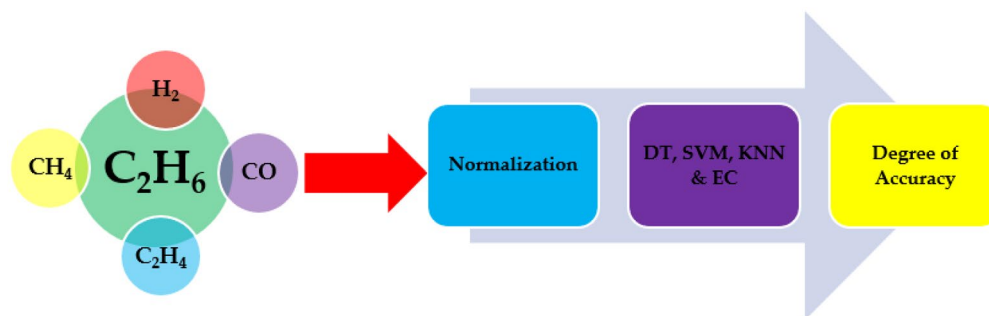
Despite that, these analysis techniques examined earlier retain their fundamental disadvantages as follows: (i) For EPS, a complete understanding root is a vital feature to guarantee the precision of analysis. Nevertheless, it is challenging to acquire a comprehensive knowledge base. Moreover, the EPS produces inadequate understanding capability; (ii) Fuzzy theory is challenging to establish a suitable link equation linking the input and output parameters<sup>94</sup>; (iii) SVM is a double-categorization procedure, that causes difficulty in terms of constructing an acquiring mechanism, choosing kernel features, and establishing variables in dual-classification challenges. Consequently, SVM has the inherent deficiency of low categorization effectiveness<sup>95,96</sup>; (iv) The execution of ELM is not balanced given that its concealed layer variable is casually selected. Compared to the fault analysis techniques discussed in Section “[Review of existing DGA approaches](#)”, the neural network has an extra general function in fault analysis of PTs due to its lack of sophistication, solid nonlinear-fitting capability, and high accuracy. For instance, the authors in Ref.<sup>97</sup> used a neural-fuzzy network to determine the initial faults in PTs, as well as to execute and examine the anticipated procedure using simulation trials. In Ref.<sup>98</sup> the author published validated research for selecting the best multi-layer perceptron (MLP) neural network simulation through comparisons of two output data kinds and three concealed layer categories. According to the trial findings, MLP neural network ratio amalgamation simplifies more accurately compared to different MLP neural network simulations.

In Ref.<sup>99</sup>, the author presented an ML-based scheme for PT fault analysis based on DGA, a bat algorithm (BA), as well as improving the probabilistic neural network (PNN). Investigation revealed that the recommended ANN-based approach was detected more precisely when compared to the Rogers ratios technique when a DGA technique centered on ANN was applied. The back propagation neural networks (BPNN) model is a highly common one amongst several neural network processes and it is being broadly utilized in various grounds of fault analysis. In particular, power electronic systems<sup>100</sup>, transformers<sup>101</sup>, batteries<sup>102,103</sup>, photovoltaic systems (PV)<sup>104,105</sup>, etc. be that as it may, the BPNN model still has several underlying deficiencies, for instance, dull merging speed and over-fitting difficulty<sup>106,107</sup>. Fortunately, a significant compilation of optimization processes has been established to optimize the BPNN version, such as Genetic algorithm (GA)<sup>108–111</sup>, Means end analysis (MEA)<sup>112</sup>, Particle Swarm Optimization (PSO)<sup>113,114</sup>, Simulated Annealing (SA)<sup>115</sup>, BA<sup>116,117</sup>, etc. Adaptive systems, including GA and MEA, are among these techniques and are currently utilized as optimizing approaches aiming for the perfect weights and limitations of ANNs.

## Proposed approach

In this work, a multi-classification model that is based on ML algorithms is presented to have an intelligible, precise, and clear understanding of DGA. Transformers are pivotal equipment in the transmission and distribution of electrical power. The failure of a particular unit during service may interrupt a massive number of consumers and disturb commercial activities in that area. Therefore, several monitoring techniques are proposed to ensure that the unit maintains an adequate level of functionality in addition to an extended useful lifespan. DGA is a technique commonly employed for monitoring the state of OITs. The understanding of DGA samples is however unsatisfactory from the perspective of evaluating incipient faults and relies mainly on the proficiency of test engineers.

The proposed model is utilized to investigate as well as assess the state and suitable gas name subscription of 138 TO samples that revealed different stray gassing characteristics in various South African substations. This is achieved by employing four ML classifiers. The advantages of the selected classifiers are discussed in Section “[Introduction](#)” of this study. The primary objective of this study requires the development of an ML-based health index (HI) model. It is suggested that HI be used to forecast the predicted output parameter, which is conceptually connected to the input characteristics, centered on the crucial assessment<sup>4</sup>. SVM can generate fresh information and categorize non-linear problems. KNN, in comparison, serves as a reliable, trained ML classifier that may be used to address categorization and prediction issues. Its primary drawback is the fact that it gets much more sluggish as the amount of data being used grows. During the normalizing step, the normalized ratio of every gas in every data collection sample is introduced. Figure 10 shows a flowchart that depicts an outline of the planned research.



**Figure 10.** Research flowchart for MC model.

### Dataset preparation

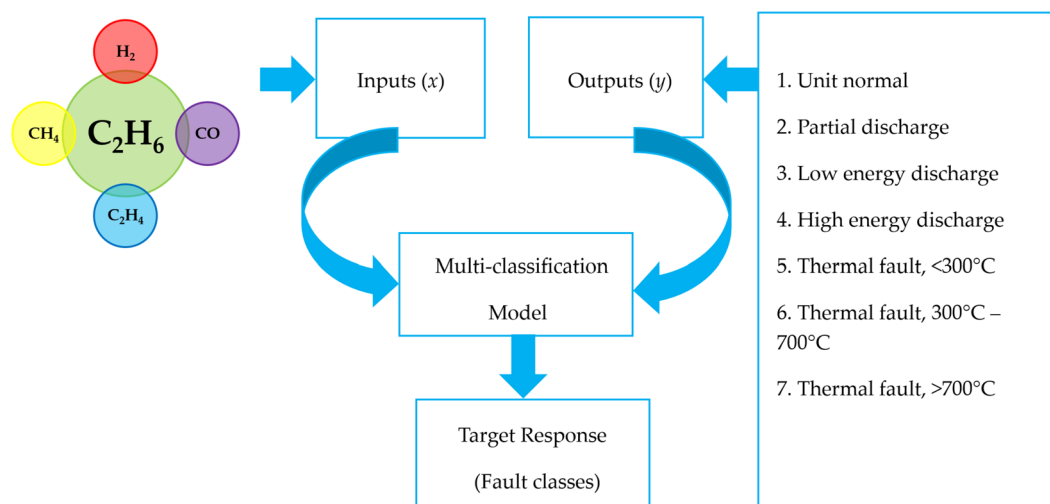
Relating to the 138 oil samples, 83 oil samples are utilized as training data, 25 oil samples as testing data, and the remaining 30 oil samples for validation purposes. The DT, SVM, KNN, and EC classifiers serve as a parameter to the HI model to forecast a trait. As a result, the presented HI computation technique will be considerably less costly. HI prognosis classifications are utilized as feature inputs with the transformer, and monitored modeling is applied. The developed HI framework must be solidly verified before it can be employed in practical applications. The dataset is loaded and distributed into the feature inputs ( $x$ ) as well as feature outputs ( $y$ ). The feature inputs are gases: ( $H_2$ ), ( $CH_4$ ), ( $C_2H_4$ ), ( $C_2H_6$ ), and ( $CO$ ) level in ppm, and feature outputs are faults category. In this present research, 83 oil samples are classified as training, 25 oil samples as evaluation, and the final 30 oil samples as verification. In Fig. 11, the function block diagram on the proposed model is illustrated for the diagnosis of various incipient transformer faults.

The proposed model generates a system response based on the feature inputs ( $x$ ) as well as targeted feature outputs ( $y$ ) absorbed into the network. The construction of a multi-classification model includes the identification of the best-performing system training techniques and parameters. In this present research, parameters are developed implicitly by considering efficiency and network reliability. The diagnostic gas contents acquired by DGA serve as the experimental base for diagnosis. The content data reflects the transformer conditions. These diagnostic gases include ( $H_2$ ), ( $CH_4$ ), ( $C_2H_4$ ), ( $C_2H_6$ ), and ( $CO$ ). To increase the effect of classification, the contents of these diagnostic gases are pre-processed using a specific data processing method, and seven features for fault diagnosis are extracted for the proposed model. From the successful implementation of the model, it can be concluded that the model has some potential advantages, which are as follows:

- It is a highly regularized strategy that is suitable for ill-posed issues.
- It offers a unique approach and has a high training curve/speed.

### Experimental setup

The training databank of 138 TO samples is used to develop the proposed model. A  $k$ -fold cross-validation method is employed in the development of the model. The  $k$ -fold cross-validation method is employed to assess



**Figure 11.** Function block diagram on the proposed model.

the model's competence with new data. The procedure has a single parameter called  $k$  that refers to the number of groups that a given data sample is to be split into. In the setup, a 30-fold cross-validation approach is employed to execute the experiment. As a result, during the 30-fold validation process, the operation is performed numerous times with varied partitions of the data findings into 30 parts. When a specific value for  $k$  is selected, it may be used in place of  $k$  in the reference to the model, such as  $k = 30$  becoming 30-fold cross-validation. It implies that the model will be evaluated several times to increase confidence in the model design. This eliminates the concept of training the model only once and not knowing if the positive result is due to luck or not. Cross-validation is a strategy for testing ML models that involves training numerous ML models on subsets of the given input data and then evaluating them on the corresponding subset. Cross-validation can be used to detect overfitting, or the failure to generalize a pattern. Performing 30-fold cross-validation generates 30 models, 30 data sources to train the models, 30 data sources to evaluate the models and 30 evaluations, one for each model.

