Article

Determination of Transformer Oil Contamination from the OLTC Gases in the Power Transformers of a Distribution System Operator

Sergio Bustamante *[ID], **Mario Manana** [ID], **Alberto Arroyo** [ID], **Alberto Laso** [ID] **and Raquel Martinez** [ID]

School of Industrial Engineering, University of Cantabria, Av. Los Castros s/n, 39005 Santander, Spain; mananam@unican.es (M.M.); arroyoa@unican.es (A.A.); lasoal@unican.es (A.L.); raquel.martinez@unican.es (R.M.)
* Correspondence: bustamantes@unican.es

check for updates

Abstract: Power transformers are considered to be the most important assets in power substations. Thus, their maintenance is important to ensure the reliability of the power transmission and distribution system. One of the most commonly used methods for managing the maintenance and establishing the health status of power transformers is dissolved gas analysis (DGA). The presence of acetylene in the DGA results may indicate arcing or high-temperature thermal faults in the transformer. In old transformers with an on-load tap-changer (OLTC), oil or gases can be filtered from the OLTC compartment to the transformer's main tank. This paper presents a method for determining the transformer oil contamination from the OLTC gases in a group of power transformers for a distribution system operator (DSO) based on the application of the guides and the knowledge of experts. As a result, twenty-six out of the 175 transformers studied are defined as contaminated from the OLTC gases. In addition, this paper presents a methodology based on machine learning techniques that allows the system to determine the transformer oil contamination from the DGA results. The trained model achieves an accuracy of 99.76% in identifying oil contamination.

Keywords: communicating OLTC; dissolved gas analysis; fault location; machine learning; maintenance management; oil insulation; OLTC contamination; power transformer

1. Introduction

Power transformers are considered to be the most important assets in power substations. Thus, their maintenance is important to ensure the reliability of the power transmission and distribution system [1,2].

Different maintenance strategies are used for assets based on the condition of and available information about the components and sub-components of the substation equipment, according to [3]. The current trend in maintenance strategies is to maintain a predictive maintenance approach based on prognostics to predict the future asset degradation [3].

Dissolved gas analysis (DGA) is the most commonly used method for managing the maintenance and establishing the health status of power transformers, with much weight given to the results [4]. The DGA method measures the gas concentrations in the oil. The gases normally measured are hydrogen (H_2), acetylene (C_2H_2), methane (CH_4), ethylene (C_2H_4), ethane (C_2H_6), carbon monoxide (CO) and carbon dioxide (CO_2). These gases are formed by the decomposition processes of insulation, which are caused by active faults. From the gas concentrations obtained in the DGA results, it is possible to identify the type of fault. Although DGA continues to be performed in the laboratory,

there is a trend toward online DGA monitoring, because it makes it possible to detect or diagnose faults that occur in the power transformer [1].

The presence of acetylene in the DGA results may indicate arcing or high-temperature thermal faults in the transformer [5–7]. Although the presence of acetylene in the DGA results does not always indicate a transformer fault, in old transformers with an on-load tap-changer (OLTC), oil or gases can be filtered from the OLTC compartment to the main transformer tank. Gas leakage between oil compartments may influence the DGA results and the identification of transformer insulation faults [6–9]. This case was shown in [10]: two faults were identified as low energy discharges (D1); later, it was determined by inspection of the equipment that it was oil contamination from the OLTC gases.

The guides [6,7] utilise the C_2H_2/H_2 ratio to give an indication of the transformer oil contamination from the OLTC gases. According to [11], the C_2H_2/H_2 ratio is used to determine the filtering of fault gases in the main tank from the OLTC compartment when the ratio is greater than or equal to two and, furthermore, when the C_2H_2 concentration is greater than or equal to 30 ppm. When electrical discharges occur in the transformer oil, the amount of acetylene is usually less than the amount of hydrogen. Since the solubility of acetylene is greater than the solubility of hydrogen, when electrical discharge is generated in the OLTC compartment, acetylene diffuses faster out of the OLTC tank. This results in the acetylene concentration in the transformer oil being greater than the hydrogen concentration. Several studies [12–14] have used this ratio to define the transformer oil contamination from the OLTC gases in several samples.

In the data processing for the calculation of the typical gas concentration, gas generation rates and fault identification are influenced by the lack of knowledge about the existence of communication with the OLTC [14]. The report uses the gas ratios included in [7] to identify faults and the C_2H_2/H_2 ratio to indicate the oil contamination from the OLTC gases.

According to [15], there are three steps to follow in the fault identification rules, including the application of the C_2H_2/H_2 ratio. The first step is to decide if the values of the gas concentrations are enough to establish that there is an active fault in the transformer oil. The second step consists of applying the C_2H_2/H_2 ratio to distinguish between contamination or not from the OLTC gases. If there is contamination from the OLTC gases, the rest of the evaluation will be performed by an expert who determines whether there is also an active fault; otherwise, proceed to Step 3, where the type of fault is identified by traditional DGA interpretation methods.

Although the C_2H_2/H_2 ratio [6,7] can be used to consider transformer oil contamination from the OLTC gases, it is necessary to take into account that high H_2 concentrations affect this ratio. A high H_2 concentration could be due to the fact that H_2 is produced in the transformer oil in almost all incipient faults. Applying this ratio with a high C_2H_2 concentration would not show the contamination of the transformer oil because of the high H_2 concentration. This situation occurred in [12]: the application of the C_2H_2/H_2 ratio did not show the communication between oil compartments; it was necessary to inspect the conservator, and two holes were found on the barrier between the main conservator and the OLTC conservator.

Machine learning (ML) techniques are widely used to assess the condition of a transformer [16–22] and identify potential faults in a transformer [23–28] using DGA, oil quality analysis (OQA), the furfuraldehyde (FFA) content of the oil, polarisation and depolarisation current (PDC) measurements and dielectric frequency domain spectroscopy (FDS). Through the application of ML algorithms, models are generated that create predictions of the transformer health status and automatically diagnose potential faults in the transformer insulation from the data provided.

According to [20,23,25], a decision tree (DT) classifier is easy to interpret because the output is a set of conditional if-then tests of the input data. In addition, it is easy to train, and it works better than other ML techniques. The accuracy of the results obtained in these studies was better than that of those obtained with the DT algorithm.

This paper presents a method for determining the contamination of the transformer oil from the OLTC gases in a group of power transformers of a distribution system operator (DSO) based on the

DGA results for the last two years, the application of the guides [6,7] and the knowledge of experts. The method proposed in this paper improves the detection of transformer oil contamination from the OLTC gases that is not determined by traditional DGA interpretation methods. Based on this transformer oil contamination classification methodology, a DT algorithm is applied to develop a prediction model that automatically recognises transformer oil contamination from the OLTC gases in DGA samples.

2. Background Theory

Power transformers are maintained based on the information available for the components and sub-components of the transformer. A risk index is used to prioritise maintenance work in a group of transformers [3,29]. This risk index is a function of the probability of failure and the consequences of failure. The probability of transformer failure is a parameter that is associated with the health index (HI) and varies over time. The consequences of failure evaluate and define the results of a failure event and do not vary if the transformer is not relocated.

