

# Interpretation of Gases Dissolved in Dielectric Oil Using Random Forests for the Detection of Anomalies in Power Transformers

## Interpretación de Gases Disueltos en Aceite Dieléctrico Mediante Bosques Aleatorios Para la Detección de Anomalías en Transformadores de Potencia

 A.S. Freire<sup>1</sup>

 J.C. Astudillo<sup>1</sup>

 C.I. Quinatoa<sup>1</sup>

 F.R. Arias<sup>1</sup>

<sup>1</sup>Universidad Técnica de Cotopaxi

E-mail: armando.freire2833@utc.edu.ec; juan.astudillo8708@utc.edu.ec;  
carlos.quinatoa7864@utc.edu.ec; fernando.arias6251@utc.edu.ec

### Abstract

The following paper presents a machine learning tool for the interpretation of anomalies in power transformers using the random forest method. Using the results of gas chromatography tests on dielectric oil from several published papers, the data set delivered by the dissolved gas analysis (DGA) in quantities of parts per million (ppm), the amount of hydrocarbon gases such as hydrogen (H<sub>2</sub>), methane (CH<sub>4</sub>), ethane (C<sub>2</sub>H<sub>6</sub>), ethylene (C<sub>2</sub>H<sub>4</sub>) and acetylene (C<sub>2</sub>H<sub>2</sub>) that serve to diagnose the internal state of the transformer is used. Due to the reduced number of collected data, there is a disadvantage to apply artificial neural networks, support vector machine, among others that need large amounts of data for each variable, but satisfactorily they are solved using random forests, because this methodology classifies better the data of smaller amount. The learning obtained by training is validated with the states obtained by the test data under IEC 60599 and IEEE C57-104, which encompass 4 diagnostics such as high energy discharge, low energy discharge, normal state and overheating, resulting in a final corroborative validation criterion for the algorithm by comparing the diagnostic results with the random forests.

**Index terms**— Dissolved gas analysis, power transformers, random forests, machine learning.

### Resumen

El siguiente documento presenta una herramienta de aprendizaje automático para la interpretación de anomalías en transformadores de potencia utilizando el método de bosques aleatorios. Mediante los resultados de ensayos de cromatografía de gases en aceite dieléctrico de varios artículos publicados, se utiliza el conjunto de datos entregados por el análisis de gases disueltos (AGD) en cantidades de partes por millón (ppm), la cantidad de gases de hidrocarburos como el hidrógeno (H<sub>2</sub>), metano (CH<sub>4</sub>), etano (C<sub>2</sub>H<sub>6</sub>), etileno (C<sub>2</sub>H<sub>4</sub>) y acetileno (C<sub>2</sub>H<sub>2</sub>) que sirven para diagnosticar el estado interno del transformador. Debido al número reducido de datos recolectados, se presenta una desventaja para aplicar redes neuronales artificiales, máquina de soporte vectorial, entre otras que necesitan grandes cantidades de datos para cada variable, pero satisfactoriamente son resueltas usando bosques aleatorios, debido a que esta metodología clasifica mejor los datos de menor cantidad. El aprendizaje obtenido por el entrenamiento se valida con los estados obtenidos por los datos de prueba bajo la norma IEC 60599 e IEEE C57-104, que engloban a 4 diagnósticos como la descarga de alta energía, descarga de baja energía, estado normal y sobrecalentamiento, obteniendo como resultado un criterio de validación final corroborativo por el algoritmo al comparar el diagnóstico de resultados con el de bosques aleatorios.

**Palabras clave**— Análisis de gases disueltos, transformadores de potencia, bosques aleatorios, aprendizaje automático.

Recibido: 25-10-2022, Aprobado tras revisión: 13-01-2023

Forma sugerida de citación: Freire, A.; Astudillo, J.; Quinatoa, C.; Arias, F. (2023). "Interpretación de Gases Disueltos en Aceite Dieléctrico Mediante Bosques Aleatorios Para la Detección de Anomalías en Transformadores de Potencia". Revista Técnica "energía". No. 19, Issue II, Pp. 90-98

ISSN On-line: 2602-8492 - ISSN Impreso: 1390-5074

Doi: <https://doi.org/10.37116/revistaenergia.v19.n2.2023.544>

© 2023 Operador Nacional de Electricidad, CENACE



## 1. INTRODUCCIÓN

El transformador de potencia es el activo más caro de una subestación [1], siempre debe mantenerse en óptimas condiciones para que la red eléctrica sea fiable y eficiente. Las fallas en transformadores de potencia causan daños importantes, provocando paradas del servicio eléctrico. El análisis de gases disueltos (AGD) se utiliza para la interpretación de fallas incipientes en aceite dieléctrico [2], además de otra prueba como el análisis físico químico como parte de un mantenimiento preventivo. Si bien es cierto, el método de AGD es utilizado por los expertos para determinar el tipo de falla interna dentro del transformador, a veces se deben valer por más muestras para garantizar un correcto resultado y esto conlleva a tiempo e inversión económica, aun así, sigue la interrogante si los datos recolectados corresponden a la falla exacta del transformador.

