A New Multi-Path Hybrid Classifier for Transformer Oil Fault Diagnostic

Youcef Benmahamed, Omar Kherif, Sofiane Chiheb, Madjid Teguar, Sherif Ghoneim and Ahmed Boubakeur

Abstract—This work aims to provide advances in diagnosis algorithms using intelligent techniques and represents an application in fault detection and classification in oil-immersed power transformers. The paper proposes a new methodology of classification using hybrid algorithms to describe an improved DGA diagnostic tool based on combining different classifiers and several input vectors. A total of six classes of electrical and thermal faults are labeled. For each fault, binary classifications are first conducted using two classifiers trained and evaluated using nine different input vectors. For this, a dataset of 501 samples is used, and the best pairs (classifier, input vector) are selected for each given binary classification. From these pairs, different hybrid classifiers are proposed. Each classifier reaches its outcome through an independent pathway, and these classifiers together form the proposed multi-path hybrid classifier. The final decision of this classifier is obtained from the decisions made at the output of each path. This application brings a global accuracy rate of up to 95% for the transformer oil diagnosis, demonstrating the proposed technique's effectiveness in the classification field. The proposed model and other conventional algorithms are compared using a small independent database of twenty elements..

***Keywords*—DGA, fault diagnosis, power transformer oil, hybrid classifier, SVM, KNN, Decision Tree.**

NOMENCLATURE

DGA	Dissolved Gas Analysis.
KNN	K-Nearest Neighbor.
SVM	Support Vector Machine.
DT	Decision Tree.

I. INTRODUCTION

The continuous and reliable operation of electricity is one of the main challenges of power companies, from generation to transmission and distribution. In this chain, power transformers play an essential role for the target. Numerous studies were conducted to investigate the impact of faults in power transformers as well as to develop and/or improve resolution techniques to prevent such faults (e.g., [1–4]). Studying more than 340 power transformers with a voltage rate from 33 to 400 kV, authors in [5] stated that insulation problems are the most common fault, accounting for 37% of power transformer failures.

Mineral insulating oil is the most common oil used in outdoor transformers. This oil is characterized by its significant dielectric strength to withstand a fairly high voltage. It also helps to reduce heat generated by transformer windings. Indeed, oil-immersed transformers' lifetime can be governed by the state of the insulation system. This latter is generally exposed to some defects arising from overheating, paper carbonization, arcing, and discharges of low or high energy [4, 6, 7]. Evaluation procedures and relevant tests of these oils can be found within the recommendations of [8, 9].

Dissolved Gas Analysis (DGA) method is one of the most effective methods used in the field of faults detection within oil-immersed power transformers [10, 11]. This method analyzes the concentration of gases liberated in the transformer oil. Different hydrocarbon gases are released due to insulating oil and paper decomposition under electrical and thermal stresses.

In general, the most important gases, in alphabetical order, are Acetylene (C_2H_2), Ethane (C_2H_6), Ethylene (C_2H_4), Hydrogen (H_2), and Methane (CH_4). A particular combination of gases characterizes each type of fault within transformer oil.

Some of the application of DGA methods can be summarized as follows; i) DGA identifies different transformer fault types, due to different thermal, electrical, and mechanical stresses on the insulating oil. Each fault produces a specific pattern of gases that can be detected through DGA. ii) DGA detects faults early before leading damage to the transformer, so it helps prevent costly repairs and downtime. iii) Monitoring changes in gas levels over time provides a trend analysis that helps predict future faults and plan maintenance activities accordingly. iv) it provides an overall assessment of the transformer's condition through analyzing various gases and their levels, which helps in determining whether the transformer is operating within normal limits or needs maintenance. v) it also helps to locate the fault within the transformer by analyzing the distribution of gases within different parts of the transformer. So, it is a powerful tool for transformer fault diagnosis [12, 13].

Manuscript received September 28, 2024; revised November 8, 2025.

Y. BENMAHAMED, M. TEGUAR and A. BOUBAKEUR are with Laboratoire de Recherche en Electrotechnique (LRE), Ecole Nationale Polytechnique (ENP), 10 Rue des Frères OUDEK, El Harrach 16200, Algiers, Algeria. (e-mail: youcef.benmahamed@g.enp.edu.dz, madjid.teguar@g.enp.edu.dz, and ahmed.boubakeur@g.enp.edu.dz),

O. KHERIF is with Advanced High Voltage Research Centre, Cardiff University, Cardiff CF24 3AA, UK. (e-mail: KherifO@cardiff.ac.uk).

S. CHIHEB is with National Higher School of Technology and Engineering, Laboratoire de Technologies des Systems Energétiques, 23005, Annaba, Algeria (s.chiheb@ensti-annaba.dz)

S. GHONEIM is with Electrical Engineering Department, College of Engineering, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia. (e-mail: s.ghoneim@tu.edu.sa).

Digital Object Identifier (DOI): 10.53907/enpesj.v5i2.286

The dissolved gases produce under the electrical, thermal, and mechanical stresses. For instance, Ethylene is related to hotspots between 150 C and 1000 C, and acetylene is associated with arcs where temperatures reach a few thousand degrees. Furthermore, partial discharges in transformer oil can result in a considerable increase in hydrogen concentration because of the ionic bombardment [13]. Six classes of faults are widely considered in this field, namely partial discharge (PD), low energy discharges (D1), high energy discharges (D2), thermal faults < 300 C (T1), thermal faults of 300 C to 700 C (T2), and thermal faults > 700 C (T3).

The literature surveys indicated some shortcomings of using DGA for transformer fault diagnosis. In some cases, the dissolved gases amount is not sufficient to refer the exact transformer fault type and it can produce false positives, where the normal aging or environmental factors increase the gas levels producing false results [15]. The nonlinearity of the data samples can also lead to incorrect of transformer fault type. DGA Data sample processes consume time, effort and cost [16]. The accuracy of DGA results influences by other factors such as the quality of the oil sample, the sampling technique used, and the laboratory analysis method [17]. Due to wrong diagnosis of the specific transformer fault types, it is difficult to determine the appropriate maintenance or repair actions [18]. Interpreting DGA results requires specialized knowledge and expertise, which may not be readily available to all users [19].