In the calculation of the transformer's HI, more condition parameters are used than for any other substation asset. The condition parameters used to calculate the HI for a power transformer are shown in Figure 1.
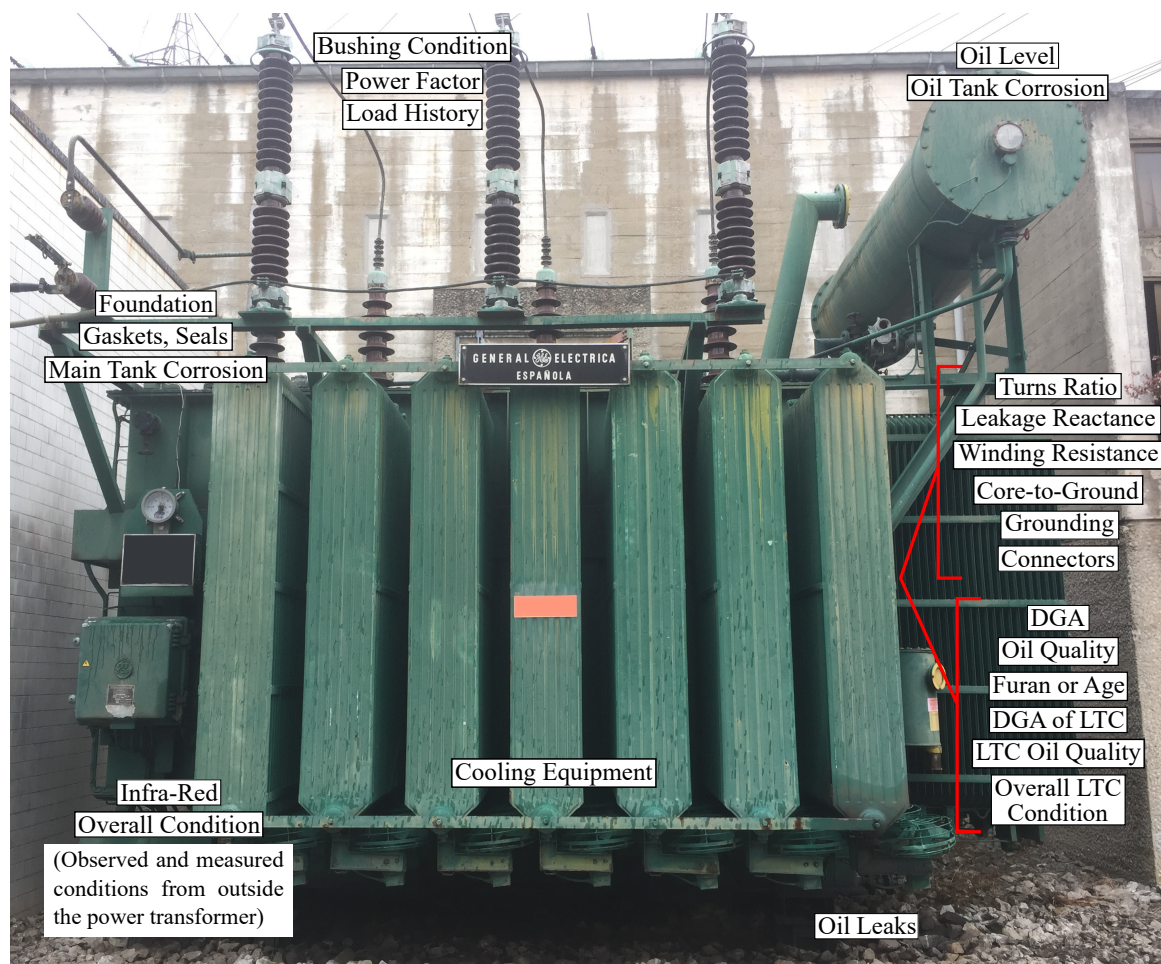

Figure 1. Locations of condition parameters used to calculate the HI of the transformer.

Based on the condition parameters shown in Figure 1, oil analyses (DGA, OQA and FFA) are the easiest to perform without de-energising the transformer and provide the most useful health information [2].

The most commonly used method to identify and diagnose transformer oil faults is DGA. The combination of gas concentrations generated in the transformer oil is a result of the nature of the fault, as well as the temperature and energy at the fault location. Faults can be identified from the gas concentrations in the DGA results using the methods for interpreting the DGA results reported in [6,7,10]. The key gases that are generated depending on the potential fault are listed in Table 1.

Table 1. Gas formation based on the type of failure [6,7].

Fault Type		H_2	CH_4	C_2H_6	C_2H_4	C_2H_2
				Gas Generated		
Thermal faults in oil (<300 °C)	T1	○	●	●	·	
Thermal faults in oil (300–700 °C)	T2	○	○	○	●	·
Thermal faults in oil (>700 °C)	T3	○			●	○
Partial discharges	PD	●	○			·
Low energy discharge—sparking	D1	●			○	●
High energy discharge—arcing	D2	●			○	●

●: major concentration; ○: secondary concentration; ·: trace concentration.

Acetylene is one of the combustible gases that can be produced when a fault occurs in a transformer [6,7]. The transformer insulation faults associated with the presence of acetylene in DGA results are high and low energy discharges and high thermal faults (Table 1). As seen in Figure 2, very high temperatures (>800 °C) followed by very rapid cooling are needed for the acetylene concentration to increase, causing it to accumulate in the oil, which occurs during arcing [5]. Traces of acetylene can also be formed at temperatures below 800 °C, as shown in Figure 2.
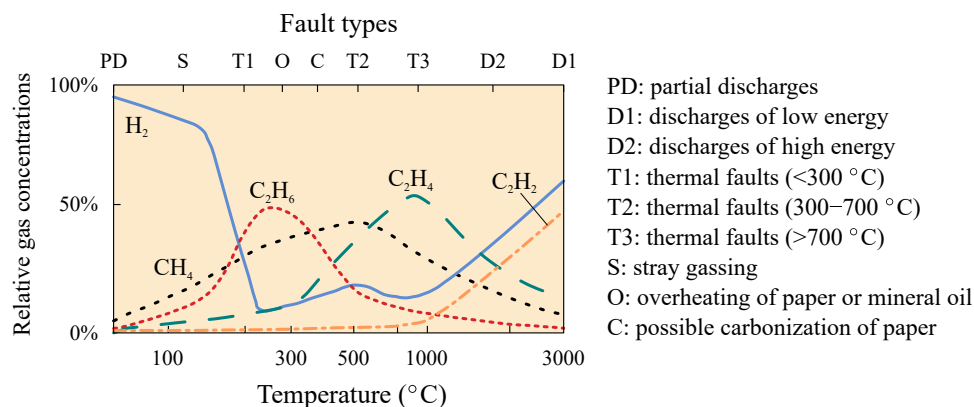

Figure 2. Gas generation based on temperature (reprinted with permission from CIGRE [5], © 2019).

The presence of acetylene in DGA results may also be due to the filtration of OLTC gases. According to [6,7], the C_2H_2/H_2 ratio can be used to determine the transformer oil contamination from the OLTC gases. When this ratio is greater than two or three, this indicates contamination.

As expected, the DGA results for transformers with oil contamination from the OLTC gases will show a higher acetylene concentration than normal. According to the guides for the interpretation of the gases generated in the transformer oil [6,7], there are different acetylene concentration limits based on whether or not there is communication between the OLTC compartment and the main tank, or even whether or not OLTC is used, as seen in Table 2. Table 2 shows the acetylene concentration limits collected in the guides [6,7]. The IEEE guide indicates the gas limit depending on the result of the O_2/N_2 ratio, also showing the 90th and 95th percentile of the typical acetylene concentration. The IEC guide indicates the 90th percentile of the typical acetylene concentration for cases where there is communication between oil compartments, called communicating OLTC, and when there is no communication between oil compartments or no OLTC is used. The acetylene concentration range when there is no communication between oil compartments or no OLTC is used is 1–20 ppm,

while the range is 60–280 ppm when there is communication between the OLTC compartment and the main tank.

Table 2. Acetylene concentration limits (ppm).

	IEEE [6]		IEC [7]	
	O_2/N_2 **Ratio** $\leq$ **0.2**	O_2/N_2 **Ratio** $>$ **0.2**	**No OLTC**	**Communicating OLTC**
90th percentile	1	2	2–20	60–280
95th percentile	2 *	7	-	-

*: This value is 4 ppm when the transformer age is >30 years.

Acetylene is generated in the OLTC during tap changes as a result of arcing between the fixed and moving parts [30,31]. Acetylene appears mainly in a non-vacuum-type OLTC, but acetylene traces can be produced in a vacuum-type OLTC [30,31].

There are several OLTC classifications according to [30–32]. The classification proposed in [30] is one of the simplest and most commonly used. It is given in Figure 3.

According to [30], the OLTC design influences the gases that can be generated due to normal operation depending on the gas generating components, as shown in Table 3.
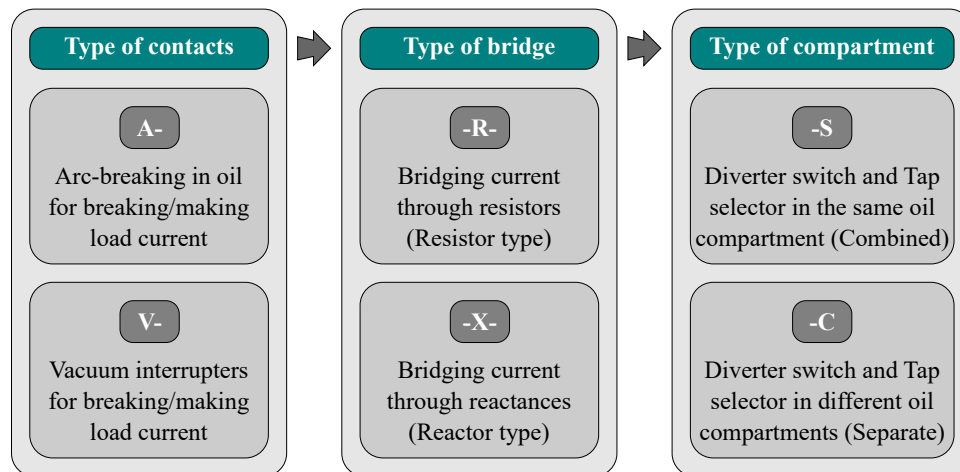

Figure 3. OLTC classification (reprinted with permission from CIGRE [30], © 2010).