In the present study, to evaluate the efficiency, several classifiers were put to work: (i) DT, (ii) SVM, (iii) KNN, and (iv) EC. Consequently, before the construction of the ML model, the settings of the proposed classifiers need to be established. Table 14 summarizes the ML setting configuration. The forward selection feature of stepwise regression<sup>118</sup> is used in the setup. Each term is either removed or included as a feature input vector according to the  $p$ -value of the present or newly entered data inputs. The  $p$ -value determines the likelihood it is to obtain a certain result when the null-hypothesis is assumed to be true. The null-hypothesis is the argument in scientific study, that no relationship exists between two sets of data or variables being trained/tested. The null-hypothesis states that any empirically observed difference is due only to chance and that no fundamental causal relationship exists, thus the word "null"<sup>119</sup>. As a result, if the null-hypothesis is considered to be true, the  $p$ -value estimates how odd the tested sample is. The likelihood of a null-hypothesis experiment is denoted as a  $p$ -value using the  $\alpha$  parameter for term addition as well as the  $\beta$  parameter<sup>120</sup> for term deletion. The  $\alpha$  and  $\beta$  parameters are the threshold values against which  $p$ -values are measured. It demonstrates how significant the observed results must be for a significance test to reject the null-hypothesis. Every data entry point indicates a different form of gas collected from DGA. The  $p$ -value following a stepwise regression analysis was used to choose the input data. The benefit of the  $p$ -value is that its parameters can be experimentally modified to achieve the best results.

### Training and testing of the ML models

As the input characteristics are supplied through the transformer HI estimation, supervisory training is adopted. 83 of the oil samples are classified as training, 25 as evaluation, and the final 30 oil samples as verification. To accommodate the narrow distribution of data, an evaluation threshold of 30 was utilized, particularly for the "Very Poor" data. In Ref.<sup>121</sup>, even though 83 of the transformer HI classifications are chosen to be utilized for training along with 25 for assessment, minimal transformer HI classifications are evaluated. As a result, obtaining more accurate data distribution is advantageous. Furthermore, a cross-validation approach is adopted for detecting overfitting or underfitting. After the model has been cross-validated, the settings can be adjusted for the next model if it does not meet the required standards. In light of this, a 30-fold cross-validation approach is carried out in a manner comparable to selection in Ref.<sup>122</sup>. The 30-fold validating approach is carried out repeatedly using different 30-part splits within the test findings. Furthermore, in this work, a comparison analysis is also conducted against the conventional DGA approaches to certify the proposed model.

### Classification accuracy

The proposed model is tested using 30-testing datasets. The precision of the classification indicates how frequently a classification algorithm is accurate. The formula for the  $S_{fn}$  is given in (2).

$$S_{fn} = \frac{P_{fn}}{\text{Number of cases of } fn} \times 100, \quad (2)$$

where  $S_{fn}$  is the proportion of valid prognosis of a certain fault type,  $fn$ , and  $P$  is the valid prognosis<sup>11</sup>.

Consistency (C) indicates the precision of the model in each dataset. This gives an enhanced metric of incorrectly categorized occurrences. The equation for the C is provided in (3).

$$C = \frac{\sum_1^{fn} S_{fn}}{\text{Number of fault types}}. \quad (3)$$

$f_n$  = type fault code ( $n=1,2,3,4,5$ ).

No	Classifier	Parameter
1	DT	Highest no. of splits: 121, Splitting criteria: Towing rule
2	SVM	Regularization = 12 norm, Box restriction setting: 896.5514 Kernel operation: Cubic, Loss = Square hinge
3	KNN	No. of neighbour = 30, Distance = Minkowski
4	EC	Ensemble scheme: AdaBoost, No. of trainees: 138, Training ratio: 0.8995, Highest no. of split up: 30

**Table 14.** Summary of MC model configuration.

The precision ( $A$ ) of the classifiers is determined by their valid prognosis ( $P$ ) in identifying distinct faults. The precision of ( $A$ ) is calculated as illustrated in (4).

$$\%A = \frac{T_{sp}}{T_{tc}} \times 100 \quad (4)$$

where  $T_{sp}$  is the total number of correct predictions and  $T_{tc}$  is the cumulative number of events<sup>76,93,100,105</sup>.

## Materials and protocols

Transformers help diversified settings that make them liable to broad failures whose outcomes are extended occurrences of power outages and disrupted commercial activities. In contrast to an overhead power line that is painless to overhaul, transformers are factory-sealed, denying technicians on-site the ability to inspect their active-part assemblies. The inception of faults of transformers during operation generates considerable hammering of revenues to power utility owners in addition to the excessive damages or replacement expenditures and the probability of an explosion. DGA is the only available scheme that provides the means to feasibly identify incipient transformer faults. Though the estimation precision of DGA schemes remains reasonably eminent, the schemes employed to understand DGA samples remain dependent on the proficiency of test engineers as opposed to precise interpretation. This work consequently sought to introduce and carry out a novel multi-classification ML-based DGA interpretation scheme that explains DGA samples strictly concerning multinomial data sets.

The oil samples used in this research came from several South African power stations. The databank contains five flammable gases initiated from distinct oil specimens collected from transformers in service. Classification Learner App in MATLAB/Simulink was employed to train the gas concentrations of 138 oil samples. The proposed MATLAB model is valid since the coding tool, MATLAB Simulink as well as the toolbox are widely utilized in the research and engineering field. The Classification Learner tool is a highly efficient platform that allows you to engage with data, choose features, establish cross-validation methods, train models, and evaluate outcomes<sup>13,20,90</sup>. It is used for routine activities such as:

- Data import and cross-validation scheme configuration.
- Data exploration as well as feature selection.
- Model training employing different classification tools.
- Model comparison and evaluation.

Researchers can develop and verify classification models by applying different techniques through the use of the Classification Learner app. Analyze the validating flaws of the simulations post-training, then select the most accurate model based on the outcomes. From the data, 83 oil samples are utilized as training data, 25 oil samples as testing data, and the remaining 30 oil samples for validation purposes. Eight classes of faults were detected from the oil data samples with a supplementary label designating a normal gas concentration. In response to 138 oil samples investigated, there is a total of 7 types of faults present as indicated in Table 15. Five gas proportions are assessed by adopting the IEC 60599:2022 guideline proposal considering the six flammable gases extracted in the different oil samples<sup>21</sup>.

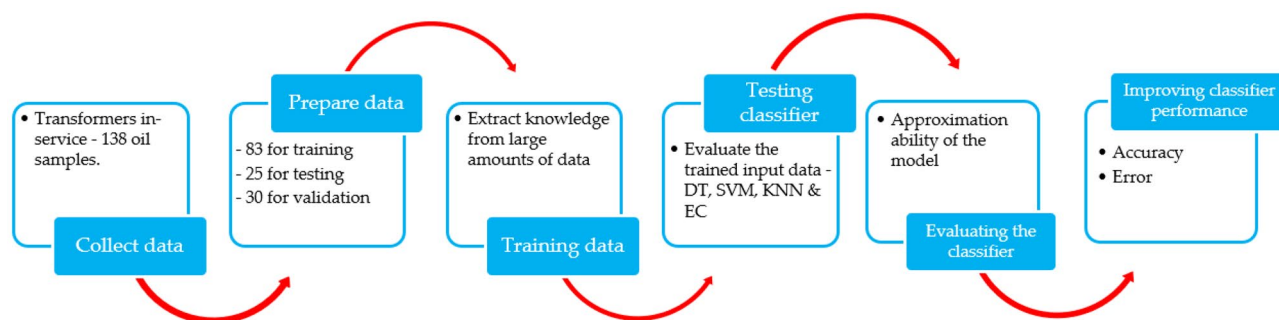
The objective of the preliminary exercise was to check the oil data samples so that faults may be detected and diagnosed. Following data pre-processing, DGA data clarification began, and a model was established utilizing the stages shown in Fig. 12. The authors provided Fig. 13, which depicts the different phases, to broaden the reader's perception.