In literature, different techniques have been developed to diagnose transformer faults. These techniques include graphical DGA methods (e.g., [20, 21]) and intelligent techniques (e.g., [22–24]). In addition, improved techniques (coupled methods) have also been created to accurately diagnose transformer faults (single and/or multiple faults) and indicate each fault's likelihood quantitatively (e.g., [23–26]). Arranging the input data of the DGA methods can affect on the DGA results. Therefore, many researchers have focused on developing input vectors that can enhance the DGA results to diagnose transformer faults correctly [4, 7, 27, 28].

This work describes an advanced classification methodology using a combination of different input vectors and various classifiers, and an application is presented to enhance the transformer fault diagnostic accuracy based on DGA. Furthermore, hybrid algorithms are proposed to improve DGA diagnostic tools. From a total dataset of 501 samples, 481 are used to train and test the proposed models, where six electrical and thermal fault classes are labeled. As main result, it was found that the global accuracy rate, reaching 95% for the power transformer diagnosis, demonstrates the effectiveness of the proposed technique.

The paper is organized as follows: the proposed methodology is presented in a Section to provide a general overview and a reference point for different applications. Section III covers the selected classifiers and input vectors for the transformer oil fault diagnosis. The dataset is also presented in this section. Results and discussions are shown in Section IV. The results obtained and compared with other classifiers are shown at the end of this section. The paper finalizes with conclusions and perspectives.

II. Proposed multi-path hybrid classifier

A. Geeral principal

In [27], the authors proposed several input vectors to train and assess a KNN classification algorithm based on a decision tree principle. The work was conducted to select the best input vector to achieve a high-accuracy diagnosis for the transformer faults. The accuracy rate has been analyzed to choose the most appropriate input vector for the proposed method. The obtained results were fascinating compared to conventional techniques of classification. Therefore, the idea is generalized in this work and introduces a new technique. This technique uses multiple classifiers, and the final decision is determined by an election (e.g., high number of appearances). Figure 1 shows the chart of the proposed technique.

The first step involves selecting an M number of different classifiers according to the desired application. The parameters associated with each classifier should be defined and adjusted for the application. Then, a set of criteria should be established to define the best pair's characterization. For instance, the simplest criterion represents the pair with the highest accuracy rate or an accuracy rate higher than a given value. Finally, a K number of criteria should be selected, and this number will affect the number of paths in the proposed multi-path hybrid classifier.

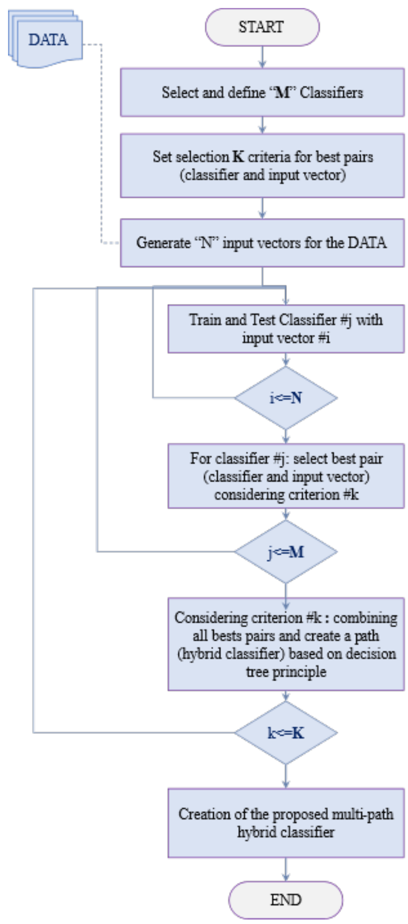


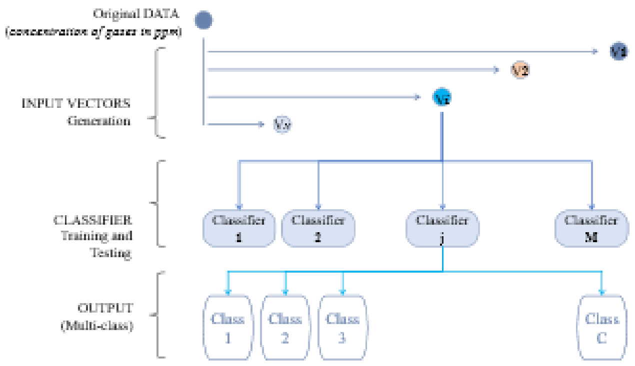
Fig. 1: A general structure of the proposed methodology.

As shown in Figure 1, the next step consists of generating different input vectors from the same dataset. The collected data generates various modified datasets, grouping the M input vectors. Taking into account the transformer oil fault diagnosis, for instance, the original data represents the first input vector, which consists of the concentration of gases in parts per million (ppm). A second input vector may be the relative concentration of each gas compared to the sum of concentrations. It should be noted that the input vector generation is a process that depends on the data where the data transformation should be established to make the data more separable and easier to use in the classification process. A detailed example of the selected input vectors is presented in Section III.

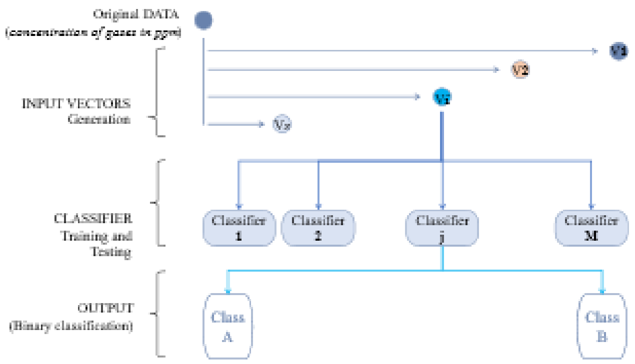
It is important to note that the selected combinations together help with the identification of various possible pathways. Therefore, the combination of these pathways represents the proposed multi-path hybrid classifier where the final decision is obtained from pre-defined criteria as shown in the forthcoming parts.

B. Pairs and paths selection

Each classifier is trained and evaluated by all the selected input vectors using the dataset defined for the testing process. Several techniques can be adopted for this phase, such as a binary classification based on the decision tree principle and multicategory classification. Figure 2 shows two possible scenarios for the classification process.