Table 3. Gas source based on OLTC design (reprinted with permission from CIGRE [30], © 2010).

OLTC Components	Gas Sources
Arc-switching contacts	High energy discharge gases
Commutation contacts, by-pass contacts	Low energy discharge gases
Vacuum interrupters	No gases
Transition resistors	Heating gases $\leq$ 300 °C (normal operation) Heating gases $>$ 300 °C (overload, fault case)
Transition reactance (preventive autotransformer, inside transformer tank)	No gases

The types of oil compartments used for the different OLTC designs are listed in Figure 3, and the pathways through which OLTC gases can contaminate the transformer insulation are shown in Figures 4 and 5, respectively. The transformer oil contamination from the OLTC gases via the gaskets and oil compartment (which is not gas tight) mainly occurs when the OLTC compartment is inside the main tank, as shown in Figure 4a–c. Contamination via the common air space occurs when the transformer and OLTC share an oil expansion tank with a common air space. This type of connection

usually applies to in-tank OLTC types because compartment-type OLTCs generally have a gas space under the tank cover [8,33].
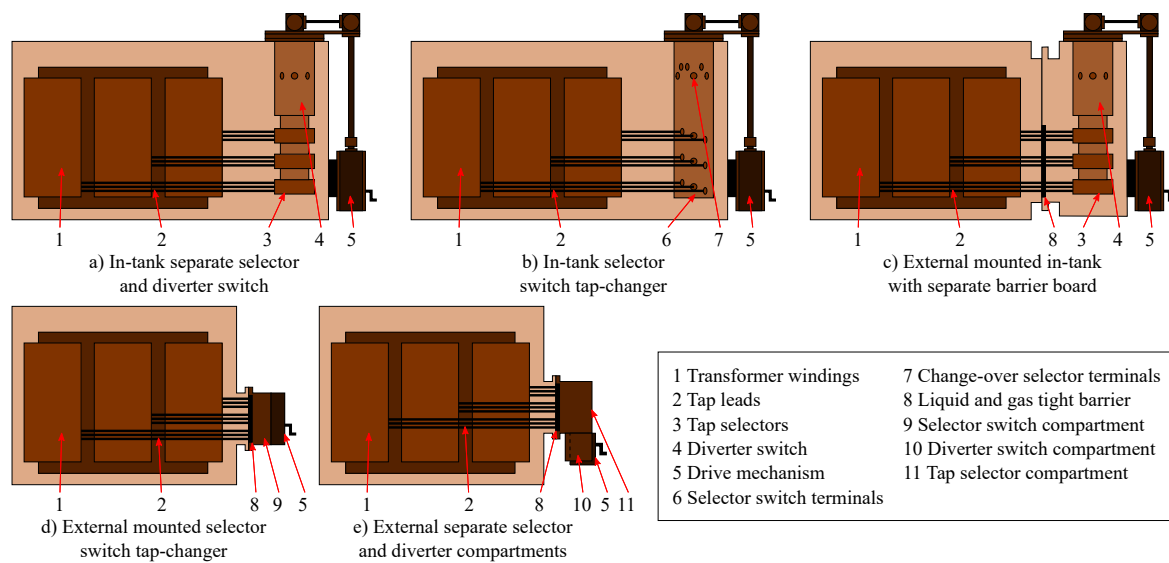

Figure 4. Types of oil compartments used by different OLTC designs [32].
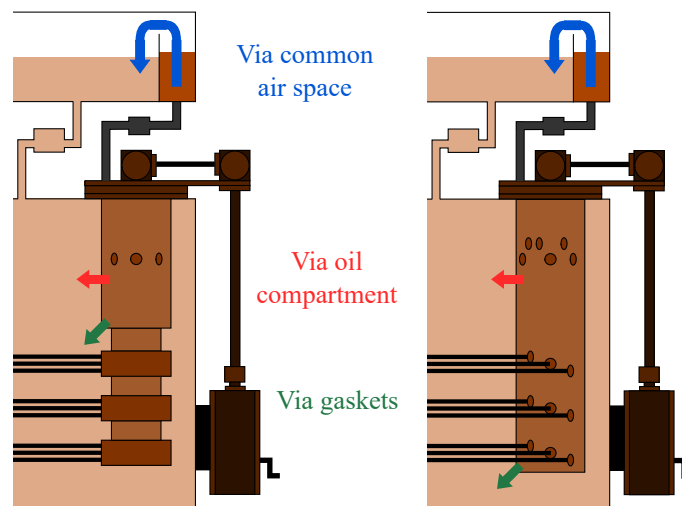

Figure 5. Different pathways for transformer oil contamination from the OLTC gases [8].

One of the most commonly used methods for identifying faults using an OLTC DGA is the Duval triangle 2 for OLTCs according to [5,30,31,34]. The Duval triangle 2 makes it possible to distinguish between normal and abnormal gas formations. The Duval triangle 2 uses the concentration ratio of three combustible gases (acetylene, ethylene and methane) to identify faults or normal operation. The faults that this method can identify, in addition to normal operation, are similar to those listed in Table 1, except for partial discharges (PD). Interpreting several DGAs of the same OLTC on the triangle helps to visualise the evolution of the gas formation in the OLTC over time. The use of the triangle for a sample of OLTCs makes it possible to visualise gas formation patterns in a population of OLTCs.

3. Study Characteristics

This study was based on 388 laboratory DGAs of 175 transformers with OLTCs performed over a period from the middle of 2017 to the middle of 2019 (two years). The transformers had age, voltage class and power rating ranges of 1–69 years, 25–400 kV and 3–450 MVA, respectively. Figure 6 shows the distribution of these transformers according to age, voltage class and power rating.

In addition, the different types of OLTCs that were installed in the power transformers are shown in Figure 6 according to the classification of Figure 3. Abbreviations that show the type of OLTC were generated from the concatenation of the letters in Figure 3: the first letter indicates the type of contacts used by the OLTC (arcing (A) or vacuum contacts (V)); the second letter indicates the type of bridge (resistor (R) or reactor type (X)); the last letter shows the type of compartment of the diverter switch and the tap selector (different (S) or the same (C)). Most of the OLTC types for the transformers studied corresponded to the in-tank types of Figure 4a,b. The newer transformers have the double compartment installation type, as shown in Figure 4e.
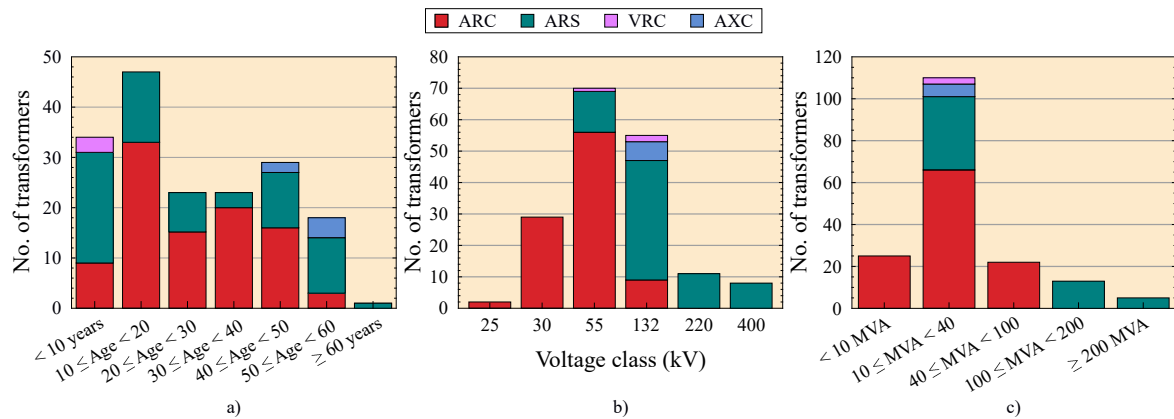

Figure 6. Distribution of power transformers according to (**a**) age, (**b**) voltage class and (**c**) power rating. A, arcing; R, resistor; V, vacuum.

