ABUAD Journal of Engineering Research and Development (AJERD)
ISSN (online): 2645-2685; ISSN (print): 2756-6811

Volume 7, Issue 1, 56-73

Detection of Incipient Faults in Power Transformers using Fuzzy Logic and Decision Tree Models Based on Dissolved Gas Analysis

Felix OLOWOLAFE [1], Kehinde Olukunmi ALAWODE[2]

[1]*System Operations Department, Osogbo Area Control Centre, Osogbo, Transmission Company of Nigeria, Nigeria*
olowofelix1982@yahoo.com

[2]*Department of Electrical and Electronic Engineering, Osun State University, Osogbo, Nigeria*
alawodekehinde@uniosun.edu.ng

Corresponding Author: olowofelix1982@yahoo.com, +2348036198789
Date Submitted: 09/01/2024
Date Accepted: 15/03/2024
Date Published: 01/04/2024

Abstract: *This paper proposes an integrated approach utilizing Fuzzy Logic and Decision Tree algorithms to diagnose early-stage faults in power transformers based on Dissolved Gas Analysis (DGA) test results of transformer insulation oil. Overcoming limitations in conventional methods such as Duval Triangle, Key Gas Analysis, Rogers Ratio, IEC Ratio, and Doernenburg Ratio, our Fuzzy Logic and Decision Tree models address issues like inaccurate diagnosis, inconsistent diagnosis, lack of decisions or out-of-code results, and time-intensive manual calculations for large DGA datasets. The Decision Tree algorithm, a machine learning technique is applied to categorize faults into thermal and electrical types. Trained with over 300 DGA samples from transformers with known faults, the models exhibit robust performance during testing with different datasets. Notably, the Duval Triangle decision tree model attains the highest accuracy among the ten developed models, achieving a 98% accuracy rate when tested with 50 samples with known faults. Moreover, Decision Tree models for KGA, Doernenburg, Rogers, and IEC also demonstrate substantial prediction accuracy at 92%, 86%, 92%, and 90% respectively underscoring the efficacy of artificial intelligence methods over traditional approaches.*

Keywords: *Dissolved Gas Analysis, Decision Tree, Fuzzy Logic, Membership Function, Incipient Faults, Conventional Methods.*

1. INTRODUCTION

Power transformers play a crucial role in power systems by enabling efficient long-distance power transmission with minimal losses [1]. Operational failures in transformers can result in inconvenience for consumers and economic losses. Swift detection of incipient faults in transformers is essential for scheduling preventive maintenance. Common faults such as thermal, arcing, and electrical discharge issues lead to contamination or deterioration of the transformer's insulation paper and mineral oil [8] often due to aging or bad operating conditions. Various diagnostic methods including Partial Discharge Measurements, Frequency Response Analysis, Furan Analysis, Thermal Monitoring, Dissolved Gas Analysis and Capacitance Measurement are available. Among these, Dissolved Gas Analysis (DGA) is widely recognized as the most suitable for fault detection in oil-filled electrical equipment. It measures the concentration of dissolved gases in mineral oil. DGA reveals potential issues through the detection of gases like CO, CH_4, H_2, C_2H_4, CO_2, C_2H_2, and C_2H_6 which result from the molecular decomposition of transformer oil and insulation paper [1]. DGA provides early warnings, diagnoses and actionable insights by assessing internal conditions. Forces inside the transformer such as thermal, electrical and mechanical stresses lead to the evolution of gases that dissolve in the oil. DGA detects individual gas concentrations in parts per million (ppm) allowing for analysis, diagnosis, and fault identification [9]. One challenge in DGA is interpreting Gas Chromatographic results. Conventional methods like Key Gas Analysis (KGA) calculate total dissolved combustible gases and rely on preset rules and standard tables for fault identification. For instance, high hydrogen gas (H_2) percentages indicate partial discharge (Table 1). However, traditional methods are unreliable, may yield inconsistent results and involve time-consuming manual calculations. This work introduces advanced methodologies including Fuzzy Logic and Decision Tree techniques to enhance reliability and speed in power transformer fault detection using DGA.

DRM employs six gases and four ratios as specified in Table 2 for diagnosing thermal and electrical faults. The diagnostic procedure entails confirming that the concentration of at least one fault gas in a ratio surpasses the required L1 concentration limit [1], [10]. As a concentration ratio method its conventional version may yield inconsistent diagnoses when compared to other ratio methods due to variations in ratio ranges. Therefore, there is a necessity to formulate an artificial intelligence model for enhanced accuracy.

Table 1: Empirical diagnostic table for KGA method.

Fault	Key Gas	Quantity of Fault Gas
Thermal fault in Oil	C_2H_4	High percentage of Ethylene gas, noticeable amount of Ethane gas
Thermal fault in Cellulose	CO	Mainly Carbon monoxide gas
Partial discharge or corona	H_2	High percentage of Hydrogen, trace amount of Methane
Arcing	C_2H_2 and H_2	Mainly Acetylene and Hydrogen gases

Table 2: Doernenburg ratio method

Fault	DRM_1=	DRM_2=	DRM_3= C_2H_2/CH_4	DRM_4=
Arcing	$0.1<DRM_1<1.0$	$DRM_2>0.75$	$DRM_3>0.3$	$DRM_4<0.4$
Partial Discharge	$DRM_1<1.0$	Negligible	$DRM_3<0.3$	$DRM_4>0.4$
Thermal Faults	$DRM_1>1.0$	$DRM_2<0.75$	$DRM_3<0.3$	$DRM_4>0.4$

Source: [13]

Similar to DRM, Roger's Ratio Method (RRM) utilizes ratios for fault diagnosis, emphasizing specific ranges over absolute concentrations of fault gases. The DGA analysis involves two tables: one for code definition and the other for fault detection rules. Consult Table 3 and Table 4 for the applicable code and fault detection rules in DGA analysis using Roger's Ratio Method. The traditional version of RRM may occasionally fail to provide a diagnosis decision due to incomplete code combinations in Table 4. This issue arises from the incomplete ratio range for each code in Table 3 attributed to its crisp nature. Consequently, there is a necessity to explore the benefits of fuzzy logic by incorporating fuzzification of the ranges.

Table 3: Code definition for Rogers ratio method

Gas Ratio	Ratio Range	Code
RG1= CH_4/H_2	RG1<0.1	5
	0.1≤RG1≤1	0
	1≤RG1≤3	1
	RG1>3	2
RG2= C_2H_6/CH_4	RG2<1	0
	RG2≥1	1
RG3= C_2H_4/C_2H_6	RG3<1	0
	1≤RG3≤3	1
	RG3>3	2
RG4= C_2H_2/C_2H_4	RG4<0.1	0
	0.1≤RG4≤3	1
	RG4>3	2

Source: [3]

Table 4: Rogers fault diagnosis table

RG1	RG2	RG3	RG4	Detected Faults
0	0	0	0	Normal
5	0	0	0	Partial Discharge (PD)
1 or 2	0	0	0	Thermal<150C
1 or 2	1	0	0	Thermal 150-200C
0	1	0	0	Thermal 200-300C
0	0	1	0	Excessively Heated Winding
1	0	1	0	Confined current in Winding
1	0	2	0	Confined current in core & tank
0	0	0	1	Low intensity discharge

Table 4: Rogers fault diagnosis table (cont'd)

RG1	RG2	RG3	RG4	Detected Faults
0	0	1 or 2	1 or 2	Arcing
0	0	2	2	Floating Potential Sparking
5	0	0	1 or 2	Tracking Partial discharge

Source: [3]

The IEC Ratio Method which is based on the International Electrotechnical Commission Standard 60599 [5] improves the Rogers ratio method by eliminating the Ethane/Methane concentration ratio due to its temperature sensitivity within a limited range of decomposition [11]. This method, derived from the Rogers ratio method, omits the C_2H_6/CH_4 ratio which has been shown to represent a limited temperature range of decomposition potentially leading to inaccurate, no-decision or out-of-code diagnoses.

