An Unsupervised Approach for Fault Diagnosis of Power Transformers

Luis Dias[a,b,*], Miguel Ribeiro[a], Armando Leitão[a,b], Luis Guimarães[a,b], Leonel Carvalho[a], Manuel A. Matos[a,b], Ricardo J. Bessa[a]

[a]*INESC Technology and Science (INESC TEC), Campus da FEUP, Rua Dr. Roberto Frias, 4200-465 Porto Portugal*
[b]*Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto Portugal*

Abstract

Electrical utilities apply condition monitoring on power transformers to prevent unplanned outages and detect incipient faults. This monitoring is often done using Dissolved Gas Analysis (DGA) coupled with engineering methods to interpret the data, however the obtained results lack accuracy and reproducibility. In order to improve accuracy, various advanced analytical methods have been proposed in the literature. Nonetheless, these methods are often hard to interpret by the decision-maker and require a substantial amount of failure records to be trained. In the context of power transformers (PT), failure data quality is recurrently questionable, and failure records are scarce when compared to non-failure records. This work tackles these challenges by proposing a novel unsupervised methodology for diagnosing PT condition. Differently from the supervised approaches in the literature, our method does not require the labeling of DGA records and incorporates a visual representation of the results in a 2-D scatter plot to assist in interpretation. A modified clustering technique is used to classify the condition of different power transformers using historical DGA data.

Finally, well-known engineering methods are applied to interpret each of the obtained clusters. The approach was validated using data from two different real-world datasets provided by a generation company and a distribution system operator. The results highlight the advantages of the proposed approach and outperformed engineering methods (from IEC and IEEE standards) and companies *legacy method*. The approach was also validated on the public IEC TC10 database, showing the capability to achieve comparable accuracy with supervised learning methods from the literature. As a result of the methodology performance, both companies are currently using it in their daily DGA diagnosis.

Keywords: Asset management, unsupervised learning, failure diagnosis, power transformers, dissolved gas analysis

---

1. Introduction

1.1. Motivation

Power transformers (PT) represent a large share of maintenance costs and capital investment in power systems. In many markets, the recent transition from a regulated to a competitive environment has dramatically increased the pressure for utilities to become more cost efficient. At the same time, the liberalization of the electricity market resulted in different operating schedules of conventional power plants, which can lead to increasingly stressful operating condition of PT, shorter maintenance cycles, and more complex actions. The early diagnosis of potential failures in PT not only increases the equipment reliability but also decreases the revenue loss due to service downtime, making a valuable contribution to the industry. In order to perform early diagnosis of potential failures, utilities rely on condition monitoring methods. The most popular and widespread methods are based on analyzing the presence/concentration of the dissolved gas in the PT oil, the element responsible for the cooling and electric insulation. Such records have been collected periodically throughout several years of operation and are a reliable source of information to anticipate future PT failure. This data analysis process is especially relevant since it embeds the operation and environmental conditions the PT is subjected to, which are specific for each country and utility[1].

During their life cycle, PTs are subjected to variable loading conditions that originate mechanical, chemical, and electrical stresses. These stresses originate different types of combustible and non-combustible gases, such as hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6), carbon monoxide (CO), carbon dioxide (CO_2), oxygen (O_2) and nitrogen (N_2) dissolved in the PT oil. Numerous techniques have been developed and improved to analyze these records, such as dissolved gas analysis (DGA), insulating oil quality, and furan analysis. Additionally, other techniques were developed to aid failure diagnosis based on specific measurement equipment[2]. While this additional equipment improves diagnosis results, the trade-off between the extra diagnosis capability and their costs is often not economically attractive. Among the techniques mentioned above, the DGA method is the most commonly used for three main reasons: a) it can provide reliable information; b) it is a non-intrusive method; and c) is a proven method capable of diagnosing PT incipient faults[3]. Furthermore, in a recent trend that proves DGA attractiveness, utilities have started to install new measurement systems to online monitor the gases concentration.

*Corresponding author
Email address: luis.m.dias@inesctec.pt (Luis Dias)

Failures can generally be categorized either as thermal failures or electrical failures. Correctly diagnosing if a PT exhibits a potential failure, the failure type and severity based on DGA is a very challenging problem as the vast literature on the topic shows.

However, even though companies regularly monitor their assets by collecting maintenance data, very few failures usually occur. Data collection and the failure occurrence may also not coincide in time, making it hard to match the corresponding data point and label it correctly. In other cases, failure data is nonexistent, which motivates the development of methodologies that use only information from dissolved gases and no labeled data about failure type.

This was the context covered by this work and faced in two real-world applications: hydro-power generation company (GenCo); distribution system operator (DSO). In both cases, the information available consisted of dissolved gases with scarce failure data. Hence, it is not suitable to apply supervised learning methods. To tackle this challenge, we propose a novel unsupervised approach that leverages unlabeled DGA data to evaluate the PT condition. Moreover, since we are using an unsupervised approach, we provide methods to interpret the results accordingly.

1.2. Related work and contributions

The methods developed by researchers over the years can be grouped into two main streams: engineering methods and analytical methods. Engineering methods have been developed by domain knowledge of the physical and chemical process occurring inside the oil tanks of the PT, and afterwards validated with empirical studies. Engineering methods have been developed by the scientific community to interpret DGA results such as Duval's Triangle (DT)[3], IEC gas ratios[4], Key Gas analysis (KGA)[5], Roger gas ratios[6], Dornenburg's gas ratios[7], Mansour Pentagon method[8] and Duval Pentagon[9]. These methods are prevalent among utilities due to their simplicity and interpretability. Such methods, whose application depends on the experience, detect fewer failures when compared to analytical methods[10]. Methods using ratios, such as the DT, excel in diagnosing the type of failure but cannot be used to separate PTs in a healthy condition from those with malfunctions. Misdiagnosing a PT in this condition can lead to unnecessary maintenance costs. A recent work[11] highlights the advantages and disadvantages of these methods.

Conversely, analytical methods based on absolute gas concentrations are better at discriminating between normal and failure condition. The amount of data collected by utilities motivated researchers to develop analytical methods using mathematical models and supervised learning techniques namely: hybrid models combining different engineering methods[10], evolutionary methods[12], Markov models[13,14], fuzzy models[15,16], multi-layered artificial neural networks[17,18], support vector machine (SVM)[19], twin support vector machine (TWSVM)[20], wavelet networks[21], Bayesian networks[22], probabilistic classifiers[23,24], nearest neighbour clustering[25] and association rules[26,27]. Some of the mathematical models and supervised learning techniques mentioned before rely on a small sample of failure data to be trained and tested, which can jeopardize their robustness and generalization capability. A recent work has tackled this issue by applying different models to the same dataset and evaluating their performance[28]. The authors concluded that, in general, nonlinear algorithms performed better than linear ones. Moreover, it was proved that SVM with Gaussian kernels, neural networks, and local linear regression are the best techniques to diagnose PT faults.

The analytical methods reviewed above are appealing for any electrical utility since all provide reliable and accurate results. However, in some real-world contexts, they may not achieve the desired results due to: (i) their black-box nature and (ii) the limited, untrustworthy, or nonexistent failure data. The first issue is related to the complexity of the mentioned approaches, which makes it hard to be interpreted by a decision-maker (DM). Although it is essential to obtain accurate results, the DM has to be able to understand and interpret them accordingly, facilitating the follow-up process. The second issue is related to the scarcity/lack or quality of failure data in some companies. Even though many companies regularly monitor their PTs by collecting DGA samples, in many contexts, none or very few PT failures have occurred, being the history of DGA records full of normal functioning PTs. Furthermore, DGA collection and the failure occurrence may not coincide in time, making it hard to match the corresponding DGA record and thus label the records correctly. These reasons make the application of supervised learning techniques challenging since their performance is better in balanced datasets (in terms of class labels) and relies on good quality labeled data. The datasets used in this study, supplied by two different companies where PT have distinct operating conditions, are examples of the previously mentioned issues. Additionally, when the DGA data was retrieved from the PT, both companies did not know their real condition. Therefore, in our study, each data point is not labeled, which limits the use of supervised learning methods.