Figure 7 shows the acetylene concentrations measured in the DGAs as a function of the sampling date. The concentration range for transformers without an OLTC or communication with the main tank, as specified in the IEC guide [7], is shown in green. The concentration range for transformers with communicating OLTC, according to [7], is shown in red.
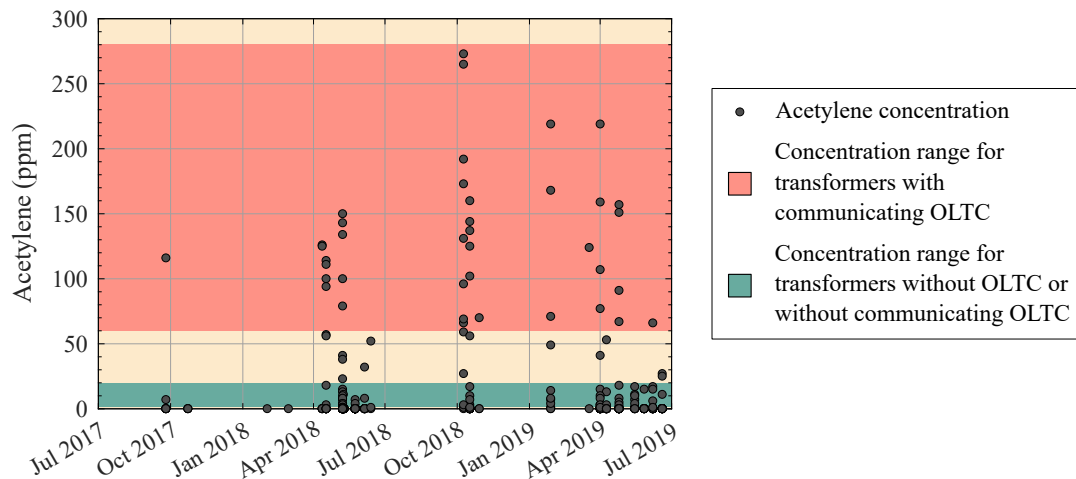

Figure 7. Acetylene concentration measurements.

Before applying the C_2H_2/H_2 ratio to the data, it was necessary to check the reliability of the data in cases that presented an abnormal acetylene concentration, following the indications given in the IEEE guide [6].

After studying the data in which there was a large increase in the acetylene concentration, the results due to gas formation in the transformer were differentiated from those of a bad sample. With this revision, it was possible to discard a result showing 1443 ppm of acetylene (not included in Figure 7), and three DGA results were considered uncertain pending the new DGA results.

Once the data due to a bad sample were removed from the dataset, the C_2H_2/H_2 ratio was applied.

After applying the ratio, five results with errors were observed and needed investigation. These errors were due to a hydrogen concentration of 0 ppm. Table 4 lists the gas concentrations of the five transformers in which no ratio result was obtained.

Table 4. Transformer data without the C_2H_2/H_2 ratio result.

Transformer No.	Sample Date	C_2H_2/H_2 Ratio	Age (Years)	Acetylene (ppm)	Hydrogen (ppm)
1	25/04/2019	—	54	0	0
2	24/05/2018	—	10	0	0
3	13/06/2018	—	44	52	0
4	27/05/2019	—	29	15	0
5	25/09/2017	—	26	0	0

Based on the trend for the rest of the DGAs related to these five transformers, three transformers were discarded because of a lack of contamination from the OLTC, and Transformers 3 and 4 were defined as contaminated from the OLTC.

Table 5 lists the DGA results for the transformers with C_2H_2/H_2 ratios greater than two. Comparing the DGA results listed in Table 5 with the rest of the DGA results related to the same transformer, it was observed that Transformers 12 and 13, despite having ratios greater than two, had a very low acetylene concentration in all of their DGA results. The acetylene concentrations of these transformers had ranges of 4–6 and 1–8 ppm, respectively. Thus, these transformers were defined as uncontaminated by the OLTC gases.

Table 5. Transformer data with the C_2H_2/H_2 ratio greater than 2.

Transformer No.	Sample Date	C_2H_2/H_2 Ratio	Age (Years)	Acetylene (ppm)	Hydrogen (ppm)
3	17/10/2018	9.33	44	56	6
3	28/01/2019	3.27	44	49	15
3	09/04/2019	2.30	44	53	23
6	19/06/2019	6.75	30	27	4
7	28/01/2019	2.54	23	71	28
8	17/04/2018	3.92	20	94	24
8	28/01/2019	2.44	20	168	69
9	07/06/2019	3.47	18	66	19
10	01/04/2019	3.06	51	107	35
10	17/10/2018	2.78	51	125	45
11	25/09/2017	2.04	44	116	57
12	07/06/2019	3	44	6	2
13	28/01/2019	4	50	8	2
14	12/04/2018	25	34	125	5
14	09/10/2018	8.27	34	273	33
15	09/10/2018	10.60	34	265	25
15	12/04/2018	6.63	34	126	19
16	25/04/2019	3.19	28	67	21
17	08/05/2018	4.33	26	13	3
17	09/10/2018	3.86	26	27	7
17	07/06/2019	3.75	26	15	4
18	15/05/2019	11	45	11	1

Based on the comparison of the DGA results listed in Table 5 with the rest of the DGA results for each transformer, twelve transformers were defined as contaminated from the OLTC gases, in addition to Transformers 3 and 4, which were previously defined as contaminated.

Table 6 lists the transformers with DGA results that exceeded 10 ppm of acetylene and C_2/H_2 ratios smaller than two. The transformers established as contaminated in Table 5 are excluded from Table 6. As listed in Table 6, the acetylene concentrations of these transformers and their ages had ranges of 11–219 ppm and 18–50 years, respectively.

Table 6. Transformer data with the C_2H_2/H_2 ratio less than 2 and a high C_2H_2 concentration.

Transformer No.	Sample Date	C_2H_2/H_2 Ratio	Age (Years)	Acetylene (ppm)	Hydrogen (ppm)
4	09/05/2018	0.55	29	11	20
19	01/04/2019	0.88	48	219	248
19	17/10/2018	0.92	48	144	157
19	17/04/2018	0.68	48	114	168
20	28/01/2019	0.90	47	219	244
20	09/10/2018	1.17	47	192	164
20	08/05/2018	0.76	47	150	198
21	01/04/2019	0.95	50	159	168
21	17/10/2018	0.85	50	137	162
21	17/04/2018	0.69	50	100	144
22	25/04/2019	1.74	49	151	87
22	08/05/2018	1.37	49	134	98
22	09/10/2018	1.31	49	131	100
23	18/03/2019	0.95	45	124	130
23	08/05/2018	0.81	45	100	123
23	09/10/2018	1.07	45	96	90
24	17/10/2018	0.65	48	102	158
24	17/04/2018	0.36	48	56	156
25	25/04/2019	1.12	49	91	81
25	08/05/2018	0.80	49	79	99
25	09/10/2018	0.85	49	69	81
26	09/10/2018	0.18	18	59	326
26	17/04/2018	0.22	18	57	264
26	01/04/2019	0.13	18	41	310
27	05/06/2018	0.94	43	32	34
27	19/06/2019	1	43	25	25
28	08/05/2018	0.25	48	23	93
29	17/04/2018	0.13	18	18	135
29	15/05/2019	0.49	18	17	35
30	25/04/2019	1.20	32	18	15
31	01/04/2019	0.06	19	15	260
31	28/01/2019	0.06	19	14	231
32	07/06/2019	0.90	29	17	19
33	19/06/2019	0.15	23	11	73

As in the previous cases, the transformers in Table 6 were compared with the rest of the DGA samples corresponding to each one. Transformers 28 and 33 had acetylene concentration ranges of 7–23 and 8–11 ppm, respectively.

Transformer 33 was discarded as contaminated because of its low acetylene concentration values and stability over time.

Based on the acetylene concentration listed in Table 6, Transformer 28 could be contaminated. All of the DGA results except for that of one sample (23 ppm) indicated low and stable acetylene concentrations. Because of this and the fact that it was connected to the power transmission network, it was defined as uncontaminated. In this way, the transformer was controlled by applying the limits of Table 2 corresponding to a transformer without OLTC to its monitoring.

From Table 6, the old transformers (Nos. 19, 20, 21, 22, 23, 24, 25 and 27) in the 43–50 year range had hydrogen and acetylene concentrations with high values. The high hydrogen values were due to gas accumulation over time because hydrogen is produced in almost all incipient faults. The historical data for each transformer were reviewed, and no fault that indicated arcing in the insulation was found. Thus, it could be said that the high acetylene values were due to OLTC gas contamination. Transformers 19, 20, 21, 22, 23, 24, 25 and 27 were defined as contaminated.