## Results

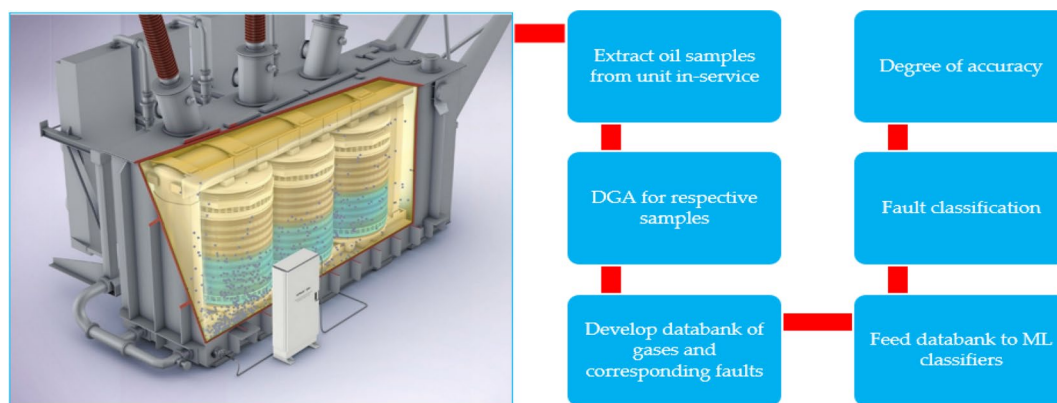
Classification Learner App in MATLAB/Simulink was utilized to train the gas concentrations of 138 oil samples extracted from various South African plants. ML classifiers in the Classification Learner App were selected and a 30-fold cross-validation was used to set the training and testing data for the model. Based on the total dataset, 83 oil samples are utilized as training data, 25 oil samples as testing data, and the remaining 30 oil samples for validation purposes. Cross-validation of 30-folds was selected, which in other words implies that the training and testing operations were repeated 30 times. Stepwise regression was used to choose the gases from the DGA

Type	Fault
PD	Partial discharge
D1	Low energy discharge
D2	High energy discharge
NF	No-Fault
T1	Thermal fault < 300 °C
T2	Thermal fault, 300 °C < T < 700 °C
T3	Thermal fault, > 700 °C
D + T	-

**Table 15.** Classification of transformer faults.



**Figure 12.** ML workflow.



**Figure 13.** TO evaluation phases.

that had the most significant feature for identifying transformer faults from the input ( $x$ ) and output ( $y$ ) data. Table 16 demonstrates the results of stepwise regression. The samples utilized to train as well as test the model have comparable traits. The  $p$ -value examines the null-hypothesis, and it possesses a likelihood of zero. An indicator with a small  $p$ -value, such as CO, which has a value of  $1.0214 \times 10^{-34}$ , is a good contributor to the model in terms of the specified characteristics. A small  $p$ -value for the gas indicates that the DGA data for that specific gas has a higher connection with the transformer fault type. The standard error is crucial for establishing the robustness of the connection between the predictive model and the reaction variable. Furthermore, standard error provides accessibility to the credibility of the  $p$ -values since it shows the numerical range that the measured numbers deviate from the prediction line. As demonstrated in Table 16, a smaller standard error correlates to a faster reaction since the model developed provides measurements of the reaction variable, which is the fault type closest to the aptness line.

The transformer states are classified into four categories: A, B, C, and D, with A signifying excellent state, B signifying fair state, C indicating that servicing is needed, and D signifying a detrimental or failing unit. This is shown in Table 17. The interpretations and constraints offered by national standards such as IEEE, ICE, and Eskom standards are used to identify these categories.

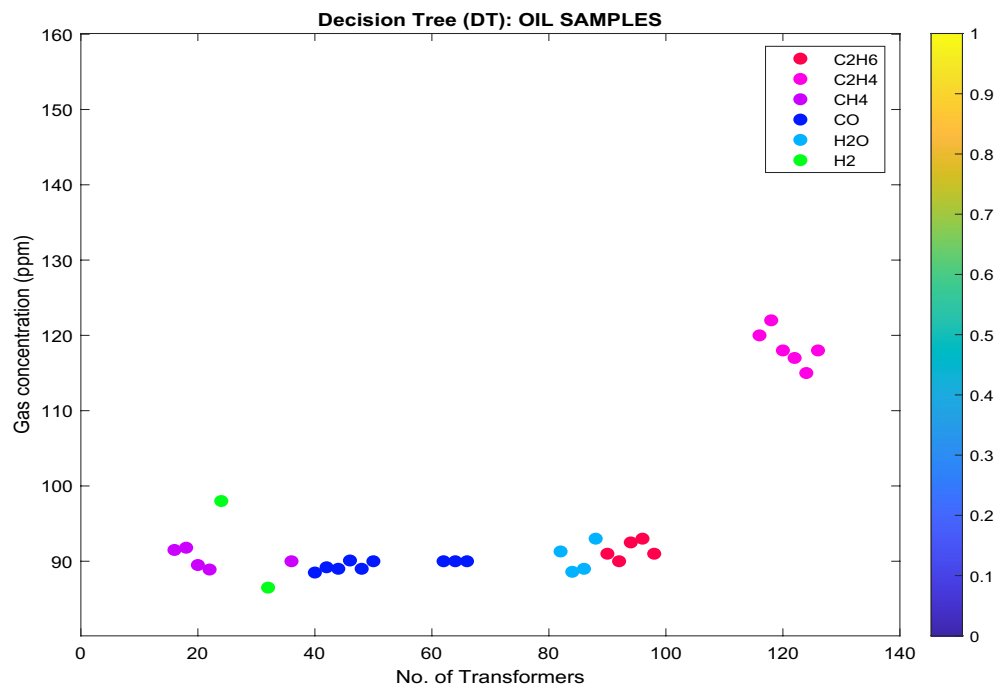
Figures 14, 15, 16 and 17 show the results of the four classifiers employed to analyze and evaluate the state and suitable gas name subscription of 138 TO samples that revealed different stray gassing characteristics in various South African substations. These gases: ( $H_2$ ), ( $CH_4$ ), ( $C_2H_2$ ), ( $C_2H_4$ ), and (CO) concentration were detected in 31

Type of gas	Training data			Testing data		
	p-value	Regression coefficient ( $\times 10^{-3}$ )	Standard error ( $\times 10^{-3}$ )	p-value	Regression coefficient ( $\times 10^{-3}$ )	Standard error ( $\times 10^{-3}$ )
$C_2H_6$	0.0398	-0.0156	0.007424	0.0378	-0.0156	0.007421
$C_2H_4$	0.9540	-0.0006	0.009493	0.9326	-0.0009	0.009492
$CH_4$	0.1683	0.0162	0.001128	0.1768	0.0016	0.001506
CO	$1.0214 \times 10^{-34}$	0.0073	0.000359	$1.0978 \times 10^{-34}$	0.0072	0.000359
$H_2$	$5.6277 \times 10^{-21}$	0.1694	0.014008	$5.3724 \times 10^{-21}$	0.1685	0.013928

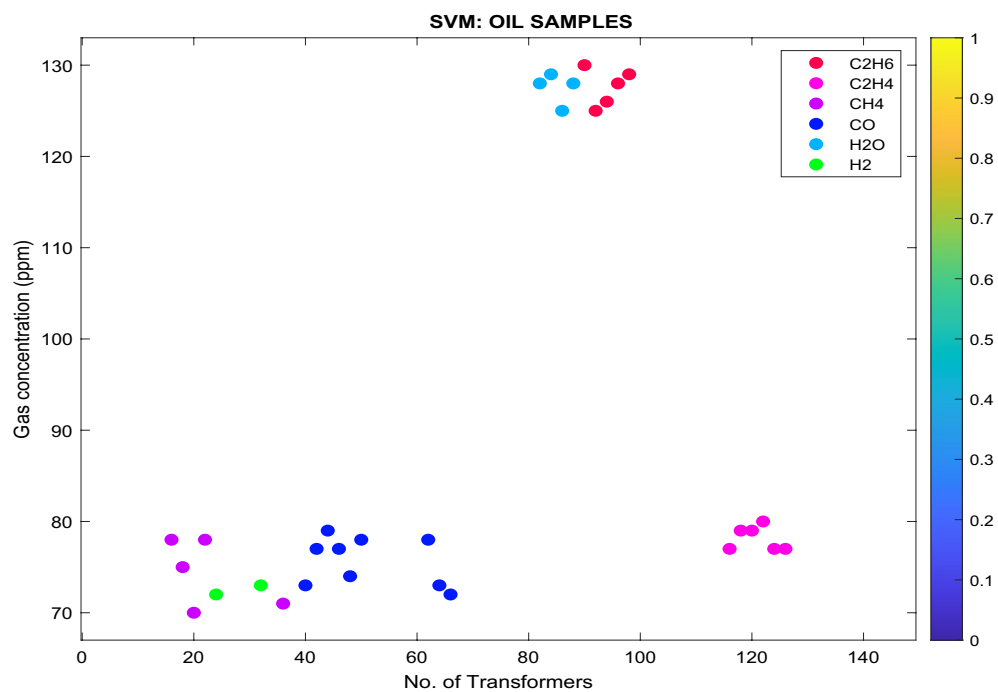
**Table 16.** Findings of feature extraction utilizing stepwise regression<sup>118</sup>.

Category	Percentage (%)	Transformer state
A	85-100	Excellent
B	70-84	Fair
C	50-69	Service
D	30-49	Detrimental/failed