(a)



(b)

Fig. 2: Best combination identification process:(a) Multiclass condition,(b) Binary condition.

A decision tree principle can be considered in both scenarios, representing a branched flowchart with two or multiple pathways for potential decisions. For a given classifier, the tree starts with a decision node, which implies a decision must be made. For the multiclass scenario in Figure 2(a), a branch is created from the decision node where the obtained decision might successively run to another decision. For the second scenario, the process is simple by applying a "one vs. all" strategy. This means that each dataset is used to train and evaluate a given classifier about two classes of faults denoted by Class 1 and Class 2 as shown in Figure 2(b).

As an application in transformer oil diagnosis, for instance, Class 1 (=Fi) might be used to refer to one of the six faults in transformer oil (i.e., PD, D1, D2, T1, T2, and T3). In this case, Class 2 (=Xi) should represent a complementary class depending on the classification (i.e., Xi = PD, D1, D2, T1, T2, and T3 – Fi). A binary classification is considered by selecting this process, which means two decisions can be obtained on the output of each classifier. It will help identify the appropriate input vector and classifier for binary classification. Figure 3 demonstrates the proposed process to determine the best combination vector-classifier to create paths for the global hybrid classifier.

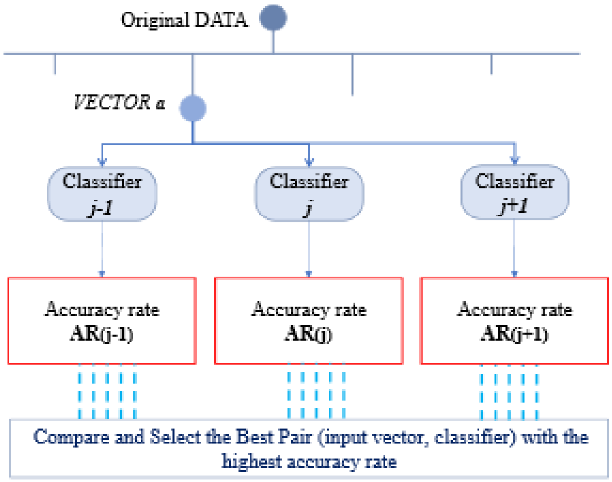


Fig. 3: Best pair selection process.

Figure 3 shows that the training and testing processes are conducted using input vectors applied to various classifiers. Indeed, many tests should be conducted so that the accuracy rate of each binary classification can be obtained as a function of the selected classifier and input vector. Therefore, this process helps identify the possible paths for global classification. Figure 4 illustrates a typical structure of the proposed technique used to determine the global accuracy rate of any path.

The best pairs selection process allows the creation of a hybrid classifier with one possible decision. This hybrid classifier, the denoted path, is not unique; different proposals can be made according to the desired accuracy rate. Eventually, each pathway reaches an outcome by passing through different classifiers and elaborating various input vectors. A combination of these decisions offers a trusted outcome that is generated from the result of the selected pathways, especially once the same outcome is obtained for different paths.

It is worth noting that the proposed multi-path hybrid classifier can be used in various classification problems, including multi-input multi-output classification (MIMO-C) systems. In addition, different scenarios can be generated from the proposed structure according to the output classes for a given situation.

III. INPUT VECTOR AND CLASSIFIER

In this work, nine input vectors have been used to train and test different classifiers, where the best combination (classifier and

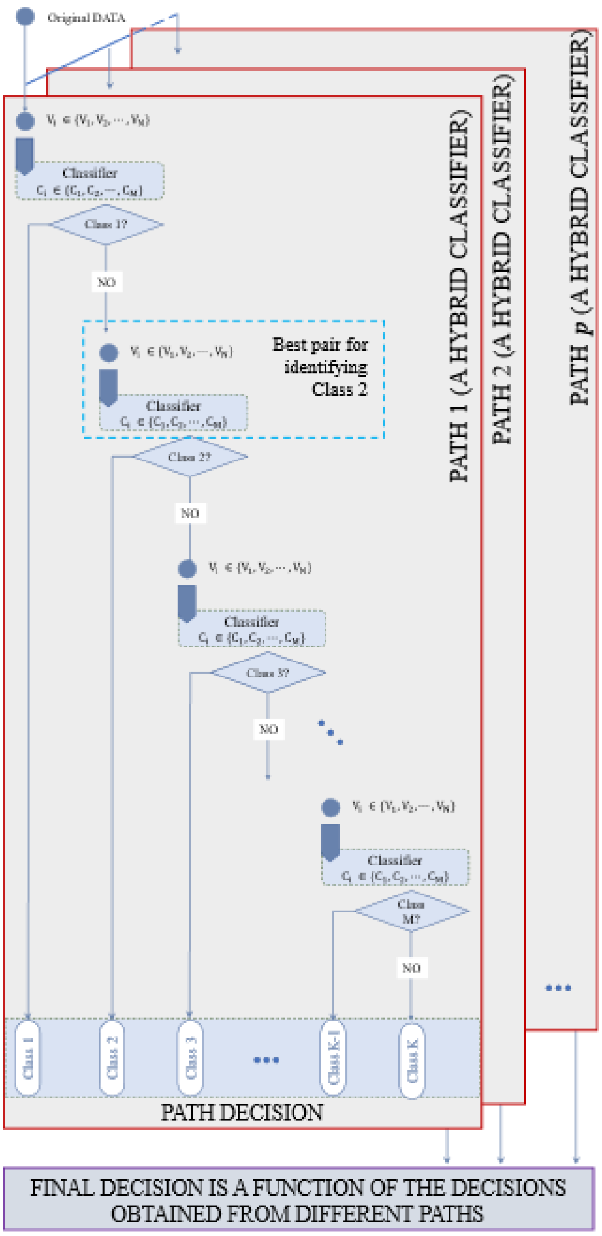


Fig. 4: General structure of the proposed classifier.

input vector) is selected for the considered binary classification. Indeed, pre-classification is first adopted to recognize the first three best combinations (input vector and classifier) that gives the highest accuracy rate where a database of 501 whose 481 samples were reserved the process of training and testing. Table I gives the distribution of the training and testing samples according to their identified fault type. Additional twenty samples (shown in Section IV.) have been used to examine the validity of the proposed classifier.