El uso del aprendizaje automático para la detección de fallas en transformadores ya se han venido realizando con poco éxito debido a la cantidad reducida de datos analizados por medio de la implementación de redes neuronales artificiales (RNA), máquina de soporte vectorial (SVM), lógica difusa, entre otros [3], [4], [5], [6], [7], los artículos citados fueron tomados para la implementación de clasificadores de bosques aleatorios debido a que la presente metodología funciona de mejor manera clasificando pocos datos.

El presente trabajo muestra una corroboración de diagnóstico a los resultados de pruebas AGD aplicados a diferentes transformadores de potencia sumergidos en aceite mineral, resultados en base a artículos científicos e investigaciones publicadas que sirven de comparación frente a la propuesta de uso de aprendizaje automático con bosques aleatorios para determinar las fallas internas en transformadores de potencia.

El siguiente documento se distribuye de la siguiente manera: La teoría utilizada para esta investigación se contempla en la sección 2, el modelo utilizado para este trabajo en la sección 3, la implementación del algoritmo utilizado para el entrenamiento en la sección 4, el análisis de resultados en la sección 5 y por último las conclusiones previstas en la sección 6.

## 2. MÉTODO DE DIAGNÓSTICO

Los gases combustibles relacionados con las fallas en transformadores se deben a la descomposición del aceite mineral, generando moléculas de gas como el hidrógeno ( $H_2$ ), el metano ( $CH_4$ ), el etano ( $C_2H_6$ ), el etileno ( $C_2H_4$ ) y acetileno ( $C_2H_2$ ) [8], el número de concentración de gas se mide en partes por millón (ppm).

Los tres principales tipos de fallas de los transformadores de potencia que pueden identificarse de forma fiable mediante una inspección visual del equipo después de que la falla se haya producido son las descargas parciales, el sobrecalentamiento térmico y el arco eléctrico [9].

Las descargas parciales y los arcos voltaicos se refieren a fallas eléctricas y corresponden al deterioro del aislamiento debido a la alta tensión eléctrica. Las fallas térmicas se refieren al deterioro del sistema de aislamiento como resultado de un aumento de la temperatura anormal. Estas subidas se producen por el sobrecalentamiento de los conductores, los cortocircuitos, el sobrecalentamiento de devanados debido a las corrientes de Foucault, conexiones sueltas y una refrigeración insuficiente [10]. De acuerdo con la norma IEC 60599 y IEEE C57-104, estos principales tipos de fallas pueden clasificarse a su vez en 6 tipos de fallas de transformadores, que se resumen en la Tabla 1.

**Tabla 1: Clasificación de fallas según la norma IEC 60599 y IEEE C57-104 [11] [12]**

Acrónimos	Fallas
PD	Descarga parcial
D1	Descarga de baja energía
D2	Descarga de alta energía
T1	Falla térmica de baja temperatura $T < 300^\circ C$
T2	Falla térmica de media temperatura $300^\circ C < T < 700^\circ C$
T3	Falla térmica de alta temperatura $T > 700^\circ C$

### 2.1. Métodos de Relación de Gases

Estos métodos son convenientes para la detección de fallas en el transformador y pueden ser analizados por los programas informáticos. Además, la concentración de un gas puede ser muy pequeña. Por lo tanto, se puede utilizar la relación de un gas con otro gas en lugar de la concentración de un gas como la Tabla 2. Las desventajas de estos métodos son que pueden no ser siempre analíticos o, en algunos casos, los resultados son inexactos.

**Tabla 2: Resultados del análisis de un gran número de transformadores sanos y averiados [13]**

Gases	Menos de 4 años de funcionamiento (ppm)	Entre 4 y 10 años de funcionamiento (ppm)	Más de 10 años de funcionamiento (ppm)
$H_2$	100-150	200-300	200-300
$CH_4$	50-70	100-150	200-300



$C_2H_6$	30-50	100-130	800-1 000
$C_2H_4$	100-150	150-200	200-400
$C_2H_2$	20-30	36-50	106-150
CO	200-300	400-500	600-700
$CO_2$	3 000-3 500	4 000-5 000	9 000-12 000

### 2.1.1 Método Dürrenberg

Este método puede utilizarse para detectar tres tipos principales de fallas, como el calentamiento, la corona de baja intensidad, la descarga parcial y el arco. El método utiliza cuatro relaciones de gas R1 ( $CH_4/H_2$ ), R2 ( $C_2H_2/C_2H_4$ ), R3 ( $C_2H_2/CH_4$ ) y R4 ( $C_2H_6/C_2H_2$ ). En primer lugar, se determinan las concentraciones de los gases en el aceite para ver si estos valores son superiores a los límites L1 permitidos o no [12], (ver Tabla 3).

Tabla 3: Límites L1 Admisibles Para el Método de la Relación de Dürrenberg [13]

Gases	Concentración L1 (ppm)
CO	350
$CH_4$	120
$H_2$	100
$C_2H_6$	65
$C_2H_4$	50
$C_2H_2$	35

Si la concentración mínima de uno de los gases  $H_2$ ,  $CH_4$ ,  $C_2H_4$  y  $C_2H_2$  supera el doble de los valores límite L1 y uno de los otros tres gases supera el L1, se considera que el transformador tiene la avería [12]. Cada una de las cuatro relaciones R1 ( $CH_4/H_2$ ), R2 ( $C_2H_2/C_2H_4$ ), R3 ( $C_2H_2/CH_4$ ) y R4 ( $C_2H_6/C_2H_2$ ) se comparan con los valores indicados en la Tabla 4.