As in the previous case, the historical data for Transformers 26, 29, 30, 31 and 32 were reviewed, and no arcing fault was found in the transformer insulation. As can be seen in Table 6, this group of transformers had low acetylene concentrations, except for Transformer 26, and a relatively stable trend. Despite this, they were classified as contaminated from OLTC gases because no arcing fault was found in the historical data.

In summary, twenty-six of the 175 transformers were defined as contaminated from the OLTC gases. Out of these, eighteen of the 26 transformers defined as contaminated were of the OLTC ARC type, and the remaining eight were of the OLTC ARS type. Figure 8 shows the distribution of these transformers according to age, voltage class and power rating. In addition, the different types of OLTCs installed in these power transformers are shown in Figure 8.
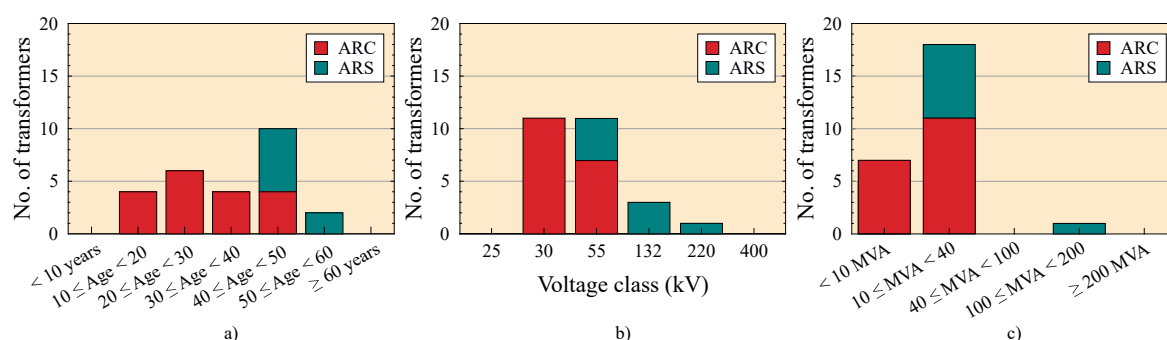

Figure 8. Distribution of power transformers defined as contaminated according to (**a**) age, (**b**) voltage class and (**c**) power rating.

It should be noted that the methodology followed in this section is an assumption of transformer oil contamination based on the study of the DGA samples. Neither the application of the ratio nor EK can guarantee this contamination. Given the DGA results, it is possible to assume the contamination in the 26 transformers and evaluate their future DGA results with the acetylene values for transformers with communicating OLTC (Table 2). It will be necessary to take into account the increases between DGA samples and check if they correlate with an increase in the number of OLTC operations in order to be able to discard the existence of faults in the transformer insulation.

From the results of the proposed method, a comparison was made with the traditional fault identification methods. The traditional fault identification methods used were the Duval triangle method (DTM) and the Duval pentagon method (DPM), as can be seen in Figure 9. The abbreviations used correspond to those shown in Table 1.

In Figure 9, the results obtained through the application of the C_2H_2/H_2 ratio are shown in red, and the results obtained through EK are represented in magenta. It is observed that most of the results are located in the areas that indicate high and low energy discharge failures, D2 and D1, respectively.

Transformers identified as contaminated through the C_2H_2/H_2 ratio are well defined in zones D1 and D2, except for one transformer whose DGA results are in zone T3 in the triangle and on the boundary between D2 and T3 in the pentagon. It can be interpreted that this transformer had a thermal fault in addition to the oil contamination from the OLTC gases.

The transformers identified as contaminated through EK are scattered over zones D1, D2 and D+T in the triangle and in zones D1 and D2 in the pentagon, because the predominant gas was no longer acetylene, making the results more distributed throughout the areas.
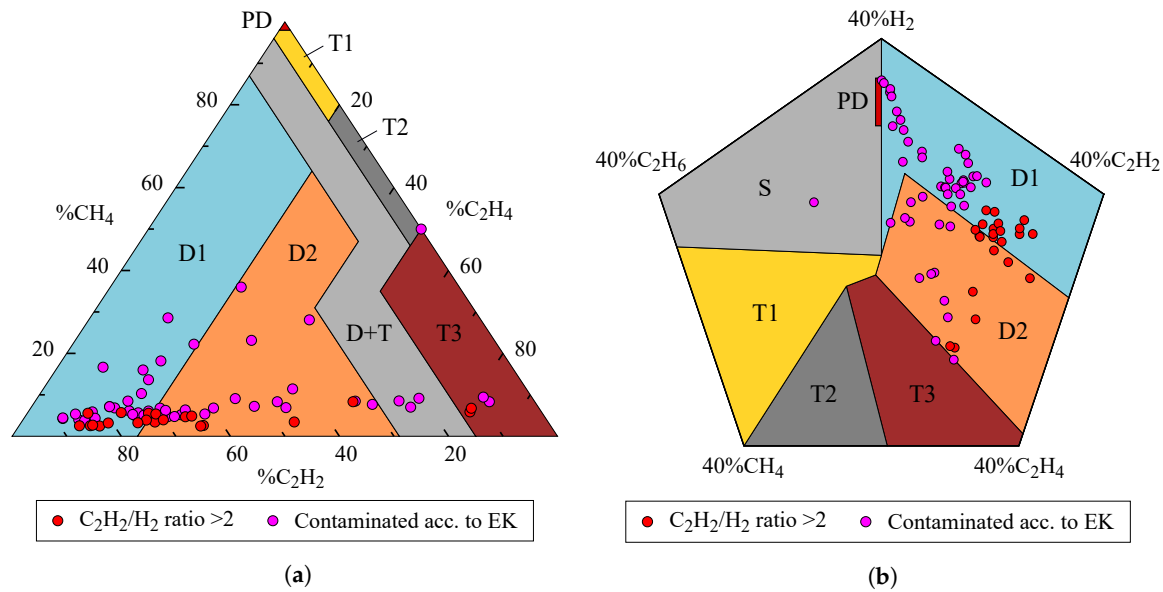
Figure 9. Fault identification through (**a**) DTM and (**b**) DPM.

4. Machine Learning Methodology

Based on the previously described results, ML techniques were used to allow the classifier algorithm to automatically detect transformer oil contamination from the OLTC gases. Figure 10 shows the flowchart of the ML methodology applied and explained in this section.
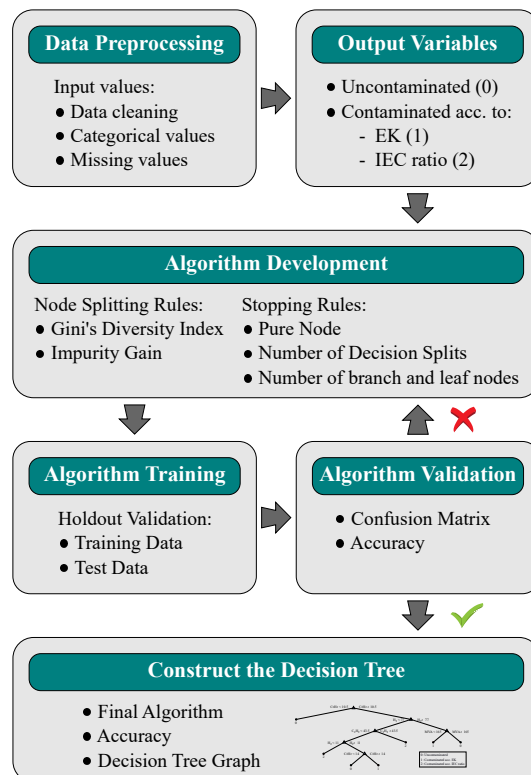

Figure 10. Machine learning methodology flowchart.

4.1. Data Preprocessing

Before starting, it was necessary to check the reliability of the data in cases that presented missing values or abnormal gas concentrations, following the indications given in the IEEE guide [6].

Then, four-hundred sixteen DGA datasets (388 DGA datasets for transformers with OLTCs and 28 DGA datasets for transformers without OLTCs) and several characteristics of the transformers were used to train the algorithm. The input variables (also known as predictors) included 12 numerical and two categorical variables. The numerical predictors included the age, power rating, voltage class and concentrations of hydrogen, acetylene, methane, ethylene, ethane, carbon monoxide, carbon dioxide, oxygen and nitrogen. The categorical predictors indicated whether or not ("Yes" or "No") the transformer was connected to the transmission network and possessed an OLTC.

