

ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE

<https://dx.doi.org/10.14482/inde.41.02.720.891>

Sistema de fusión de datos para la clasificación de gases licuados del petróleo a través de técnicas de aprendizaje de máquina

Data fusion system for classification of liquefied petroleum gases through machine learning techniques

DANIEL FERNANDO QUINTERO BERNAL *

MIGUEL ÁNGEL JARAMILLO BEDOYA **

HERNÁN ALBERTO QUINTERO VALLEJO ***

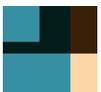
WILLIAM PRADO MARTÍNEZ ****

* Estudiante, Universidad Tecnológica de Pereira, Facultad de Ciencias Básicas. Grupo de investigación Mecabot. Orcid: <https://orcid.org/0000-0002-0745-6452>. Magíster en Instrumentación Física. dfqb@utp.edu.co

** Estudiante, Universidad Tecnológica de Pereira, Facultad de Ciencias Básicas, Grupo de investigación Mecabot. Orcid: <https://orcid.org/0000-0002-5333-0391>. Magíster en Instrumentación Física. mjaramillobe@unal.edu.co

*** Docente, Universidad Tecnológica de Pereira, Facultad de Tecnología. Grupo de investigación Mecabot. Orcid: <https://orcid.org/0000-0001-9016-9747>. Magíster en Enseñanza de la Física. Carrera 27 #10-02, barrio Álamos, Pereira, Risaralda, Colombia. Celular: 3127289354. hernan.quintero@utp.edu.co

**** Docente, Universidad Tecnológica de Pereira, Facultad de Tecnología. Grupo de investigación Mecabot. Orcid: <https://orcid.org/0000-0003-2764-0166>. Magíster en Enseñanza de la Física. prado@utp.edu.co



Resumen

La Fusión de Datos (FD) es un proceso que permite la combinación de información de diversas fuentes con un propósito determinado. En este documento se presenta el diseño y construcción de un sistema FD, particularmente una nariz electrónica (NE), para un dispositivo como el descrito en la patente “Equipo portátil para evitar la contaminación vehicular en las estaciones de servicio” con resolución número 23016. Dicho equipo es un instrumento para clasificación de Gases Licuados derivados del Petróleo (GLP), cuya aplicación se encuentra dirigida a reconocer Aceite Combustible Para Motores (ACPM)¹ o gasolina, en el proceso de abastecimiento vehicular. Estos gases están almacenados en el depósito de combustible de los automóviles y, a través de su identificación, es posible tomar las acciones requeridas para evitar la contaminación de un vehículo con un carburante diferente al que este utiliza. Se implementó un experimento que evaluó el funcionamiento de la NE mediante técnicas de clasificación. En el procesamiento de la información entregada por los sensores del prototipo fueron utilizadas dos metodologías de aprendizaje de máquina supervisadas, K-Nearest Neighbor (KNN) y Naive Bayes, para diferenciación de GLP. Gracias a una prueba de validez, se determinó que la exactitud de las técnicas implementadas fue de 1, por lo tanto, el equipo está en capacidad de discriminar adecuadamente los gases de gasolina y ACPM en condiciones ambientales dinámicas.

Palabras clave: clasificación, fusión de datos, gases licuados del petróleo, nariz electrónica, técnicas de aprendizaje de máquina.

¹ Los términos dísel, diésel, petrodísel, gasóleo y gasoil hacen referencia al mismo tipo de hidrocarburo.

Abstract

Data fusion (DF) is a process that allows the combination of information from various sources for a specific purpose. A DF system, particularly an electronic nose (EN), was designed and built for a device as described in the patent “Portable equipment to avoid vehicle pollution in service stations” with resolution number 23016. It is an instrument for the classification of liquefied petroleum gases (LPG), the application of which is aimed at recognizing diesel fuel or gasoline, in the supply process. These gases are stored in the fuel tank of cars, and by identifying them take the actions required to prevent the contamination of vehicles with a fuel different from that used by the latter. A system that supports the operation of the EN was implemented. In the processing of the information provided by the sensors of the prototype the methodologies of machine learning, K-Nearest Neighbor and Naive Bayes, for lpg differentiation were used. Through a validity test, it was determined that the accuracy of the implemented techniques was 1, therefore, they are ideal methodologies for the classification of diesel fuel and gasoline in dynamic environments.

Keywords: Classification, data fusion, electronic nose, liquefied petroleum gases, machine learning techniques.

Origen de subvenciones o apoyos recibidos

Proyecto de investigación inscrito en la base de datos de la Vicerrectoría de Investigación, Innovación y Extensión de la Universidad Tecnológica de Pereira [1]. La información relevante se detalla a continuación:

—Código CIE: 9-19-1.

—Nombre del Proyecto: *Desarrollo del prototipo que aplica sobre la patente titulada: equipo portátil para evitar la contaminación vehicular en las estaciones de servicio.*

—Objetivo general: desarrollar el prototipo de un equipo capaz de identificar el tipo de combustible que usa un automóvil y que genere una señal eléctrica que dé cuenta de si existe o no correspondencia del carburante leído con el que se ha programado en el dispositivo.

—Objetivos específicos: 1) Diseñar e implementar el circuito electrónico basado en microcontroladores, que reciba las señales de los sensores asociados al sistema y controle los actuadores del prototipo. 2) Seleccionar el algoritmo para la identificación del tipo de gas y el control de las señales de salida e implementarlos en el microcontrolador. 3) Diseñar e implementar la caja contenedora que soporta la electrónica y sensórica del prototipo, que permita su instalación sobre la pistola de dosificación de combustible. 4) Validar la operación del prototipo con pruebas de laboratorio y de campo, usando en este último, vehículos tipo diésel y a gasolina.