Tabla 4: Diagnóstico de Fallas por el Método de la Relación de Dürrenberg [12]

Tipo de falla	$R_1 \left( \frac{CH_4}{H_2} \right)$	$R_2 \left( \frac{C_2H_2}{C_2H_4} \right)$	$R_3 \left( \frac{C_2H_2}{CH_4} \right)$	$R_4 \left( \frac{C_2H_6}{C_2H_2} \right)$
Descomposición térmica	<1	<0,75	<0,3	>0,4
Corona	<0,1	Insignificante	<0,3	>0,4
Arco	$0,1 < R_1 < 1$	>0,75	>0,3	<0,4

### 2.1.2 Relación de Rogers

Este método es en realidad el método mejorado de Dürrenberg. Esta técnica tiene a menudo una precisión de más del 80% en los gases solubles, se utilizan dos relaciones de las cuatro relaciones introducidas por Dürrenberg, incluyendo R1 ( $CH_4/H_2$ ) y R2 ( $C_2H_2/C_2H_4$ ) con dos nuevas relaciones ( $C_2H_4/C_2H_6$ ) y ( $C_2H_6/CH_4$ ). La identificación de los fallas se realiza mediante los códigos mostrados en la Tabla 5 [8].

Tabla 5: Códigos de Relación de Gases para el Método de relación de Rogers [12]

Relación de gases	Rango	Código	Rango	Código
$\left( \frac{C_2H_6}{CH_4} \right)$	<1	0	$\geq 1,0$	1
$\left( \frac{C_2H_4}{C_2H_6} \right)$	<1	0	$\geq 1,0 ; < 3,0$	1
	$\geq 3,0$	2		
$\left( \frac{C_2H_2}{C_2H_4} \right)$	<0,5	0	$> 0,5 ; < 3,0$	1
	$\geq 3,0$	2		
$\left( \frac{CH_4}{H_2} \right)$	$\leq 0,1$	5	$> 0,1 ; < 1,0$	0
	$\geq 1,0 ; < 3,0$	1	$\geq 3,0$	2

Utilizando los códigos de relación de gases presentados en la Tabla 5, pueden identificarse doce tipos de fallas diferentes de acuerdo con la Tabla 6 [11].

Tabla 6: Tipos de fallas Diferentes [11]

Tipo de falla	$\left( \frac{CH_4}{H_2} \right)$	$\left( \frac{C_2H_6}{CH_4} \right)$	$\left( \frac{C_2H_4}{C_2H_6} \right)$	$\left( \frac{C_2H_2}{C_2H_4} \right)$
Descomposición normal	0	0	0	0
Descarga parcial	5	0	0	0
Calentamiento leve inferior a 150 °C	1-2	0	0	0
Calentamiento suave (150-200 °C)	1-2	1	0	0
Calentamiento suave (200-300 °C)	0	1	0	0
Calentamiento del conductor	0	0	1	0
Corriente circulante del bobinado	1	0	1	0
Corriente de circulación del tanque y del núcleo	1	0	2	0
Arco	0	0	0	1
Arco en plena carga	0	0	1-2	1-2
Continúa el arco	0	0	2	2

### 2.1.3 Método de relación IEC

Debido a que la relación ( $C_2H_6/CH_4$ ) sólo muestra un rango limitado de degradación del aceite, es limitado [11]. Las tres relaciones de gas restantes tienen diferentes rangos de temperatura en comparación con el método de Rogers [11]. El código de la relación de gases se indica en Tabla 7 y las fallas se dividen en 9 categorías diferentes como se muestra en la Tabla 8.

Tabla 7: Códigos de Relación de Gases para el Método de Relación IEC [11]

Relación de gases	Rango	Código	Rango	Código
	<0,1	0	$0,1 \leq x < 3,0$	1

$\left(\frac{C_2H_2}{C_2H_4}\right)$	$\geq 3,0$	2		
$\left(\frac{CH_4}{H_2}\right)$	$< 0,1$	1	$0,1 \leq x < 1,0$	0
	$\geq 1,0$	2		
$\left(\frac{C_2H_4}{C_2H_6}\right)$	$< 1,0$	0	$1,0 \leq x < 3,0$	1
	$\geq 3,0$	2		

En [11] se introducen otras dos relaciones de gas para identificar las fallas específicas, se recomienda que el valor  $(C_2H_2/H_2)$  sea superior a 3 para detectar la contaminación debida al funcionamiento del cambiador de tomas. Del mismo modo, que la relación  $(O_2/N_2)$  es inferior a 0,3, el fenómeno de calentamiento inusual u oxidación del aceite es determinado.