**Table 17.** Transformer categories.

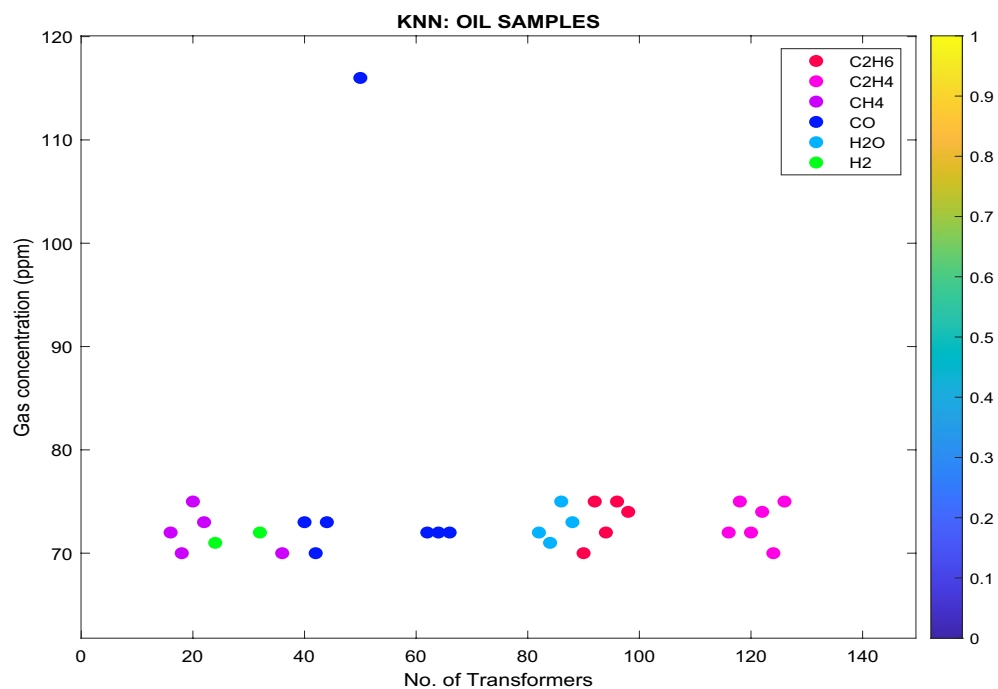


**Figure 14.** Analyzed databank using DT classifier.

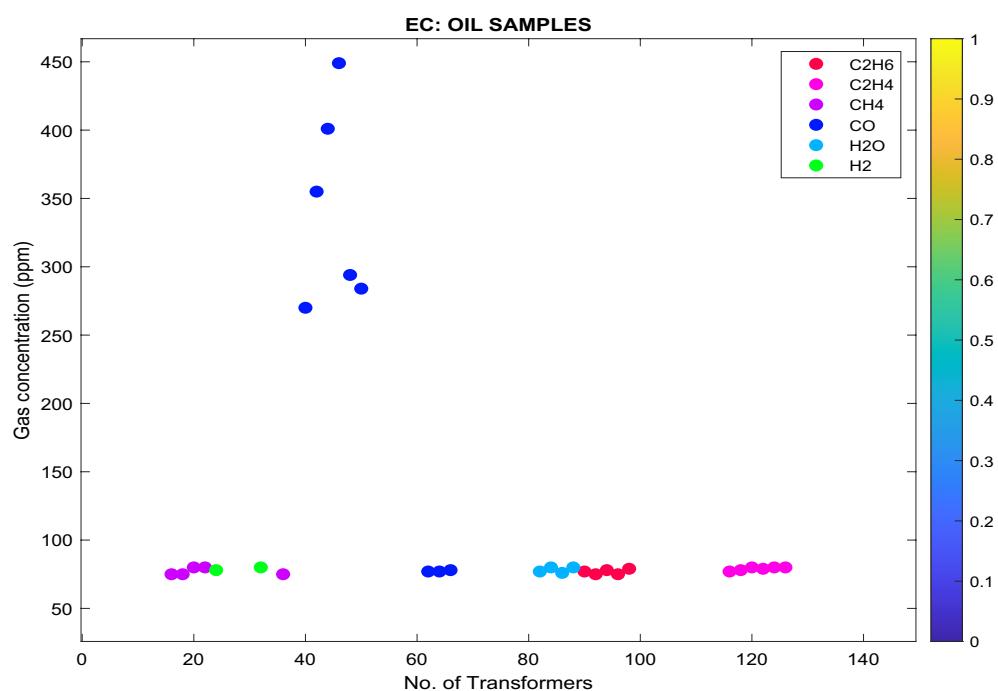


**Figure 15.** Analyzed databank using SVM classifier.





**Figure 16.** Analyzed databank using KNN classifier.



**Figure 17.** Analyzed databank using EC classifier.

transformers. These were based on transformers that were significantly affected by faults. The same dataset was used to train the classifiers, and it was performed to determine how well each classifier performs when underfitting or overfitting occurs. This occurs when the model cannot determine a meaningful relationship between the input ( $x$ ) and output ( $y$ ) data. Underfit models are more probable if they have not been trained for the proper amount of time on a large number of data points.

Therefore, the following observations were made, and suggestions were made based on the findings and consultation with the IEC 60599:2022 and Eskom standard:

In Fig. 14:

- Transformers (in pink) are classified as state 2 since the  $C_2H_4$  concentration is between 101 and 700 ppm, as specified in Table 7. These transformers are classified as Class C, as specified in Table 17. The recommendations are: (i) exercise caution; (ii) analyze for individual gases to find cause; (iii) determine load dependence.
- Transformers (in red) are classified as state 2 since the  $C_2H_6$  concentration is between 66 and 100 ppm. These transformers are classified as Class B. The recommendations are: (i) exercise caution; (ii) analyze for individual gases to find cause; (iii) determine load dependence.
- Transformers (in purple) are classified as state 4 since the  $CH_4$  concentration is < 120 ppm. These transformers are classified as Class A. The recommendation is to: (i) No action is required.
- The moisture is 25 ppm (light blue), which is greater than 15 ppm (Eskom Specification—Ref: 240-75661431.). These transformers are classified as Class C. The recommendation is to (i) TO needs refinement/service to enhance the oil state and resampling.
- Transformers (dark blue) are classified as state 1 since the CO concentration is < 350 ppm. These transformers are classified as Class A. The recommendations are: (i) No action is required.
- Transformers (in green) are classified as state 1 since the  $H_2$  concentration is < 100 ppm. These transformers are classified as Class A. The recommendations are: No action is required.

In Fig. 15:

- Transformers (in pink) are classified as state 2 since the  $C_2H_4$  concentration is between 51 and 100 ppm, as specified in Table 7. These transformers are classified as Class B, as specified in Table 17. The recommendations are: (i) exercise caution; (ii) analyze for individual gases to find cause; (iii) determine load dependence.
- Transformers (in red) are classified as state 3 since the  $C_2H_6$  concentration is between 101 and 150 ppm. These transformers are classified as Class C. The recommendations are: (i) plan service/maintenance; (ii) analyze for individual gases to find cause; (iii) remove if possible.
- Transformers (in purple) are classified as state 1 since the  $CH_4$  concentration is < 120 ppm. These transformers are classified as Class A. The recommendation is to: (i) No action is required.
- The moisture is 25 ppm (light blue), which is greater than 15 ppm (Eskom Specification—Ref: 240-75661431.). The recommendation is to (i) TO needs refinement to enhance the oil state and resampling.
- Transformers (dark blue) are classified as state 1 since the CO concentration is < 350 ppm. These transformers are classified as Class A. The recommendations are: (i) No action is required.
- Transformers (in green) are classified as state 1 since the  $H_2$  concentration is < 100 ppm. These transformers are classified as Class A. The recommendations are: No action is required.

In Fig. 16:

- Transformers (in pink) are classified as state 2 since the  $C_2H_4$  concentration is between 51 and 100 ppm, as specified in Table 7. These transformers are classified as Class B, as specified in Table 17. The recommendations are: (i) exercise caution; (ii) analyze for individual gases to find cause; (iii) determine load dependence.
- Transformers (in red) are classified as state 2 since the  $C_2H_6$  concentration is between 66 and 100 ppm. These transformers are classified as Class B. The recommendations are: (i) exercise caution; (ii) analyze for individual gases to find cause; (iii) determine load dependence.
- Transformers (in purple) are classified as state 1 since the  $CH_4$  concentration is < 120 ppm. These transformers are classified as Class A. The recommendation is to: (i) No action is required.
- The moisture is 25 ppm (light blue), which is greater than 15 ppm (Eskom Specification—Ref: 240-75661431.). The recommendation is to (i) TO needs refinement to enhance the oil state and resampling.
- Transformers (dark blue) are classified as state 1 since the CO concentration is < 350 ppm. These transformers are classified as Class A. The recommendations are: (i) No action is required.
- Transformers (in green) are classified as state 1 since the  $H_2$  concentration is < 100 ppm. These transformers are classified as Class A. The recommendations are: No action is required.

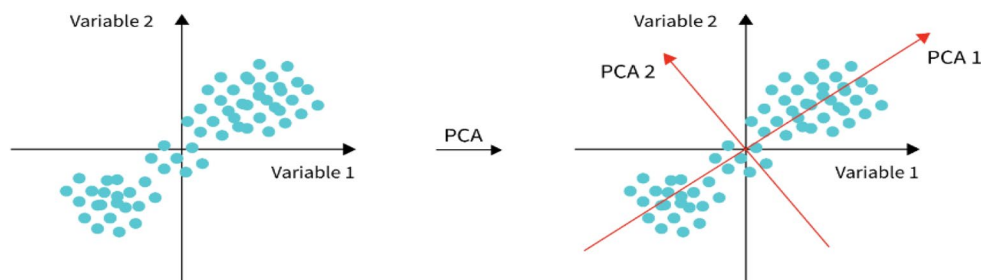
In Fig. 17:

- Transformers (in pink) are classified as state 2 since the  $C_2H_4$  concentration is between 51 and 100 ppm, as specified in Table 7. These transformers are classified as Class B, as specified in Table 17. The recommendations are: (i) exercise caution; (ii) analyze for individual gases to find cause; (iii) determine load dependence.
- Transformers (in red) are classified as state 2 since the  $C_2H_6$  concentration is between 66 and 100 ppm. These transformers are classified as Class B. The recommendations are: (i) exercise caution; (ii) analyze for individual gases to find cause; (iii) determine load dependence.
- Transformers (in purple) are classified as state 1 since the  $CH_4$  concentration is < 120 ppm. These transformers are classified as Class A. The recommendation is to: (i) No action is required.
- The moisture is 25 ppm (light blue), which is greater than 15 ppm (Eskom Specification—Ref: 240-75661431.). The recommendation is to i) TO needs refinement to enhance the oil state and resampling.
- 3 × Transformers (dark blue) are classified as state 1 since the CO concentration is < 350 ppm. These transformers are classified as Class A. The recommendations are: (i) No action is required. 3 × Transformers (dark blue) are classified as state 2 since the CO concentration is between 351 and 570 ppm. These transformers are classified as Class B. The recommendations are: (i) No action is required.