As the Authors in [18] detailed, the 481 samples were collected from different sources where gas concentration values were observed in service for faulty equipment inspected in service. The database is mainly collected from relevant publications (e.g., [29]) and Egyptian Electric Utility (Report [30]). Data are collected from [31] and [32] for the twenty samples used for validation.

A holdout method is used for the dataset decomposition based on the well-known decomposition (2/3 for training and 1/3 for testing). Although the dataset is unbalanced, the authors randomly assign sample sets for each class by 2/3 in the training phase, and by 1/3 in the testing phase. In this way, one can ensure that the existence of each class in the training and testing processes – proportionally by its initial dimension. This scenario helps with the reduction of the potential risk of each classifier. Therefore, 321 samples have been randomly selected for the training phase and 160 samples for the testing. Among other data mining methods, three classifiers have been considered as follows:

Table. I
DISTRIBUTION OF SAMPLES OVER FAULT TYPES

Symbol	Training	Testing	Total
PD	32	16	48
D1	53	26	79
D2	84	42	126
T1	63	32	95
T2	32	16	48
T3	57	28	85
Total	321	160	481

- Support Vector Machine (SVM)
- K-Nearest Neighbor (KNN)
- Decision tree (DT)

As mentioned previously, a binary classification has been independently developed for all of the classifiers. In order to simplify the comparison process and avoid going inside the algorithms, MATLAB toolboxes have been used for the application. These toolboxes are exploited to use the SVM, KNN, NB, DT classifiers. Details of each are available online on [33–35].

It is well-known that many interpretative methods based on DGA were reported to detect the incipient fault nature within an oil-immersed power transformer. These mainly include, as input vectors, the concentration of the dissolved gases in ppm, relative concentration of gases in percentage, IEC ratios, Rogers four-ratios, Dornenburg ratios, Duval triangle coordinates, Duval pentagon coordinates, a combination of Rogers and Dornenburg ratios, and a combination of Duval triangle-pentagon coordinates as follows:

V1 Concentrations of the gases in parts per million
V2 Percentage to the total sum
V3 IEC ratios
V4 Rogers four-ratios
V5 Dornenburg ratios
V6 Duval triangle coordinates
V7 Duval pentagon coordinates
V8 Rogers and Dornenburg ratios
V9 Duval triangle-pentagon coordinates

It should be noted that more details about these input vectors and their formulation are reported in a previous work [25].

Table. II

CLASSIFICATION RESULTS FOR TWO DIFFERENT CLASSIFIERS FOR 10 SAMPLES OF PD CLASS

SAMPLE	ACTUAL CLASS	PREDICTED CLASS																	
		KNN									SVM								
		V1	V2	V3	V4	V5	V6	V7	V8	V9	V1	V2	V3	V4	V5	V6	V7	V8	V9
PD-1	PD	PD	PD	PD	D1	PD	PD	PD	PD	PD	PD	PD	D2	PD	PD	D1	PD	D2	PD
PD-2	PD	PD	PD	PD	PD	PD	T3	PD	T1	PD	PD	PD	PD	PD	T1	D1	PD	PD	PD
PD-3	PD	PD	PD	PD	PD	PD	T3	PD	PD	PD	PD	PD	PD	PD	PD	T3	PD	D2	PD
PD-4	PD	PD	PD	PD	T1	PD	T1	PD	T1	PD	PD	PD	PD	PD	T3	PD	T3	T2	PD
PD-5	PD	PD	PD	PD	PD	PD	PD	PD	T3	PD	PD	PD	D2	PD	PD	D1	PD	D2	PD
PD-6	PD	PD	PD	PD	PD	T1	T3	PD	T1	PD	PD	PD	PD	PD	PD	D1	PD	D2	PD
PD-7	PD	PD	PD	D1	PD	PD	T3	PD	PD	PD	PD	PD	PD	PD	PD	T3	PD	D2	PD
PD-8	PD	PD	PD	PD	PD	PD	T3	PD	T1	PD	PD	PD	PD	PD	T2	T3	PD	PD	PD
PD-9	PD	PD	PD	PD	PD	PD	T3	PD	T1	PD	PD	PD	PD	PD	PD	T1	PD	D2	PD
PD-10	PD	PD	PD	T1	PD	PD	T2	PD	D2	PD	T1	T1	PD	PD	T3	T3	T3	T1	PD

Table. III

CLASSIFICATION RESULTS FOR TWO DIFFERENT CLASSIFIERS FOR 10 SAMPLES OF D2 CLASS

SAMPLE	ACTUAL CLASS	PREDICTED CLASS																	
		KNN									SVM								
		V1	V2	V3	V4	V5	V6	V7	V8	V9	V1	V2	V3	V4	V5	V6	V7	V8	V9
D2-1	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D1	D2	D2	D2	D2	T1	D2	D2	D2
D2-2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	T3	D2	D2	D2
D2-3	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	T3	D2	D2	D2
D2-4	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D1	D2	D2	D2	D2	D2	D2	D2	D2
D2-5	D2	D2	D2	D2	D2	PD	D2	D2	T1	D2	D2	D2	D2	D2	D2	D2	D2	D2	T1
D2-6	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D1	D2	D2	D2	D2	T3	T2	D2	D2
D2-7	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D1	D2	D2	D2	D2	T3	T2	D2	D2
D2-8	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D1	D2	D2	D2	D2	T1	D2	D2	D2
D2-9	D2	D2	D2	D2	D2	PD	D2	D2	PD	D2	D2	D2	D2	D2	D2	T1	D2	D2	D2
D2-10	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2	D2

IV. Results and discussions

This section provides the obtained results when applying the proposed classification methodology. A binary classification is presented, and the corresponding results are discussed to identify the hybrid classifiers. Secondly, an example of a multi-path hybrid classifier is presented and discussed. Finally, the obtained results along with a comparison with other classifiers are shown.

A. Binary classification and pair selection

The proposed methodology's first step is finding the best pairs (classifier, input vector) from each binary classification. A set of 481 samples has been used for the training and testing stages, two-thirds of the samples were reserved for training phase and one-third for testing utilizing the MATLAB' functions. Table II illustrates a brief description and example of the binary classification results using two different classifiers (KNN and SVM) and considering the nine input vectors (V1 to V9). The classification results in this table consider only the PD class of faults where ten samples are selected arbitrary to examine the existence of best pairs.