As far as we know, few analytical approaches in the literature do not require failure data to diagnose potential internal failures in the PT. Therefore, the present work aims at overcoming the previously identified shortcomings, and it produces the following original contributions: 1) combine data-driven techniques and IEC/IEEE standardized procedures for PT diagnosis in order to extract features with practical engineering knowledge that facilitate the interpretation of results by the DM; 2) presents a fully unsupervised approach capable of tackling the case in which the match between DGA records and

PTs failure status is not available. These contributions are tested and validated using data from two utility companies. Furthermore, the proposed approach is applied in the public IEC TC10 dataset[29] to assess the generalization capability and benchmark the performance against state-of-the-art supervised algorithms. It is important to underline that the proposed methodology is for PT diagnosis and serves as a basis for advanced maintenance policies (which are out of the scope in this paper).

1.3. Structure of the paper

The remainder of this paper is structured as follows. Section 2 introduces the conceptual framework for assessing the PT condition. Section 3 describes in detail the proposed methodology. Section 4 presents the application of the proposed approach on the DGA datasets supplied by two different companies. Then, in Section 5 the approach is validated with IEC TC10 dataset and compared with various state-of-the-art methods. Finally, Section 6 presents the conclusions and highlights the prospects of future work.

2. Conceptual Framework

The goal of the proposed methodology is to diagnose PT condition solely based on collected data, without requiring any prior knowledge regarding the PT condition. This kind of approach is useful in cases where companies store data without the label characterizing the PT condition when the sample was collected. The proposed approach is divided into two parts and can be replicated to other electrical assets: i) construct the data-driven model, and ii) use the model to classify the equipment condition. The main output of the methodology is the label of the PT condition, which indicates if it is in normal operating condition or not. In the case of faulty condition, the methodology provides information regarding the failure type and the resulting severity. The process of the unsupervised method proposed in this paper is summarized in Figure 1.

Whenever new DGA data is available, the first step is to analyze if the data-driven model responsible for classifying the PT condition is valid. This step is always initiated in case the model has to be created. The goal is to retain the DGA gases with a significant impact on the PT condition. To do so, it is measured the amount of information in each gas (H_2, CH_4, C_2H_2, C_2H_4, C_2H_6, CO, CO_2, O_2 and N_2) to select the gases that add information to the model. Consequently, an entropy measure is used to reduce the problem complexity and to eliminate gases that do not add significant value to the data-driven model. After this procedure, the gases selected are combined to create new features. For each selected gas, a weight is computed based on the entropy metric. The goal is to give increasing importance to gases that contain more information. Compared to other methods defining weights based on expert knowledge or subjective judgments,

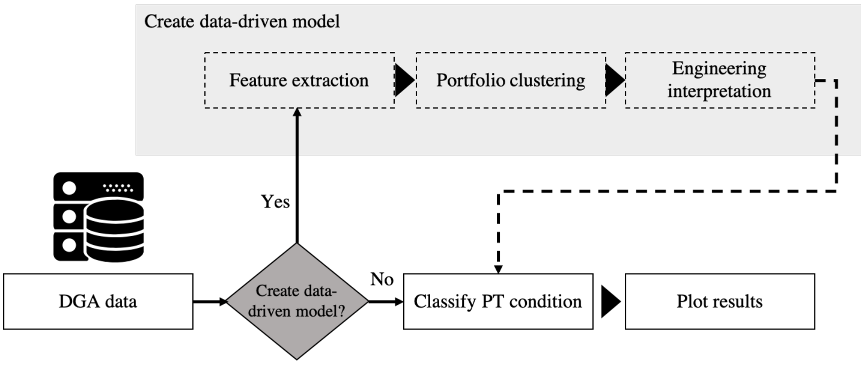

Figure 1: Diagram of the conceptual framework

this entropy-based approach defines the weights based on the information content of each variable (or in other words, its discriminative value). This step is fully unsupervised and purely data-driven.

The computed weights are then used in Principal Component Analysis (PCA) to help the aggregation gases with similar variation pattern. This reduces the problem complexity even further and improves the model's performance. To enhance the interpretability of the results, the gases are aggregated into two components, principal component 1 (PC^1) and principal component 2 (PC^2), allowing a visual representation of the results. Note that this aggregation will result in a marginal loss of information when compared to the individual use of each select gas. In the next step, named "Portfolio Clustering", two clustering approaches are applied using the K-means algorithm (KMA) to generate cluster centroids based on the degree of similarity of the gases concentration (PT condition). The first approach uses the entropy weights to cluster DGA records, while the second uses the transformed variables of the PCA to perform the clustering. The objective is to separate DGA records into those with a normal condition from the ones with an abnormal condition (incipient fault). Both alternatives are represented in Figure 2.

The E-PCA-C approach applies a weighted PCA approach where the obtained PC^1 and PC^2 are used to create the clusters. The E-C approach employs a weighted KMA where the selected gases and their respective weights are used to find the clusters. The purpose of applying both approaches is to understand the advantages of reducing the problem complexity and using the most significant features. Simpler models are preferred to complex models due to their superior validity, according to the principle of Occam's razor[30].

After the two stages described above, it is still not possible to classify the condition

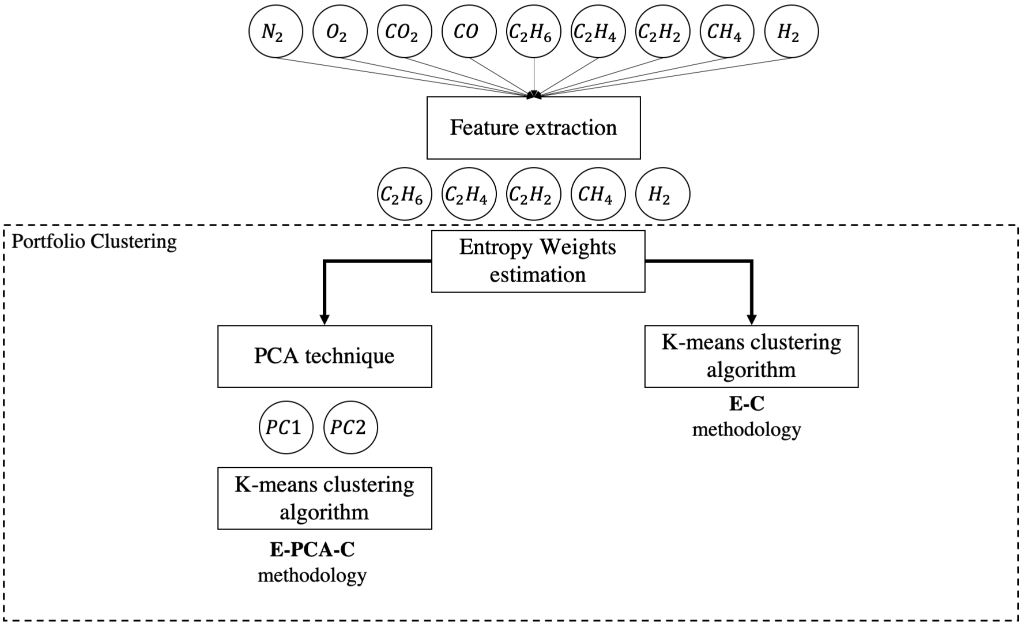

Figure 2: Clustering alternatives applied on the proposed approach

for the PT belonging to each cluster, which is a disadvantage compared to supervised methods. To overcome this issue, engineering methods are used in the "Engineering interpretation" step. KGA[5] is used to label clusters with a normal condition or incipient failure, and for the latter, it allows to establish the level of severity. This labeling procedure is complemented with the DT[3], which classifies the type of failure (thermal or electrical) verified in each cluster labeled with incipient failure. The goal is to classify the PT condition based on the type of failure (thermal or electrical) and failure severity. This labeling procedure is detailed in section 3.3. The data-driven model is completed when the clusters are constructed and classified (or labeled). During this step, no additional data is introduced into the developed approach.