- Transformers (in green) are classified as state 1 since the H<sub>2</sub> concentration is < 100 ppm. These transformers are classified as Class A. The recommendations are: No action is required.

The ML classification outcomes of 138 oil samples without using the principal component analysis (PCA) are presented in the study. PCA is a well-known feature-harvesting method in mathematical research. This method removes key elements by linear conversion and obtains minimum sizes to demonstrate original data<sup>72</sup>. As a result, PCA is utilized to minimize parameter sizes, remove duplicate data, optimize classifier construction with minimal data loss, as well as enhance classification efficiency<sup>29</sup>. PCA includes the following stages: (i) obtaining the dataset; (ii) normalizing the dataset; (iii) calculating the correlation array; and (iv) interpreting the correlation array<sup>55</sup>. Figure 18 depicts the principle.

The classification learner app utilized for ML classification in MATLAB software provides a diverse set of sub-classifiers. The purpose of integrating these classifiers was to achieve optimal accuracy, and then employ the classifier/s that produce significant results for condition monitoring of the transformer. The results of the classifiers will differ due to some classifiers experiencing overfitting/underfitting as a result of (i) training data size being too small or not containing enough data samples to accurately represent all possible input data values; (ii) when the training data contains large amounts of irrelevant information, known as noisy data; and (iii) when the model trains for too long on a single sample set of data. Moreover, each classifier has different strengths and attributes. The training, validation, and testing process usually takes time, and the amount of time varies depending on the size of the data set and the competence of the classifier. Therefore, the longer it takes to train the classifier, the lower the accuracy, hence the results will be slightly different. The results of the ML classifiers are shown in Table 18. It is worth noting from the results that the highest degree of accuracy is 87.7%, which was produced by Bagged Trees, followed by Fine KNN with 86.2%, and the third in rank is Quadratic SVM with 84.1%.



**Figure 18.** Principal analysis component (PCA) principle.

Classifier	Type	Accuracy (%)	Prediction speed (objects/sec.)	Training time (sec.)
DT	Fine tree	82.6	360	79.946
	Medium tree	82.6	360	68.674
	Coarse tree	82.6	390	80.457
SVM	Linear SVM	82.6	53	70.575
	Quadratic SVM	84.1	55	59.946
	Cubic SVM	82.6	71	84.655
	Fine Gaussian SVM	82.6	73	84.433
	Medium Gaussian SVM	82.6	100	103.5
	Course Gaussian SVM	79	100	143.28
KNN	Fine KNN	86.2	230	77.51
	Medium KNN	82.6	240	79.36
	Coarse KNN	77.5	250	159.39
	Cosine KNN	82.6	240	81.82
	Cubic KNN	81.9	350	111.81
	Weighted KNN	81.9	390	111.42
EC	Boosted trees	77.5	540	116.2
	Bagged trees	87.7	34	59.93
	Subspace discriminant	79	38	139.66
	Subspace KNN	81.9	28	64.11
	RUSBoosted trees	82.6	46	67.58

**Table 18.** ML DGA classification outcomes.

Technique	Correct fault identification (%)
Proposed multi-classification model	82.6; 84.1; 86.2, 87.7
CIGRE approach <sup>123</sup>	80.76
Doernenburg approach <sup>124</sup>	79.03
KG approach <sup>125</sup>	75.8
Nomograph approach <sup>126</sup>	68.69
IEC approach <sup>127,128</sup>	82.25
Duval triangle approach <sup>129,130</sup>	84.11
Roger's ratio approach <sup>41,131</sup>	71.15

**Table 19.** Comparison between the diagnostic accuracy of several DGA approaches and the proposed multi-classification model for 138 oil samples.

Table 19 illustrates the comparison results of the proposed model and seven DGA approaches, namely: the CIGRE approach, Doernenburg approach, KG approach, Nomograph approach, IEC approach, Duval triangle approach, and Roger's ratio approach. The accuracy of the proposed multi-classification model, consisting of DT, SVM, KNN, and EC is 82.6%, 84.1%, 82.6%, and 87.7% respectively. These were the outcomes of the study. Therefore, when these findings are compared to DGA approaches, they demonstrate a considerable increase in the proportion of accurate fault-type estimation, which is above 20%.

## Conclusions

In this work, a novel multi-classification model that is based on ML algorithms was proposed to have an intelligible, precise, and perfect understanding of DGA. The proposed model was used to analyze 138 TO samples that revealed different stray gassing characteristics in various South African substations. This was achieved by employing four ML classifiers. Experimental evidence using DT classifier viz. Fine Tree, Medium Tree, and Coarse Tree suggested that these classifiers are feasible in classifying stray gas characteristics from normal TO with a degree of accuracy of 82.6%. Additionally, the test results from the SVM classifier comprised Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM, and Course Gaussian SVM conclude that these classifiers are viable in classifying stray gassing specificities with the degree of accuracy from 79 to 84.1%. Further, experimental findings from KNN employing Fine KNN, Medium KNN, Coarse KNN, Cosine KNN, Cubic KNN, and Weighted KNN appear to indicate that these classifiers are feasible in classifying stray gassing properties with a degree of accuracy from 77.5 to 86.2%. Lastly, experimental information Boosted Trees, Bagged Trees, Subspace Discriminant, Subspace KNN, and RUS Boosted Trees indicate that these classifiers are feasible in classifying stray gassing peculiarities with a degree of accuracy from 77.5 to 87.7%. The findings can be explained that there are different DG concentrations in stray gassing phenomena from normal TO. In this work, it was demonstrated that the gas concentrations of transformer H<sub>2</sub>, CH<sub>4</sub>, CO, C<sub>2</sub>H<sub>4</sub>, and C<sub>2</sub>H<sub>6</sub> can be used to discriminate stray gassing phenomena from normal TO and their differences can be classified with the highest degree of accuracy of 87.7% by the Bagged Trees classifier over other ML classifiers. It is advisable that the training, testing, and validation of DGA oil samples be extended and tested several times to validate the findings in this work.

For forthcoming studies, the findings amassed in this work can be employed as a benchmark in developing a portable device that utilizes ML algorithms herein for the validation of DGA results. Another proposal is that additional research is undertaken in applying DTs to formulate new stray gassing limits for the various gas concentrations. In this fashion, the production of DGs will be easily interpreted with proper permissible levels thereby utility owners and consequently, consumers do not have to suffer from prolonged power outages. Furthermore, the authors will test the proposed model based on ANN algorithms and compare it to the results presented in this work to determine which approach yields more accurate statistics.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 18 September 2023; Accepted: 26 December 2023

Published online: 03 January 2024

## References

1. Poonnoy, N., Suwanasri, C. & Suwanasri, T. Neural network approach to dissolved gas analysis for fault analysis in power transformers. In *2022 International Electrical Engineering Congress (iEECON)* (eds Poonnoy, N. *et al.*) 1–4 (IEEE, 2022).
2. Jia, J. *et al.* Validity evaluation of transformer DGA online monitoring data in grid edge systems. *IEEE Access* **8**, 60759–60768 (2020).
3. Wang, L., Littler, T. & Liu, X. Gaussian process multi-class classification for transformer fault diagnosis using dissolved gas analysis. *IEEE Trans. Dielectr. Electr. Insul.* **28**(5), 1703–1712 (2021).
4. Gouda, O. E., El-Hoshy, S. H. & Ghoneim, S. S. Enhancing the diagnostic accuracy of DGA techniques based on IEC-TC10 and related databases. *IEEE Access* **9**, 118031–118041 (2021).
5. Cui, H. *et al.* Impact of load ramping on power transformer dissolved gas analysis. *IEEE Access* **7**, 170343–170351 (2019).