4.2. Calculation of Output Variables

The output variable (known response) used in the training of the algorithm was the classification developed in Section 3, but the responses classified as contaminated were divided into two categories according to the C_2H_2/H_2 ratio and expert knowledge (EK). Based on the division of the responses classified as contaminated into two classes and the responses of uncontaminated transformers, a total of three classes were defined, as can be seen in Figure 11. The responses of the 28 observations of transformers without OLTCs were defined as uncontaminated, which helped the algorithm to learn to classify them using the ML techniques.
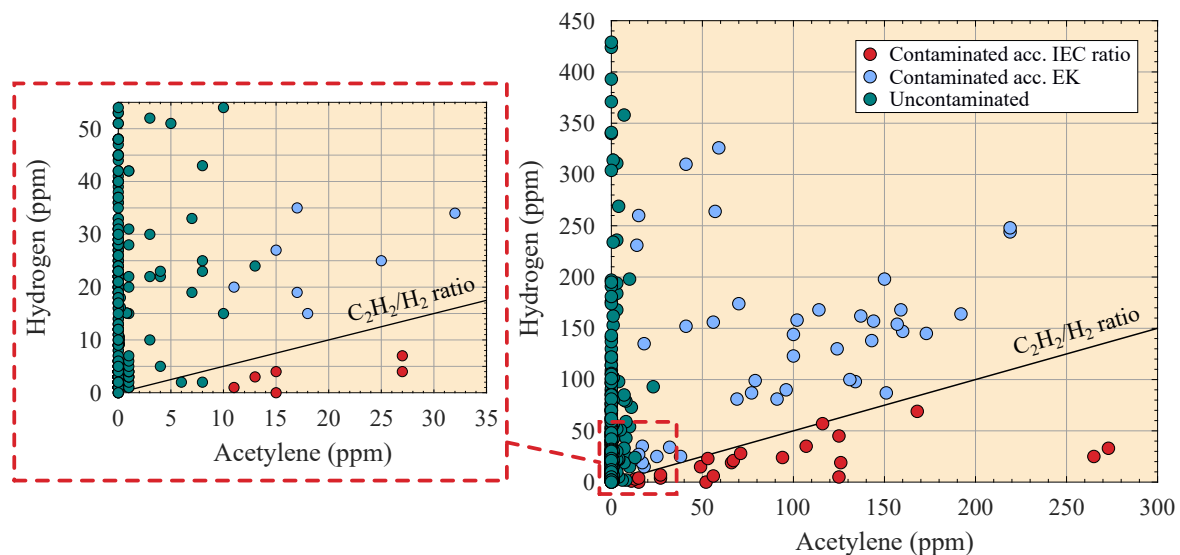

Figure 11. Responses used for algorithm training.

4.3. Algorithm Development

The DT classifier was the algorithm selected to be trained based on the accuracy results collected in [20,23,25]. The software used to develop the classifier algorithm was MATLAB R2018b [35].

The developed algorithm obtains an adjusted binary classification DT based on the input variables and responses [36,37], in order to predict responses for new data. The obtained binary tree divides branching nodes based on the values of the attributes.

Based on the predictors and responses explained in Sections 4.1 and 4.2, the dataset used in the algorithm development was as follows:

$$
\begin{aligned}
T &= \{T_1, T_2, ..., T_m\} \\
x_i &= \{x_1, x_2, ..., x_p\} \\
S &= \{S_1, S_2, ..., S_m\}
\end{aligned}
\tag{1}
$$

where T is the set of all observations, m is the number of observations, x_i is the set of values of a predictor, p is the number of predictors and S is the set of responses to be predicted; each response takes one of the three classes defined in Section 4.2.

The calculation procedure to develop the optimal DT, starting from the node t that contains the set of all observations T, follows the next steps.

1. Calculate the impurity of node t.

    Gini's diversity index (I_t) measures the impurity of node t and can be written as:

    $$I_t = 1 - \sum_i p^2(i) \tag{2}$$

    where the sum is over the classes i at the node, and $p(i)$ is the probability of class i at the node. A pure node (a node with just one class) has a Gini index of zero; otherwise, the Gini index is positive.

2. Calculate the probability at node t.

    The probability that an observation is in node t is given as:

    $$P(T) = \sum_{j \in T} w_j \tag{3}$$

    where w_j is the weight of observation j and T is the set of all observations at node t. If no different weights are used, $w_j = 1/n$, where n is the sample size.

3. Sort predictor elements in ascending order.

    Each ordered element of the predictor x_i can be a candidate to split the node.

4. Calculate the impurity gain (ΔI).

    In order to determine the best way to split node t using x_i, the ΔI over all splitting candidates is calculated. For all splitting candidates of x_i, the algorithm splits node t into left (t_L) and right (t_R) nodes, each with its set of observations, T_L and T_R, respectively. Then, the ΔI is calculated as follows:

    $$\Delta I = P(T)I_t - P(T_L)I_{t_L} - P(T_R)I_{t_R} \tag{4}$$

    where $P(T_L)$ and $P(T_R)$ are the probabilities that an observation is at node t_L and t_R, respectively, and I_{t_L} and I_{t_R} are the impurities at the child nodes.

5. Selection of splitting node.

    The algorithm selects the splitting candidate that produces the largest ΔI.

6. Once the splitting node has been selected, the child nodes (t_L and t_R) become parent nodes (node t). Then, the previous steps are recursively repeated to split the new parent nodes until pure nodes are achieved or the stopping rules are reached.

Following the previous steps, a very deep DT can be generated with many small leaves, achieving a low training error, but the test error is usually high. To avoid deep growth, stopping rules should be imposed. The stopping rules used were:

*   The maximum number of decision splits was 50.
*   The minimum number of branch node observations was 10.
*   The minimum number of leaf node observations was one.

4.4. Algorithm Training and Validation

In order to verify the accuracy of the algorithm [38], the dataset was divided into two groups, training data and test data, using a percentage ratio of 90:10 or 50:50 to observe how the amount of training data influenced the accuracy of the algorithm. Because the observations in the training data differed in each execution of the algorithm, even with the same ratio, the algorithm was executed five times with the same ratio.

Accuracy (A) was obtained in each execution of the algorithm to validate the trained model. A is calculated as:

$$A(\%) = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \cdot 100 \tag{5}$$

where TP is the number of predicted true positives, TN is the number of predicted true negatives, FP is the number of predicted false positives, and FN is the number of predicted false negatives.

The accuracy results based on the percentages of training and test data are listed in Table 7. In addition, Table 7 lists the confusion matrices obtained in each case. It is important to keep in mind that a transformer misclassified as contaminated (false positive) would be evaluated with high C_2H_2 limits (Table 2). Thus, its status assessment would go unnoticed in terms of C_2H_2.

Table 7. Decision tree classifier accuracy results and confusion matrices from different training and test data.

Run No.	Class	90% Training and 10% Test			80% Training and 20% Test			70% Training and 30% Test			60% Training and 40% Test			50% Training and 50% Test		
	T \ P	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
1	0	35	0	0	70	1	0	104	1	0	139	2	0	174	1	2
	1	0	4	0	0	8	0	0	10	2	0	14	3	0	18	2
	2	0	0	2	0	0	4	0	3	4	0	2	6	0	4	7
		A (%): 100			A (%): 98.80			A (%): 95.16			A (%): 95.78			A (%): 95.67		
2	0	35	0	0	70	0	0	104	1	0	140	1	0	174	2	0
	1	0	4	0	0	8	0	0	13	0	1	16	0	0	18	3
	2	0	0	2	0	0	5	0	3	3	0	2	6	0	2	9
		A (%): 100			A (%): 100			A (%): 96.77			A (%): 97.59			A (%): 96.63		
3	0	35	0	0	71	0	0	104	1	0	141	1	0	174	2	0
	1	0	4	0	0	8	0	0	12	0	0	13	3	3	14	4
	2	0	0	2	0	0	4	0	0	7	0	2	6	0	1	10
		A (%): 100			A (%): 100			A (%): 99.19			A (%): 96.39			A (%): 95.19		
4	0	34	1	0	69	1	0	105	0	0	140	1	0	175	1	0
	1	0	4	0	0	8	0	0	13	0	0	16	0	0	20	0
	2	0	0	2	0	0	5	0	0	6	0	3	6	1	4	7
		A (%): 97.56			A (%): 98.80			A (%): 100			A (%): 97.59			A (%): 97.11		
5	0	35	0	0	70	0	0	105	0	0	142	0	0	174	2	0
	1	0	4	0	0	8	0	1	11	1	1	15	0	0	21	0
	2	0	0	2	0	0	5	1	1	4	1	2	5	0	4	7
		A (%): 100			A (%): 100			A (%): 96.77			A (%): 97.59			A (%): 97.11		

T: true class; P: predicted class; A: accuracy.