Table 5: IEC ratio codes

Concentration Ratios	Range	Code
	$IEC_1<0.1$	0
$IEC_1 = C_2H_2/C_2H_4 = R_1$	$0.1 \leq IEC_1 \leq 3.0$	1
	$IEC_1>3.0$	2
	$IEC_2<0.1$	1
$IEC_2=CH_4/H_2 = R_2$	$0.1 \leq IEC_2 \leq 1.0$	0
	$IEC_2>1.0$	2
	$IEC_3<1.0$	0
$IEC_3=C_2H_4/C_2H_6 = R_3$	$1.0 \leq IEC_3 \leq 3.0$	1
	$IEC_3>3.0$	2

Source: [5]

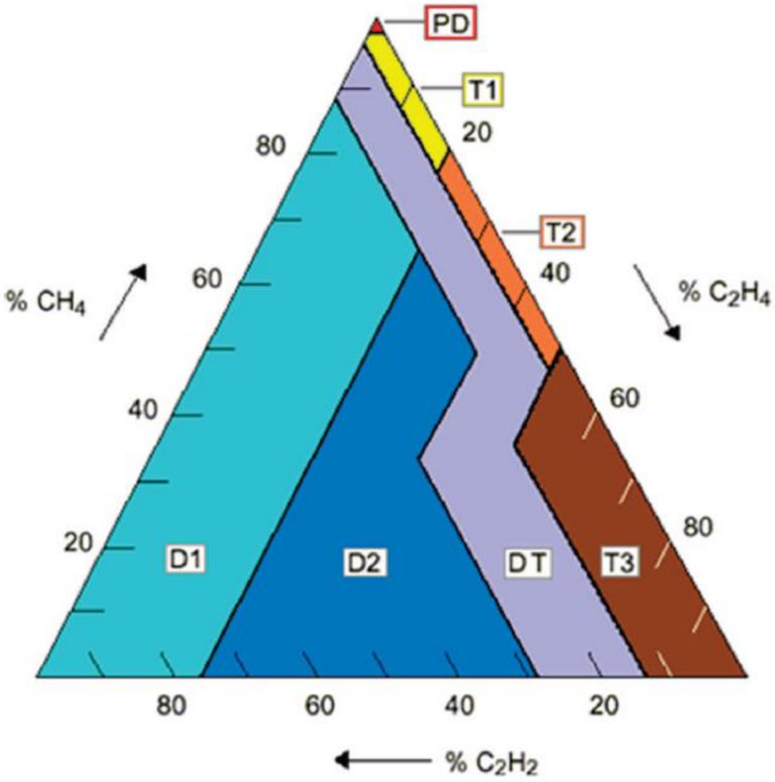
Table 6: IEC fault diagnosis table

R1	R2	R3	Diagnosed faults
0	0	0	No fault
0	1	0	Low Intensity Corona
1	1	0	High Intensity Corona
1or 2	0	1or 2	Sparking
1	0	2	Arcing
0	0	1	Overheating<150°C
0	2	0	Overheating 150-300 °C
0	2	1	Overheating 300-700°C
0	2	2	Overheating>700°C

Source: [5]

The Duval Triangle Method utilizes graphical representation in Dissolved Gas Analysis (DGA) for early detection of faults in transformers. Acetylene (C_2H_2), methane (CH_4) and ethylene (C_2H_4) percentage concentrations are represented on the three sides of a triangle, calculated based on the gas proportions in the mixture. Figure 1 illustrates this method in DGA analysis presenting seven fault zones. Compared to ratio-based approaches, the Duval Triangle Method demonstrates superior precision in fault detection in power transformers. However, being a graphical interpretation method it can be time-consuming for a large number of transformers necessitating the development of its AI model.

This study shows that applying Decision Tree to the Duval triangle method can create an effective AI model capable of detecting all known incipient faults in power transformers like electrical, thermal and discharges with higher sensitivity and accurate precision. Also, the decision tree models outperforming its fuzzy logic counterparts for all the five conventional methods as seen in Table 13. Similarly, in the household of the Fuzzy Logic, Fuzzy Duval Triangle is the most effective model in the detection of electrical faults like arcing and sparking and also thermal faults of different temperature range while other fuzzy models are effective in detecting thermal faults. To this end, one can reliably use the output of the fuzzy models in their respective area of good fault detection and sensitivity to further train the decision tree models, being a machine learning technique that relies on good training data for 100% accuracy in the long run.



PD	Partial Discharge
D1	Discharges of low energy
D2	Discharges of high energy
DT	Combination of thermal faults and discharges
T1	Thermal faults not exceeding 300°C
T2	Thermal faults exceeding 300°C but not exceeding 700°C
T3	Thermal faults exceeding 700°C

Figure 1: Duval triangle fault zones. Source: [12]

In the following sections, the methodologies through which the models were developed are discussed for a few of the conventional methods. The models were tested with 50 DGA dataset with known faults and run on a CORE i5 10TH GEN Laptop computer. The results obtained were tabulated and presented graphically. For each fault condition, a model's precision and recall were calculated and discussed. Conclusions were then drawn from the results and recommendations were made.

2. METHODOLOGY

In this study, fuzzy logic and decision tree models were implemented for each of the five conventional methods. Decision tree is a machine learning approach based on a supervised learning.

2.1 Fuzzy Logic Method Overview

Fuzzy logic is employed for incipient fault characterization due to its capability to manage uncertainty and imprecision, allows continuous output between 0 and 1 based on input values. The methodology is implemented using MATLAB's fuzzy logic toolbox and utilizes low, medium, and high membership functions for seven gases. Input variables are fuzzified and a rule base is established with the output representing the fault condition. The final output is obtained through center-of-gravity (COG) defuzzification.

2.2 Fuzzy Logic for IEC Method

The Simulink circuit for the IEC Method (Figure 2) includes gas concentrations in parts per million (ppm) for five gases CH_4, H_2, C_2H_6, C_2H_2 and C_2H_4 obtained after carrying out DGA test on a suspected faulty transformer's oil sample which serves as input for the Simulink circuit and processed through dividers and a Bus selector according to Table 5 rules. The resulting values serve as input for the fuzzy logic system, and the output display provides values based on defined rules and membership functions. According to Table 7, the fault condition inside the transformer is 'Sparking' which has value range of 2.8, 3, and 3.2. This technique employs three input membership functions represented by variables R1 to R3, with fault conditions defined as the output membership function. In the fuzzy logic system, three inputs and one output are specified. Refer to Table 7 for details on the membership functions assigned to the input and output variables, where L stands for low, M is medium and H is high.

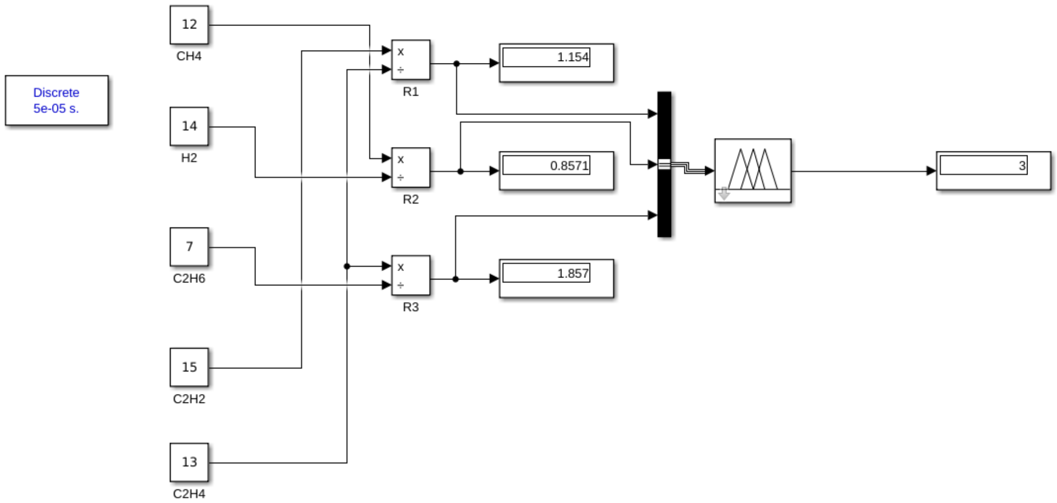
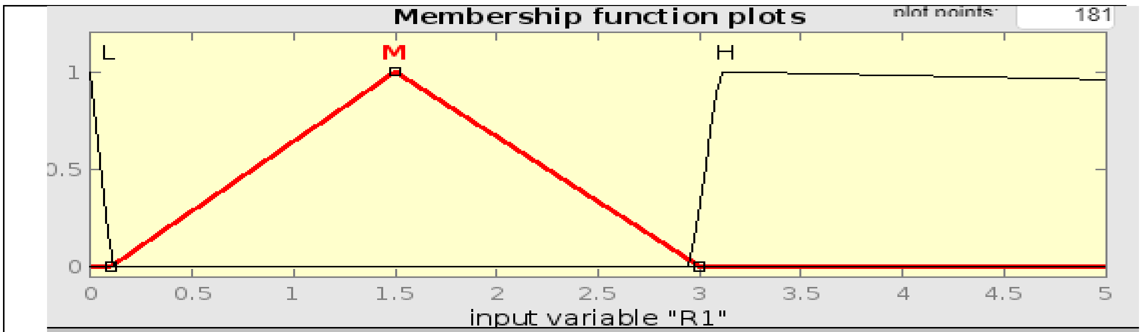