INTRODUCCIÓN

FD es el estudio de técnicas eficientes para convertir la información, obtenida de diferentes fuentes, en una manera automática o semiautomática de manera que brinde un soporte efectivo en la toma de decisiones humanas o computadas [2]. Además, cuando se lleva a cabo una integración de información heterogénea —diversas modalidades², dimensionalidades y temporalidades— para obtener una mejor representación del objeto de estudio, FD suele conocerse con el nombre de Fusión de Datos Multi-Modal (FDMM) [3]. Según Kaur *et al.* en los años recientes se ha visto un crecimiento en la demanda de artículos científicos relacionados con FD debido a la necesidad de algoritmos de alto rendimiento; dada la capacidad que presentan las técnicas de aprendizaje de máquina —para estimar, clasificar, reducir la dimensionalidad y agrupar [3]—, se espera que mejoren el rendimiento de los algoritmos de FD [4].

² Una modalidad hace referencia a un canal que guarda la información de un sistema en particular [3].

Sin embargo, muchas investigaciones de FD, basado en técnica de aprendizaje de máquina, carecen de análisis acerca de la estabilidad de sus desarrollos frente a condiciones adversas³ [2]; de hecho, Ma *et al.* indicaron que las investigaciones enfocadas en FD 2D, como las imágenes de luz visible y luz infrarroja, típicamente suelen ignorar el ruido. Así mismo, los autores anotaron que el desarrollo de sistemas de FD también deberían enfocarse en las necesidades de aplicaciones concretas, ya que dichas investigaciones comúnmente se centran en mantener la información útil de los datos [5]. Aunado a lo anterior, Bokade *et al.* manifestaron la imperiosa necesidad del desarrollo de enfoques basado en técnicas de FD que aborden un número mayor de modalidades. Finalmente, los autores señalaron que no existe una base teórica lo suficientemente bien desarrollada para FDMM que permita determinar el mejor enfoque para la integración de un conjunto de datos disponible y lograr un propósito deseado en un dominio de aplicación, lo que evidencia un tema de investigación candente [3].

NE es una aplicación tecnológica, perteneciente al área de FDMM, la cual se define como un instrumento compuesto, principalmente, por un conjunto de sensores químicos ajustados para sensar diferentes tipos de gases, y dotado de un sistema de reconocimiento de patrones, capaz de identificar olores simples o complejos [6].

Existen variadas aplicaciones de NE desarrolladas en áreas afines al conocimiento teórico e industrial. Z. Wang P. *et al.* utilizaron en [7] este tipo de tecnología para detectar gases que podrían ser incoloros —tales como hidrógeno, metano, acetileno y propano—, con lo cual resolvieron problemas de sensibilidad cruzada del sensor por un proceso de entrenamiento a través de Back Propagation in Neural Network (BPNN). J. Palacín *et al.* utilizaron en [8] una matriz de 16 sensores digitales para detectar, de manera no específica, la calidad del aire y, por medio del procesamiento de datos: a) lograron reducir la dimensional al aplicar Principal Component Analysis (PCA) y b) por el método de KNN, consiguieron la clasificación de dos compuestos volátiles objetivo (etanol y acetona). S. Kiani *et al.* muestran en [9] un análisis de las posibles técnicas de muestreo más usadas según la literatura en NE, mediante el cual exponen las ventajas y desventajas en la utilización en términos de elección de técnica, tipo de muestra, sensibilidad, concentración, entre otras. En el campo de los hidrocarburos las NE tienen una figuración importante; es así que M. Ferreiro G. *et al.* indican en [10] cómo estos instrumentos son usados de manera alterna para hacer una rápida caracterización de diferentes muestras derivadas del petróleo —incluidos la gasolina, el diésel, los solventes aromáticos y el etanol— como una opción de los métodos utilizados en la actualidad, como la cromatografía.

³ El término de condiciones adversas hace referencia a las perturbaciones propias y externas de un sistema.

tografía de gases y la espectrofotometría [11], las cuales tienen algunas limitaciones en contraste con NE. H. A. Q. Vallejo muestra en [12] “la implementación de un dispositivo para la supervisión del suministro vehicular de gasolina o ACPM”, por lo que en [13] presenta la patente “Equipo portátil para evitar la contaminación vehicular en las estaciones de servicio” con resolución número 23016 de la Superintendencia de Industria y Comercio. Por otra parte, el proceso de detección llevado a cabo por los sensores provee una señal electrónica que debe ser filtrada, analizada y fusionada para extraer la información requerida. Así pues, existen diversas técnicas usadas para tal fin como las empleadas por Z. Wang, P. [7] y S. Kiani *et al.* [9], que detallan el uso de redes neuronales con el propósito de identificar diferentes clases de aromas. Wang, H. [14] y S. Bedoui, H. [15] usan máquinas de soporte vectorial como alternativa de redes neuronales artificiales para la identificación de diferentes tipos de compuestos, con lo cual superan las desventajas que presenta el entrenamiento de redes neuronales artificiales.

El abastecimiento de combustible vehicular en Colombia y algunos países de Latinoamérica es realizado manualmente por el asistente de la estación de servicio (islero) [16], el cual debe identificar el combustible requerido para todos los vehículos que ingresan a la estación. En algunos casos este reconocimiento se facilita si el vehículo cuenta con avisos visuales que indiquen el tipo de combustible; sin embargo, la mayoría de ellos no cuentan con esta ayuda. En razón a esto, existe la posibilidad de que el operario equivoque el tipo de carburante empleado y cause daños al sistema de inyección de combustible y al motor, así como contaminación ambiental, pues la mezcla de diésel y gasolina no puede ser reutilizada [12].

A nivel internacional, las estadísticas muestran numerosos casos de repostaje erróneo de combustible [17]; en las situaciones en las que el proceso fue realizado por un funcionario de la estación de servicio, se han presentado en algunos casos demandas legales, indemnizaciones económicas a los dueños de los vehículos y gastos derivados para la disposición final del combustible contaminado.

Lo anterior pone de manifiesto la necesidad de resolver un problema que genera pérdidas económicas y daños ambientales. Bajo esta perspectiva, el presente trabajo muestra una propuesta tecnológica que: 1) genera una aplicación concreta en el área de FD en la cual se consideran condiciones medio ambientales (perturbaciones), 2) permite identificar el tipo de combustible utilizado por un vehículo, 3) tiene el potencial de llegar a alertar y evitar el suministro del combustible de manera equivocada y 4) sirve de base para posteriores propuestas de implementación en el área de FDMM.