The created data-driven model is applied in the "Classify PT condition" step to group the PTs based on their shortest Euclidean distance to the centroids of the previously defined clusters. This step diagnosis the PTs condition based on the assigned cluster centroid. Thus, each PT is assigned to a condition equal to the centroid of its cluster. In the final step, "Plot results", a 2D plot is generated using PC^1 and PC^2, which allows a visual interpretation of the results. The goal is to facilitate the DM interpretation while quickly visualizing the PT condition. As mentioned before, this approach can be applied to future DGA records without having to create a new data-driven model from scratch and does not require labeled data (i.e. failure condition label and normal condition label).

The key concept behind it is to group PT based on their current estimated condition. The following section provides a detailed explanation regarding the analytical models and engineering methods applied in each step of the methodology.

3. Data-driven Fault Diagnosis Methodology

Consider a dataset of N DGA observations indexed by i. Each observation is composed of G gas variables and corresponding concentrations: x^g is the vector of observations for gas g, x_i the concentrations of the different G gases at sample i and x_i^g the concentration of gas g at sample i.

3.1. Feature extraction

Entropy is a concept used in information theory to quantify the information found in a given set of observations. Observations that are more unlikely to occur will provide more information when compared to ones that occur frequently. Entropy $H(x^g)$ will be used to evaluate the amount of information that a specific x^g holds. This measure can be obtained using the sum of probabilities of a given x_i^g from a set of N observations multiplied by its negative natural logarithm:

$$H(x^g) = -\sum_{i=1}^{N} x_i^g ln(x_i^g) \tag{1}$$

The lower the probability of a given x_i^g when picked at random, the higher the amount of information the set of N observations holds. The key idea is that the more distinct values are present in a given x^g, the more information it contains. Therefore, entropy is maximized when each x_i^g, from a set of N observations, has an equal probability of occurring.

In the proposed methodology, a modified entropy metric, proposed in[31] for weighting attributes importance in a decision-aid scenario is adopted, which allow the calculation of the entropy $e(d^g)$ of each g gas based on the available information. Given the vector of values $d^g = (d_1^g, ..., d_n^g)$ scaled with the maximum value found in each set x^g (Eq. 2) and the sum of all values D_g of vector d^g (Eq. 3),

$$d_i^g = \frac{x_i^g}{arg\, max\, x^g} \tag{2}$$

$$D_g = \sum_{i=1}^{N} d_i^g \tag{3}$$

the entropy measure of a given x^g is given by the natural logarithm ln and the respective scaled values d^g

$$e(d^g) = -f \sum_{i=1}^{N} \frac{d_i^g}{D_g} ln \frac{d_i^g}{D_g} \tag{4}$$

where $f > 0$, defined by the inverse natural logarithm $(ln(N))^-1$ is constrained to the following two intervals: i) $0 \leq d_g^j \leq 1$ and ii) $e(d_g) \geq 0$. If all d_i^g become identical for a given x^g, then $d_i^g/D_g = 1/N$, and $e(d^g)$ assumes its maximum value, that is, $e_{max} = ln(N)$. Thus, setting $f = 1/e_{max}$ implies $0 \leq e(d_g) \leq 1$ for all d^g. Such normalization is needed for comparative purposes. The total entropy of the dataset is defined as:

$$H = \sum_{g=1}^{G} e(d^g) \tag{5}$$

Note that when $e(d_g)$ increases, less information is conveyed by x^g. Actually, if $e(d_g) = e_{max} = ln(N)$, x^g would not provide any useful information at all and can be removed. The proposed approach selects a specific x^g that is above a previously defined threshold T_g (this work defines $T_g = 0.10$).

3.2. Portfolio clustering

After selecting the set of gases to use, the KMA is applied to cluster all x_i observations and to construct the respective cluster centroids. Two clustering alternatives, as shown in Figure 2, are considered in the proposed data-driven approach: i) weighted clustering coupled with entropy using the selected x_i observations (E-C) or ii) PCA integrated with entropy to cluster x_i using the L created principal components PC^l (E-PCA-C). The main goal is to compare these two alternatives and understand which representation delivers the best results.

The KMA is a clustering method that aims to segment the set of N points projected in m dimensional space separated into a finite number of K clusters that minimizes the within-cluster sum of square distance[32]. A given x_i is assigned to the k cluster with the nearest mean distance. The final result is a partitioning of all x_i into a set of clusters. This problem can be formulated as follows:

$$min \sum_{k=1}^{K} \sum_{x \in k} ||x_i - \overline{x}_k||^2 \tag{6}$$

where $\overline{x}_k$ is the mean of all points assigned to cluster k. Moreover, this procedure is equivalent to a pairwise squared deviations minimization for all points in cluster k:

$$\underset{k}{\operatorname{argmin}} \sum_{k=1}^{K} \frac{1}{2|k|} \sum_{x,y \in k} ||x - y||^2 \tag{7}$$

The proposed methodology uses Lloyd's iterative algorithm[33]. This algorithm minimizes the mean squared Euclidean distance between all x_i observations and the assigned cluster k. The implementation can be described as follows:

1. The user selects the number of clusters K.

2. Create K centroids projected in a M dimensional place. The initial cluster location, defined by their centroids C_k, is selected at random.

3. Assign the point x_i to cluster k based on the shortest Euclidean distance to centroid C_k.

4. Calculate new location for C_k using the mean of all observations $\overline{x_i}$ in cluster k.

5. Repeat steps 2 to 4 until Eq. 6 is minimized or the centroid position for each k cluster is stable (the cluster position has marginal modifications).

In both E-C and E-PCA-C, each x_i^g is scaled into a new variable y_i^g using the z-score method, to prevent an undesired bias in the Euclidean distance calculation. In E-C, the y_i is assigned to cluster k based on the weighted shortest Euclidean distance. This allows to give more relevance to gases that have more impact on the PT condition, leading to a better y_i allocation to each created cluster.

$$E_d(C_k, y_i) = \sqrt{\sum_{g=1}^{G} w_g * (C_k^g - y_i^g)^2} \tag{8}$$

In the proposed approach, the level of importance of each selected x^g is differentiated based on the entropy measure $e(d^g)$. The weights w_g are reversely related to $e(d^g)$, thus $1 - e(d^g)$ is used rather than $e(d^g)$ to normalize and assure that $0 \leq w_g \leq 1$ and $\sum_{g=1}^{G} w_g = 1$. The weights are scaled using the following equation:

$$w_g = \frac{1 - e(d^{\hat{g}})}{G - \sum\limits_{\hat{g}=1}^{G} e(d^{\hat{g}})} \tag{9}$$

After assigning all y_i to a cluster, the new position for each centroid C_k is computed. This new position will be centered by all N_k observations y_i inserted in cluster k. The weighted average approach is used again to calculate the new position:

$$C_k^g = \frac{\sum\limits_{y_i^g \in k} y_i^g}{N_k} \tag{10}$$

A similar procedure is followed for approach E-PCA-C. However, the PCA technique is used to transform the selected x^g into components. PCA is a linear transformation process in which the x_i^g values are projected into a new set of orthogonal axis (also called principal components) accordingly to the direction that maximizes the original information represented by each component[34]. The linear transformation is given by the product between the vector of coefficients β_L with vector x^g:

$$PC^l = \beta_l^T x^g = \sum_{g=1}^{G} \beta_{lg} x^g \tag{11}$$

such as $\beta_l^T \beta_l = 1$. The entropy is used together with the scaled y_i^g value to allow the PCA to incorporate more information from more relevant x^g. The transformed individual value Y_i^g is defined as follows:

$$Y_i^g = w_g y_i^g \tag{12}$$

This linear transformation will create new L uncorrelated principal components PC^l. The first component has the largest possible variance, thus accounting as much as possible for all the variability present in the data. The following components will account for less variability when compared to the preceding ones. The proposed approach reduces x^g into two principal components. Afterwards, KMA is applied to aggregate the PT based on these two components, and the respective Euclidean distance becomes:

$$E_d(C_k, PC^l) = \sqrt{\sum_{l=1}^{P} (C_k^l - PC_i^l)^2} \tag{13}$$

The new position for each centroid C_k is calculated as follows:

$$C_k^i = \frac{\sum\limits_{PC_i^j \in k} PC_i^j}{N_k} \tag{14}$$

3.3. Engineering interpretation

The final step of the methodology is to interpret the obtained results. First, the data is projected into a 2D plot using the two PC that aggregate the selected x_i^g. The purpose of the graphical representation of the results is twofold: i) enable a quick assessment of the PT condition and ii) define the boundary of each condition on the obtained graphical representation. With the 2D plot, it is possible to visualize the assignment of each x_i to a particular cluster k. To enhance the results interpretation, we display C_k, which is depicted with a larger point size than the x_i.