Figure 2: MATLAB simulink of IEC method

Table 7: Membership functions for IEC method

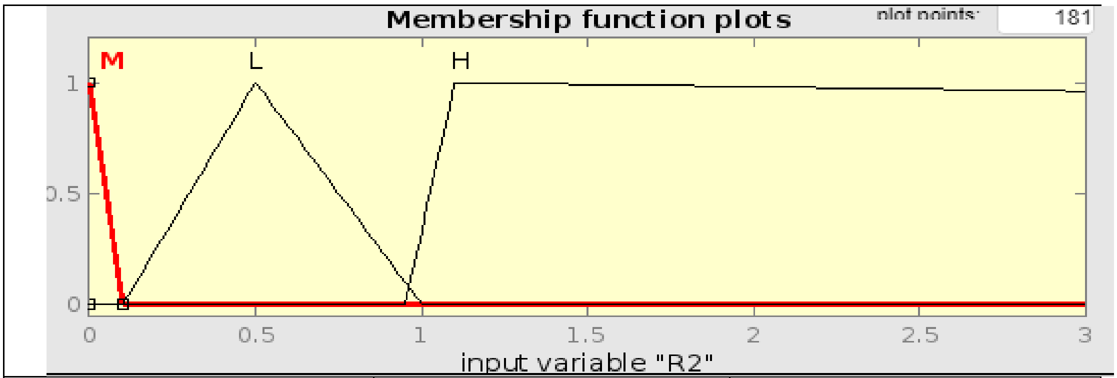
Inputs	Membership Functions	Range
R1	L	[-10 0 0.1]
	M	[0.1 1.5 3]
	H	[2.95 3.1 50]
R2	L	[0.1 0.5 1]
	M	[-100 0 0.1]
	H	[0.95 1.1 50]
R3	L	[-100 0.9 1.1]
	M	[1 2 3]
	H	[2.95 3.1 50]
Output	Normal	[0.8 1 1.2]
	Partial Discharge	[1.8 2 2.2]
	Sparking	[2.8 3 3.2]
	Arcing	[3.8 4 4.2]
	Thermal Fault (<150C)	[4.8 5 5.2]
	Thermal Fault 150 - 300C)	[5.8 6 6.2]
	Thermal Fault (>300C)	[6.8 7 7.2]

The plots for the membership functions of inputs R1 and R2 of IEC method are given in Figure 3 and 4.



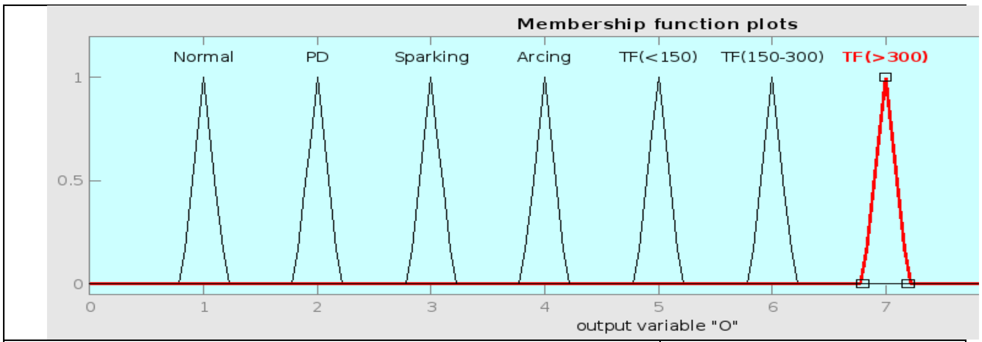
L	M	H
[-10 0 0.1]	[0.1 1.5 3]	[2.95 3.1 50]

Figure 3: Membership function of R1 in IEC method



L	M	H
[0.1 0.5 1]	[-100 0 0.1]	[0.95 1.1 50]

Figure 4: Membership function of R2 in IEC method



Normal	[0.8 1 1.2]
Partial Discharge	[1.8 2 2.2]
Sparking	[2.8 3 3.2]
Arcing	[3.8 4 4.2]
Thermal Fault (<150C)	[4.8 5 5.2]
Thermal Fault 150 - 300C)	[5.8 6 6.2]
Thermal Fault (>300C)	[6.8 7 7.2]

Figure 5: Membership Function of output variable in IEC Method

Utilizing membership functions, ten rules are formulated to generate an output for each input condition. Figure 6 displays the rules editor for the IEC method, and Figure 7 illustrates the rule viewer.

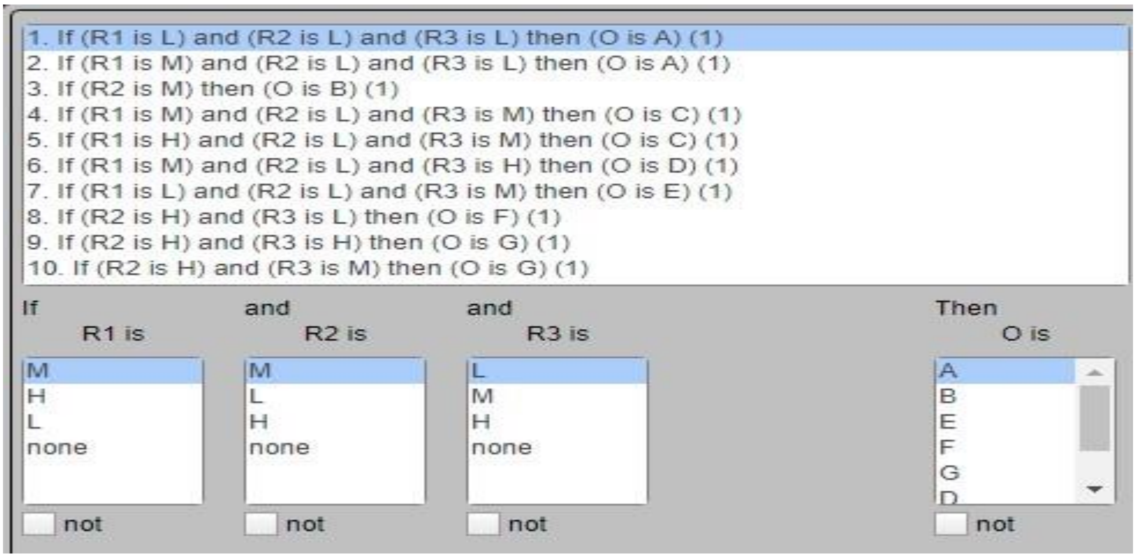


Figure 6: Fuzzy logic rules for IEC method

From the Above rules in figure 6 executed by the Fuzzy Inference System (FIS), rule number one is calculated as follows: Rule 1: If Ratio 1 (R1) is LOW and Ratio 2 (R2) is LOW and Ratio 3 (R3) is LOW, THEN Output (O) is A where A has been designated for a particular fault condition with value range in the output membership functions as shown in figure 5.

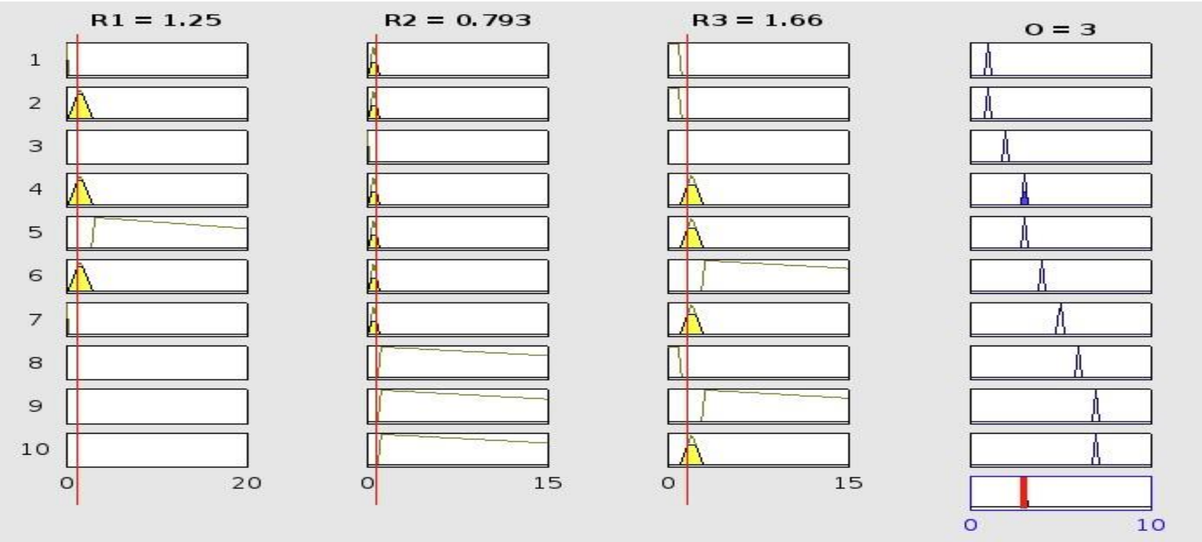


Figure 7: Rule viewer for IEC method

Fuzzy logic models for Rogers, Doernenburg, KGA and Duval Triangle were also developed using the same process as that of IEC above.