From the results obtained, one can clearly see that the classifier selection and the input vector are key factors in the classification purpose. For the same classifier, different decisions are obtained when considering different input vectors. The correct decision is obtained from both classifiers when using the input vector V9. Vectors V1 and V2 allow obtaining 100% accuracy with KNN against 90% with the SVM classifier.

Likewise, one can see that vector V4 is effective with SVM classifier since all obtained decisions are correct, which is not the case when considering the KNN classifier. Regarding correct decisions, the KNN classifier shows a ratio of 70/90 against 63/90 for the SVM classifier. It can be explained by the fact that the KNN classifier is more suitable for such a classification than the SVM and the selected input vectors.

Overall, the results can be summarized in the fact that the pair (classier, input vector) considerably impact the obtained decision of a classification. In order to examine such conclusions, different samples are considered where the same analysis is conducted. Table III gives the obtained results using the two classifiers and considering the nine input vectors for D2 faults.

Compared to the results in Table II, the decisions for the second classification in Table III show a higher accuracy rate when using the D2 fault. It may imply that conclusions on a specific pair (classifier, input vector) applied to a given class of faults cannot be generalized to other classes. For instance, input vector V6 with KNN classifier gives an accuracy rate of 100% for D2 classification against 20% for PD one.

Regarding the SVM classifier, one can see from Tables II and III that input vector V8 is suitable for distinguishing D2 from the PD fault. Therefore, each classification should be associated with one or more pairs (classifier, input vector) to provide a high accuracy classification decision. The testing data consists of 160 samples, which all group the six defined classes of faults. Table IV calculates the accuracy rate for the two classifiers using the nine input vectors.

The results show accuracy rates varying between 0 and 100% for different classifiers and input vectors. For instance, one can identify the pair (KNN, V9) as the best for classifying the PD fault while the pair (SVM, V6) is the worst for the same fault class. Regarding T1 fault, the best pair is (SVM, V2) and the worst is (KNN, V8). Based on the overall results, one can conduct binary classifications considering tree decision principle where the best pair shall be used in each stage to provide the best decision at the end of the classification process.

It is worth noting that the classification can be made with different input vector/classifier combinations. Therefore, not only a single path that could exist, but one can define several paths. Thus, a multi-path hybrid classifier should be considered as described in the following section.

B. Multi-path hybrid classifier

Based on the results, numerous paths can be proposed using combinations between faults classified in Table IV in descending order.

Table. IV
ACCURACY RATE FOR DIFFERENT INPUT VECTORS

	Accuracy rate using KNN (%)					
	PD	D1	D2	T1	T2	T3
V1	93.75	80.77	90.48	87.50	75.00	96.43
V2	93.75	76.92	97.62	87.50	75.00	96.43
V3	75.00	50.00	78.54	93.75	75.00	85.71
V4	62.50	53.85	78.54	71.88	81.25	85.71
V5	87.50	11.54	71.43	03.13	06.25	10.71
V6	37.50	23.08	85.71	78.13	37.50	71.43
V7	93.75	73.08	95.24	90.63	75.00	92.86
V8	18.75	11.54	88.10	00.00	00.00	14.29
V9	100	76.92	90.48	87.50	93.75	92.86
	Accuracy rate using SVM (%)					
V1	87.50	88.46	83.33	87.50	43.75	96.43
V2	93.75	80.77	90.48	96.88	87.50	89.29
V3	93.75	11.54	35.71	71.88	37.50	89.29
V4	25.00	73.08	28.57	90.63	00.00	00.00
V5	75.00	46.15	83.33	75.00	62.75	00.00
V6	00.00	00.00	88.10	84.38	00.00	82.14
V7	87.50	73.08	92.86	90.63	68.75	92.86
V8	68.75	03.85	97.62	71.88	87.50	00.00
V9	93.75	76.92	92.86	93.75	81.25	92.86

The straightforward way is to consider the same pairs for the classification. It means that one can create a multi-path classifier without hybridization in each single path (i.e., the same classifier and input vector for a given path). Eighteen classifications have been considered using two classifiers and nine different input vectors. In this case, the corresponding accuracy rates are summarized in Table V.

Table. V
ACCURACY RATE FOR DIFFERENT CLASSIFICATION PATHS

	ACCURACY RATE (%)	
	KNN	SVM
V1	88.13	83.75
V2	89.38	89.38
V3	77.50	48.75
V4	73.13	33.75
V5	32.50	60.00
V6	61.88	57.50
V7	88.13	86.25
V8	29.39	56.25
V9	89.38	90.00

As can be seen in Table V, some input vectors give good accuracy rates for both classifiers (e.g., V1, V2, V7 and V9) whilst a low accuracy rate is obtained if one changes the classifier for some input vectors (e.g., V3 with KNN and V3 with SVM). The best results in the considered case are obtained for an input vector V9 and SVM classification. In addition, using V2 with both classifiers results in an accuracy rate of 89.38%. The same result is obtained from different paths when the input vector V9 is used with KNN classification. Overall, better results may be obtained if one can study furthermore this hybridization.

From the results in Table V, one can create a multi-path classification where the final decision is taken from those decisions calculated at the output of several paths.

Table. VI
ACCURACY RATE FOR DIFFERENT MULTI-PATH CLASSIFIERS

	SPECIFICATION		ACCURACY
	CLASSIFIER	INPUT	RATE (%)
Path 1	SVM	V9	-
Path 2	SVM	V2	-
Path 3	KNN	V1	-
Multi-path 1–3	-	-	90.00
Path 4	KNN	V7	-
Path 5	KNN	V9	-
Multi-path 1–5	-	-	92.50
Path 6	SVM	V1	-
Path 7	KNN	V2	-
Multi-path 1–7	-	-	91.88
Path 8	SVM	V7	-
Path 9	KNN	V3	-
Multi-path 1–9	-	-	91.88

As a simple example, the number of paths is increased, and the final decision is made for different scenarios. Four different multipath classifiers have proposed where the number of paths is increased from classifier to another as follows:

- Multi-path Classifier 1: Paths 1, 2 and 3.
- Multi-path Classifier 2: Paths 1, 2, 3, 4 and 5.
- Multi-path Classifier 3: Paths 1, 2, 3, 4, 5, 6 and 7.
- Multi-path Classifier 4: Paths 1, 2, 3, 4, 5, 6, 7, 8 and 9.