Figure 3 illustrates an example of the aforementioned 2D plot applied in the GenCo dataset using the E-C methodology. From the depicted results, it is possible to distinguish the different set of k clusters instantly. Furthermore, the results show that most clusters are either concentrated on the PC^1 axis or the PC^2 axis. These initial results show the capacity of the proposed method in detecting the two types of failure modes in PT: thermal failures and electrical failures. Nevertheless, while it is possible to represent the set of k clusters graphically, a clear interpretation of the results is necessary since an unsupervised approach was followed.

First, the KGA is applied to label each created C_k concerning failure severity[5]. The method is used to rank each C_k from best (condition 1) to worst condition (condition 4) and to label each C_k with the "normal condition" or with the "failure condition". This is done by analysing the concentration of each gas coupled with the Total Dissolved Combustible Gas (TDCG) indicator, which is calculated as follows:

$$TDCG = \sum_{g=1}^{G} x_i^g \tag{15}$$

Each C_k is assigned to a given condition based on the thresholds defined in Table 1 applied to its centroid[4] . For instance, if the TDCG exceeds 720, but it is bellow 1920 (value obtained by the KGA), then the PT is assigned with "Condition 2". The same applies to any gas above the defined limits. A C_k with a worse condition will have a higher failure probability when compared to another one with a lower condition level. Every C_k classified with "Condition 1" will group PT with normal condition state.

Table 1: IEEE Key Gas analysis thresholds

TDCG (ppm)	H_2 (ppm)	CH_4 (ppm)	C_2H_2 (ppm)	C_2H_4 (ppm)	C_2H_6 (ppm)	CO (ppm)	CO_2 (ppm)	Status
≤ 720	< 100	< 120	< 1	< 50	< 65	< 350	< 2500	[C1]-Condition 1
≤ 1920	< 700	< 400	< 9	< 100	< 100	< 570	< 4000	[C2]-Condition 2
≤ 4630	< 1800	< 1000	< 35	< 200	< 150	< 1400	< 10000	[C3]-Condition 3
≥ 4630	> 1800	> 1000	> 35	> 200	> 150	> 1400	> 10000	[C4]-Condition 4

Nonetheless, the labeling of C_k is incomplete since the failure type is unknown. The analysis is completed with the DT method, which allows diagnosing the failure type (electrical or thermal) present in each C_k not labeled with "normal condition". This approach is widely used in practice since it was showed to be accessible and good at classifying PT incipient faults. Furthermore, the results can be plotted in a Triangle representation, which facilitates results interpretation. The method makes use of the percentage of three gases dissolved on the PT incipient oil (CH_4, C_2H_2 and C_2H_4):

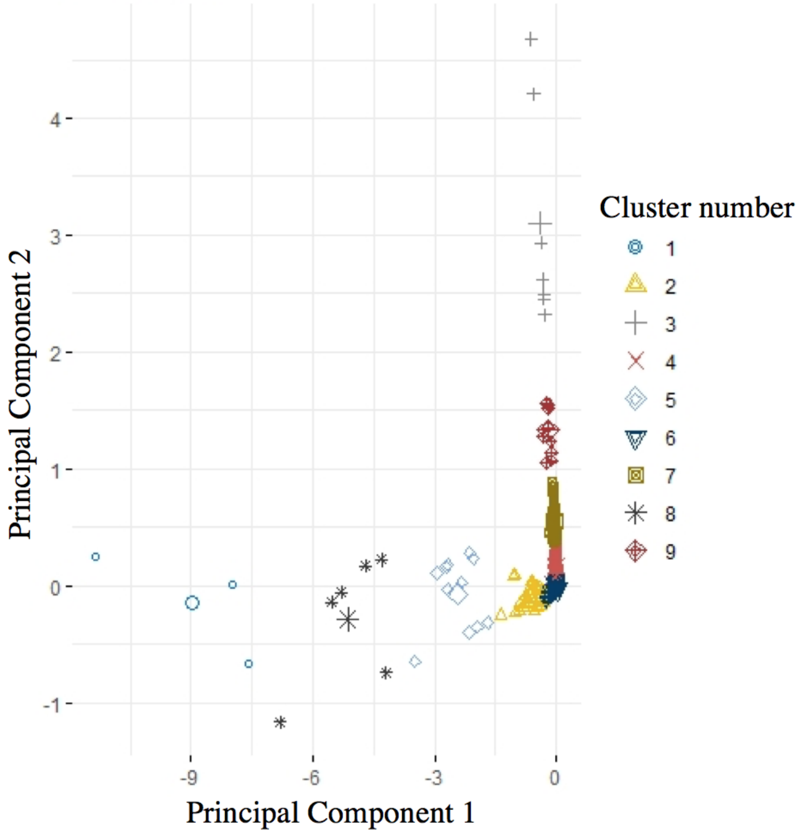

Figure 3: Graphical representation of the results using clustering alternative E-C applied for dataset X

$$\%CH_4 = \frac{CH_4}{CH_4 + C_2H_2 + C_2H_4} \tag{16}$$

$$\%C_2H_2 = \frac{C_2H_2}{CH_4 + C_2H_2 + C_2H_4} \tag{17}$$

$$\%C_2H_4 = \frac{C_2H_2}{CH_4 + C_2H_2 + C_2H_4} \tag{18}$$

The percentage of each gas is then combined to classify the PT with a given fault state. The seven possible fault states are presented in Table 2, together with the respective type of failure. The limits were defined in a study conducted in[3]. Each C_k will be assigned with a fault state based on the obtained coordinates for the centroid. Thermal failure T is assigned to C_k if the centroid is assigned to a fault state included in T. Otherwise, E is assigned to C_k.

Table 2: Duval Triangle possible fault states

Failure type	Fault state	$\%CH_4$	$\%\ C_2H_2$	$\%C_2H_6$
[E]-Electrical	Partial Discharge (PD)	>98%		
	Low energy discharge (D1)		>13%	<23%
	High energy discharge (D2)		13%-29%	23%-38%
			>29%	>23%
[T]-Thermal	Thermal fault <300ºC (T1)		<4%	<20%
	Thermal fault 300-700ºC (T2)		<4%	20%-50%
	Thermal fault >700ºC (T3)		<15%	>50%
[E&T]-Electrical & Thermal	Thermal & electrical fault (DT)	unlabeled value		

In the following section, numerical results are presented for real-world datasets, and the proposed method is compared against state-of-the-art methods using the IEC TC10 dataset[29].

4. Model validation

This section validates the proposed unsupervised method (described in section 3) on real-world data. We start by describing the distinct datasets provided by two utilities: i) dataset X belongs to a GenCo encompassing HV/MV PT, ii) dataset Y belongs to a DSO including HV/MV PT and iii) dataset X&Y results by merging both datasets. In these datasets, the vast majority of the records (approx. 99.36%) are not labeled. This is due to the fact that both companies did not perform tests to confirm the PT real condition.