2.3 Decision Tree-Based Machine Learning Method

Utilizing the decision tree algorithm for DGA data analysis enables predictions of the transformer's internal health status. The process involves:

- **i.** Data Collection: Gather DGA test data from transformers with known fault conditions, sourced from the IEC TC 10 database.
- **ii.** Feature Selection: Identify and select relevant features with a strong correlation to fault diagnosis from the dataset.
- **iii.** Data Split: Divide the dataset into two sets for training and testing the model. The training set constructs the decision tree, while the testing set evaluates its performance.
- **iv.** Training the Decision Tree: Construct the decision tree using the training set, with the algorithm determining optimal split points for features.
- **v.** Evaluate the Decision Tree: Assess performance using the testing set, comparing predictions with actual fault diagnoses. Common metrics include accuracy, precision, and recall.
- **vi.** Predict Fault Diagnosis: Once trained and evaluated, the decision tree can predict fault diagnoses for new transformers based on DGA results. Inputting DGA data into the decision tree yields predicted fault diagnoses, as illustrated in Figure 8.

3. RESULTS AND DISCUSSION

Fuzzy Logic and decision tree models were implemented in MATLAB/Simulink on an Intel Core i5 CPU for the five traditional methods, resulting in a total of ten developed models. These models underwent testing with a dataset of 50 Dissolved Gas Analysis (DGA) oil samples obtained from faulty transformers with known fault conditions. The results, presented in Figure 9, indicate the percentage accuracy of each model. Evaluation of overall accuracy involved comparing calculated faults with actual faults in the 50 samples, encompassing thermal faults of varying temperature ranges, arcing, and discharges. The Duval Triangle decision tree model achieved the highest accuracy at 98%, correctly identifying the faults in 49 out of 50 instances. Other decision tree models also demonstrated notable accuracies, as depicted in Figure 9. In contrast, the Key Gas method (KGA) fuzzy model exhibited the lowest accuracy among the ten models at 34%.

3.1 Fuzzy Logic

Fuzzy Logic employs rules and membership functions for fault calculation. Assigning a specific number to each fault allows a comparison between actual faults and fuzzy output. Table 8 showcases a dataset comprising 50 DGA results from the analysis of oil samples taken from faulty transformers with known faults.

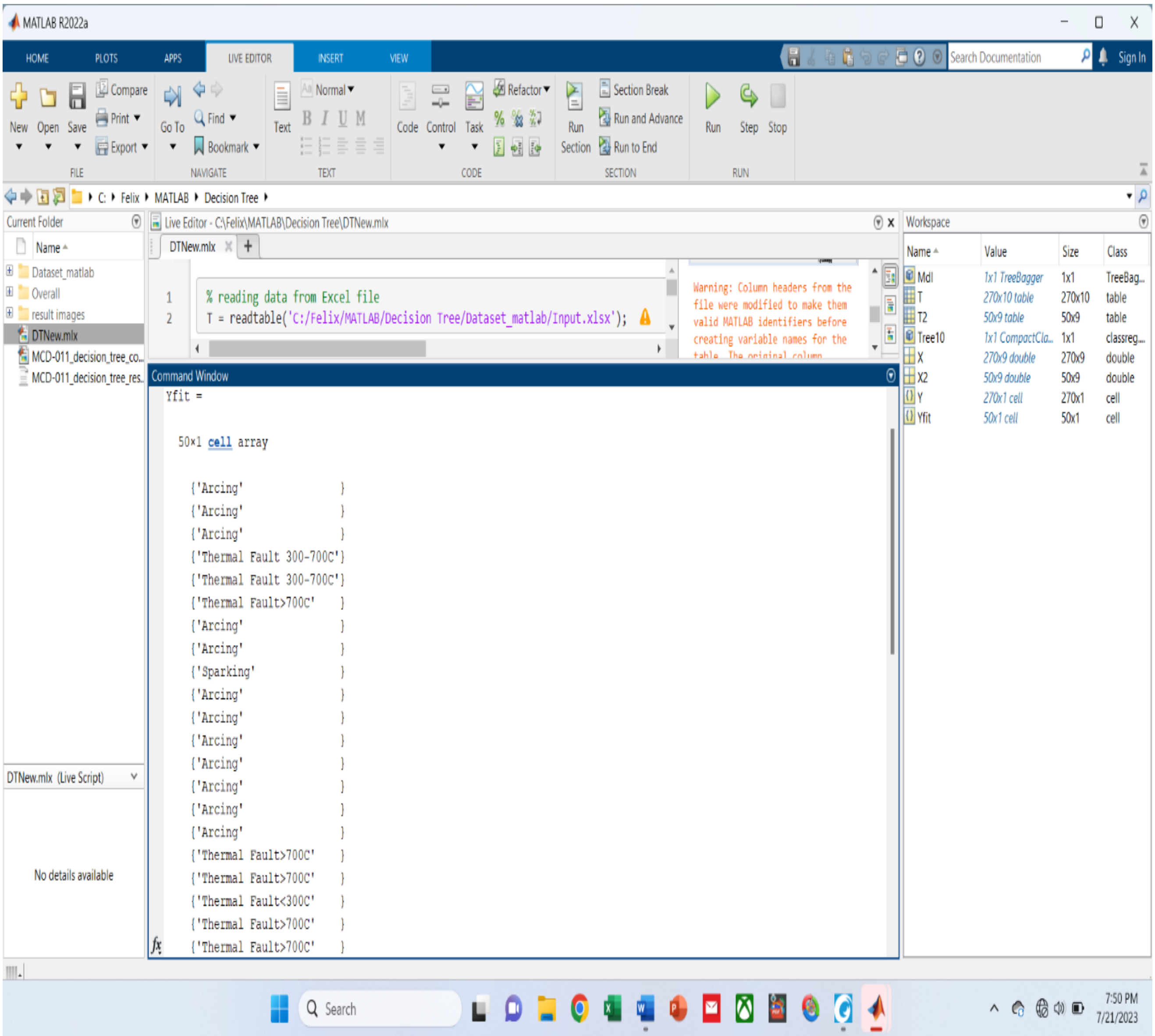


Figure 8: Some of the results of 50 DGA data samples with pre-known faults as predicted by decision tree algorithm

Table 8**:** DGA results of 50 faulty transformer oil samples with known faults

H_2	H_2O	CO_2	CO	C_2H_4	C_2H_6	CH_4	C_2H_2	TCDG	Actual Faults
1176		3400	299	2931	1178	3426	0	9010	Thermal Fault
43		2326	718	139	65	116	0	1081	Thermal Fault
54		4497	1358	101	23	143	0	1679	Thermal Fault
19		1114	229	62	27	47	0	384	Thermal Fault
114		5675	1309	574	80	241	0	2318	Thermal Fault

Table 8**:** DGA results of 50 faulty transformer oil samples with known faults (cont'd)

H_2	H_2O	CO_2	CO	C_2H_4	C_2H_6	CH_4	C_2H_2	TCDG	Actual Faults
99		1625	304	386	73	288	0	1150	Thermal Fault
1374		2819	783	5376	628	2648	298	11107	Thermal Fault
34		3197	645	136	100	83	0	998	Thermal Fault
20		2576	712	17	61	132	0	942	Thermal Fault
13		5197	1046	16	83	138	0	1296	Thermal Fault
37		926	132	57	19	96	0	341	Thermal Fault
46		2678	184	150	27	147	0	554	Thermal Fault
18		2633	390	51	20	63	0	542	Thermal Fault
20		1854	571	77	45	19	0	732	Thermal Fault
25		1814	424	58	42	49	0	598	Thermal Fault
127		2024	0	32	0	24	81	264	Arcing
441		1123	161	224	43	207	261	1337	Arcing
217		1544	176	458	14	286	884	2035	Arcing
48		0	0	75	3	43	81	250	Arcing
318		0	0	583	57	337	641	1936	Arcing
200		308	138	60	9	30	98	535	Arcing
678		1909	768	89	31	70	237	1873	Arcing
762		5346	459	54	38	93	126	1532	Arcing
440		1232	428	62	31	522	183	1666	Arcing
127		2024	0	32	0	24	81	264	Arcing
441		1123	161	224	43	207	261	1337	Arcing
217		1544	176	458	14	286	884	2035	Arcing
210		1070	167	102	12	43	187	721	Arcing
678		2211	216	108	92	368	163	1625	Arcing
1498		3176	487	395	323	395	92	3190	Arcing

Table 8: DGA results of 50 faulty transformer oil samples with known faults (cont'd)

H_2	H_2O	CO_2	CO	C_2H_4	C_2H_6	CH_4	C_2H_2	TCDG	Actual Faults
160		0	0	1	3	10	1	175	Sparking
195		3674	607	79	127	660	22	1690	Sparking
113		871	32	57	20	24	61	307	Sparking
78				13	11	20	28	150	Sparking
305				161	33	100	541	1140	Sparking
1230				233	27	163	692	2345	Sparking
645				110	13	86	317	1171	Sparking
95				11		10	39	155	Sparking
595				89	9	80	244	1017	Sparking
240		0	0	28	5	20	96	389	Sparking
893		2207	350	18	6	724	1	1992	Partial Discharge
441		492	302	62	73	678	0	1556	Partial Discharge
92600			6400	0.05	0.05	10200	0.05	109200.2	Partial Discharge
26788			704	27	2111	18342	0.05	47972.05	Partial Discharge
441		492	302	62	73	678	0	1556	Partial Discharge
893		2207	350	18	6	724	1	1992	Partial Discharge
234		0	230	10.1	162	25	0.8	661.9	Partial Discharge
235		0	231`	10	165	25	1	667	Partial Discharge
225		0	220	7.4	115	23	0.4	590.8	Partial Discharge
239		0	225	10.9	149	25	0.76	649.66	Partial Discharge

3.2 Fuzzy Duval Triangle

The Duval Triangle method evaluates four distinct faults, excluding Normal. These faults are High Energy Discharge (2), Mixed Thermal and Electrical Fault (3), Thermal Fault <300°C (5), Thermal Fault 300°C–700°C (6), and Thermal Fault >700°C (7). High Energy Discharge and Thermal Faults are the most prevalent, each occurring 15 times. Among the fifty instances, fuzzy logic yielded correct predictions 41 times, resulting in 82% accuracy for the Duval Triangle method. Table 9 shows the Duval fuzzy output and the actual fault for each oil sample.