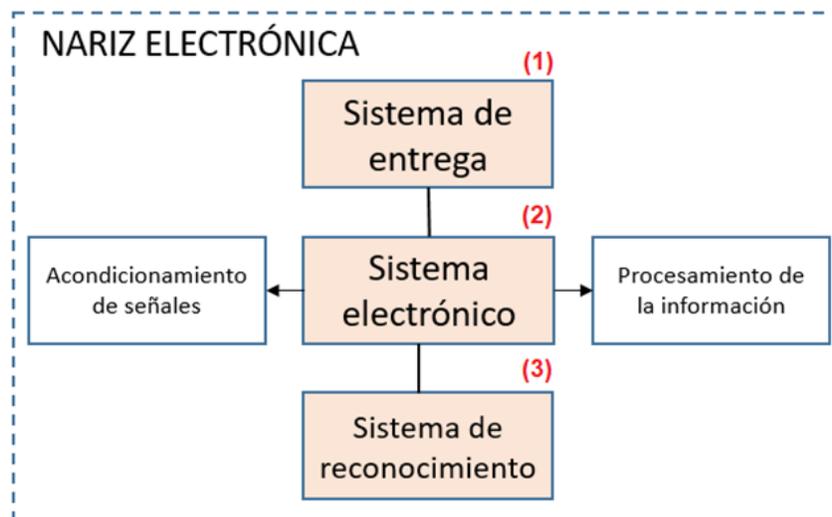
Concretamente, la investigación mostrada en este documento, la cual está basada en los desarrollos de la patente con resolución número 23016 de la Superintendencia de Industria y Comercio [13], está dirigida a detectar y reconocer GLP, en este caso, gasolina y ACPM, que son fundamentalmente una combinación de moléculas orgánicas

de propano y butano con trazas de otros compuestos [18], los cuales están contenidos en el depósito de combustible de los automóviles. A la par se practica una medición de las condiciones ambientales del entorno (temperatura, presión atmosférica y humedad relativa). La fusión de la información para la clasificación de los gases se hizo a través de dos métodos supervisados de aprendizaje de máquina: el KNN y Naive Bayes [19]. Las técnicas de análisis supervisadas, mencionadas anteriormente, fueron seleccionadas con base en las clases a las que se dirige la clasificación.

MATERIALES Y MÉTODOS

De acuerdo con la distribución que constituye una NE (Figura 1), el prototipo desarrollado tiene los siguientes componentes [20]:

- Sistema de entrega, que se compone de lo siguiente: a) elementos de conducción del gas conformado por bomba de aspiración, tuberías, acoples entre otros y b) un arreglo de sensores que permite la detección de diferentes tipos de gases, de acuerdo con la aplicación particular en la que se utilizará la NE.
- Sistema electrónico, el cual está constituido por diferentes elementos eléctricos y electrónicos, que facultan la obtención de la huella digital de las variables físico-químicas asociadas a la aplicación.
- Sistema de reconocimiento de patrones, en el cual los datos obtenidos por el sistema electrónico son finalmente analizados con técnicas de reconocimiento de patrones.

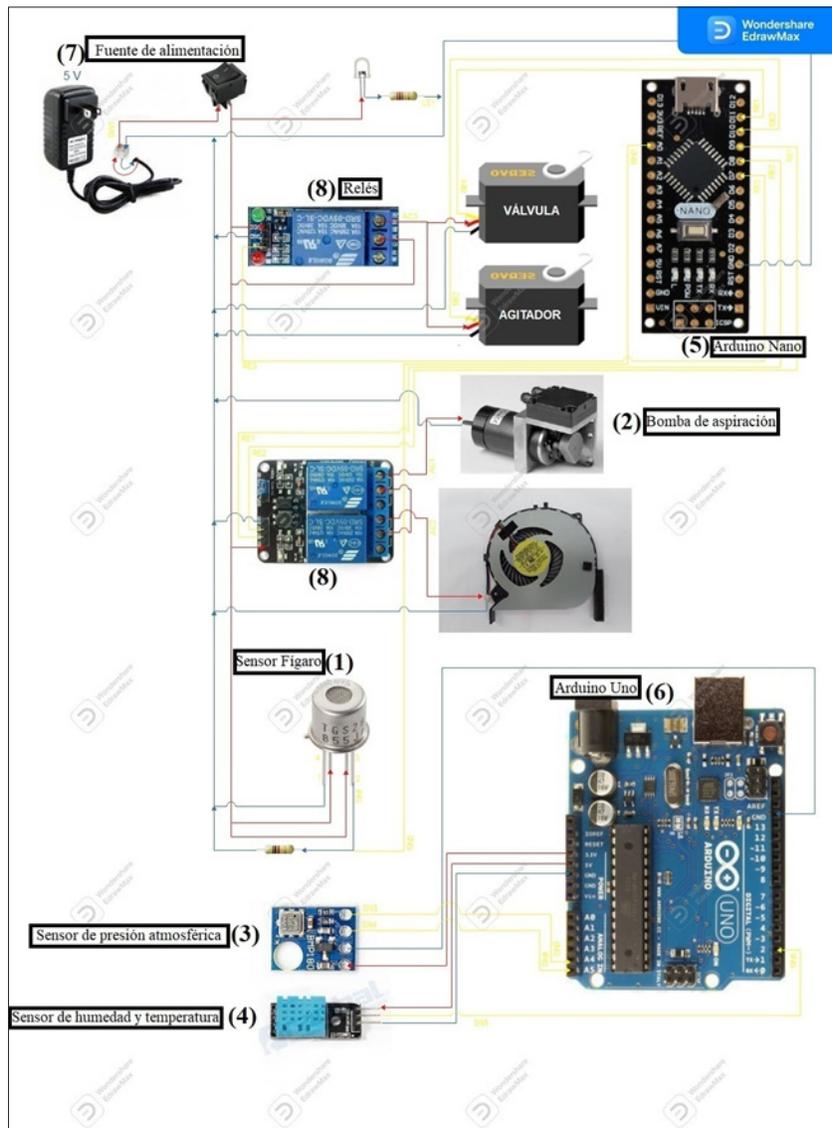


Fuente: los autores.