**Tabla 8: Clasificación de Averías Mediante los Códigos de Relación de la IEC [11]**

Tipo de falla	$\left(\frac{C_2H_2}{C_2H_6}\right)$	$\left(\frac{CH_4}{H_2}\right)$	$\left(\frac{C_2H_4}{C_2H_6}\right)$
Erosión normal, no hay falla	0	0	0
Descarga parcial con baja energía	insignificante	1	0
Descarga parcial con alta energía	1	1	0
Descarga de alta energía	1-2	0	1-2
Descarga de baja energía	1	0	2
Falla térmica ( $T \leq 150\text{ }^\circ\text{C}$ )	0	0	1
Falla térmica ( $150\text{ }^\circ\text{C} < T \leq 300\text{ }^\circ\text{C}$ )	0	2	0
Falla térmica ( $300\text{ }^\circ\text{C} < T \leq 700\text{ }^\circ\text{C}$ )	0	2	1
Falla térmica ( $700\text{ }^\circ\text{C} < T$ )	0	2	2

### 2.1.4 Método del triángulo de Duval

La tabla 9 corresponde a las categorías de fallas dictadas por Duval, lo que significa que hay 6 tipos de fallas en el triángulo de Duval y hay 7 tipos de fallas en el triángulo de Duval modificado [13], este último es el utilizado actualmente por las normativas IEC 60599 e IEEE C57-104.

**Tabla 9: Categorización de Fallas por el Triángulo de Duval [11]**

Triángulo de Duval modificado		Triángulo de Duval	
Tipo de falla	Código de falla	Tipo de falla	Código de falla
Descarga parcial	PD	Arco de alta energía	a
Arco de baja energía	D1	Arco de baja energía	b
Arco de alta energía	D2	Corona	c
Combinación de fallas eléctricas y térmicas	DT	Puntos calientes ( $T < 200\text{ }^\circ\text{C}$ )	d
Fallas térmicas ( $T < 300\text{ }^\circ\text{C}$ )	T1	Puntos calientes ( $200\text{ }^\circ\text{C} < T < 400\text{ }^\circ\text{C}$ )	e
Fallas térmicas ( $300\text{ }^\circ\text{C} < T < 700\text{ }^\circ\text{C}$ )	T2	Puntos calientes ( $T > 400\text{ }^\circ\text{C}$ )	f
Fallas térmicas ( $700\text{ }^\circ\text{C} < T$ )	T3		

El triángulo de Duval utiliza los gases  $CH_4$ ,  $C_2H_4$  y  $C_2H_2$  en los que sus generaciones corresponden a un aumento de los niveles de energía en el transformador. Los tres lados del triángulo se expresan en términos de  $(x, y, z)$  que son, respectivamente, las concentraciones relativas de  $CH_4$ ,  $C_2H_4$  y  $C_2H_2$  en porcentaje [13]. Suponiendo que A, B y C son las concentraciones relativas de  $CH_4$ ,  $C_2H_4$  y  $C_2H_2$  en términos de ppm respectivamente, entonces  $x, y$  y  $z$  son como se menciona a continuación:

$$x = 100 * \frac{A}{A+B+C} \tag{1}$$

$$y = 100 * \frac{B}{A+B+C} \tag{2}$$

$$z = 100 * \frac{C}{A+B+C} \tag{3}$$

Hay que tener en cuenta que  $x, y$  y  $z$  están en el rango de 0-100% y las coordenadas con  $(x_0, y_0, z_0)$ , sólo especifican un punto dentro del triángulo. El tipo de falla también está determinado por la zona en la que se encuentra el punto  $(x_0, y_0, z_0)$  [13].

Es importante señalar que la mayor parte del aceite mineral utilizado en los transformadores no produce ninguna cantidad medible de gases a temperaturas inferiores a  $300\text{ }^\circ\text{C}$  [1], sin embargo, algunos aceites minerales producen  $H_2$  y  $CH_4$  a una temperatura muy baja ( $100\text{ }^\circ\text{C}$ ) al principio de su vida útil, este hecho se produce en el primer año de servicio del transformador y es un proceso irreversible. La concentración de estos dos gases alcanza un nivel constante después de un tiempo de trabajo del transformador. Los valores de estos dos gases para el aceite deben ser considerados con el fin de evitar las interpretaciones erróneas de los resultados.

### 3. APRENDIZAJE AUTOMÁTICO

El aprendizaje automático consiste en codificar programas que ajustan automáticamente su rendimiento en función de su exposición a la información de los datos. Este aprendizaje se consigue mediante un modelo parametrizado con parámetros ajustables automáticamente en función de diferentes criterios de rendimiento. Las técnicas de aprendizaje automático pueden dividirse, a grandes rasgos, en dos grandes clases, aunque a menudo se añade una más [14]. Estas son las clases:

- Aprendizaje supervisado
- Aprendizaje no supervisado
- Aprendizaje por refuerzo

#### 3.1. Aprendizaje Supervisado

Algoritmos que aprenden a partir de un conjunto de ejemplos etiquetados para generalizar al conjunto de todas las entradas posibles. Ejemplos de técnicas de aprendizaje supervisado: regresión logística, máquinas de vectores de apoyo, árboles de decisión, bosques aleatorios, etc. [15].

##### 3.1.1 Bosques aleatorios

Los bosques aleatorios es un algoritmo de aprendizaje automático muy preciso, mucho más robusto que los árboles de decisión y capaz de modelar enormes espacios de características [16].