From Table 7, it is observed that a larger training group resulted in fewer false positives and negatives.

Since the accuracy results shown in Table 7 were correct, the trained algorithm was validated; otherwise, it would be necessary to modify the parameters defined in the algorithm development to improve the algorithm accuracy.

4.5. Construct the DT

The final training of the algorithm [38] was performed with all the observations, achieving an accuracy of 99.76% for the responses. As a result, an adjusted binary classification decision tree was obtained based on the predictors and responses contained in the dataset, as shown in Figure 12.

As expected, Figure 12 shows that the most important predictor was the acetylene concentration, followed by the hydrogen concentration and power rating.
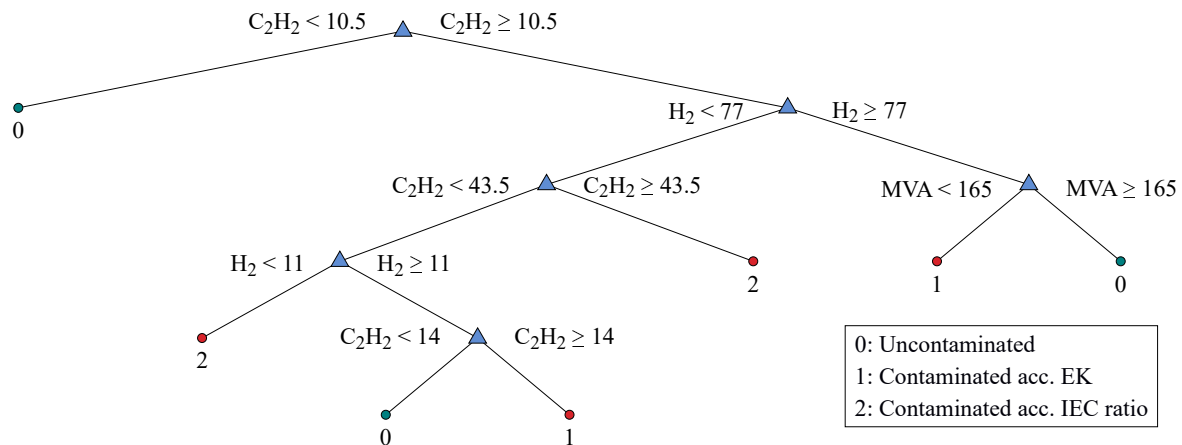

Figure 12. Decision tree obtained from the trained algorithm (concentrations in ppm and power rating in MVA).

The DGA sample that was poorly predicted by the trained algorithm belonged to Transformer 4. This transformer had two DGA samples that were used in the algorithm training (Tables 4 and 6). The responses of the DGA samples and the wrong prediction of the Transformer 4 are shown in Figure 13. The wrong prediction was defined as uncontaminated by the trained algorithm, while it was defined as contaminated according to EK, since its acetylene and hydrogen concentrations were 11 and 20 ppm, respectively. The wrong prediction was in the vertical zone generated between the responses of uncontaminated and contaminated transformers according to EK (Figures 11 and 13); therefore, the trained algorithm could not determine the transformer contamination accurately.
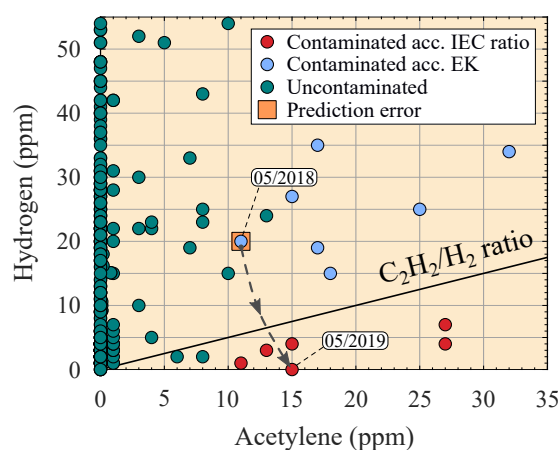
Figure 13. Transformer 4 responses and prediction error.

As new DGA samples are obtained, the trained algorithm will be run using the predictors explained in Section 4.1. The algorithm will output the predictions of whether or not there is contamination. These results will be compared to previous transformer oil contamination definitions. In cases where the algorithm is not correct, each sample must be studied individually to know if it is an error or the transformer has gone from being uncontaminated to being contaminated from OLTC

gases (applying the methodology explained in Section 3). From the results of the application of the trained algorithm, it will be improved if necessary.

5. Conclusions

This paper presents the application and expert knowledge of the C_2H_2/H_2 ratio collected in [6,7] in order to determine transformer oil contamination from the OLTC gases in a group of power transformers of a DSO. The 175 power transformers studied based on 388 DGA results have different ages, voltage classes, and power ratings.

Based on the application of the ratio, thirteen transformers are defined as contaminated, and one transformer is classified as uncontaminated due to low acetylene values, in the 8–11 ppm range. The expert interpretations of the DGA results for 13 power transformers define them as contaminated based on a study of the rest of the DGA results for each of them.

As expected, the trend in the results is that old transformers are contaminated by the OLTC gases, and sixty-three percent of the contaminated transformers have ages in the range of 30–60 years. Based on these results, it is observed that the youngest contaminated transformers have their tap selector and diverter switch in different compartments, whereas the oldest contaminated transformers have their tap selector and diverter switch in the same compartment.

It should be noted that the methodology performed in this study is an assumption of transformer oil contamination based on the study of the DGA samples. Neither the application of the ratio, nor the EK can guarantee this contamination. The evaluation of the future DGA results of the transformers defined as contaminated will have to be performed taking into account the number of operations of the OLTCs, in order to discard the existence of faults in the insulation of the transformer.

In addition, this paper proposes the application of ML techniques in the developed classification methodology. The proposed approach achieves an effective technique to distinguish whether or not there is transformer oil contamination from the OLTC gases using a DT. The algorithm is trained using different percentages of training and test data to obtain the optimal DT.

Based on the proposed approach, a trained classifier is obtained to make it possible to perform the classification automatically. The results of the trained classifier versus the classifications made in this study show an accuracy of 99.76%.

From the classification of transformers as contaminated or uncontaminated performed in this study, the calculation results for the limit values of the C_2H_2 concentration [39] are improved and divided into two groups, depending on the OLTC communication between the main tank and the OLTC compartment, similar to the IEC classification.

In future work, the model trained from the ML techniques will be used with new DGA results. Its accuracy will be checked, and the model will be improved if necessary.

Abbreviations

The following abbreviations are used in this manuscript:

CH_4	Methane
C_2H_2	Acetylene
C_2H_4	Ethylene
C_2H_6	Ethane
D1	Discharges of low energy
D2	Discharges of high energy
DGA	Dissolved gas analysis
DPM	Duval pentagon method
DSO	Distribution system operator
DT	Decision tree
DTM	Duval triangle method
EK	Expert knowledge
FDS	Frequency domain spectroscopy
FFA	Furfuraldehyde
H_2	Hydrogen
HI	Health index
ML	Machine learning
OLTC	On-load tap-changer
OQA	Oil quality analysis
PD	Partial discharges
PDC	Polarisation and depolarisation current
T1	Thermal faults ($<300\ ^\circ$C)
T2	Thermal faults (300–$700\ ^\circ$C)
T3	Thermal faults ($>700\ ^\circ$C)