After describing the real-world data, the construction of the data-driven models and numerical results is discussed. The section ends with a comparison between the current approach used by both companies and the proposed unsupervised approach.

4.1. Dataset description

The real world datasets contain concentration readings of 9 DGA gases (H_2, CH_4, C_2H_2, C_2H_4, C_2H_6, CO, CO_2, O_2 and N_2). The oil samples are collected on an annual basis and analysed by an independent laboratory where the gas concentration is measured in *ppm*. All the oil samples are retrieved from the same location inside the PT. Dataset X consists of 3580 DGA records retrieved from 375 HV/MV PT, while dataset Y has 11952 DGA records with 684 HV/MV PT. Additionally, dataset X and dataset Y were combined to create dataset X&Y in order to check if the results are improved with a larger amount of data. Dataset X&Y consists of 15532 DGA records with 1059 PT.

The GenCo was able to provide some information regarding electrical tests where it was confirmed the existence of internal failures for part of the observations. These tests are reliable for failure detection but much more costly to perform when compared to DGA tests. From these data, it is possible to label some of the provided DGA records with two types of labels: 1) “internal fault” or 2) “normal condition”. Unfortunately, no information is provided regarding the type of internal failure (thermal or electrical failure). Moreover, the PT have different power ratings, and not all of them are from the same manufacturer. This implies that failure detection might be more complex due to their different reliability and DGA concentrations. By having this new information, it was possible to identify the existence of 16 “internal fault” and 84 “normal condition” records referring to 100 PT on dataset X. No additional information was inserted in dataset Y.

The proposed approach was applied in each dataset utilizing the developed clustering alternatives: E-PCA-C and E-C. This will result in six data-driven models that will be used to assess each PT condition. Each of the six generated data-driven models (three for each clustering alternative) is tested using dataset X since it has labeled records. These records were not included in each of the considered datasets since the goal is to evaluate the generalization capability of the methodology. Presently, both utilities use expert knowledge and two engineering methods to diagnose PT condition: i) KGA is used to identify DGA records with failure condition, namely records with $TDCG > 720$, and ii) the DT to diagnose the type of failure. The goal of the study is to see if it is possible to improve the current performance.

4.2. Construction of the diagnosis model

The proposed methods are tested and validated on a separate dataset with confirmed condition labels provided by the GenCo in dataset X. The first phase is the selection of gases to be included in the data-driven model that groups PT based on their condition.

As shown in Table 3, only 5 gases (H_2, CH_4, C_2H_2, C_2H_4, C_2H_6) from the original DGA gases are selected when applying the methodology.

Table 3: Normalized entropy results for the considered datasets

Entropy	Dataset X	Dataset Y	Dataset X&Y
$\mathbf{H_2}$	0.20	0.13	0.16
$\mathbf{CH_4}$	0.18	0.19	0.19
$\mathbf{C_2H_2}$	0.13	0.19	0.12
$\mathbf{C_2H_4}$	0.19	0.29	0.24
$\mathbf{C_2H_6}$	0.16	0.15	0.15
CO	0.03	0.02	0.03
CO_2	0.06	0.02	0.04
O_2	0.02	0.01	0.04
N_2	0.03	0.00	0.04
Total records	**3580**	**11952**	**15532**

These gases are selected using a previously defined threshold of $T_g > 0.10$ since the goal is to remove gases that have a low amount of information (i.e. entropy metric) compared to the remaining set. Table 3 shows that, the set of gases selected in all three datasets is the same (H_2, CH_4, C_2H_2, C_2H_4, C_2H_6), which confirms the robustness of the results. Furthermore, note that the selected gases are in line with other engineering methods found in the literature[28]. The assigned weights vary between datasets since this variability is assumed to be caused by the fact that the dataset comes from utilities operating in different energy businesses (generation and electrical grids) and with PT from different manufacturers.

After selecting the gases, the approach described in section 3.2 is applied to cluster the PTs based on their assigned C_k. Nine possible states ($k = 9$) are considered in the study since it provided the best results for all three datasets. Each cluster refers to the normal state condition, thermal failure, and electrical failure with different condition severity. However, further analysis is required at this stage to label each created C_k correctly. In Table 4, the percentage of explained variance of both components used to plot Figure 4, approximately account for 80% of the total variance in each dataset. The results confirm that it is acceptable to aggregate the selected gases into two principal components without a significant loss of information.

Engineering methods are used to label each C_k, and consequently, all PT DGA records assigned to cluster k. Each C_k is labeled concerning the condition and failure type. The used label for each DGA records is found in Table 1 and in Table 2, respectively. The labeling results are shown in Tables 5 and 6, which shows respectively, the condition severity and corresponding failure type assigned to each cluster.

As shown in Table 5, C_1 is the cluster of PT with a normal condition N in all

Table 4: Results obtained from the principal component analysis for PC^1 and PC^2

	Dataset X		**Dataset Y**		**Dataset X&Y**	
	PC^1	PC^2	PC^1	PC^2	PC^1	PC^2
H_2	0.04	0.00	0.12	0.97	0.09	0.96
CH_4	0.42	0.08	0.60	0.06	0.56	0.01
C_2H_2	0.14	0.99	0.09	0.18	0.08	0.25
C_2H_4	0.82	0.06	0.63	0.07	0.69	0.09
C_2H_6	0.37	0.13	0.47	0.10	0.45	0.09
Explained Variance	**79%**		**82%**		**78%**	

three datasets, according to the KGA. The assignment of N to C_1 is also made evident by the number of DGA records included, which represent more than 90% of the total records. After labeling the respective cluster with N, the remaining clusters are defined by their severity level coupled with the highlighted failure type. Each C_k is assigned to a thermal failure T or electrical failure F with a given severity level. The results obtained highlight that the number of DGA records decreases with the increase in the severity level. This observation is coherent since the failure probability increases with severity level[4]. Therefore, it is expected that the number of records present in each dataset is less for clusters with an increasing severity level.

Table 5: Applied methodology results for each dataset and the corresponding condition

	Dataset X				**Dataset Y**				**Dataset X&Y**			
Cluster K	E-C	KGA	E-PCA-C	KGA	E-C	KGA	E-PCA-C	KGA	E-C	KGA	E-PCA-C	KGA
1	3386	C1	3333	C1	11181	C1	10869	C1	14705	C1	13656	C1
2	76	C2	119	C2	254	C2	712	C2	327	C2	1295	C2
3	7	C3	59	C2	260	C2	214	C2	211	C2	382	C2
4	82	C2	9	C3	180	C2	59	C2	191	C2	97	C2
5	5	C3	7	C4	26	C3	34	C2	27	C3	31	C3
6	11	C3	33	C2	32	C3	22	C3	44	C3	36	C3
7	6	C4	11	C3	11	C4	22	C4	14	C4	16	C4
8	3	C4	6	C4	7	C4	13	C4	11	C4	11	C4
9	4	C4	3	C4	1	C4	7	C4	2	C4	8	C4
Total	3580				11952				15532			

The next step after labeling each cluster with a given severity level is to determine the failure type. Table 6 shows for each C_k the assigned failure type, as well as the DGA records distribution. Each x_i was individually labeled using the DT method to understand the assigned condition label to C_k. The outcome of this procedure shows that the failure type assigned to each C_k is aligned with the DGA records failure distribution. For most of C_k, the DGA records lead to a single type of failure. Nonetheless and in most cases, C_2 is the cluster centroid that does not have a definite assignment to a single failure type. In this case, C_2 is a cluster that captures the transition from a normal to a failure condition.