FIGURA 1. MÓDULOS DE LA NE.

Sistema electrónico

En la Figura 2, se presenta un esquema general del sistema eléctrico y electrónico que se desarrolló para la NE, en el cual se distinguen los siguientes elementos: un sensor Fígaro (1), una bomba de aspiración (2), un sensor de presión atmosférica (3), un sensor de humedad y temperatura (4), sistemas electrónicos digitales para la adquisición de los datos entregados por los sensores (5) y, para el manejo de las operaciones de mezclado y apertura de válvulas (6), una fuente de alimentación eléctrica (7) y relés para el manejo de los motores que tiene el sistema (8).



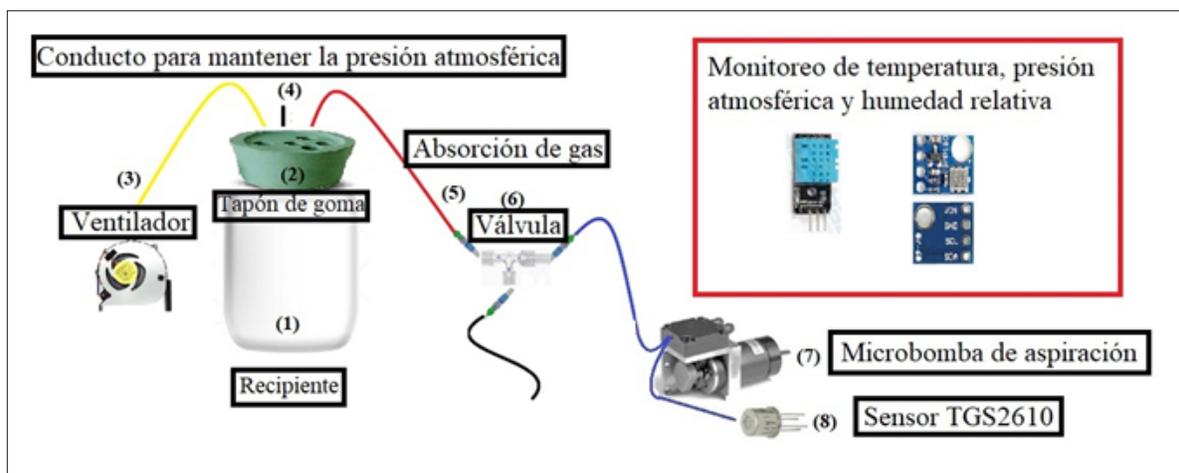
Fuente: los autores.

FIGURA 2. ESQUEMA ELECTRÓNICO DEL SISTEMA.

Sistema de entrega

El sistema de entrega es aquel que permite recoger y direccionar el vapor del combustible para llevarlo al sensor de gas; adicionalmente, este sistema permite conducir aire limpio para hacer el purgado o limpieza después de haber tomado una muestra.

En la Figura 3 se expone un esquema del sistema de entrega que se implementó para la NE, en el que se puede detallar: un recipiente (1) con capacidad aproximada de 25 ml, un tapón de goma (2) con tres orificios —en el que se encuentra el conducto para hacer burbujear del combustible por medio de un ventilador (3), el conducto que mantiene el interior del recipiente a presión atmosférica (4) y el conducto por el cual se aspira el vapor del carburante (5)—, una válvula de tres vías (6) que permite seleccionar el vapor del combustible o aire limpio para la purga del sistema de medición y una micro bomba de aspiración NMP015B (7) que hace circular el gas por el sensor Fígaro TGS2610 (8).



Fuente: los autores.

FIGURA 3. SISTEMA DE ENTREGA.

El proceso para la toma de las muestras a través del sistema de entrega de la Figura 3 fue automatizado mediante una aplicación desarrollada en LabVIEW, con base en la siguiente secuencia: agitar el recipiente (1) que contenía la muestra de combustible para generar vapores, abrir la válvula de tres vías (6) y encender la microbomba (7) de manera que los gases del recipiente sean conducidos hasta el sensor (8), hacer la toma de datos —humedad relativa, temperatura, presión atmosférica y concentración de gas— y purgar el sistema de entrega conmutando la válvula de tres vías (6) para que entre aire del ambiente; finalmente, detener y preparar al sistema para la siguiente operación.

Sistema de reconocimiento

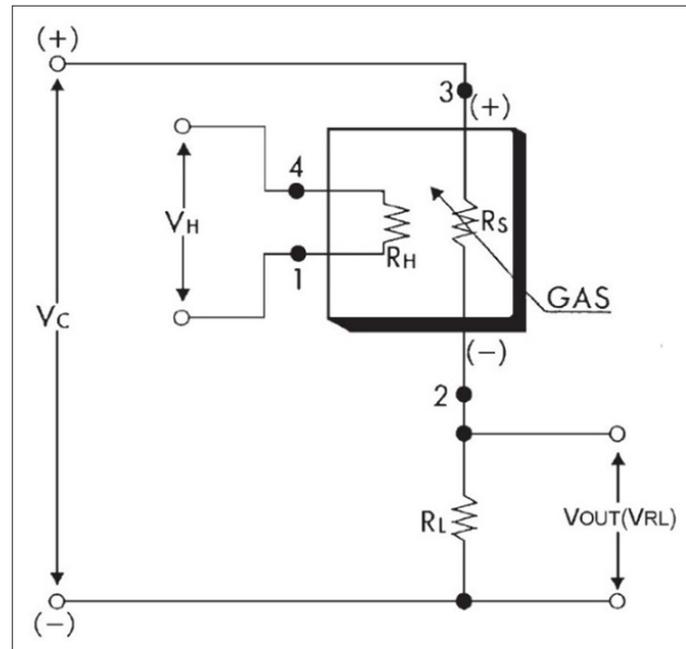
Este sistema está constituido por una tarjeta de desarrollo Arduino Nano la cual se encuentra enlazada con la aplicación desarrollada en el software LabVIEW; mediante este conjunto se leen y almacenan los datos que entrega el sensor de gas conectado a la tarjeta de desarrollo. El tiempo de muestreo usado fue de 20 ms y en cada lectura se tomaron, adicionalmente, los datos de presión, temperatura y humedad del gas, cuyos sensores se conectaron a una segunda tarjeta de desarrollo Arduino (UNO). Esta distribución fue realizada con el fin de no sobrecargar el procesamiento de las tarjetas de desarrollo.