Multipath classifier 1 groups three different paths (Paths 1, 2 and 3). Path 1 is the classification using SVM classifier with the input vector V9. Path 2 is the classification using the SVM and V2.

In general, the described paths and the obtained accuracy rates are given in Table VI where the final decision for each multipath classifier is selected from the outputs of the different paths.

In this table, the criterion for selecting final decision consists of selecting the decision with higher recurrence, which is obtained from the paths' decisions. In case of equality or difficulties in the selection, the criterion consists of prioritizing the decision of the best path (usually the first one has higher accuracy rate compared to the others).

It should be noted that the selection criteria can affect the output of the multipath classifiers. It could be another paper's subject since different factors may contribute to the final decision. For instance, it was found that the number of paths used in the multipath classifier considerably impacts the final decision. Figure 5 shows the obtained accuracy rate as a function of the number of paths.

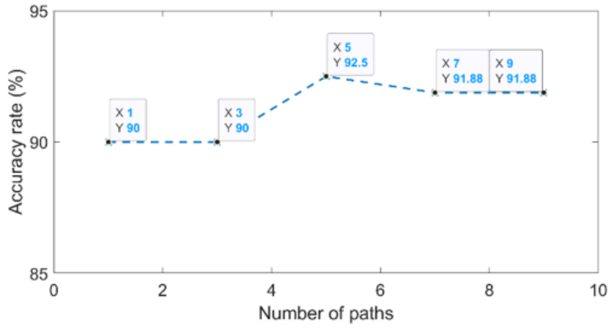


Fig. 5: Accuracy rate as a function of number of paths.

The results indicate that the number of the selected paths is a sensitive factor for creating a multi-path hybrid classifier. A higher number of paths may reduce the effectiveness of the classifiers.

Therefore, an appropriate selection should be adopted where optimal choices of classification algorithm with appropriate input data should be considered to diagnose transformer faults better.

Table. VII
COMPARISON BETWEEN THE PROPOSED ALGORITHM AND OTHER RELEVANT METHODS

	H2	CH4	C2H6	C2H4	C2H2	ACT	IEC-60599	Roger's 4 ratios	Duval	TKNN	TSVM	DT-V4	MPH
1	1230	163	27	692	233	D2	D2	UD	D2	D1	D2	D2	D2
2	120	10	30	25	5	D1	PD	UD	T3	D1	PD	D1	D1
3	3700	6400	2400	10	7690	T2	UD	UD	D1	T2	T3	T1	T2
4	6	2990	29990	67	26076	T1	UD		D1	T1	T1	T1	T1
5	34	21	4	56	49	D2	D2	D2	D2	D2	D2	D2	D2
6	120	140	30	0	120	T1	UD	UD	D1	T1	T2	T1	T1
7	240	17	0	5	40	PD	UD	UD	D1	PD	PD	PD	PD
8	6454	2313	121	6432	2159	D2	D2	UD	D2	D2	D2	D2	D2
9	650	53	20	0	34	PD	UD	PD	D1	PD	PD	PD	PD
10	125	680	290	20	900	T3	UD	UD	D1	T3	T3	T1	T3
11	1076	95	71	231	4	PD	UD	UD	T3	PD	PD	PD	PD
12	140	95	10	80	60	D2	D2	D2	D2	D2	D2	D2	D2
13	300	700	280	36	1700	T3	UD	UD	D1	T3	T1	T3	T3
14	960	4000	1290	6	1560	T2	UD	UD	D1	T2	T1	T2	T2
15	1450	940	211	61	322	T1	UD	UD	D1	T1	T1	T1	T1
16	2500	10500	4790	6	13500	T2	UD	UD	D1	T2	T1	T2	T2
17	305	100	33	541	161	D1	D2	UD	D2	D2	D2	D1	D2
18	796	999	234	31	1599	T3	UD	UD	D1	T3	T1	T3	T3
19	37800	1740	249	8	8	PD	PD	PD	PD	PD	PD	PD	PD
20	33046	619	58	0	2	PD	UD	PD	PD	PD	PD	PD	PD
							5/20	5/20	6/20	18/20	12/20	18/20	19/20
ACCURACY RATE (%)							25	25	30	90	55	90	95

ACT: ACTUAL CLASS OF FAULT UD: UNDETERMINED

Therefore, an appropriate selection should be adopted where optimal choices of classification algorithm with appropriate input data should be considered to diagnose transformer faults better.

C. Independant data and validation

A dataset of 20 new samples tests different classifiers for the validation and comparison stages, including conventional ones (IEC-60566 method, Roger's ratios and Duval triangle).

A total of seven classifiers are considered as shown in Table VII which gives the decisions of different classifiers along with the accuracy rates. The classifier MPH is a multi-path hybrid classifier that combines three paths–TKNN (as best path), TSVM and DT-V4. TKNN is a hybridization between the decision tree and the KNN classifier whilst TSVM is similar hybridization using SVM with Decision tree principle as studied in a previous work [26]. DT-V4 is a classification using DT classification using the input vector V4. It should be noted that the selection here is just an example and an infinite number of choices can be considered.

The results show that the MPH improved the diagnosis results (95 %) compared to other conventional techniques and single-path classifiers. Therefore, the proposed methodology can help classify faults in oil-immersed power transformers by providing a better accuracy rate than conventional diagnostic techniques.

V. Conclusion

A study on transformer oil diagnosis using DGA has been made in this paper using 501 samples to provide an advance in the field. The pioneer classifiers SVM and KNN have been used with different input vectors to understand the pair "classifier, input vector" effect on the diagnostic accuracy. For a given sample, the decision is based on the use of a vote on the results of the two algorithms (SVM and KNN) through the injection of several input vectors. Analyzing the results, classification paths were considered, where multiple paths were combined to form a new multi-path hybrid classifier. This strategy can be more practical for improving the diagnosis of power transformer oil than using the classical way when employing a single classifier and a unique input vector. Enrichment in the input vector crafted the classifier to reduce the percentage of misdiagnosis and the hybridization with another classifier made a strong decision on the state of the sample. Hybridizing several algorithms and input vectors can effectively diagnose the power transformer fault.