Table 6: Obtained results for each dataset clustering and the correspondent cluster type of failure

Cluster K	**Dataset X**						**Dataset Y**						**Dataset X&Y**					
	E-C			E-PCA-C			E-C			E-PCA-C			E-C			E-PCA-C		
	N	E	T	N	E	T	N	E	T	N	E	T	N	E	T	N	E	T
1	3386	0	0	3333	0	0	11181	0	0	10869	0	0	14705	0	0	13656	0	0
2	0	76	0	0	96	23	0	254	0	0	120	592	0	19	308	0	461	834
3	0	7	0	0	59	0	0	19	241	0	26	188	0	211	0	0	18	364
4	0	0	82	0	9	0	0	0	180	0	50	9	0	0	191	0	2	95
5	0	5	0	0	7	0	0	23	3	0	0	34	0	24	3	0	28	3
6	0	0	11	0	3	30	0	0	32	0	20	2	0	0	44	0	0	36
7	0	0	6	0	0	11	0	11	0	0	0	22	0	14	0	0	16	0
8	0	3	0	0	0	6	0	0	7	0	13	0	0	0	11	0	0	11
9	0	0	4	0	0	3	0	1	0	0	0	7	0	2	0	0	0	8
Total	3580						11952						15532					

A 2D plot is used to better support the analysis of the obtained results. Figure 4 depicts an example of the results obtained for dataset Y using the E-PCA-C clustering method. If one correlates the results shown in Table 6 and the ones depicted in Figure 4, it is possible to conclude that cluster $k = 1$, which is the one referring to the normal condition, corresponds to the origin of the 2D plot. On the other hand, the remaining clusters with a higher PC^l value will correspond to clusters with worse condition. Considering the results shown in Table 4, this means that the original x_i will have an overall DGA higher concentration (ppm).

Furthermore, as showed previously in Figure 3, the distribution of the clusters are mostly defined by a single PC^l. According to the failure distribution shown in Table 6 and the depicted results in Figure 4, the visual distribution appears to indicate that each PC^l is pointing to a given failure mode. This is clear in the results presented in Table 4, where it is possible to observe that PC^1 incorporates more information of gases related to thermal failures (CH_4, C_2H_4 and C_2H_6), while PC^2 incorporates gases information related to electrical failures (H_2 and C_2H_2). Hence, it is possible to conclude that component PC^1 refers to thermal failures diagnoses, while PC^2 is associated with electrical failures. This behaviour is consistent across the three datasets, which reinforces the generalization capability of the proposed methodology.

4.3. Results evaluation for utilities data

After labeling and interpreting the created clusters, the data-driven models are tested and validated on the labeled records supplied by the GenCo in dataset X. As mentioned in section 4.1, the total number of PT in this dataset is 100, with 16 confirmed incipient failures while the remaining ones are confirmed not to have any failure. Note that the supplied failure records had no information regarding the type of failure. Therefore, for this study, it is considered that any data point outside cluster $k = 1$ corresponds to a failure state. In this case, the model's ability to detect incipient failures without raising too many false alarms is evaluated. The following performance metrics are used: a)

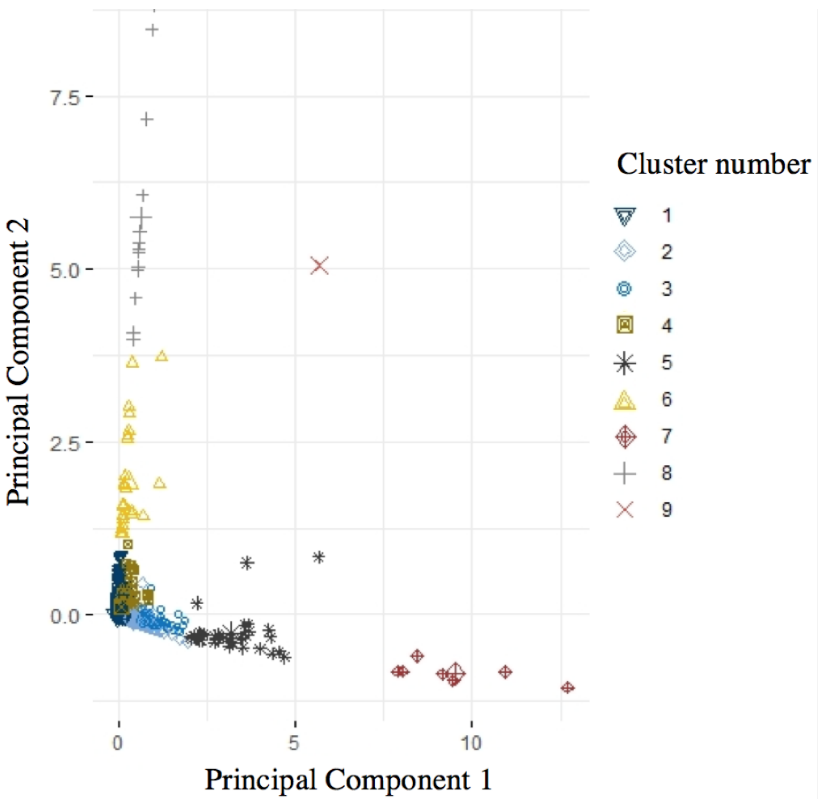

Figure 4: Graphical representation of the results using clustering method E-PCA-C applied for dataset Y

Precision (Eq. 19), which measures the number of True Positives (TP) divided by the total number of TP and False Positives (FP); and b) Recall (Eq. 20), which represents the proportion of failures correctly identified considering also the False Negatives (FN).

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

By definition, a TP is counted when the data-driven model detects the incipient failure. On the other hand, FP is used to count the number of times that each data-driven model incorrectly assesses the PT with incipient failure. In contrast, FN is used to count the number of non-identified incipient failures. The goal is to understand how each data-driven model can detect the PT failure without too many FP (increases accuracy) while reducing FN (increases recall).

In order to compute Precision and Recall, the two proposed clustering approaches, namely E-PCA-C and E-C, are directly applied to the previously described datasets (Dataset X, Dataset Y, and Dataset X&Y). In total will be generated six distinct data-driven models with their respective cluster centroids. The obtained results are compared with the *legacy method* used by the GenCo and the DSO in practice, which combines the DT method with the KGA. This means that the PT is diagnosed with a potential failure using the DT if the absolute concentration of the gases exceeds the thresholds defined in Table 1 for "Condition 1"; otherwise, the PT is considered to be in normal condition. These two methods are combined since the DT is unable to label the PT with "normal" condition and also to reduce the number of incorrectly identified failures (FP). Therefore, the use of the KGA improves the performance of the DT method since it separates PT with a normal condition and potential failure. The comparison results between the proposed approach and the *legacy method* are presented in Table 7.

Table 7: Results for the test dataset provided by the GenCo

Models		**Training Dataset**	**Precision**		**Recall**	
E-C	E-PCA-C	Dataset X	92%	83%	63%	63%
		Dataset Y	100%	93%	63%	81%
		Daset X&Y	100%	94%	63%	100%
Legacy Method		-	16%		44%	

The results in Table 7 show that the proposed approach proves that it can achieve better results regarding both metrics. All the variations of the novel data-driven model can improve precision since the number of FP (false failure detection) is reduced, meaning a lower number of false alarms. Furthermore, an improvement in failure detection (i.e.

higher Recall) is shown, meaning that more failures are detected. Additionally, the results prove that the clustering alternative E-C can obtain better results concerning precision in regards to E-PCA-C. This is justified by the loss of information that occurs when PCA is used instead of the original gases. Therefore, it is expected to have less FP in E-C than E-PCA-C. Nevertheless, the analyses of the Recall indicator reveals more failures are detected in the latter approach. This highlights the trade-off between Recall and Precision in both clustering alternatives.

Nonetheless, E-PCA-C is better suited for this problem since failing to detect failures is more expensive. It was also possible to conclude that the best results come from a knowledge constructed with dataset X&Y and the worse performance came from the model constructed with dataset X (which has the smallest number of observations). This shows that combining additional PT data, even from different utilities, contributes to improve classification accuracy.