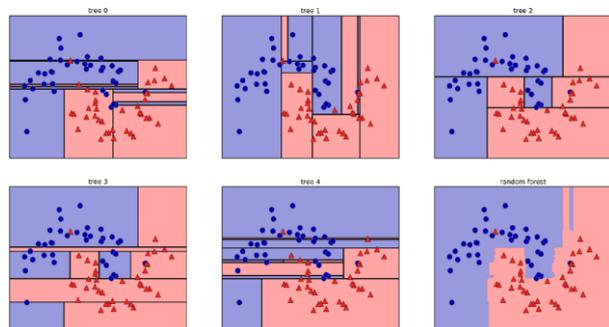


Figura 1: Límites de Decisión Encontrados por Cinco Árboles de Decisión Aleatorios y el Límite de Decisión Obtenida al Promediar sus Probabilidades Predichas. Fuente: [17]

En la Fig. 1 se visualiza el proceso de aleatoriedad compuesto por cinco árboles a un conjunto de datos, se puede ver claramente que los límites de decisión aprendidos por los árboles son bastante diferentes, cada uno de ellos comete algunos errores, ya que algunos de los puntos de entrenamiento que aparecen aquí no se incluyeron realmente en los conjuntos de entrenamiento de los árboles, debido al muestreo bootstrap que realiza el algoritmo.

El bosque aleatorio se ajusta menos que cualquiera de los árboles por separado, al encontrar más rápido la clasificación de cada muestra tomada y proporciona un límite de decisión mucho más intuitivo al resultado final.

### 4. APLICACIÓN DE BOSQUE ALEATORIO AL AGD

Para la aplicación de algoritmos de bosques aleatorios en el análisis de gases disueltos en aceite dieléctrico se designa 4 estados de diagnóstico como se observa en la Tabla 10.

Tabla 10: Diagnóstico de Resultados AGD

ESTADO	DESIGNACIÓN
Descarga Alta Energía	1
Descarga Baja Energía	2
Estado Normal	3
Sobrecalentamiento	4

Trabajar con datos numéricos resulta más fácil para el algoritmo, por tal motivo se designa para este trabajo los estados de la Tabla 10 como números del 1 al 4, que se interpreta como el resultado de diagnóstico de AGD.

#### 4.1. Análisis de Datos

Para el entrenamiento del algoritmo se realiza la recolección de datos de pruebas de AGD realizados a varios transformadores mediante bibliografía recolectada para esta investigación, obteniendo 128 datos para el entrenamiento [3], [4], [5], [6], (ver Fig. 2) y 64 datos de prueba [7], [3], (ver Fig. 3) con 5 principales gases a evaluar para garantizar el aprendizaje del algoritmo mediante la aplicación de bosques aleatorios.

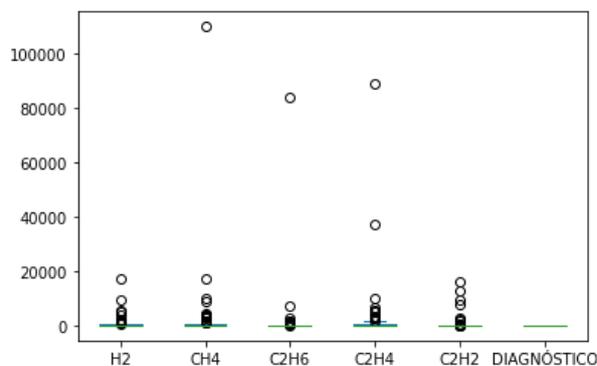


Figura 2: Distribución de Datos de Entrenamiento Mediante Diagramas de Caja

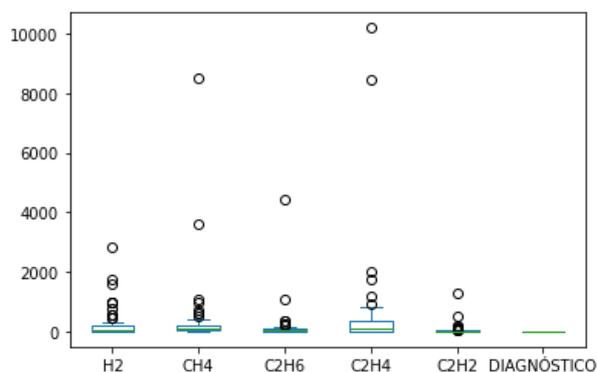


Figura 3: Distribución de Datos de Prueba Mediante Diagramas de Caja

En la Tabla 11 se observa la distribución de datos recolectados para su entrenamiento y prueba, con la cantidad de gases por cada estado de diagnóstico analizado.

**Tabla 11: Distribución de Datos por Categorías de Entrenamiento**

		DIAGNÓSTICO	H <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>2</sub>
Datos de entrenamiento	Descarga Alta Energía		39	39	39	39	39
	Descarga Baja Energía		18	18	18	18	18
	Estado Normal		9	9	9	9	9
	Sobrecalentamiento		62	62	62	62	62
			62	62	62	62	62
Datos de prueba	Descarga Alta Energía		10	10	10	10	10
	Descarga Baja Energía		13	13	13	13	13
	Estado Normal		10	10	10	10	10
	Sobrecalentamiento		31	31	31	31	31
			31	31	31	31	31