El software de LabVIEW permitió generar una base de datos con una marca de tiempo de cada lectura, en una hoja de Excel; adicionalmente, se utilizó la librería “Diligent LINX” (Control Arduino, Raspberry Pi, BeagleBone), con el fin de facilitar la comunicación con el Arduino Nano como tarjeta de adquisición.

Otras librerías utilizadas fueron: módulo LabVIEW MathScript [21], que maximiza la utilización de la CPU y aumenta el rendimiento de datos sin cambiar la velocidad de transmisión. Para el Arduino UNO, encargado de las lecturas de los sensores DHT11 y BMP180, se utilizaron las librerías: Adafruit-BMP085-Library [22], DHT-sensor-library [23], Adafruit_Sensor [24] y Wire [25].

Por otro lado, para la detección del gas se usó un sensor marca Fígaro, con referencia TGS2610, que está diseñado para la detección de gases combustibles incluidos los GLP y que se clasifica como un sensor pasivo, resistivo y químico [26], el cual convierte la muestra de gas en una variación de resistencia eléctrica. Su distribución de pines y dimensiones puede observarse en [27]. Adicionalmente, las principales características eléctricas del sensor se pueden visualizar en [16], [27].

El circuito eléctrico para el montaje del sensor (Figura 4) fue realizado con base en los esquemas de la hoja de especificaciones del fabricante.



Fuente: información obtenida de referencia [27].

FIGURA 4. CIRCUITO ELÉCTRICO DEL SENSOR FÍGARO TGS2610.

En razón a que la sensibilidad del sensor Fígaro está dada por la relación entre los cambios de concentración de gas y las variaciones de la resistencia R_s [28], y con el fin de lograr los rangos de voltaje a la salida del sensor requeridos en el diseño electrónico (0 V-5 V), este dispositivo se calibró con la resistencia de carga (R_L) en un entorno que contenía la concentración del gas objetivo (ecuación 1), según como indica el fabricante [27].

$$V_{RL} = \left(\frac{R_L}{R_S + R_L} \right) V_C \quad (E1)$$

Se hicieron mediciones de ACPM y gasolina (gases objetivo), a un volumen y temperatura controlado, con diversos valores de resistencia (R_L), a saber: 1,0 k Ω , 1,5 k Ω y 2,0 k Ω . Experimentalmente se estableció un valor a R_L de 1,0 k Ω , que garantiza mediciones de gasolina sin saturación.

Toma de muestras

Los datos que se utilizaron en el experimento fueron obtenidos considerando tres ciudades con características de altitud, humedad y temperatura diferentes, y con variaciones importantes de estas. Estas mediciones se hicieron para adquirir datos en ambientes diversos, con la finalidad de obtener una base de datos rica multimodal.

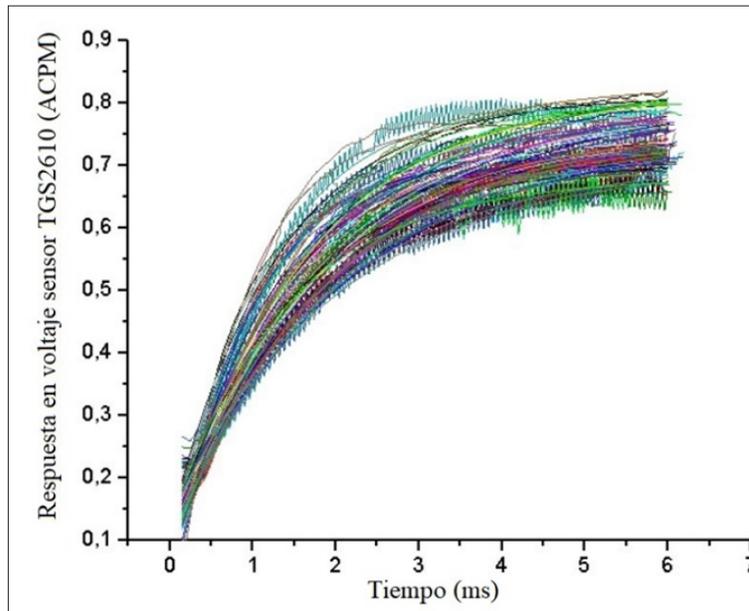
Las ciudades escogidas fueron Manizales (Caldas), Cartago (Valle del Cauca) y Pereira (Risaralda); en cada una se tomaron 58 muestras por combustible, ACPM y gasolina, para un total de 348 ensayos. Posteriormente, la información obtenida de los ensayos se organizó manualmente en seis archivos correspondientes a cada ciudad y combustible. El tiempo requerido para la toma de datos de cada ensayo fue de 45 segundos y entre ensayos de 1 minuto. Toda la operación fue controlada con la aplicación desarrollada en LabVIEW.

Preprocesamiento de los datos

Los seis archivos fueron procesados posteriormente en el programa ORIGIN, con el fin de determinar la ecuación matemática que mejor representara la respuesta del sensor frente a las concentraciones de ACPM y gasolina.