The precision and recall of this Duval model for each fault condition are computed as given in Equations 1 to 8.

$$\text{Precision for Partial discharge } P_{PD} = \frac{\text{correct PD predictions (2)}}{\text{correct PD predictions (2)+false PD predictions (0)}} = \frac{2}{2+0} = 100\% \qquad (1)$$

$$\text{Precision for Sparking } P_{SP} = \frac{\text{correct SP predictions (9)}}{\text{correct SP predictions (9+false SP predictions (3)}} = \frac{9}{9+3} = 75\% \qquad (2)$$

$$\text{Precision for Arcing } P_{ARC} = \frac{\text{correct ARC predictions (15)}}{\text{correct ARC predictions (15)+false ARC predictions (6)}} = \frac{15}{15+6} = 71.4\% \qquad (3)$$

$$\text{Precision for Thermal } P_{THM} = \frac{\text{correct THM predictions (15)}}{\text{correct THM predictions (15)+false THM predictions (9)}} = \frac{15}{15+9} = 62.5\% \qquad (4)$$

$$\text{Recall for Partial discharge } R_{PD} = \frac{\text{correct PD predictions (2)}}{\text{correct PD predictions (2)+failed PD predictions (8)}} = \frac{2}{2+8} = 20\% \qquad (5)$$

$$\text{Recall for Sparking } R_{SP} = \frac{\text{correct SP predictions (9)}}{\text{correct SP predictions (9)+failed SP predictions (1)}} = \frac{9}{9+1} = 90\% \qquad (6)$$

$$\text{Recall for Arcing } R_{ARC} = \frac{\text{correct ARC predictions (15)}}{\text{correct ARC predictions (15)+failed ARC predictions (0)}} = \frac{15}{15+0} = 100\% \qquad (7)$$

$$\text{Recall for Thermal } R_{THM} = \frac{\text{correct THM predictions (15)}}{\text{correct THM predictions (15)+failed THM predictions (0)}} = \frac{15}{15+0} = 100\% \qquad (8)$$

Table 9 Fuzzy-duval triangle

Fuzzy Outputs	**Actual Faults**	**Fuzzy Outputs**	**Actual Faults**	**Fuzzy Outputs**	**Actual Faults**
Low Thermal 5	Thermal Fault	Arcing 2	Arcing	Arcing 2	Sparking
Low Thermal 5	Thermal Fault	Arcing 2	Arcing	Arcing 2	Sparking
Low Thermal 5	Thermal Fault	Arcing 2	Arcing	Arcing 2	Sparking
Low Thermal 5	Thermal Fault	Arcing 2	Arcing	Sparking 1	Sparking
Low Thermal 5	Thermal Fault	Arcing 2	Arcing	Arcing 2	Sparking
Low Thermal 5	Thermal Fault	Sparking 1	Arcing	Sparking 1	Sparking
High Thermal 7	Thermal Fault	Sparking 1	Arcing	Low Thermal 5	Partial Discharge
Low Thermal 5	Thermal Fault	Arcing 2	Arcing	Low Thermal 5	Partial Discharge
Low Thermal 5	Thermal Fault	Arcing 2	Arcing	Partial Discharge 4	Partial Discharge
Low Thermal 5	Thermal Fault	Arcing 2	Arcing	Partial Discharge 4	Partial Discharge
Low Thermal 5	Thermal Fault	Arcing 2	Arcing	Low Thermal 5	Partial Discharge
Low Thermal 5	Thermal Fault	Sparking 1	Arcing	Low Thermal 5	Partial Discharge
Low Thermal 5	Thermal Fault	Arcing &Thermal 3	Arcing	Medium Thermal 6	Partial Discharge
Low Thermal 5	Thermal Fault	Arcing &Thermal 3	Sparking	Medium Thermal 6	Partial Discharge
Low Thermal 5	Thermal Fault	Low Thermal 5	Sparking	Medium Thermal 6	Partial Discharge
Arcing 2	Arcing	Arcing 2	Sparking	Medium Thermal 6	Partial Discharge
Arcing 2	Arcing	Arcing 2	Sparking		

Table 10: Percentage precision and recall of the model for each fault condition

Faults	%Precision	%Recall
Partial Discharge	100	20
Thermal	62.5	100
Arcing	71.4	100
Sparking	75	90

The same 50 dataset in Table 8 was also used to test the fuzzy models for Rogers ratio, IEC ratio, Doernenburg ratio and the Key Gas Analysis methods. Each model's overall accuracy, precision and recall per fault were also calculated.

3.3 Decision Tree:

Distinct decision tree codes are formulated for each of the five methods, employing the Decision Tree approach. The algorithm is trained with IEC TC 10 database data and tested using the data presented in Table 8. After training, new gas values are input to identify potential faults through the TreeBagger algorithm. This algorithm employs bootstrap aggregation, constructing each tree in the ensemble on a bootstrap copy of the input data. Input and target variables are organized into separate arrays, both utilized in the tree algorithm to establish rules. Table 11 displays the prediction of the Duval Triangle decision tree model for the dataset in Table 8.

Table 11: Decision Tree-Duval triangle

Decision Tree	Actual Fault	Decision Tree	Actual Fault	Decision Tree	Actual Fault
Thermal Fault	Thermal Fault	Arcing	Arcing	Arcing	Sparking
Thermal Fault	Thermal Fault	Arcing	Arcing	Arcing	Sparking
Thermal Fault	Thermal Fault	Arcing	Arcing	Arcing	Sparking
Thermal Fault	Thermal Fault	Arcing	Arcing	Sparking	Sparking
Thermal Fault	Thermal Fault	Arcing	Arcing	Arcing	Sparking
Thermal Fault	Thermal Fault	Arcing	Arcing	Arcing	Sparking
Thermal Fault	Thermal Fault	Arcing	Arcing	Partial Discharge	Partial Discharge
Thermal Fault	Thermal Fault	Arcing	Arcing	Partial Discharge	Partial Discharge
Thermal Fault	Thermal Fault	Arcing	Arcing	Partial Discharge	Partial Discharge
Thermal Fault	Thermal Fault	Arcing	Arcing	Partial Discharge	Partial Discharge
Thermal Fault	Thermal Fault	Arcing	Arcing	Partial Discharge	Partial Discharge
Thermal Fault	Thermal Fault	Arcing	Arcing	Partial Discharge	Partial Discharge
Thermal Fault	Thermal Fault	Arcing	Arcing	Partial Discharge	Partial Discharge
Thermal Fault	Thermal Fault	Partial Discharge	Sparking	Partial Discharge	Partial Discharge
Thermal Fault	Thermal Fault	Arcing	Sparking	Partial Discharge	Partial Discharge
Arcing	Arcing	Arcing	Sparking	Partial Discharge	Partial Discharge
Arcing	Arcing	Arcing	Sparking		

From Table 11, the Duval Triangle Decision Tree algorithm accurately identified faults in 49 out of 50 instances, resulting in 98% accuracy for the Decision Tree model. Precision and recall of this model for each fault condition are computed as given in Equations 9 to 16.

$$\text{Precision for Partial discharge } \mathrm{P_{PD}} = \frac{\textbf{correct PD predictions (10)}}{\textbf{correct PD predictions (10)+false PD predictions (1)}} = \frac{\mathbf{10}}{\mathbf{10+1}} = 90.9\% \tag{9}$$

$$\text{Precision for Sparking } \mathrm{P_{SP}} = \frac{\textbf{correct SP predictions (9)}}{\textbf{correct SP predictions (9)+false SP predictions (0)}} = \frac{\mathbf{9}}{\mathbf{9+0}} = 100\% \tag{10}$$

Precision for Arcing $P_{ARC} = \frac{\textbf{correct ARC predictions (15)}}{\textbf{correct ARC predictions (15)+false ARC predictions (8)}} = \frac{\mathbf{15}}{\mathbf{15+8}} = 65.2\%$ (11)

Precision for Thermal $P_{THM} = \frac{\textbf{correct THM predictions (15)}}{\textbf{correct THM predictions (15)+false THM predictions (0)}} = \frac{\mathbf{15}}{\mathbf{15+0}} = 100\%$ (12)