References

[1] S. Ghoneim, T.A. Farrag, A.A. Rashed, E.M. El-Kenawy, and A. Ibrahim, "Adaptive Dynamic Meta-Heuristics for Feature Selection and Classification in Diagnostic Accuracy of Transformer Faults," IEEE Access, vol. 9, 2021, pp. 78324–78340, doi: 10.1109/ACCESS.2021.3083593.

[2] Recent Developments in DGA Interpretation, CIGRE Technical Brochure 296, Paris, France, Jun. 2006.

[3] E. Altayef, F. Anayi, M. S. Packianather and O. Kherif, "On the Effects of Lamination Artificial Faults in a 15 kVA Three-Phase Transformer Core," IEEE Access, vol. 10, pp. 19348-19355, 2022, doi: 10.1109/ACCESS.2022.3151367.

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

[5] M. Kezunovic and Y. Guo, "Modeling and simulation of the power transformer faults and related protective relay behavior," IEEE Trans. Power Del., vol. 15, no. 1, pp. 44–50, Jan. 2000.

[6] X. Ji, Y. Zhang, and Q. Liu, "Insulation Condition Assessment of Power Transformers Employing Fused Information in Time and Space Dimensions, Electric Power Components and Systems", Electric Power Components and Systems, vol. 48, no. 1-2, pp. 213-223, Mar. 2020.

[7] Y. Benmahamed, O. Kherif, M. Teguar, A. Boubakeur and S. Ghoneim, "Accuracy Improvement of Transformer Faults Diagnostic Based on DGA Data Using SVM-BA Classifier", MDPI Energies, vol. 14, no. 10, pp. 2970, 2021, doi: 10.3390/en14102970.

[8] Mineral Oil-Impregnated Electrical Equipment in Service—Guide to the Interpretation of Dissolved and Free Gases Analysis, IEC Publication 60599, 2007.

[9] "IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers," in IEEE Std C57.104-2019 (Revision of IEEE Std C57.104-2008), vol., no., pp.1-98, 1 Nov. 2019.

[10] M. Duval, 'Dissolved gas analysis: It can save your transformer,' IEEE Elect. Insul. Mag., vol. 5, no. 6, pp. 22–27, Nov./Dec. 1989.

[11] K. Kaur, D. Bhalla and J. Singh, "Fault Diagnosis for Oil Immersed Transformer Using Certainty Factor," in IEEE Transactions on Dielectrics and Electrical Insulation, vol. 31, no. 1, pp. 485-494, Feb. 2024, doi: 10.1109/TDEI.2023.3307513.

[12] T. Manoj, C. Ranga, Sherif S. M. Ghoneim, U. Mohan Rao, Saad A. Mohamed Abdelwahab, "Alternate and Effective Dissolved Gas Interpretation to Understand the Transformer Incipient Faults", IEEE Transactions on Dielectrics and electrical insulation, Jan. 2023.

[13] Rahman Azis Prasojo, Muhammad Akmal A Putra, Meyti Eka Apriyani, Anugrah Nur Rahmanto, Sherif SM Ghoneim, Karar Mahmoud, Matti Lehtonen, Mohamed MF Darwish, "Precise transformer fault diagnosis via random forest model enhanced by synthetic minority over-sampling technique", Electric Power Systems Research, Volume, 220, Pages 109361, 2023.

[14] N. Bakar, A. Abu-Siada, and S. Islam, "A review of dissolved gas analysis measurement and interpretation techniques," IEEE Elect. Insul. Mag., vol. 30, no. 3, pp. 39–49, May/Jun. 2014.

[15] Prasojo, R.A.; Gumilang, H.; Suwarno; Maulidevi, N.U.; Soedjarno, B.A. A Fuzzy Logic Model for Power Transformer Faults' Severity Determination Based on Gas Level, Gas Rate, and Dissolved Gas Analysis Interpretation. Energies 2020, 13, 1009. https://doi.org/10.3390/en13041009.

[16] Sherif S. M. Ghoneim, Karar Mahmoud, Matti Lehtonen, And Mohamed M. F. Darwish, "Enhancing Diagnostic Accuracy of Transformer Faults Using Teaching-learning-Based Optimization", IEEE Access, VOLUME 9, 2021, pp. 30817- 30832.

[17] Ward, S.A.; El-Faraskoury, A.; Badawi, M.; Ibrahim, S.A.; Mahmoud, K.; Lehtonen, M.; Darwish, M.M.F. Towards Precise Interpretation of Oil Transformers via Novel Combined Techniques Based on DGA and Partial Discharge Sensors. Sensors 2021, 21, 2223. https://doi.org/10.3390/s21062223.

[18] B. Zhao, M. Yang, H.R. Diao, B. An, Y.C. Zhao, Y.M. Zhang, A novel approach to transformer fault diagnosis using IDM and naive credal classifier, International Journal of Electrical Power Energy Systems, Volume 105, 2019, Pages 846-855.

[19] Abu-Siada, A. Improved Consistent Interpretation Approach of Fault Type within Power Transformers Using Dissolved Gas Analysis and Gene Expression Programming. Energies 2019, 12, 730. https://doi.org/10.3390/en12040730

[20] M. Duval, “A review of faults detectable by gas-in-oil analysis in transformers,” in IEEE Electrical Insulation Magazine, vol. 18, no. 3, pp. 8-17, May-June 2002.

[21] M. Duval and L. Lamarre, “The Duval Pentagon- Anew Complementary Tool for the Interpretation of Dissolved Gas Analysis in Transformers,” IEEE Elec. Insul. Mag., Vol. 30, No. 6, pp. 9-12, 2014.

[22] M. Zhang and W. Chen, “Fault Diagnosis of Power Transformer Based on SSA—MDS Pretreatment,” in IEEE Access, vol. 10, pp. 92505-92515, 2022.

[23] J. Li, G. Li, C. Hai, and M. Guo, “Transformer Fault Diagnosis Based on Multi-Class AdaBoost Algorithm,” in IEEE Access, vol. 10, pp. 1522-1532, 2022.