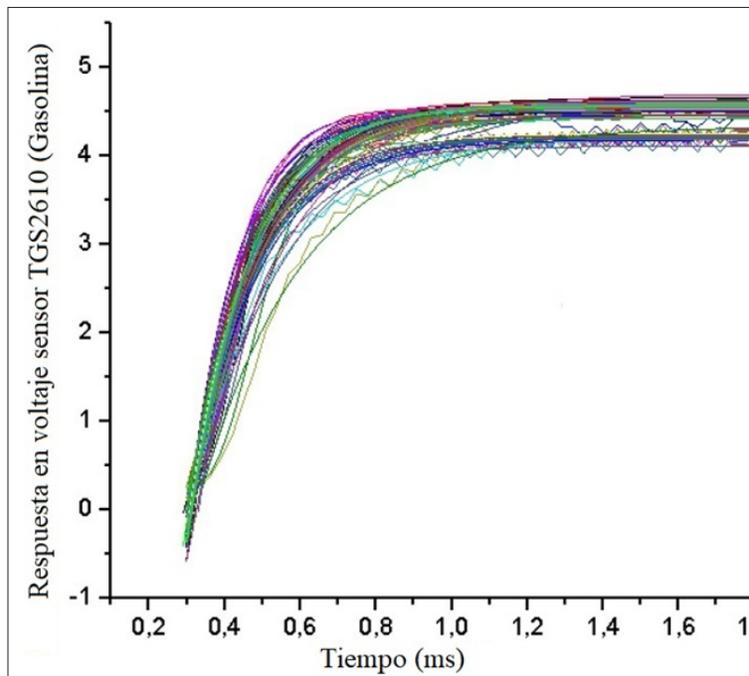
Para eliminar las variaciones de alta frecuencia y mejorar la obtención de la ecuación matemática, se utilizó el filtro de suavizado tipo Savitzky-Golay incluido en la herramienta “Smooth” de ORIGIN. El filtro Savitzky-Golay tiende a preservar las características de los datos y el perfil general, como la altura y el ancho del pico, realizando una regresión polinómica local alrededor de cada punto y creando un nuevo valor suavizado. Los parámetros utilizados para el filtro fueron: tamaño de la ventana 5, condición límite ninguna y orden polinomial 2.

En las Figuras 5 y 6 se presentan las curvas suavizadas para la respuesta en voltaje del sensor Fígaro con respecto al tiempo para los ensayos realizados.



Fuente: los autores.

FIGURA 5. RESPUESTAS SUAVIZADAS PARA LOS DATOS DE ACPM.



Fuente: los autores.

FIGURA 6. RESPUESTAS SUAVIZADAS PARA LOS DATOS DE GASOLINA.

Después del análisis, se obtuvo que la ecuación que mejor se ajustaba al comportamiento de los datos, tanto para gasolina como para ACPM, es una ecuación exponencial de la forma que se presenta en la ecuación 2.

$$y = y_0 + A \cdot e^{R_0 \cdot x} \quad (E2)$$

Donde x es el tiempo en milisegundos, y_0 representa la característica de la respuesta estacionaria, y los parámetros A y R_0 se asocian a la respuesta transitoria.

En la Tabla 1 se presenta información de ejemplo en la forma como se estructuró la base de datos multimodal que alimentó posteriormente el algoritmo de reconocimiento de patrones. En esta se observan los atributos obtenidos del preprocesamiento en el programa ORIGIN — y_0 , R_0 y A —, los atributos correspondientes a las variables ambientales medidas durante los ensayos —temperatura, presión y humedad—, un atributo discreto relacionado con la ciudad y la clase objetivo (relacionada con el tipo de combustible).

TABLA 1. MUESTRAS DE LA BASE DE DATOS

Yo	A	Ro	Temperatura (°C)	Presión (kPa)	Humedad (%)	Ciudad	Combustible
3,84	-2,24	-4,7	22,5	78,61	66	Manizales	Gasolina
4,63	-15,74	-3,67	26,95	85,49	72	Pereira	Gasolina
4,66	-29,57	-5,79	31,7	90,84	44	Cartago	Gasolina
0,52	-0,33	-0,6	24,6	78,53	61	Manizales	ACPM
0,76	-0,75	-1,41	28,55	85,48	72	Pereira	ACPM
0,811	-0,69	-0,76	31,65	90,76	42	Cartago	ACPM

Fuente: los autores.

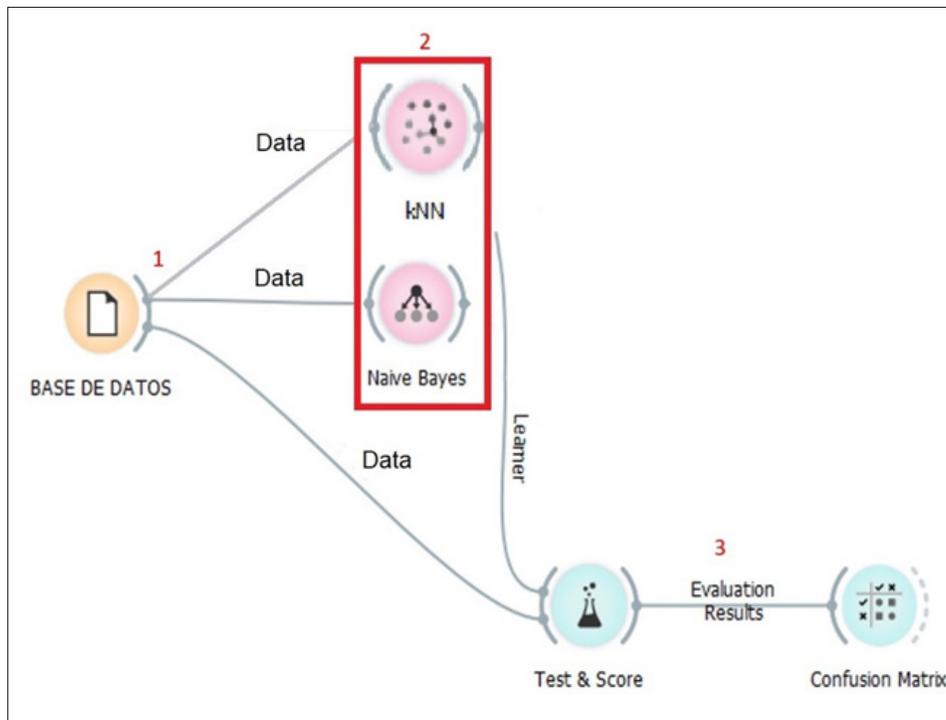
Procesamiento de los datos

Para la FD se utilizó la herramienta de software ORANGE de uso libre, la cual está basada en el lenguaje de programación Python. Dos técnicas supervisadas de análisis de datos fueron empleadas —KNN y Naive Bayes—, y la base de datos que alimenta el software fue organizada de acuerdo con la configuración que el programa requiere.