Recall for Partial discharge $R_{PD} = \frac{\textbf{correct PD predictions (10)}}{\textbf{correct PD predictions (10)+failed PD predictions (0)}} = \frac{\mathbf{10}}{\mathbf{10+0}} = 100\%$ (13)

Recall for Sparking $R_{SP} = \frac{\textbf{correct SP predictions (9)}}{\textbf{correct SP predictions (9)+failed SP predictions (1)}} = \frac{\mathbf{9}}{\mathbf{9+1}} = 90\%$ (14)

Recall for Arcing $R_{ARC} = \frac{\textbf{correct ARC predictions (15)}}{\textbf{correct ARC predictions (15)+failed ARC predictions (0)}} = \frac{\mathbf{15}}{\mathbf{15+0}} = \mathbf{100\%}$ (15)

Recall for Thermal $R_{THM} = \frac{\textbf{correct THM predictions (15)}}{\textbf{correct THM predictions (15)+failed THM predictions (0)}} = \frac{\mathbf{15}}{\mathbf{15+0}} = \mathbf{100\%}$ (16)

Table 12: Percentage precision and recall for each fault condition

Faults	**%Precision**	**%Recall**
Partial Discharge	90.9	100
Thermal	100	100
Arcing	65.2	100
Sparking	100	90

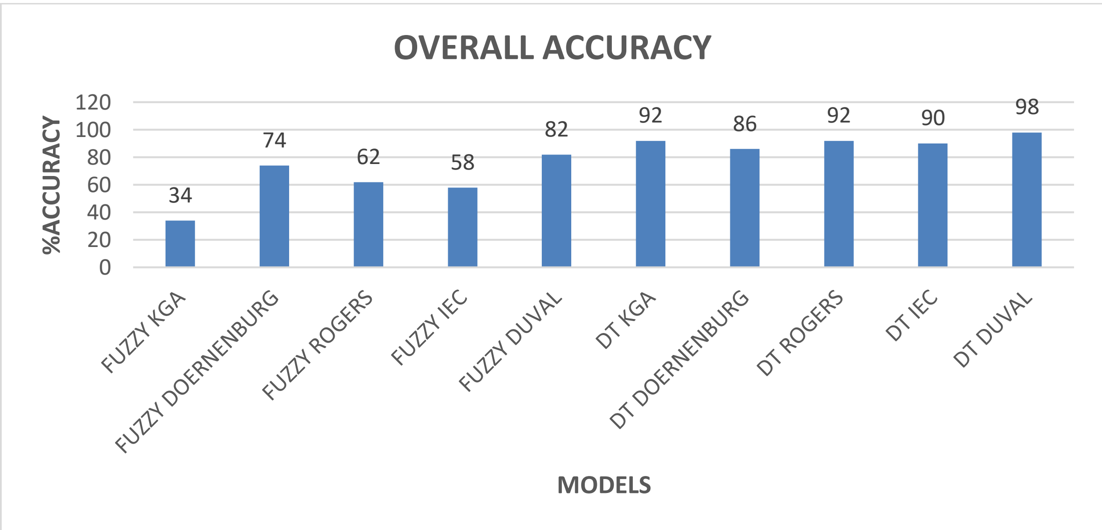


Figure 9: Comparison of overall accuracy for the fuzzy logic and the decision tree models

3.4 Performance Metrics

Performance metrics assess the effectiveness of a classification machine learning model. Precision and recall are commonly used metrics. In transformer fault detection, precision gauges the accuracy of fault detection, while recall measures the model's sensitivity to specific fault conditions. Depending on the application, there can be a trade-off between precision and recall. In the context of detecting incipient faults in transformers, recall is deemed more desirable. Figure 10 through 17 display fault percentage precision and recall for each expert model. The preferred model is one with the highest recall, aiming to identify as many positive cases (faults) as possible, even at the expense of some false positives. It's important to note that classifying sparking as arcing is considered correct, as both are discharges but of varying energy levels.