[24] H. Zhou, L. Lu, G. Wang and Z. Su, "A New Validity Detection Method of Online Status Monitoring Data for Power Transformer," in IEEE Access, vol. 12, pp. 16095-16104, 2024, doi: 10.1109/ACCESS.2024.3358399.

[25] I. B. M. Taha, A. Hoballah and S. S. M. Ghoneim, “Optimal ratio limits of rogers’ four-ratios and IEC 60599 code methods using particle swarm optimization fuzzy-logic approach,” in IEEE Transactions on Dielectrics and Electrical Insulation, vol. 27, no. 1, pp. 222-230, Feb. 2020.

[26] Y. Wu, X. Sun, Y. Zhang, X. Zhong and L. Cheng, “A Power Transformer Fault Diagnosis Method-Based Hybrid Improved Seagull Optimization Algorithm and Support Vector Machine,” in IEEE Access, vol. 10, pp. 17268-17286, 2022.

[27] O. Kherif, Y. Benmahamed, M. Teguar, A. Boubakeur and S. S. M. Ghoneim, “Accuracy Improvement of Power Transformer Faults Diagnostic Using KNN Classifier with Decision Tree Principle,” in IEEE Access, vol. 9, pp. 81693-81701, 2021.

[28] Sherif S. M. Ghoneim, Ibrahim B. M. Taha, Rizk Fahim, Saad A. Mohamed Abdelwahab, “Effect of Data Transformation on the Diagnostic Accuracy of Transformer Faults and the Performance of the Supervised Classifiers”, International Journal of Renewable Energy Research, Vol.12, No.2, June, 2022.

[29] M. Duval, and A. DePablo, "Interpretation of turn-to-turn insulation fault by dissolved gas analysis," in IEEE Transactions on Dielectrics and Electrical Insulation, vol. 25, no. 4, pp. 1560-1566, August 2018, doi: 10.1109/TDEI.2018.007477.

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

[31] J. Dukarm and F. Jakob “Thermodynamic estimation of transformer fault severity” IEEE/PES Transmission and Distribution Conference and Exposition (TD), Dallas, TX, USA, pp. 1-1, 2016

[32] J. Soni and D. Suthar “An Experimental Analysis to Check Accuracy of DGA Using Duval Pentagonal Method in Power Transformer”, ICRISET2017 the Int. Conf. on Research and Innovations in Science, Engineering and Technology, vol 1, pp. 394–401, Aug. 2017

[33] https://www.mathworks.com/help/stats/fitcsvm.html

[34] https://www.mathworks.com/help/stats/fitcknn.html

[35] https://www.mathworks.com/help/stats/fitctree.html

VI. Biography

YOUCEF BENMAHAMED received the engineer’s and master’s degrees in power electronics engineering, in 2014, and the Ph.D. degree in high voltage techniques from the École Nationale Polytechnique (ENP) of Algiers, in 2019. He is currently an associate professor with the ENP. His research interests include diagnosis, artificial intelligence, and optimization techniques.

OMAR KHERIF Omar obtained a M/Eng. and a Ph.D. in electrical engineering from the Ecole Nationale Polytechnique (ENP, Algiers) in 2015 and 2019, respectively. Currently, he holds the position of Discipline Manager (Earthing Lightning Protection Systems) at RINA Tech UK Limited. He was the Head of Design Services with Kingsmill Industries (UK) Limited, following the successful completion of a KTP Project involving Cardiff University and Kingsmill Industries (UK) Limited. Dr Kherif is an active member of IEEE, IEEE SIGHT, GEL/81(BSI), IET, CNHT, ICAEE, CISTEM and EMTP/ATP-EUG. His main interests are in high voltage engineering, electromagnetic transients, earthing and lightning protection solutions.

SOFIANE CHIHEB received his M.S. degree from the Badji Mokhtar University of Annaba, in 2013 and his Ph. D. degree in electrical engineering from the Ecole Nationale Polytechnique (ENP), in 2018. He is currently an associate professor with the National Higher School of Technology and Engineering (ENSTI) Annaba - Algeria. His research interests include high voltage techniques, grounding systems, power grids and renewable energies.

MADJID TEGUAR received the degree in electrical engineering, the master’s degree, and the Ph.D. degree in high voltage engineering from the École Nationale Polytechnique (ENP) of Algiers, in 1990, 1993, and 2003, respectively. He is currently a Professor of Electrical Engineering with ENP. His research interests include insulation systems, insulation coordination, earthing of electrical energy systems, and polymeric cables insulation.

SHERIF S. M. GHONEIM received the B.Sc. and M.Sc. degrees from the Faculty of Engineering at Shoubra, Zagazig University, Egypt, in 1994 and 2000, respectively, and the Ph.D. degree in electrical power and machines from the Faculty of Engineering, Cairo University, in 2008. Since 1996, he has been teaching with the Faculty of Industrial Education, Suez Canal University, Egypt. From 2005 to 2007, he was a Guest Researcher with the Institute of Energy Transport and Storage (ETS), University of Duisburg–Essen, Germany. He joined Taif University as an Associate Professor at the Electrical Engineering Department, Faculty of Engineering. His research interests include grounding systems, dissolved gas analysis, breakdown in SF6 gas, and AI technique applications.

AHMED BOUBAKEUR was born in Biskra, Algeria, in 1952. He received the engineer’s degree in electrical engineering from the École Nationale Polytechnique (ENP) of Algiers, in 1975, and the Ph.D. degree in technical sciences from the Institute of High Voltage Engineering, Warsaw University of Technology, Poland, in 1979. He is currently a Professor with the ENP of Algiers, where he has been giving lecturers and supervising research in the field of high voltage engineering, since 1982. His research interests include discharge phenomena, insulators pollution, lightning, polymeric cables insulation, transformer oil ageing, neural networks and fuzzy logic application in HV insulation diagnosis, and electric field calculation and measurement. He is also a member of IEEE DEIS and the Algerian HV Power Systems Association ARELEC (National Algerian Comity of CIGRE and ENP Elders Association ADEP). He has been a member of the Editorial Board and an Associate Editor of IET Science, Measurement & Technology (SMT).