References

1. Wu, X.; He, Y.; Duan, J. A Deep Parallel Diagnostic Method for Transformer Dissolved Gas Analysis. *Appl. Sci.* **2020**, *10*, 1329. [CrossRef]
2. Bohatyrewicz, P.; Płowucha, J.; Subocz, J. Condition Assessment of Power Transformers Based on Health Index Value. *Appl. Sci.* **2019**, *9*, 4877. [CrossRef]
3. CIGRE. *Ageing High Voltage Substation Equipment and Possible Mitigation Techniques*; WG A3.29. Technical Brochure No. 725; CIGRE: Paris, France, 2018.
4. Azmi, A.; Jasni, J.; Azis, N.; Kadir, M.A. Evolution of transformer health index in the form of mathematical equation. *Renew. Sustain. Energy Rev.* **2017**, *76*, 687–700. [CrossRef]
5. CIGRE. *Advances in DGA Interpretation*; JWG D1/A2.47. Technical Brochure No. 771; CIGRE: Paris, France, 2019.
6. IEEE. *IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers*; IEEE Std C57.104-2019 (Revision of IEEE Std C57.104-2008); IEEE: Piscataway, NJ, USA, 2019; pp. 1–98. [CrossRef]
7. IEC. *Mineral Oil-Filled Electrical Equipment in Service—Guidance on the Interpretation of Dissolved and Free Gases Analysis*; IEC 60599 ed.3.0; IEC: Geneva, Switzerland, 2015.
8. Frotscher, R. Common oil expansion tank for transformers and tap-changers. *Transform. Mag.* **2017**, *4*, 20–25.
9. N'cho, J.; Fofana, I.; Hadjadj, Y.; Beroual, A. Review of Physicochemical-Based Diagnostic Techniques for Assessing Insulation Condition in Aged Transformers. *Energies* **2016**, *9*, 367. [CrossRef]
10. Duval, M.; dePabla, A. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases. *IEEE Electr. Insul. Mag.* **2001**, *17*, 31–41. [CrossRef]
11. Mollmann, A.; Pahlavanpour, B. New guidelines for interpretation of dissolved gas analysis in oil-filled transformers. *Electra* **1999**, *186*, 31–51.

12. Samsudin, R.; Ramli, A.Q.; Berhanuddin, A.; Zaidey, Y. Field experience of transformer untanking to identify electrical faults and comparison with Dissolved Gas Analysis. In Proceedings of the 2009 3rd International Conference on Energy and Environment (ICEE), Malacca, Malaysia, 7–8 December 2009; pp. 299–305. [CrossRef]

13. Jadav, R.B.; Saha, T.K.; Ekanayake, C. Transformer diagnostics using dissolved gas analysis and polarisation and depolarisation current measurements—A case study. In Proceedings of the AUPEC 2011, Brisbane, Australia, 25–28 September 2011; pp. 1–6.

14. Istad, M.K.; Foros, J.; Tveten, E.G.; Kolstad, M.L.; Solvang, E.; Rosenlund, G.H.; Hagner, S.; Henderson, M.G.; Coullon, J.L. *Use Case Testing—SAMBA WP5 Report*; Technical Report; SINTEF Rapport: Trondheim, Norway, 2019.

15. Werle, P.; Radigk, C.; Sorgatz, W.; Hahn, M.; Wasserberg, V. Comparison of Different DGA (Dissolved Gas Analysis) Methods for the Condition Assessment of Power Transformers. In Proceedings of the International Conference Large Power Transformers-Modern Trends in Application, Operation & Maintenance, New Delhi, India, 12–13 October 2006.

16. Alqudsi, A.; El-Hag, A. Application of Machine Learning in Transformer Health Index Prediction. *Energies* **2019**, *12*, 2694. [CrossRef]

17. Qi, B.; Zhang, P.; Rong, Z.; Wang, J.; Li, C.; Chen, J. Rapid Transformer Health State Recognition Through Canopy Cluster-Merging of Dissolved Gas Data in High-Dimensional Space. *IEEE Access* **2019**, *7*, 94520–94532. [CrossRef]

18. Kadim, E.; Azis, N.; Jasni, J.; Ahmad, S.; Talib, M. Transformers Health Index Assessment Based on Neural-Fuzzy Network. *Energies* **2018**, *11*, 710. [CrossRef]

19. Liu, Y.; Song, B.; Wang, L.; Gao, J.; Xu, R. Power Transformer Fault Diagnosis Based on Dissolved Gas Analysis by Correlation Coefficient-DBSCAN. *Appl. Sci.* **2020**, *10*, 4440. [CrossRef]

20. Benhmed, K.; Shaban, K.B.; El-Hag, A. Cost effective assessment of transformers using machine learning approach. In Proceedings of the 2014 IEEE Innovative Smart Grid Technologies—Asia (ISGT ASIA), Kuala Lumpur, Malaysia, 20–23 May 2014; pp. 328–332. [CrossRef]

21. Ashkezari, A.D.; Ma, H.; Saha, T.K.; Ekanayake, C. Application of fuzzy support vector machine for determining the health index of the insulation system of in-service power transformers. *IEEE Trans. Dielectr. Electr. Insul.* **2013**, *20*, 965–973. [CrossRef]

22. Akhavanhejazi, M.; Gharehpetian, G.; Faraji-dana, R.; Moradi, G.; Mohammadi, M.; Alehoseini, H. A new on-line monitoring method of transformer winding axial displacement based on measurement of scattering parameters and decision tree. *Expert Syst. Appl.* **2011**, *38*, 8886–8893. [CrossRef]

23. Guo, C.; Dong, M.; Wu, Z. Fault Diagnosis of Power Transformers Based on Comprehensive Machine Learning of Dissolved Gas Analysis. In Proceedings of the 2019 IEEE 20th International Conference on Dielectric Liquids (ICDL), Roma, Italy, 23–27 June 2019; pp. 1–4. [CrossRef]

24. Huang, X.; Wang, X.; Tian, Y. Research on Transformer Fault Diagnosis Method based on GWO Optimized Hybrid Kernel Extreme Learning Machine. In Proceedings of the 2018 Condition Monitoring and Diagnosis (CMD), Perth, Australia, 23–26 September 2018; pp. 1–5. [CrossRef]

25. Han, Y.; Zhao, D.; Hou, H. Oil-immersed Transformer Internal Thermoelectric Potential Fault Diagnosis Based on Decision-tree of KNIME Platform. *Procedia Comput. Sci.* **2016**, *83*, 1321–1326. [CrossRef]

26. Samantaray, S.; Dash, P. Decision Tree based discrimination between inrush currents and internal faults in power transformer. *Int. J. Electr. Power Energy Syst.* **2011**, *33*, 1043–1048. [CrossRef]

27. Fei, S.-W.; Zhang, X.-B. Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Syst. Appl.* **2009**, *36*, 11352–11357. [CrossRef]

28. Ganyun, L.; Haozhong, C.; Haibao, Z.; Lixin, D. Fault diagnosis of power transformer based on multi-layer SVM classifier. *Electr. Power Syst. Res.* **2005**, *74*, 1–7. [CrossRef]

29. CIGRE. *Condition Assessment of Power Transformers*; WG A2.49. Technical Brochure No. 761; CIGRE: Paris, France, 2019.

30. CIGRE. *DGA in Non-Mineral Oils and Load Tap Changers and Improved DGA Diagnosis Criteria*; WG D1.32. Technical Brochure No. 443; CIGRE: Paris, France, 2010.

31. IEEE. *IEEE Guide for Dissolved Gas Analysis in Transformer Load Tap Changers*; IEEE Std C57.139-2015 (Revision of IEEE Std C57.139-2010); IEEE: Piscataway, NJ, USA, 2016; pp. 1–45. [CrossRef]

32. IEC. *Tap-Changers—Part 2: Application Guide*; IEC 60214-2 ed.1.0; IEC: Geneva, Switzerland, 2004.

33. Frotscher, R. Hermetical sealing of transformers and tap-changers. *Transform. Mag.* **2017**, *4*, 24–29.

34. Duval, M. The duval triangle for load tap changers, non-mineral oils and low temperature faults in transformers. *IEEE Electr. Insul. Mag.* **2008**, *24*, 22–29. [CrossRef]

35. MATLAB. *version 9.5.0.1033004 (R2018b)*; The MathWorks Inc.: Natick, MA, USA, 2019.

36. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Wadsworth and Brooks: Monterey, CA, USA, 1984.

37. Coppersmith, D.; Hong, S.J.; Hosking, J.R.M. Partitioning Nominal Attributes in Decision Trees. *Data Min. Knowl. Discov.* **1999**, *3*, 197–217. [CrossRef]

38. Bustamante, S.; Manana, M.; Arroyo, A.; Martinez, R.; Laso, A. Developed Algorithms—Transformer oil contamination from the OLTC gases. *Mendeley Data* **2020**. [CrossRef]

39. Bustamante, S.; Manana, M.; Arroyo, A.; Martinez, R.; González, A.; Rodríguez, J.I. Case Study—Calculation of DGA Limit Values and Sampling Interval in Power Transformers. In Proceedings of the 2019 6th International Advanced Research Workshop on Transformers (ARWtr), Cordoba, Spain, 7–9 October 2019; pp. 64–68. [CrossRef]