5. Model Benchmark

In this section, the proposed approach is benchmarked against both engineering and analytical methods revised in section 1.2. To do so, we utilize the publicly available IEC TC10 dataset[29]. This public dataset is composed by 133 log-transformed DGA records related to seven gas concentrations (H_2, CH_4, C_2H_2, C_2H_4, C_2H_6, CO, CO_2), which were collected by different utilities. In this case, the exact information about the PT condition is available. In total, the dataset has 33 PTs diagnosed with a normal condition and 100 confirmed failures.

- 74 Electrical failures (56% of the total records)
- 26 Thermal Failures (20% of the total records)
- 33 Normal DGA records (24% of the total records)

The fundamental goal is to prove that the proposed approach is capable of outperforming the engineering methods and achieving similar results compared to supervised methods. The results shown in Table 8 compare the engineering methods with the models obtained by applying our approach to the datasets from the previous section. This is a fair comparison since the engineering methods do not change between different datasets. Thus, this comparison allows for a validation of the reproducibility of the obtained models.

Regarding the comparison with the state-of-the-art supervised methods, the results of the direct application of the proposed methodology to the IEC TC10 database are shown in Table 9. In order to validate generalization capability of the proposed approach, it is also presented the best performing data-driven model on the IEC TC10 database, which was previously generated on the real-world data (GenCo and DSO).

Comparing engineering methods

To compare the proposed approach with the engineering methods, we use the same indicators mentioned in the previous section (Precision and Recall). Nevertheless, a few adjustments are necessary since, in the IEC TC10 dataset, the type of internal failure is known. In this validation process, Precision and Recall are not used with engineering methods unable to classify PT with a normal condition, namely DT, Dornenburg, and IEC. This is due to the fact that Precision will be equal to the proportion of failures found in the dataset (75%) and Recall will have a perfect score (100%). Henceforth, Precision and Recall are excluded from the engineering methods mentioned above. Nevertheless, since these engineering methods (DT, Dornenburg, and IEC) are able to classify the type of failure (thermal or electrical), we propose an additional indicator to evaluate the correct assessment (CA) of the PT condition. Henceforth, we propose an additional indicator, Accuracy (Eq. 21), which is defined as follows:

$$Accuracy = \frac{CA}{CA + FA} \tag{21}$$

A correct assessment (CA) will be counted whenever the classified condition, either by the engineering methods or the proposed approach, matches the actual condition described in the IEC TC10[29] database (normal condition, thermal failure, and electrical failure). Conversely, a failed assessment (FA) is counted whenever there is a mismatch between the classified and actual conditions. With Accuracy, the goal is to evaluate if the method is capable of correctly diagnosing the three considered states, on top of the detection of the failure state. Table 8 presents the obtained results for each method.

Table 8: Results evaluation for the IEC TC10[29] dataset

Models	**Precision**	**Recall**	**Accuracy**
Dataset X (E-PCA-C)	82%	**92%**	**75%**
Dataset X (E-C)	83%	86%	75%
Dataset Y (E-PCA-C)	76%	82%	56%
Dataset Y (E-C)	**89%**	78%	74%
Dataset X&Y (E-PCA-C)	72%	86%	41%
Dataset X&Y (E-C)	88%	77%	73%
Duval Triangle	(-)	(-)	**74%**
Rogers	85%	59%	61%
IEC	(-)	(-)	37%
Dornenburg	(-)	(-)	62%
Key Gas	**91%**	**76%**	20%

Looking at Precision, it is possible to conclude that the most precise engineering method is the KGA, which is marginally better than our most accurate data-driven model. However, this method has much lower Accuracy (20%) when compared to our proposed approach (74%). Regarding Recall, the proposed approach, when compared to the engineering methods, can detect the majority of failures, which is aligned with the results from section 4.3. From the Accuracy metric, it is possible to see that our novel approach achieves results similar to the ones obtained by DT (i.e. the most accurate engineering method), which is remarkable for an unsupervised approach. However, the model using the clustering method E-PCA-C has low Accuracy. As mentioned before, this is justified by the loss of information when using only two principal components, which affects the correct failure type assessment.

Moreover, the same conclusions are obtained for the comparison between data-driven models. E-PCA-C generally is better for failure detection but worse regarding Precision and Accuracy when compared to E-C. However, the best data-driven model for this case is created in dataset X, providing a better recall. Nonetheless, the results show again that the precision is increased for datasets with more observations.

Comparing analytical methods

To fully assess the performance of our novel approach, we also present a comparison with state-of-the-art supervised learning methods reported in the literature[28]. In the reported work, the authors benchmark the state-of-the-art supervised learning methods using a similar Accuracy metric (Eq. 21). The metric is not exactly the same since the authors only consider two possible condition states (failure and normal condition), instead of considering the reported three states in the IEC TC10 dataset (normal condition, thermal, and electrical failure). Hence, to enable the benchmark of the proposed approach, it is used the methodology adopted in[28] to compute the reported Accuracy metric. To do so, the proposed approach is directly applied to 80% of the data (IEC TC10 dataset Training Dataset) and then evaluated in the remaining 20% (IEC TC10 dataset Test Dataset). Additionally, the CA is counted whenever the predicted condition (failure or normal condition) matches the real condition. It is important to note that the two failure types are aggregated into one failure condition. Table 9 shows the benchmarking results.

The results demonstrate that it is possible to obtain a comparable performance between the proposed unsupervised approach and the supervised learning methods. When the methodology is applied to the IEC TC10 dataset using E-PCA-C clustering, it is possible to the approach even outperform two supervised methods, namely C-4.5 and SVM linear. Furthermore, when the methodology is not directly applied to IEC 10 training data (i.e. dataset X&Y is used for constructing the data-driven model), the obtained accuracy (80%) is competitive, proving the generalization potential of the novel methodology.

Table 9: Results comparison with the supervised learning methods reported in the literature[28]

Models	**Accuracy**
k-NN	91%
SVM gaussian	90%
Neural Networks log.	89%
SVM quadratic	88%
Low Dimensional scaling (LDS)	88%
Test Dataset IEC TC10 (E-PCA-C)	86%
C-4.5	85%
SVM linear	85%
Dataset X&Y (E-C)	80%

6. Conclusions

This paper presented a novel unsupervised approach capable of diagnosing the PT condition. This approach was developed with two goals: solving a problem faced by electric utilities where records are not labeled and at the same time, build a white-box approach easily interpretable by the DM. The proposed unsupervised methodology showed competitive results when compared against state-of-the-art engineering methods (from IEC and IEEE standards) and supervised learning methods, in metrics such as Precision, Recall, and Accuracy. Moreover, it is proved that it is possible to improve the current *legacy method* used by the generation company and a distribution network operator. The results showed that it is possible to obtain more precise failure detection and accurate PT condition diagnosis. Finally, this work adds two novelties: i) the graphical representation and interpretation of the results for unsupervised DGA analysis, ii) and the advantage of combining the entropy metric with KMA and PCA to enhance the clustering procedure. The advantage of the proposed novelties are shown in two real-world datasets.

The graphical representation supports the DM in quickly understanding the PT condition since it exhibits a clear separation between three states: normal condition, thermal failure, and electrical failure. Furthermore, it also distinguishes the PT condition severity in each state. This will potentially enable utilities to use and analyse online DGA data and immediately understand the degradation process of the PTs condition. Additionally, utilities with a low number of PT, and consequently, fewer data can benefit from the proposed data-driven models and its replication potential. Finally, the two principal components, constructed during the feature extraction phase, can be used in a similar manner like a health index (see[35,36] for more details about this concept) since they allow to quantify and monitor the PT condition while providing failure information.

Nonetheless, other unsupervised approaches should be applied to validate the obtained results, as well as different dimension reduction techniques. One example of a future study is to use one class SVM instead of Key gas analysis to separate PT in a normal condition from PT with potential incipient failure. It would be interesting to understand how one can benefit from using different unsupervised analytical techniques in a multiple-models framework. In turn, this would potentially further improve the proposed approach performance.