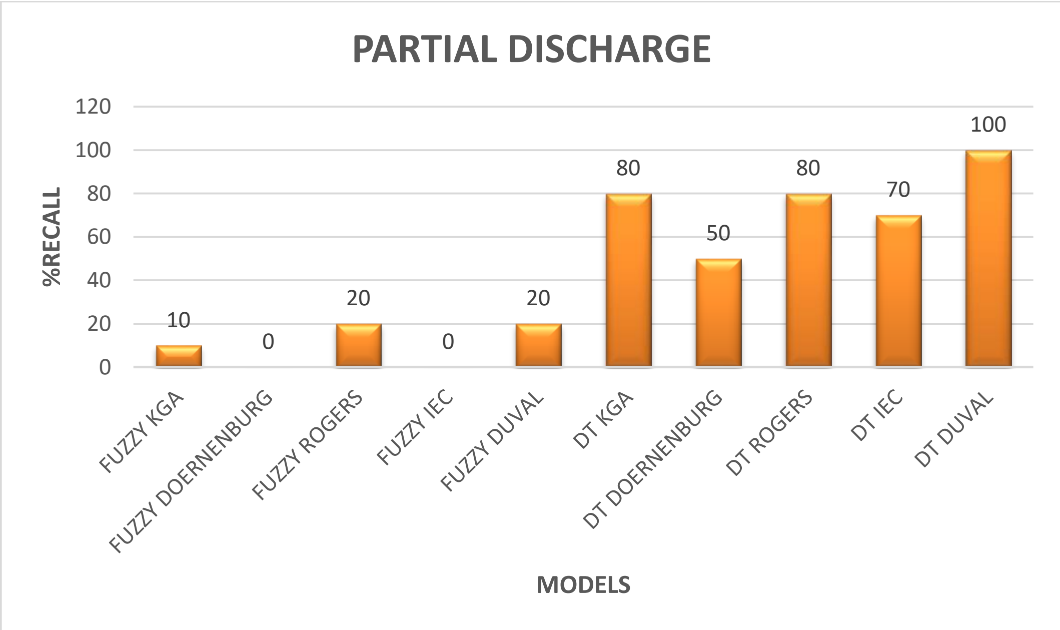


Figure10: Percentage recall of each model for partial discharge

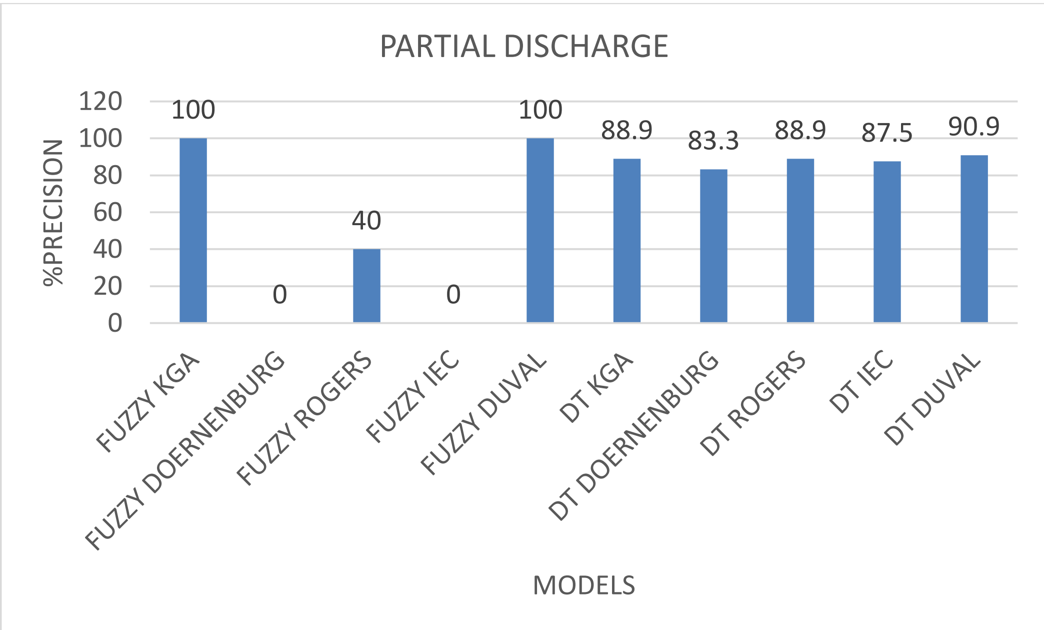


Figure11: Percentage precision of each model for partial discharge

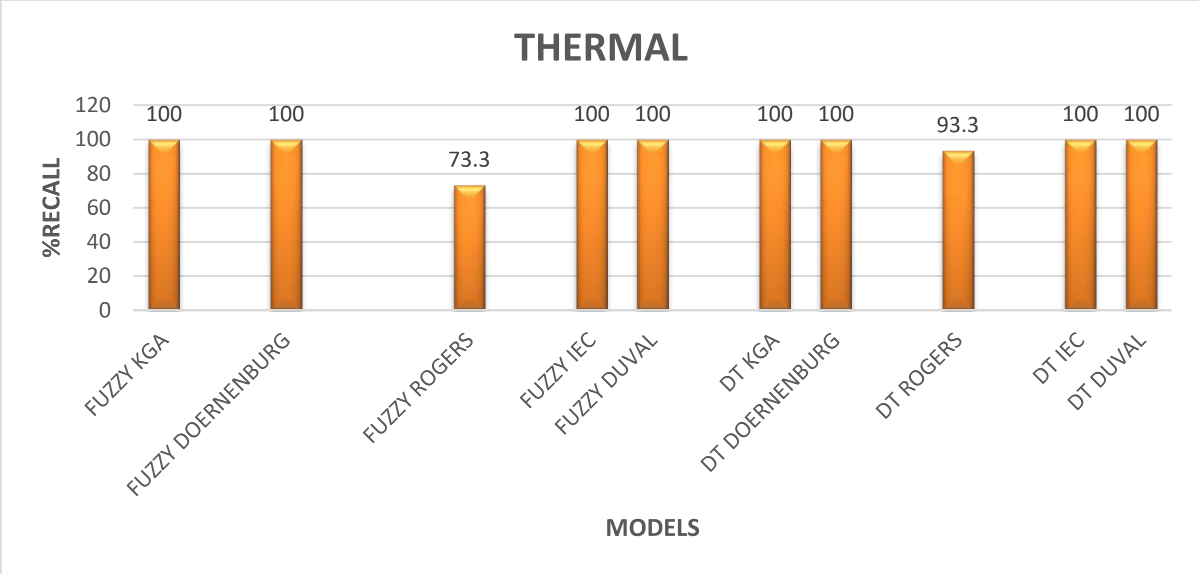


Figure12: Percentage recall of each model for thermal fault

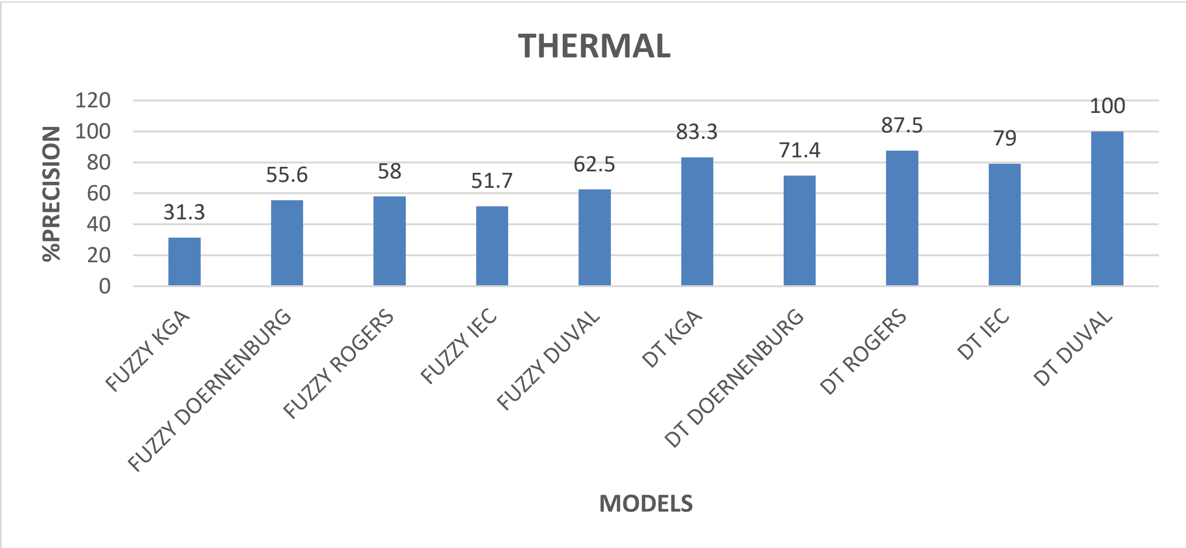


Figure13: Percentage precision of each model for thermal fault

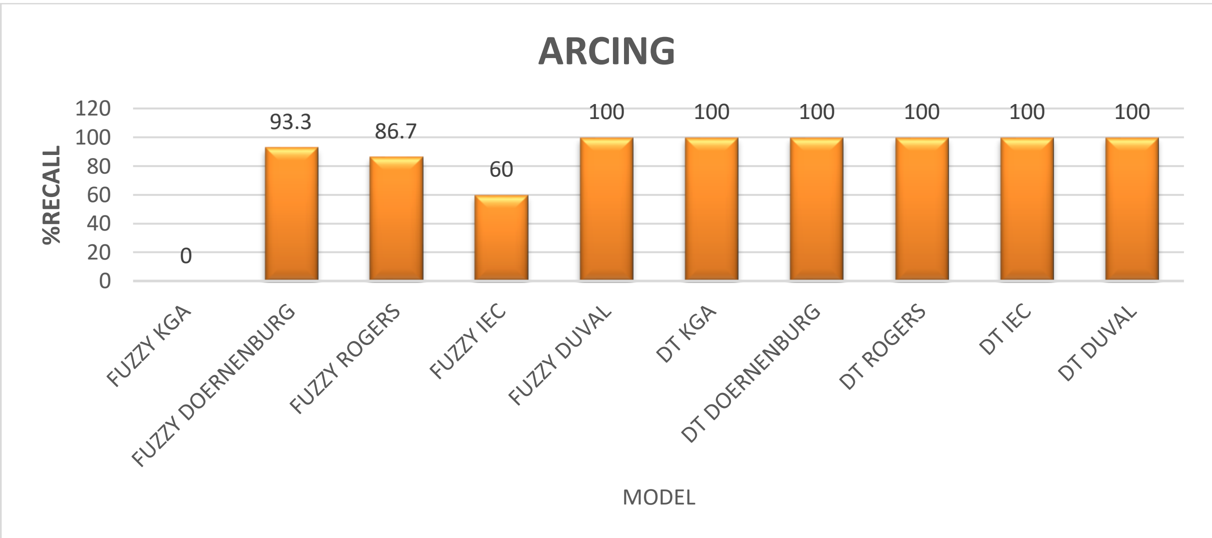


Figure14: Percentage recall of each model for arcing

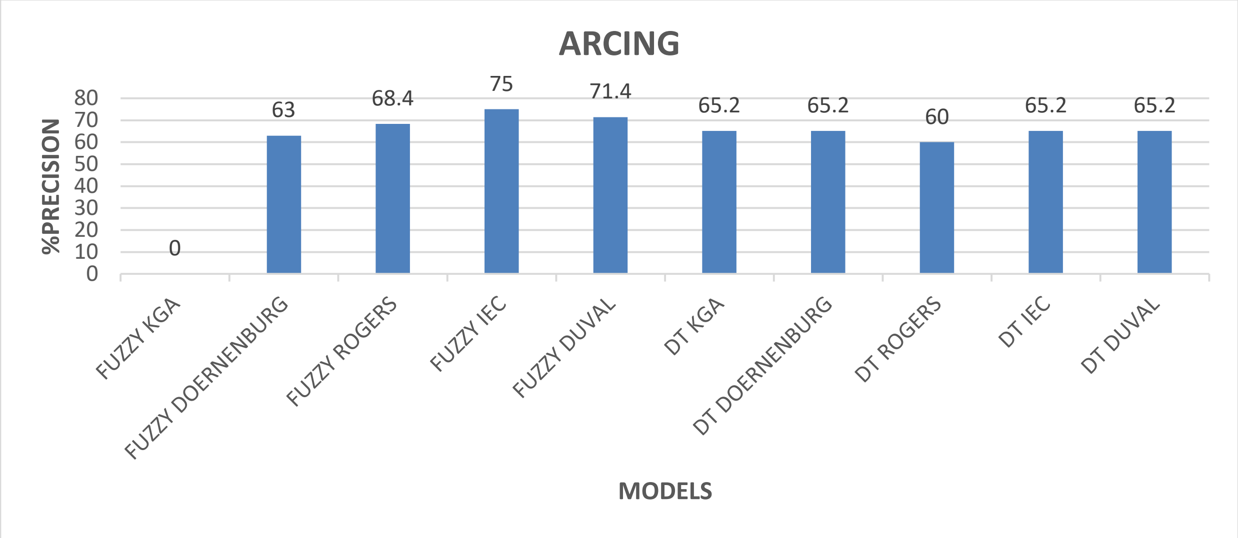


Figure15: Percentage precision of each model for arcing

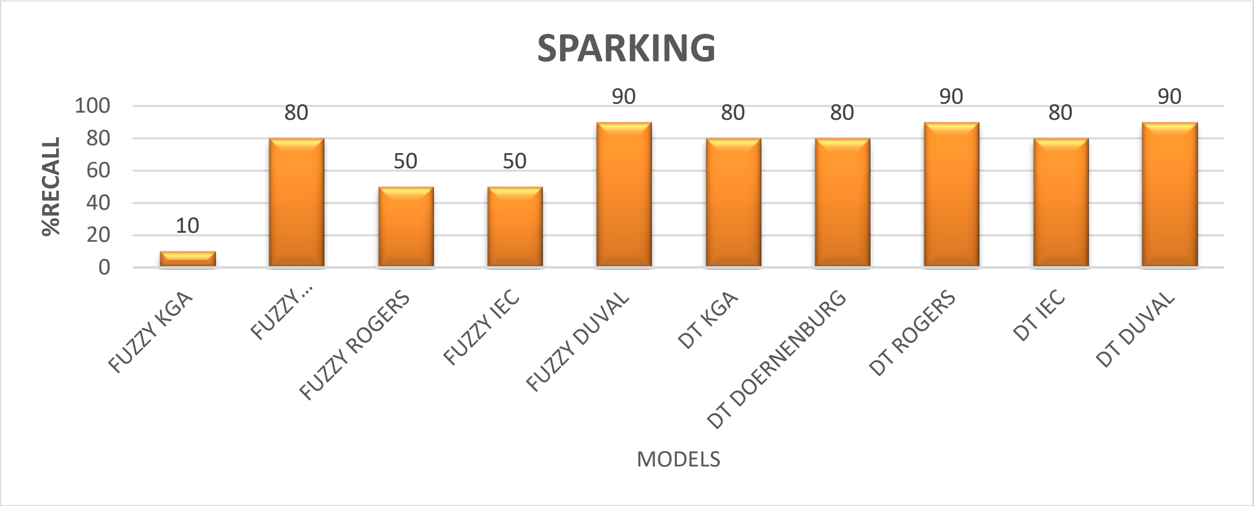


Figure16: Percentage recall of each model for sparking

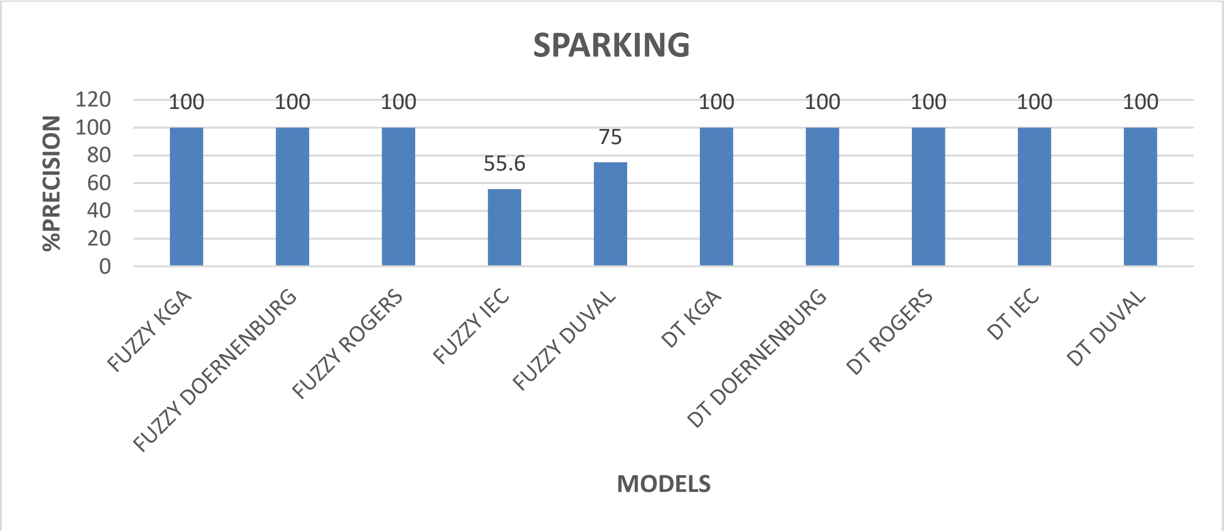


Figure17: Percentage precision of each model for sparking

Table 13 below shows the faults identified with the highest capability by each of the developed models. It is observed from the table that the Duval Triangle Decision tree model is the best of all of the models in detecting transformer's incipient faults. It is also observed from the recall charts in Figure 10, Figure 12, Figure 14 and Figure 16 that the Decision tree models of all the five methods are better than the Fuzzy Logic counterpart as they are sensitive to more faults and can predict them with good accuracy.

Table 13. Faults identified by each model

Models	Identified Faults
Fuzzy KGA	Thermal Faults
Fuzzy Doernenburg	Thermal Faults
Fuzzy Rogers	Thermal Faults
Fuzzy IEC	Thermal Faults
Fuzzy Duval Triangle	Arcing, Sparking, Thermal Faults
DT KGA	Thermal Faults, Arcing
DT Doernenburg	Thermal Faults, Arcing
DT Rogers	Arcing, Sparking
DT IEC	Thermal Faults, Arcing
DT Duval Triangle	Partial Discharge, Thermal Faults, Arcing, Sparking

From Table 13, it can be seen that Duval triangle DT model is more sensitive to transformer faults viz: arcing, sparking, thermal and discharges than all other models.

4. CONCLUSION

This paper employs Fuzzy logic and decision tree algorithms for DGA analysis, evaluating their performance on a dataset of 50 DGA samples with known faults. The Duval triangle decision tree model emerges as the most accurate for all known transformer's internal fault detection like electrical, thermal and discharges in terms of accurate prediction and sensitivity. The acceptable accuracy of other models underscores and confirms the efficacy of artificial intelligence methods over traditional approaches. Similarly, in the household of the Fuzzy Logic, Fuzzy Duval Triangle is the most effective model in the detection of thermal faults and electrical faults like arcing and sparking while other fuzzy models are effective in detecting thermal faults. Moreover, Decision Tree models exhibit increased accuracy, suggesting potential enhancements through re-training with a more precise dataset possibly from the output of the fuzzy models in their respective area of accurate fault detection and sensitivity (specialisation). Future implementations may benefit from developing models based on these conventional methods using advanced AI techniques such as deep learning for higher computational capability.