Se seleccionaron dos técnicas de aprendizaje de máquina supervisado, debido a que los atributos introducidos en la base de datos eran conocidos —características de los gases objetivo y mediciones de las condiciones ambientales— como también las clases para la clasificación —gasolina y ACPM— [20]; la estructura empleada en ORANGE, para la técnica KNN, se puede observar en la Figura 7.

En la Figura 7 se pueden distinguir principalmente tres partes: (1) la base de datos en la que fue ingresada la información recolectada en las diferentes ciudades, (2) el bloque de configuración del algoritmo KNN en el que se parametriza el número de vecinos y la medida de la distancia a utilizar, y (3) Prueba y Puntaje-Matriz de confusión que almacena la respuesta de la matriz de confusión.

Los parámetros utilizados en el algoritmo KNN fueron: la medida de la distancia euclidiana (dado que el orden de la magnitud de todos los atributos es semejante) y un número de vecinos igual a cinco. Para la visualización de los datos se agregaron las opciones *Scatter Plot*, *Data Table* y *Distributions*.



Fuente: los autores.

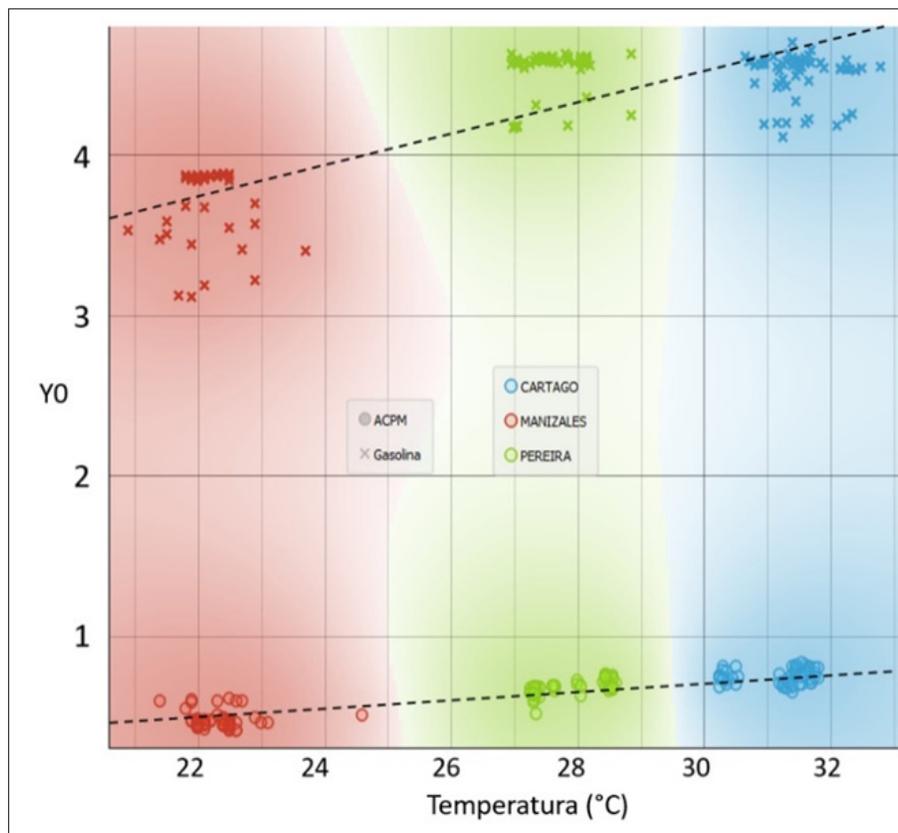
FIGURA 7. ESTRUCTURA EN ORANGE PARA KNN Y NAIVE BAYES.

La estructura del Naive Bayes parametrizada en ORANGE se puede observar en la Figura 7, la cual presenta el bloque del algoritmo Naive Bayes en la ecuación 2. Adicionalmente, se garantizó que la base de datos estuviera balanceada con el fin de evitar que el algoritmo se inclinará hacia la clase con más información.

RESULTADOS

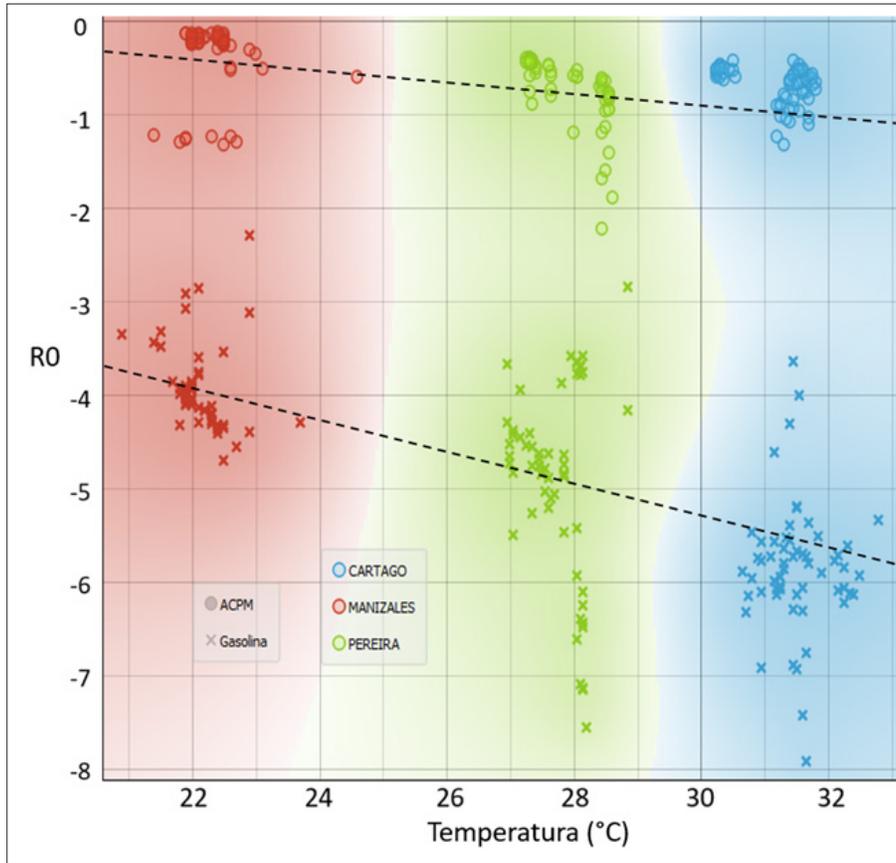
Las variaciones de las condiciones ambientales para los ensayos se encuentran en los siguientes rangos: temperatura entre 20,9 °C a 32,8 °C, presión atmosférica entre 76,43 kPa a 90,84 kPa y humedad relativa del aire entre 40% y 75%.