6. Emara, M. M., Peppas, G. D. & Gonos, I. F. Two graphical shapes based on DGA for power transformer fault types discrimination. *IEEE Trans. Dielectr. Electr. Insul.* **28**(3), 981–987 (2021).
7. Huang, X., Yuan, Y. and Li, J. A review of transformer fault diagnosis based on information system theory and machine learning (2023).
8. Islam, N. *et al.* Power transformer health condition evaluation: A deep generative model aided intelligent framework. *Electr. Power Systems Res.* **218**, 109201 (2023).
9. Kang, Y., Park, H., Smit, B. & Kim, J. A multi-modal pre-training transformer for universal transfer learning in metal-organic frameworks. *Nat. Mach. Intell.* **5**(3), 309–318 (2023).
10. Faiz, J. & Soleimani, M. Assessment of computational intelligence and conventional dissolved gas analysis methods for transformer fault diagnosis. *IEEE Trans. Dielectr. Electr. Insul.* **25**(5), 1798–1806 (2018).
11. Buchacz, T., Buchacz, J. & Duval, M. Stray gassing of oil in HV transformers. *IEEE Trans. Dielectr. Electr. Insul.* **28**(5), 1729–1734 (2021).
12. Maccartney, G. R., Rappaport, T. S., Samimi, M. K. & Sun, S. Millimeter-wave omnidirectional path loss data for small cell 5G channel modeling. *IEEE Access* **3**, 1573–1580 (2015).
13. Taha, I. B., Ibrahim, S. & Mansour, D.-E.A. Power transformer fault diagnosis based on DGA using a convolutional neural network with noise in measurements. *IEEE Access* **9**, 111162–111170 (2021).
14. Duval, M. & Buchacz, J. Identification of arcing faults in paper and oil in transformers—Part I: Using the Duval pentagons. *IEEE Electr. Insul. Mag.* **38**(1), 19–23 (2021).
15. Yang, D., Qin, J., Pang, Y. & Huang, T. A novel double-stacked autoencoder for power transformers DGA signals with an imbalanced data structure. *IEEE Trans. Ind. Electron.* **69**(2), 1977–1987 (2021).
16. Camponogara, E., Jia, D., Krogh, B. H. & Talukdar, S. Distributed model predictive control. *IEEE Control Syst. Mag.* **22**(1), 44–52 (2002).
17. Tra, V., Duong, B.-P. & Kim, J.-M. Improving the diagnostic performance of a power transformer using an adaptive over-sampling method for imbalanced data. *IEEE Trans. Dielectr. Electr. Insul.* **26**(4), 1325–1333 (2019).
18. Guo, Z., Li, C., Rong, Z. & Qi, B. DGA warning values for PD defects obtained from normal data analysis on a transformer experimental platform. *IEEE Trans. Dielectr. Electr. Insul.* **28**(5), 1759–1766 (2021).
19. Wang, X., Wang, Z., Liu, Q. & Dyer, P. Dissolved gas analysis of thermal faults in transformer liquids simulated using the immersed heating method. *IEEE Trans. Dielectr. Electr. Insul.* **25**(5), 1749–1757 (2018).
20. Ali, M. S., Omar, A., Jaafar, A. S. A. & Mohamed, S. H. Conventional methods of dissolved gas analysis using oil-immersed power transformer for fault diagnosis: A review. *Electr. Power Syst. Res.* **216**, 109064 (2023).
21. Raj, R. A., Sarathkumar, D., Andrews, L. J. B. & Venkatachary, S. K. Key gases in transformer oil—an analysis using self organizing map (SOM) neural networks. In *2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT)* (eds Raj, R. A. *et al.*) 642–647 (IEEE, 2023).
22. Han, X. *et al.* A transformer condition recognition method based on dissolved gas analysis features selection and multiple models fusion. *Eng. Appl. Artif. Intell.* **123**, 106518 (2023).
23. Shutenko, O. & Kulyk, O. Recognition of low-temperature overheating in power transformers by dissolved gas analysis. *Electr. Eng.* **104**(4), 2109–2121 (2022).
24. Soni, R. & Mehta, B. Diagnosis and prognosis of incipient faults and insulation status for asset management of power transformer using fuzzy logic controller & fuzzy clustering means. *Electr. Power Syst. Res.* **220**, 109256 (2023).
25. Ekojono, Prasojo, R. A., Apriyani, M. E. & Rahmanto, A. N. Investigation on machine learning algorithms to support transformer dissolved gas analysis fault identification. *Electr. Eng.* **104**(5), 3037–3047 (2022).
26. Banshwar, A., Kumar Sharma, N., Pathak, M., Sharma, B. B. & Kumar, S. An investigation to different methods of health assessment in power transformers. In *Advances in Mechanical and Materials Technology: Select Proceedings of EMSME 2020* (eds Govindan, K. *et al.*) 749–756 (Springer, 2022).
27. Velásquez, R. M. A. Support vector machine and tree models for oil and Kraft degradation in power transformers. *Eng. Fail. Anal.* **127**, 105488 (2021).
28. Liu, J. *et al.* Classifying transformer winding deformation fault types and degrees using FRA based on support vector machine. *IEEE Access* **7**, 112494–112504 (2019).
29. Wu, Y., Sun, X., Dai, B., Yang, P. & Wang, Z. A transformer fault diagnosis method based on hybrid improved grey wolf optimization and least squares-support vector machine. *IET Gen. Transm. Distrib.* **16**(10), 1950–1963 (2022).
30. Alqudsi, A. & El-Hag, A. Application of machine learning in transformer health index prediction. *Energies* **12**(14), 2694 (2019).
31. Moradzadeh, A. & Pourhossein, K. Application of support vector machines to locate minor short circuits in transformer windings. In *2019 54th International Universities Power Engineering Conference (UPEC)* (eds Moradzadeh, A. & Pourhossein, K.) 1–6 (IEEE, 2019).
32. Guo, Y., Mustafaoglu, Z. & Koundal, D. Spam detection using bidirectional transformers and machine learning classifier algorithms. *J. Comput. Cogn. Eng.* **2**(1), 5–9 (2023).
33. Liu, Y., Li, J., Li, Z., Qiao, L., Chen, S. and Lv, X. Transformer fault diagnosis model based on iterative nearest neighbor interpolation and ensemble learning. In *Proc. of the 2019 2nd International Conference on Data Science and Information Technology*, 140–146 (2019).
34. Li, Z. *et al.* Fault diagnosis of transformer windings based on decision tree and fully connected neural network. *Energies* **14**(6), 1531 (2021).
35. Xu, Y., Zhou, C., Geng, J., Gao, S. & Wang, P. A method for diagnosing mechanical faults of on-load tap changer based on ensemble empirical mode decomposition, Volterra model and decision acyclic graph support vector machine. *IEEE Access* **7**, 84803–84816 (2019).
36. Zhong, M. *et al.* Dissolved gas in transformer oil forecasting for transformer fault evaluation based on HATT-RLSTM. *Electr. Power Syst. Res.* **221**, 109431 (2023).
37. Shahbazi, N., Bagheri, S. & Gharehpetian, G. Identification and classification of cross-country faults in transformers using K-NN and tree-based classifiers. *Electr. Power Syst. Res.* **204**, 107690 (2022).
38. Sinduja, M., Maheswari, R. & Vigneshwaran, B. Transformer oil quality assessment using machine learning techniques. In *2022 International Conference on Computer Communication and Informatics (ICCCI)* (eds Sinduja, M. *et al.*) 1–5 (IEEE, 2022).
39. En, B. 60599: 2016. Mineral oil-filled electrical equipment in service. Guidance on the interpretation of dissolved and free gases analysis [Electronic resource]. *British Standards Institute* (2016).
40. Ma, H., Saha, T. K., Ekanayake, C. & Martin, D. Smart transformer for smart grid—Intelligent framework and techniques for power transformer asset management. *IEEE Trans. Smart Grid* **6**(2), 1026–1034 (2015).
41. Soto, A. R., Lima, S. L. & Saavedra, O. R. Incipient fault diagnosis in power transformers by DGA using a machine learning ANN-mean shift approach. In *2019 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)* (eds Soto, A. R. *et al.*) 1–6 (IEEE, 2019).
42. Patekar, K. D. & Chaudhry, B. DGA analysis of transformer using Artificial neural network to improve reliability in Power Transformers. In *2019 IEEE 4th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON)* (eds Patekar, K. D. & Chaudhry, B.) 1–5 (IEEE, 2019).