REFERENCES

[1] Abu-Siada, A. & Islam, S. (2012). A novel online technique to detect power transformer winding faults, *IEEE Transactions on Power Delivery*, 2(27), 849–857.
[2] Ma, H., Ekanayake, C. & Saha, T.K. (2012). Power transformer fault diagnosis under measurement originated uncertainties, *IEEE Transactions on Dielectrics and Electrical Insulation*, 6(19), 1982–1990.
[3] Rogers, R.R. (1978). IEEE and IEC codes to interpret incipient faults in transformers using gas in oil analysis, *IEEE transactions on electrical insulation*, 5, 349–354.
[4] Abu-Siada, A., Hmood, S. & Islam, S. (2013). A new fuzzy logic approach for consistent interpretation of dissolved gas-in-oil analysis, *IEEE Transactions on Dielectrics and Electrical Insulation*, 6(20), 2343–2349.
[5] IEC, I.E.C 60599 (2015). Mineral Oil-Impregnated Electrical Equipment in Service Guide to the Interpretation of Dissolved and Free Gases Analysis, *International Electrotechnical Commission: Geneva, Switzerland.*
[6] Yang, X., Nielsen, S. & Ledwich, G. (2017). Frequency domain spectroscopy measurements of oil-paper insulation for energized transformers, *IEEE Transactions on Dielectrics and Electrical Insulation*, 3(24), 1657–1664.
[7] Zhao, B., Yang, M., Diao, H.R., An, B., Zhao, Y.C. & Zhang, Y.M. (2019). A novel approach to transformer fault diagnosis using IDM and naive credal classifier, *International Journal of Electrical Power & Energy Systems*, (105), 846–855.
[8] Arshad, M. (2005). Remnant life estimation model using fuzzy logic for power transformer asset management, *Curtin University of Technology*.
[9] Lefeng, C. & Tao, Y.U. (2018). Typical scenario analysis of equilibrium stability of multi-group asymmetric evolutionary games in the open and ever-growing electricity market, *Proceedings of the Chinese Society for Electrical Engineering*, 19(38), 5687–5703.

[10] Alawode, K.O. & Olowolafe, F. (2023). A Review of Methods for Diagnosing Incipient Faults in Power Transformers using Dissolved Gas Analysis, *UNIOSUN Journal of Engineering and Environmental Sciences (UJEES)*, 5(1).
[11] Gouda, O.E., El-Hoshy, S.H. & ElTamaly, H.H. (2018). A new proposed three ratios technique for the interpretation of mineral oil transformers based dissolved gas analysis, *IET Gener. Transm. Distrib.*, 12(11), 2650– 2661.
[12] Yousuf D. Almoallem, Ibrahim B.M. Taha, Mohamed I. Mosaad, Lara Nahma, Ahmed Abu-Siada. (2021). Application of Logistic Regression Algorithm in the Interpretation of Dissolved Gas Analysis for Power Transformers, *Electronics*, 10(10), 1206.
[13] Mohammed El-Amine Senoussaoui, M. Mostefa Brahami, M.N. Brahami, I.S Bousmaha. (2019). Comparative study of DGA Interpretation methods, *4th International Conference on Electrical Engineering*.
[14] Duval, M. (2003). New techniques for dissolved gas-in-oil analysis, *IEEE Electrical Insulation Magazine*, 2(19), 6-15.