Future work will consist in the following: a) test different unsupervised analytical techniques; b) forecasting the principal components and future condition of the PT; c) generate synthetic DGA data with recent advances in deep learning, e.g., by using generative adversarial networks (GANs); d) use the proposed methodology to online monitor the condition of PTs and evaluate its diagnosis capability.

Acknowledgements

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020.

References

[1] Satish V Kulkarni and S A Khaparde. *Transformer engineering: design, technology, and diagnostics*. CRC Press, 2016.

[2] Rohit Kumar Arora, Michel Duval, Alfonso DePablo, Transformers Committee, others, R. R. Rogers, Michel Duval, M Wang, A.J. Vandermaar, and K.D. Srivastava. Review of condition assessment of power transformers in service. *IEEE Electrical Insulation Magazine*, 18(6):12–25, 2002.

[3] M. Duval. Dissolved gas analysis: It can save your transformer. *IEEE Electrical Insulation Magazine*, 5(6):22–27, 1989.

[4] Rohit Kumar Arora. Different DGA techniques for monitoring of transformers. *International Journal of Electronics and Electrical Engineering*, 1(4):299–303, 2013.

[5] Transformers Committee and others. IEEE guide for the interpretation of gases generated in oil-immersed transformers. *Institute of Electrical & Electronics Engineers, Inc., NY*, 1992.

[6] RR Rogers. Ieee and iec codes to interpret incipient faults in transformers, using gas in oil analysis. *IEEE transactions on electrical insulation*, (5):349–354, 1978.

[7] E Dornenburg and W Strittmatter. Monitoring oil-cooled transformers by gas-analysis. *Brown Boveri Review*, 61(5):238–247, 1974.

[8] Diaa-Eldin A Mansour. A new graphical technique for the interpretation of dissolved gas analysis in power transformers. In *2012 Annual Report Conference on Electrical Insulation and Dielectric Phenomena*, pages 195–198. IEEE, 2012.

[9] Michel Duval and Laurent Lamarre. The duval pentagon-a new complementary tool for the interpretation of dissolved gas analysis in transformers. *IEEE Electrical Insulation Magazine*, 30(6):9–12, 2014.

[10] A. Abu-Siada and S. Hmood. A new fuzzy logic approach to identify power transformer criticality using dissolved gas-in-oil analysis. *International Journal of Electrical Power and Energy Systems*, 67:401–408, 2015.

[11] Jawad Faiz and Milad Soleimani. Assessment of computational intelligence and conventional dissolved gas analysis methods for transformer fault diagnosis. *IEEE Transactions on Dielectrics and Electrical Insulation*, 25(5):1798–1806, 2018.

[12] X.R. Chai H.A. Illias and A.H.A. Bakar. Hybrid modified evolutionary particle swarm optimisation-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis. *Measurement*, 90:94–102, August 2016.

[13] Heping Jia, Gregory Levitin, Yi Ding, and Yonghua Song. Reliability analysis of standby systems with multi-state elements subject to constant transition rates. *Quality and Reliability Engineering International*, 35(1):318–328, 2019.

[14] Zhenglin Liang and Ajith Parlikad. A markovian model for power transformer maintenance. *International Journal of Electrical Power & Energy Systems*, 99:175 – 182, 2018.

[15] Hasmat Malik, Rajneesh Sharma, and Sukumar Mishra. Fuzzy reinforcement learning based intelligent classifier for power transformer faults. *ISA Transactions*, 101:390 – 398, 2020.

[16] Ricardo M Arias Velásquez and Jennifer V Mejia Lara. Principal components analysis and adaptive decision system based on fuzzy logic for power transformer. *Fuzzy Information and Engineering*, 9(4):493–514, 2017.

[17] Vladimiro Miranda, Adriana R.Garcez Castro, and Shigeaki Lima. Diagnosing faults in power transformers with autoassociative neural networks and mean shift. *IEEE Transactions on Power Delivery*, 27(3):1350–1357, 2012.

[18] Sherif SM Ghoneim, Ibrahim BM Taha, and Nagy I Elkalashy. Integrated ann-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis. *IEEE Transactions on Dielectrics and Electrical Insulation*, 23(3):1838–1845, 2016.

[19] Khmais Bacha, Seifeddine Souahlia, and Moncef Gossa. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. *Electric Power Systems Research*, 83(1):73–79, 2012.

[20] Fang Yuan, Jiang Guo, Zhihuai Xiao, Bing Zeng, Wenqiang Zhu, and Sixu Huang. A transformer fault diagnosis model based on chemical reaction optimization and twin support vector machine. *Energies*, 12(5):960, 2019.

[21] Ricardo Manuel Arias Velásquez, Jennifer Vanessa Mejía Lara, and Andres Melgar. Converting data into knowledge for preventing failures in power transformers. *Engineering Failure Analysis*, 101:215 – 229, 2019.

[22] Jose Ignacio Aizpurua, Victoria M Catterson, Brian G Stewart, Stephen DJ McArthur, Brandon Lambert, Bismark Ampofo, Gavin Pereira, and James G Cross. Power transformer dissolved gas analysis through bayesian networks and hypothesis testing. *IEEE Transactions on Dielectrics and Electrical Insulation*, 25(2):494–506, 2018.

[23] Xiaohui Yang, Wenkai Chen, Anyi Li, Chunsheng Yang, Zihao Xie, and Huanyu Dong. Ba-pnn-based methods for power transformer fault diagnosis. *Advanced Engineering Informatics*, 39:178–185, 2019.

[24] B. Zhao, M. Yang, H.R. Diao, B. An, Y.C. Zhao, and Y.M. Zhang. A novel approach to transformer fault diagnosis using idm and naive credal classifier. *International Journal of Electrical Power & Energy Systems*, 105:846 – 855, 2019.

[25] Md Mominul Islam, Gareth Lee, and Sujeewa Nilendra Hettiwatte. A nearest neighbour clustering approach for incipient fault diagnosis of power transformers. *Electrical Engineering*, 99(3):1109–1119, 2017.

[26] Lilia Tightiz, Morteza Azimi Nasab, Hyosik Yang, and Abdoljalil Addeh. An intelligent system based

on optimized anfis and association rules for power transformer fault diagnosis. *ISA transactions*, 2020.

[27] Bo Qi, Peng Zhang, Zhihai Rong, and Chengrong Li. Differentiated warning rule of power transformer health status based on big data mining. *International Journal of Electrical Power & Energy Systems*, 121:106150, 2020.

[28] Yann LeCun and Piotr Mirowski. Statistical machine learning and dissolved gas analysis: A review. *IEEE Transactions on Power Delivery*, 27(4):1791–1799, 2012.

[29] Michel Duval and Alfonso DePablo. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases. *IEEE Electrical Insulation Magazine*, 17(2):31–41, 2001.

[30] David JC MacKay and David JC Mac Kay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[31] Milan Zeleny. *Multiple Criteria Decision Making*. McGraw-Hill, 1982.

[32] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics*, 28(1):100–108, 1979.

[33] M. R. Garey, D. S. Johnson, and Hans S. Witsenhausen. The complexity of the generalized Lloyd - Max problem (Corresp.). *IEEE Transactions on Information Theory*, 28(2):255–256, 1982.

[34] Karl Pearson. LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572, 1901.

[35] Ali Naderian Jahromi, Ray Piercy, Stephen Cress, Jim R. R. Service, and Wang Fan. An approach to power transformer asset management using health index. *IEEE Electrical Insulation Magazine*, 25(2):20–34, March 2009.

[36] J.I. Aizpurua, B.G. Stewart, S.D.J. McArthur, B. Lambert, J.G. Cross, and V.M. Catterson. Improved power transformer condition monitoring under uncertainty through soft computing and probabilistic health index. *Applied Soft Computing*, 85:105530, 2019.