43. Hussein, A. R., Dakhil, A. M., Rashed, J. R. & Othman, M. Intelligent expert system for diagnosing faults and assessing quality of power transformer insulation oil by DGA method. *Misan J. Eng. Sci.* **1**(1), 47–57 (2022).
44. Badawi, M. *et al.* Reliable estimation for health index of transformer oil based on novel combined predictive maintenance techniques. *IEEE Access* **10**, 25954–25972 (2022).
45. Cui, H., Jia, P., Peng, X. & Li, P. Adsorption and sensing of CO and C<sub>2</sub>H<sub>2</sub> by S-defected SnS<sub>2</sub> monolayer for DGA in transformer oil: A DFT study. *Mater. Chem. Phys.* **249**, 123006 (2020).
46. Wang, L., Littler, T. & Liu, X. Hybrid AI model for power transformer assessment using imbalanced DGA datasets. *IET Renew. Power Gen.* **17**, 1912–1922 (2023).
47. Demirci, M., Gözde, H. & Taplamacioglu, M. C. Improvement of power transformer fault diagnosis by using sequential Kalman filter sensor fusion. *Int. J. Electr. Power Energy Syst.* **149**, 109038 (2023).
48. Shutenko, O., Proskurnia, O. & Abramov, V. Comparative analysis of risks which are accompanied by the use of typical and boundary gases concentrations for the diagnostics of high voltage transformers. *Energetika* <https://doi.org/10.6001/energetika.v64i3.3806> (2018).
49. Jiang, J. *et al.* Dynamic fault prediction of power transformers based on lasso regression and change point detection by dissolved gas analysis. *IEEE Trans. Dielectr. Electr. Insul.* **27**(6), 2130–2137 (2020).
50. Dhini, A., Faqih, A., Kusumoputro, B., Surjandari, I. & Kusiak, A. Data-driven fault diagnosis of power transformers using dissolved gas analysis (DGA). *Int. J. Technol.* **11**(2), 388–399 (2020).
51. Li, S. *et al.* Interpretation of DGA for transformer fault diagnosis with complementary SaE-ELM and arctangent transform. *IEEE Trans. Dielectr. Electr. Insul.* **23**(1), 586–595 (2016).
52. Nemeth, B., Laboncz, S. & Kiss, I. Condition monitoring of power transformers using DGA and fuzzy logic. In *2009 IEEE Electrical Insulation Conference* (eds Nemeth, B. *et al.*) 373–376 (IEEE, 2009).
53. Zhang, Y. *et al.* Power transformer fault diagnosis considering data imbalance and data set fusion. *High Voltage* **6**(3), 543–554 (2021).
54. El-kenawy, E.-S.M. *et al.* Feature selection and classification of transformer faults based on a novel meta-heuristic algorithm. *Mathematics* **10**(17), 3144 (2022).
55. Mollmann, A. & Pahlavanpour, B. New guidelines for interpretation of dissolved gas analysis in oil-filled transformers. *Electra* **186**, 31–51 (1999).
56. Bohatyrewicz, P. & Banaszak, S. Assessment criteria of changes in health index values over time—A transformer population study. *Energies* **15**(16), 6078 (2022).
57. Jongen, R., Morshuis, P., Smit, J., Janssen, A. & Gulski, E. A statistical approach to processing power transformer failure data. In *19th International Conference on Electricity Distribution* Vol. 4 (eds Jongen, R. *et al.*) (Cired, 2007).
58. Bustamante, S. *et al.* Dissolved gas analysis equipment for online monitoring of transformer oil: A review. *Sensors* **19**(19), 4057 (2019).
59. Bakar, N. A., Abu-Siada, A. & Islam, S. A review of dissolved gas analysis measurement and interpretation techniques. *IEEE Electr. Insul. Mag.* **30**(3), 39–49 (2014).
60. Kim, Y., Lee, S., Seo, H., Jung, J. & Yang, H. Development of dissolved gas analysis (DGA) expert system using a new diagnostic algorithm for oil-immersed transformers. In *2012 IEEE International Conference on Condition Monitoring and Diagnosis* (eds Kim, Y. *et al.*) 365–369 (IEEE, 2012).
61. Zhao, A.-X., Tang, X.-J., Zhang, Z.-H. & Liu, J.-H. The DGA interpretation method using relative content of characteristic gases and gas-ratio combinations for fault diagnosis of oil-immersed power transformers. In *Proceedings of 2014 International Symposium on Electrical Insulating Materials* (eds Zhao, A.-X. *et al.*) 124–127 (IEEE, 2014).
62. Faiz, J. & Soleimani, M. Dissolved gas analysis evaluation in electric power transformers using conventional methods a review. *IEEE Trans. Dielectr. Electr. Insul.* **24**(2), 1239–1248 (2017).
63. Endo, F. *et al.* Analysis of gasses generated by electrical discharges in low viscosity silicone oil. In *2008 International Conference on Condition Monitoring and Diagnosis* (eds Endo, F. *et al.*) 272–275 (IEEE, 2008).
64. Oehlmann, K. & Mao, L. Evaluation of distribution network transformer dissolved gas analysis (DGA) data. In *PES T&D 2012* (eds Oehlmann, K. & Mao, L.) 1–4 (IEEE, 2012).
65. Abbasi, A. R. Fault detection and diagnosis in power transformers: A comprehensive review and classification of publications and methods. *Electr. Power Syst. Res.* **209**, 107990 (2022).
66. IEC. *Mineral Oil-Impregnated Electrical Equipment in Service-Guide to the Interpretation of Dissolved and Free Gases Analysis* (CEI, 2015).
67. B. EN60599, "and IEC 60599: 1999 *Mineral oil-impregnated electrical equipment in service guide to the interpretation of dissolved and free gases analysis*," *British Standard Institute (BSI) and International Electrotechnical Commission (IEC)* (1999).
68. Singh, J., Singh, S. & Singh, A. Distribution transformer failure modes, effects and criticality analysis (FMECA). *Eng. Fail. Anal.* **99**, 180–191 (2019).
69. Fernández, I., Ortiz, A., Delgado, F., Renedo, C. & Perez, S. Comparative evaluation of alternative fluids for power transformers. *Electr. Power Syst. Res.* **98**, 58–69 (2013).
70. I. E. Commission, "International Standard IEC 60599: 1999+ A1: 2007, Mineral oil-impregnated electrical equipment in service—a guide to the interpretation of dissolved and free gases analysis (1999).
71. Landucci Júnior, A. Manutenção pré-corretiva em transformadores de potência: um novo conceito de manutenção (2017).
72. Babukutty, S. & Khule, S. Dissolved gas analysis: A proactive approach for recognition of faults in transformer using MATLAB GUI. *J. Adv. Res. Signal Process. Appl.* **1**(2), 14–19 (2021).
73. Modi, M. M. & Patel, R. A. Intelligent approach to interpret incipient faults of power transformer from DGA database. *J. Inst. Eng. (India) Ser. B* **104**, 1–8 (2023).
74. J. Costa, D. Silva, and P. Branco, "Large-Power Transformers: Time Now for Addressing Their Monitoring and Failure Investigation Techniques. *Energies* 2022, 15, 4697, ed: s Note: MDPI stays neutral about jurisdictional claims in published ... (2022).
75. Robalino, D. M. & Mahajan, S. M. Dissolved gas analysis of a thermally overloaded oil-immersed current transformer. In *2010 IEEE/PES Transmission and Distribution Conference and Exposition: Latin America (T&D-LA)* (eds Robalino, D. M. & Mahajan, S. M.) 645–652 (IEEE, 2010).
76. Merita, A. K., Sharma, R. & Chauhan, S. Study and diagnosis of key gases to assess the condition of power transformers. In *2011 3rd International Conference on Electronics Computer Technology* Vol. 4 (eds Merita, A. K. *et al.*) 343–347 (IEEE, 2011).
77. Irungu, G., Akumu, A. & Munda, J. Fault diagnostics in oil-filled electrical equipment: Review of duval triangle and the possibility of alternatives. In *2016 IEEE Electrical Insulation Conference (EIC)* (eds Irungu, G. *et al.*) 174–177 (IEEE, 2016).
78. Putra, M. A. A., Prasoj, R. A., Novfowan, A. D. & Neelmani, N. Dissolved gas analysis of generator step up transformer in Grati power plant using random forest based method. *J. Telecommun. Netw. (Jurnal Jaringan Telekomunikasi)* **13**(1), 51–58 (2023).
79. Duval, M. A review of faults detectable by gas-in-oil analysis in transformers. *IEEE Electr. Insul. Mag.* **18**(3), 8–17 (2002).
80. Bulut, M. & Özcan, E. A new approach to determine maintenance periods of the most critical hydroelectric power plant equipment. *Reliab. Eng. Syst. Saf.* **205**, 107238 (2021).
81. Saroja, S., Haseena, S. & Madhavan, R. Dissolved gas analysis of transformer: An approach based on ML and MCDM. *IEEE Trans. Dielectr. Electr. Insul.* <https://doi.org/10.1109/TDEI.2023.3271609> (2023).