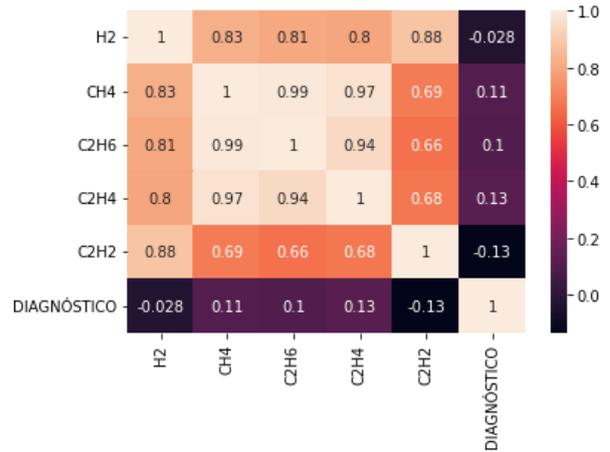
La categoría 4 (Sobrecalentamiento) es la mayoritaria, seguida por la categoría 1 (Descarga de alta energía). Las categorías con menos datos (de entrenamiento y prueba) son las categorías 3 y 2 respectivamente.

**Tabla 12: Análisis Exploratorio de Datos de Categorías de Entrenamiento**

	H <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>2</sub>	DIAGNÓSTICO
Conteo	128,00	128,00	128,00	128,00	128,00	128
Media	526,27	1 496,65	848,54	1 647,41	449,19	2,73
Desviación estándar	1 836,46	9 876,91	7 442,85	8 542,81	2 115,82	1,34
Mínimo	0,00	0,15	0,00	0,00	0,00	1
25%	15,68	7,18	3,37	5,16	0,01	1
50%	106,00	71,00	28,85	77,50	6,33	3
75%	275,50	335,50	88,50	637,00	67,75	4
Máximo	17 000,00	110 000,00	84 000,00	89 000,00	16 000,00	4

Al analizar los datos de entrenamiento (ver Tabla 12) la desviación estándar para cada una de las características se observan valores altos con respecto a su valor promedio, lo que indica que cada característica no se encuentra concentrada en un rango específico sino más bien distribuida en un amplio rango de valores. De manera que es más viable el uso de modelos tipo bosques aleatorios.

En la Fig. 4 se presenta la matriz de correlación de los 5 tipos de gases analizados.

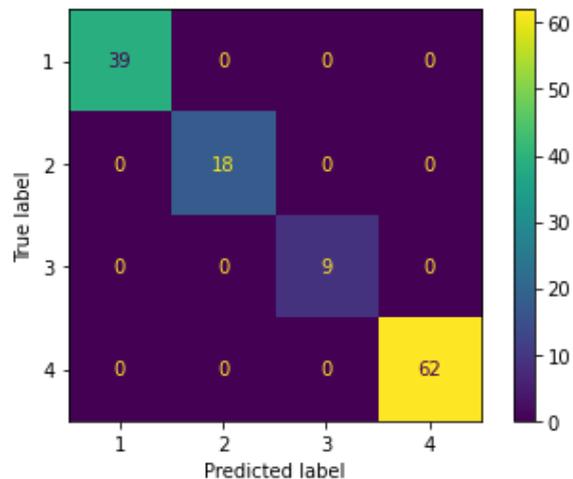


**Figura 4: Matriz de Correlación de Datos de Entrenamiento (Mapa de calor)**

Al realizar un análisis de correlación entre las características se encuentra que hay una alta correlación (cercana a 1) entre los pares de características CH<sub>4</sub> – C<sub>2</sub>H<sub>6</sub>, CH<sub>4</sub> – C<sub>2</sub>H<sub>4</sub> y C<sub>2</sub>H<sub>4</sub> – C<sub>2</sub>H<sub>6</sub>. Esto indica que probablemente no todas las características resultarán igualmente relevantes al momento de elegir la categoría a la que pertenece cada dato.

## 5. ANÁLISIS DE RESULTADOS

Teniendo en cuenta el tamaño limitado del set de entrenamiento, se optó por usar el enfoque de validación cruzada para la selección del modelo y evaluación. El modelo se realiza mediante un proceso sistemático de selección de hiperparámetros que arrojan el mejor desempeño posible (recall) para cada categoría evaluando el desempeño de clasificación comparando el valor verdadero y el valor predicho. En la Fig. 5 se muestra la matriz de confusión de datos de entrenamiento.



**Figura 5: Matriz de Confusión de Datos de Entrenamiento**

Para el modelo de bosque aleatorio se realizó un análisis de importancia de features que permite determinar aquellas características que resultan más relevantes al momento de la clasificación (ver Fig. 6).

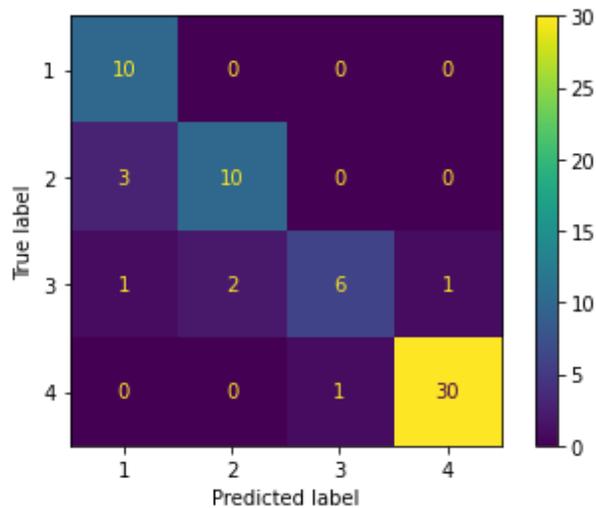


Figura 6: Matriz de Confusión de Datos de Validación de Resultados

Los resultados de recall alcanzados por el algoritmo de entrenamiento frente a los datos de prueba se obtiene el siguiente desempeño de asertividad (ver Tabla 13).