82. Baker, E., Nese, S. V. & Dursun, E. Hybrid condition monitoring system for power transformer fault diagnosis. *Energies* **16**(3), 1151 (2023).
83. Manoj, T. & Ranga, C. An intelligence expert system for quality health monitoring of power transformers using testing attributes. *Eng. Res. Express* **5**(2), 025033 (2023).
84. Zhang, D. *et al.* Transformer maintenance decision based on condition monitoring and fuzzy probability hybrid reliability assessment. *IET Gen. Transm. Distrib.* **17**(4), 976–992 (2023).
85. Altayef, E., Anayi, F., Packianather, M., Benmahamed, Y. & Kherif, O. Detection and classification of lamination faults in A 15 kVA three-phase transformer core using SVM, KNN and DT algorithms. *IEEE Access* **10**, 50925–50932 (2022).
86. Han, X. *et al.* A novel power transformer fault diagnosis model based on harris-hawks-optimization algorithm optimized kernel extreme learning machine. *J. Electr. Eng. Technol.* **17**(3), 1993–2001 (2022).
87. Behkam, R., Karami, H., Naderi, M. S. & Gharehpetian, G. Condition monitoring of distribution transformers using frequency response traces and artificial neural network to detect the extent of windings axial displacements. In *2022 26th International Electrical Power Distribution Conference (EPDC)* (eds Behkam, R. *et al.*) 18–23 (IEEE, 2022).
88. Jiang, J., Chen, R., Chen, M., Wang, W. & Zhang, C. Dynamic fault prediction of power transformers based on hidden Markov model of dissolved gases analysis. *IEEE Trans. Power Deliv.* **34**(4), 1393–1400 (2019).
89. Ma, X., Hu, H. & Shang, Y. A new method for transformer fault prediction based on multi-feature enhancement and refined long short-term memory. *IEEE Trans. Instrum. Meas.* **70**, 1–11 (2021).
90. Li, J., Li, G., Hai, C. & Guo, M. Transformer fault diagnosis based on multi-class AdaBoost algorithm. *IEEE Access* **10**, 1522–1532 (2021).
91. Rodríguez, J., Contreras, J. & Gaytán, C. Evaluation and interpretation of dissolved gas analysis of soybean-based natural ester insulating liquid. *IEEE Trans. Dielectr. Electr. Insul.* **28**(4), 1343–1348 (2021).
92. Laayati, O. *et al.* Toward smarter power transformers in microgrids: A multi-agent reinforcement learning for diagnostic. In *Digital Technologies and Applications: Proceedings of ICDTA'23, Fez, Morocco* Vol. 2 (eds Motahhir, S. & Bossoufi, B.) 640–649 (Springer, 2023).
93. Li, J., Hai, C., Feng, Z. & Li, G. A transformer fault diagnosis method based on parameters optimization of hybrid kernel extreme learning machine. *IEEE Access* **9**, 126891–126902 (2021).
94. Boonsaner, N., Chanchaoensook, P., Bunnag, C., Suwantaweek, A. & Vongphanich, K. The study and analysis of oil-immersed power transformer by using artificial neural network for designing program apply in the industry of testing oil-immersed transformers. In *2020 8th International Conference on Condition Monitoring and Diagnosis (CMD)* (eds Boonsaner, N. *et al.*) 274–277 (IEEE, 2020).
95. Suksagoonpanya, S. *et al.* Dissolved gas analysis of palm oil compared with mineral oil from different types of breakdown voltage. In *2020 8th International Conference on Condition Monitoring and Diagnosis (CMD)* (eds Suksagoonpanya, S. *et al.*) 330–333 (IEEE, 2020).
96. Shutenko, O. & Kulyk, O. Comparative analysis of the defect type recognition reliability in high-voltage power transformers using different methods of DGA results interpretation. In *2020 IEEE Problems of Automated Electrodrive. Theory and Practice (PAEP)* (eds Shutenko, O. & Kulyk, O.) 1–6 (IEEE, 2020).
97. Syafruddin, H. & Nugroho, H. P. Dissolved gas analysis (DGA) for diagnosis of fault in oil-immersed power transformers: A case study. In *2020 4th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)* (eds Syafruddin, H. & Nugroho, H. P.) 57–62 (IEEE, 2020).
98. Patel, R. & Thakur, H. Integrated fuzzy approach for incipient fault detection in power transformers. In *2016 IEEE Industrial Electronics and Applications Conference (IEACon)* (eds Patel, R. & Thakur, H.) 211–218 (IEEE, 2016).
99. Rai, A. S. & Maheshan, C. Artificial bee colony algorithm based interpretation of dissolved gas analysis. In *2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT)* (eds Rai, A. S. & Maheshan, C.) 106–110 (IEEE, 2019).
100. Mohamad, F., Hosny, K. & Barakat, T. Incipient fault detection of electric power transformers using fuzzy logic based on Roger's and IEC method. In *2019 14th International Conference on Computer Engineering and Systems (ICCES)* (eds Mohamad, F. *et al.*) 303–309 (IEEE, 2019).
101. Ghoneim, S., Mansour, D., Bedir, I. & Alharthi, M. A decision transformer fault diagnostics system based on dissolved gas analysis. In *2019 21st International Middle East Power Systems Conference (MEPCON)* (eds Ghoneim, S. *et al.*) 76–80 (IEEE, 2019).
102. Zeng, B. *et al.* A transformer fault diagnosis model based on hybrid grey wolf optimizer and LS-SVM. *Energies* **12**(21), 4170 (2019).
103. Benmahamed, Y., Teguair, M. & Boubakeur, A. Diagnosis of power transformer oil using PSO-SVM and KNN classifiers. In *2018 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)* (eds Benmahamed, Y. *et al.*) 1–4 (IEEE, 2018).
104. Ghosh, A. & Banerjee, S. A comparison between classical and advanced controllers for a boost converter. In *2018 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)* (eds Ghosh, A. & Banerjee, S.) 1–6 (IEEE, 2018).
105. Zope, N., Ali, S. I., Padmanaban, S., Bhaskar, M. S. & Mihet-Popa, L. Analysis of 132kV/33kV 15MVA power transformer dissolved gas using transport-X Kelman Kit through Duval's triangle and Roger's Ratio prediction. In *2018 IEEE International Conference on Industrial Technology (ICIT)* (eds Zope, N. *et al.*) 1160–1164 (IEEE, 2018).
106. Hamid, M. *et al.* Dissolved gas analysis (DGA) of vegetable oils under electrical stress. In *2017 International Conference on High Voltage Engineering and Power Systems (ICHVEPS)* (eds Hamid, M. *et al.*) 29–34 (IEEE, 2017).
107. Wattakapaiboon, W. & Pattanadech, N. The state of the art for dissolved gas analysis based on interpretation techniques. In *2016 International Conference on Condition Monitoring and Diagnosis (CMD)* (eds Wattakapaiboon, W. & Pattanadech, N.) 60–63 (IEEE, 2016).
108. Su, C. Q. A new fuzzy logic method for transformer incipient fault diagnosis. In *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (ed. Su, C. Q.) 324–327 (IEEE, 2016).
109. Wannapring, E., Suwanasri, C. & Suwanasri, T. Dissolved gas analysis methods for distribution transformers. In *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (eds Wannapring, E. *et al.*) 1–6 (IEEE, 2016).
110. Ilias, H. A. & Chai, X. R. Hybrid modified evolutionary particle swarm optimization-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis. *Measurement* **90**, 94–102 (2016).
111. Ghoneim, S. S., Taha, I. B. & Elkalashy, N. I. Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis. *IEEE Trans. Dielectr. Electr. Insul.* **23**(3), 1838–1845 (2016).
112. Bhalla, D., Bansal, R. K. & Gupta, H. O. Transformer incipient fault diagnosis based on DGA using fuzzy logic. In *India International Conference on Power Electronics 2010 (IICPE2010)* (eds Bhalla, D. *et al.*) 1–5 (IEEE, 2011).
113. Malik, H. & Mishra, S. Feature selection using RapidMiner and classification through a probabilistic neural network for fault diagnostics of power transformer. In *2014 Annual IEEE India Conference (INDICON)* (eds Malik, H. & Mishra, S.) 1–6 (IEEE, 2014).
114. Nagpal, T. & Brar, Y. S. Neural network based transformer incipient fault detection. In *2014 International Conference on Advances in Electrical Engineering (ICAEE)* (eds Nagpal, T. & Brar, Y. S.) 1–5 (IEEE, 2014).

115. Gómez, N., Wilhelm, H., Santos, C. & Stocco, G. Dissolved gas analysis (DGA) of natural ester insulating fluids with different chemical compositions. *IEEE Trans. Dielectr. Electr. Insul.* **21**(3), 1071–1078 (2014).
116. Ahmed, M., Geliel, M. & Khalil, A. Power transformer fault diagnosis using a fuzzy logic technique based on dissolved gas analysis. In *21st Mediterranean Conference on Control and Automation* (eds Ahmed, M. *et al.*) 584–589 (IEEE, 2013).
117. Mehta, A. K., Sharma, R., Chauhan, S. & Saho, S. Transformer diagnostics under dissolved gas analysis using support vector machine. In *2013 International Conference on Power, Energy, and Control (ICPEC)* (eds Mehta, A. K. *et al.*) 181–186 (IEEE, 2013).
118. Ghunem, R. A., Assaleh, K. & El-Hag, A. H. Artificial neural networks with stepwise regression for predicting transformer oil furan content. *IEEE Trans. Dielectr. Electr. Insul.* **19**(2), 414–420 (2012).
119. Neyman, J. & Pearson, E. S. IX. On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Charact.* **231**(694–706), 289–337 (1933).
120. Afrasiabi, S., Afrasiabi, M., Parang, B. & Mohammadi, M. Designing a composite deep learning based differential protection scheme of power transformers. *Appl. Soft Comput.* **87**, 105975 (2020).
121. Rajesh, K. N., Rao, U. M., Fofana, I., Rozga, P. & Paramane, A. Influence of data balancing on transformer DGA fault classification with machine learning algorithms. *IEEE Trans. Dielectr. Electr. Insul.* **30**(1), 385–392 (2022).
122. Shamsudin, M. S., Yakub, F., Shapii, M. I., Mohmad, A. & Abd Hamid, N. A. Investigation on online DGA monitoring to determine transformer health index using machine learning. *J. Phys. Conf. Ser.* **2128**(1), 012024 (2021).
123. Singh, S., Raj, N., Gera, R. & Dinkar, A. National conference on the failure of major equipment's of sub-station-case studies 23rd–24th February 2023, New Delhi. *Water Energy Int.* **65**(12), 55–58 (2023).
124. Siregar, Y. & Lumbanraja, T. J. H. Analysis of interference methods on transformers based on the results of dissolved gas analysis tests. *Int. J. Electr. Comput. Eng* <https://doi.org/10.11591/ijec.v13i4.pp3672-3685> (2023).
125. Manoj, T., Ranga, C., Ghoneim, S. S., Rao, U. M. & Abdelwahab, S. A. M. Alternate and effective dissolved gas interpretation to understand the transformer incipient faults. *IEEE Trans. Dielectr. Electr. Insul.* <https://doi.org/10.1109/TDEI.2023.3237795> (2023).
126. Standard, I. *Mineral Oil-Impregnated Electrical Equipment in SERVICE-GUIDE to the Interpretation of Dissolved and Free Gases Analysis* (IEC, 2015).
127. Taha, I. B., Ghoneim, S. S. & Zaini, H. G. Improvement of Rogers four ratios and IEC Code methods for transformer fault diagnosis based on Dissolved Gas Analysis. In *2015 North American Power Symposium (NAPS)* (eds Taha, I. B. *et al.*) 1–5 (IEEE, 2015).
128. Taha, I. B., Ghoneim, S. S. & Duaywah, A. S. Refining DGA methods of IEC Code and Rogers four ratios for transformer fault diagnosis. In *2016 IEEE Power and Energy Society General Meeting (PESGM)* (eds Taha, I. B. *et al.*) 1–5 (IEEE, 2016).
129. Duval, M. Dissolved gas analysis: It can save your transformer. *IEEE Electr. Insul. Mag.* **5**(6), 22–27 (1989).
130. Duval, M. & Lamarre, L. The Duval pentagon—a new complementary tool for the interpretation of dissolved gas analysis in transformers. *IEEE Electr. Insul. Mag.* **30**(6), 9–12 (2014).
131. International Electrotechnical Commission. Interpretation of the analysis of gases in transformers and other oil-filled electrical equipment in service. *Publication 599*, 21–25 (1978).

## Acknowledgements

The authors would like to acknowledge the University of Pretoria for wholly supporting the study for academic purposes.

## Author contributions

Conceptualization, M.K.N.; methodology, M.K.N.; software, M.K.N.; validation, M.K.N., and M.N.G.; formal analysis, M.K.N.; investigation, M.K.N.; writing—original draft preparation, M.K.N.; writing—review and editing, M.K.N. and M.N.G.; visualization, M.K.N.; supervision, M.N.G. All authors have read and agreed to the published version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.K.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024