Tabla 13: Resultados de Validación por Categorías

Recall categoría 1:	100,00%
Recall categoría 2:	76,90%
Recall categoría 3:	60,00%
Recall categoría 4:	96,80%

Una vez completado el entrenamiento se realiza la validación de prueba con las 64 muestras de AGD con el respectivo diagnóstico de falla aplicando el algoritmo de bosques aleatorios (ver Fig. 7).

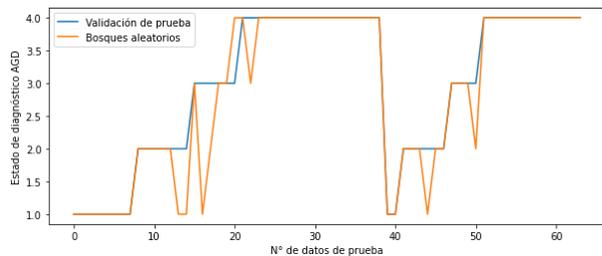


Figura 7: Resultados de Entrenamiento Mediante la Validación de Datos de Prueba y Aplicación de Bosques Aleatorios

## 6. CONCLUSIONES Y RECOMENDACIONES

Se usa el algoritmo de bosques aleatorios debido a que tiene un mayor desempeño al trabajar con pocos datos de entrenamiento 128 datos y 64 datos de prueba, el resultado obtenido de recall en cada categoría es un 100% en entrenamiento para los 4 estados, mediante la validación de prueba se obtiene un 100% para el diagnóstico de descarga de alta energía, 77% para descarga de baja energía, 60% para estado normal y un 97% para el estado de sobrecalentamiento. Obteniendo 78 resultados acertados y 8 datos incorrectos en su validación.

El modelo de bosque aleatorio es el más adecuado para realizar la clasificación con los sets de datos proporcionados, sin embargo, se debe tener en cuenta que el modelo tiene algo de overfitting, debido a que tiene un mejor desempeño con el set de entrenamiento que con el de prueba. Esto es de esperar dado al reducido número de datos de entrenamiento disponibles no permite un mejor desempeño del algoritmo. De igual forma se debe tener en cuenta que por contar con un set desbalanceado, se obtuvieron mejores desempeños para las categorías 1 y 4 y que el desempeño disminuye para las categorías minoritarias (2 y 3).

A futuro se sugiere recolectar más datos de entrenamiento y prueba, lo que redundará probablemente en un mejor desempeño del modelo, el algoritmo permite predecir un diagnóstico de falla mediante las pruebas de análisis de gases disueltos en transformadores de aceite dieléctrico, método diferente a los presentados en la sección 2 que generalmente se utilizan, el objetivo de este trabajo es corroborar los resultados mediante la aplicación de aprendizaje automático con el uso de bosques aleatorios.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] A. Naderian, S. Cress, R. Piercy, F. Wang, and J. Service, "An approach to determine the health index of power transformers," *Conf. Rec. IEEE Int. Symp. Electr. Insul.*, pp. 192–196, 2008, doi: 10.1109/ELINSL.2008.4570308.
- [2] A. D. Ashkezari, H. Ma, T. Saha, and C. Ekanayake, "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.*, vol. 20, no. 3, pp. 965–973, 2013, doi: 10.1109/TDEI.2013.6518966.
- [3] G. Lv, H. Cheng, H. Zhai, and L. Dong, "Fault diagnosis of power transformer based on multi-layer SVM classifier," *Electr. Power Syst. Res.*, vol. 75, no. 1, pp. 9–15, 2005, doi: 10.1016/j.epr.2004.07.013.
- [4] D. V. S. S. Sarma and G. N. S. Kalyani, "ANN approach for condition monitoring of power transformers using DGA," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. C, pp. 444–447, 2004, doi: 10.1109/tencon.2004.1414803.
- [5] C. P. Hung and M. H. Wang, "Diagnosis of incipient faults in power transformers using CMAC neural network approach," *Electr. Power Syst. Res.*, vol. 71, no. 3, pp. 235–244, 2004, doi: 10.1016/j.epr.2004.01.019.
- [6] K. Shrivastava and A. Choubey, "A novel association rule mining with IEC ratio based dissolved gas analysis for fault diagnosis of power transformers," *Int. J. Adv. Comput. Res.*, vol. 2, no. 2, 2012.

- [7] S. S. M. Ghoneim and I. B. Taha, "Artificial Neural Networks for Power Transformers Fault Diagnosis Based on IEC Code Using Dissolved Gas Analysis," *Int. J. Control. Autom. Syst.*, vol. 4, no. 2, pp. 18–21, 2015.
- [8] S. Chakravorti, D. Dey, and B. Chatterjee, *Recent Trends in the Condition Monitoring of Transformers: Theory, Implementation and Analysis*, vol. 67. 2013. doi: 10.1007/978-1-4471-5550-8.
- [9] M. Duval and A. DePablo, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases," *IEEE Electr. Insul. Mag.*, vol. 17, no. 2, pp. 31–41, 2001, doi: 10.1109/57.917529.
- [10] U. Djillali, L. D. E. Sidi, B. E. L. Abbes, S. Mohammed, and E. Amine, "Contributions des techniques intelligentes au diagnostic industriel des transformateurs de puissance," 2019.
- [11] IEC 60599, *Mineral oil-filled electrical equipment in service – Guidance on the interpretation of dissolved and free gases analysis*, 3rd ed. 2015.
- [12] IEEE Std C57.104, *IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers*, vol. 1991, no. 3. 1992. [Online]. Available: <http://ieeexplore.ieee.org/stampPDF/getPDF.jsp?tp=&arnumber=29023%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:IEEE+Guide+for+the+Interpretation+of+Gases+Generated+in+Oil-Immersed+Transformers#0>
- [13] B. Vahidi and A. Teymouri, *Quality Confirmation Tests for Power Transformer Insulation Systems*. 2019. doi: 10.1007/978-3-030-19693-6.
- [14] I. Vasilev, D. Slater, G. Spacagna, P. Roelants, and V. Zocca, *Python Deep Learning*. 2019.
- [15] L. Igual and S. Seguí, *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications*. 2017.
- [16] A. Prinzie and D. Van Den Poel, "Random multiclass classification: generalizing random forests to random MNL and random NB," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4653 LNCS, pp. 349–358, 2007, doi: 10.1007/978-3-540-74469-6\_35.
- [17] A. Pajankar and A. Joshi, *Introduction to Machine Learning with Scikit-learn*. 2022. doi: 10.1007/978-1-4842-7921-2\_5.
- [18] L. E. O. Breiman, "Random Forests," pp. 5–32, 2001.



**Armando Freire Freire.**- Nació en Latacunga, Ecuador en 1994. Curso sus estudios secundarios en el Instituto Tecnológico Ramón Barba Naranjo. Recibió su título de Ingeniero Eléctrico de la Universidad Técnica de Cotopaxi en 2020. Actualmente desempeña el cargo de operador de la subestación eléctrica Novacero planta Lasso y cursa sus estudios de cuarto nivel en la Universidad Técnica de Cotopaxi en la carrera de electricidad mención en sistemas eléctricos de potencia. Sus campos de investigación están relacionados con las energías renovables, calidad de energía, coordinación de protecciones, mantenimiento y operación de equipos de potencia.



**Juan Astudillo Muñoz.**- Nacido en Ambato el 30 de noviembre de 1981, sus estudios secundarios los realizó en el Instituto Superior Tecnológico Docente Guayaquil, obteniendo el título de Bachiller Técnico en Electricidad, en el año 2000 ingresa a la Escuela Politécnica Nacional para en el 2006 obtener el título de Ingeniero Eléctrico. Del 2013 al 2015 obtiene su título de cuarto nivel como Master en la Administración de la Energía y sus fuentes renovables en el Instituto Tecnológico y de Estudios Superiores de Monterrey, México. Ha desempeñado varios cargos tanto en la empresa privada como pública, como Coordinador de Manejo Eléctrico en NOVACERO planta Lasso; Operador del Centro de Control de Transmisión, Ingeniero de Estudios Eléctricos, Supervisor de Operación de la Zona Norte en CELEC EP-TRANSELECTRIC, docente a tiempo parcial en las facultades de Ingeniería Eléctrica de la Escuela Politécnica Nacional, y de la Universidad Técnica de Cotopaxi, y en la actualidad como CEO de ELECSATEC CIA. LTDA.



**Carlos Quinatoa Caiza.**- Nació en Tanicuchi, Ecuador en 1988. Colegio. Gral. Marco Aurelio Subía. Ingeniero en Sistemas Eléctricos de Potencia de la Universidad Técnica de Cotopaxi, Master en Ciencias de la Ingeniería Eléctrica de la Universidad Tecnológica de Pereira y Aspirante a Doctor en Ciencias de la Ingeniería Eléctrica de la Universidad Central de Venezuela, estudios realizado mediante la obtención de la beca Alemana DAAD, miembro activo de la Instituto de Ingeniero Eléctrico y Electrónicos IEEE, investigador junior de la IEEE, consejero de la Rama Estudiantil IEEE-UTC, coordinador de la Maestría en Electricidad Mención

Sistemas Eléctricos de Potencia de la UTC, docente investigador de ingeniería en electricidad y maestría en electricidad de la UTC.



**Fernando Arias Atiaja.-** Nació el 02 de septiembre de 1980. Curso los estudios tecnológicos en el Instituto Tecnológico Superior Ramón Barba Naranjo; Título de Tecnólogo en Mantenimiento Eléctrico y Control Industrial. Recibió su título de Ingeniero

Eléctrico de la Universidad Técnica de Cotopaxi en 2021. Sus campos de investigación están relacionados con instrumentación, eficiencia energética y protección ambiental.