Las gráficas de dispersión en las Figuras 8, 9 y 10 muestran de manera común el comportamiento de los atributos γ_0 , R_0 y A con respecto a la temperatura, y la Figura 11, el atributo A con respecto a la presión atmosférica. Las regiones sombreadas en rojo, verde y azul, corresponden a las ciudades de Manizales, Pereira y Cartago, respectivamente. Adicionalmente, los datos marcados con círculos corresponden a ACPM, y los marcados con equis son gasolina.



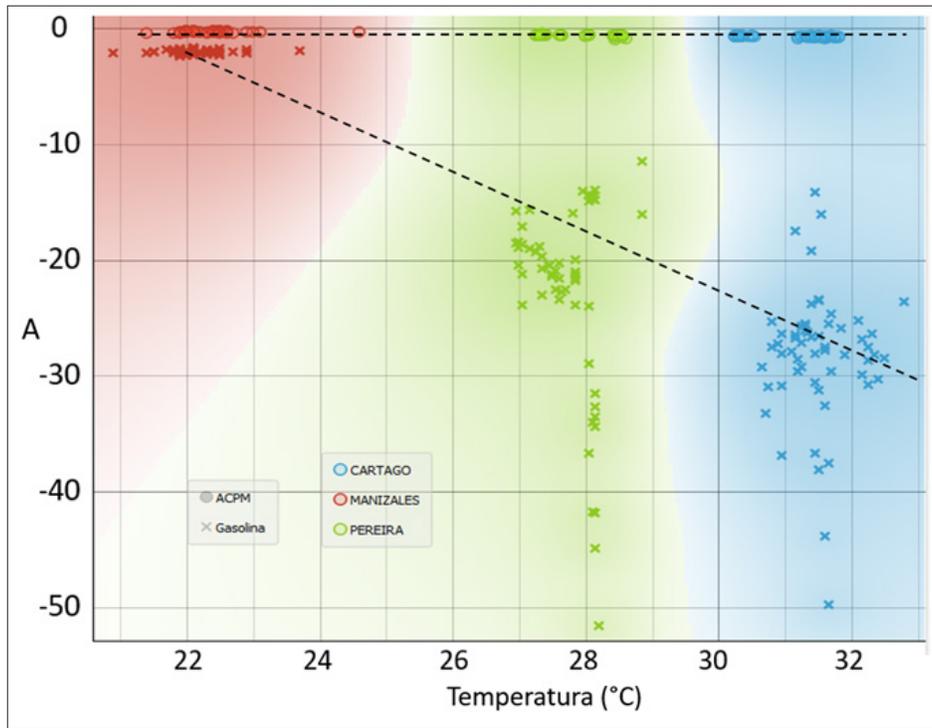
Fuente: los autores.

FIGURA 8. GRÁFICA DE DISPERSIÓN PARA YO Y TEMPERATURA.



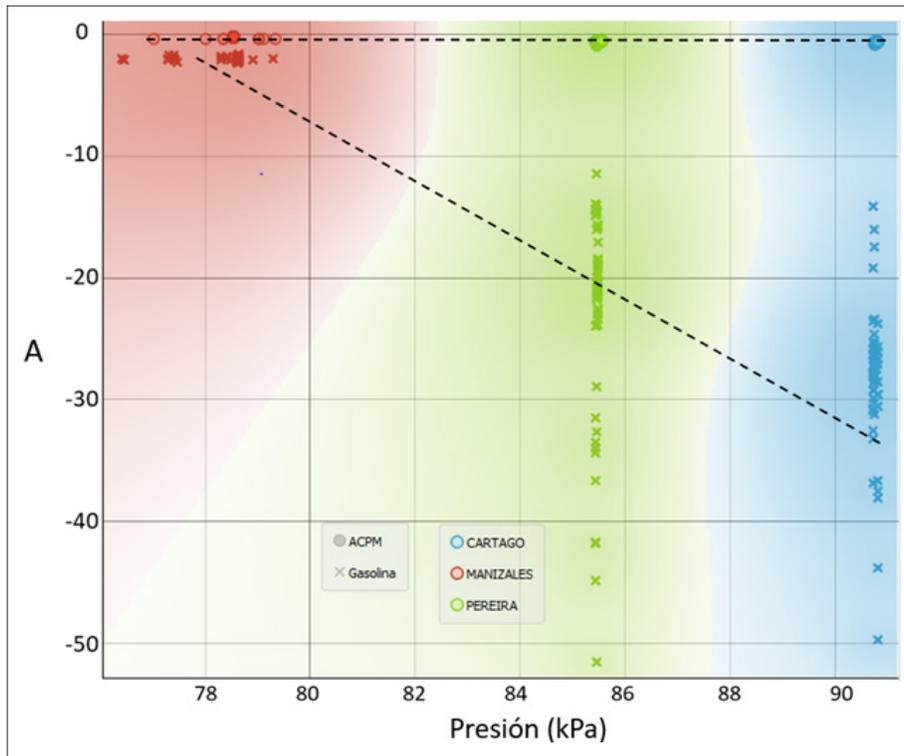
Fuente: los autores.

FIGURA 9. GRÁFICA DE DISPERSIÓN DE RO Y TEMPERATURA.



Fuente: los autores.

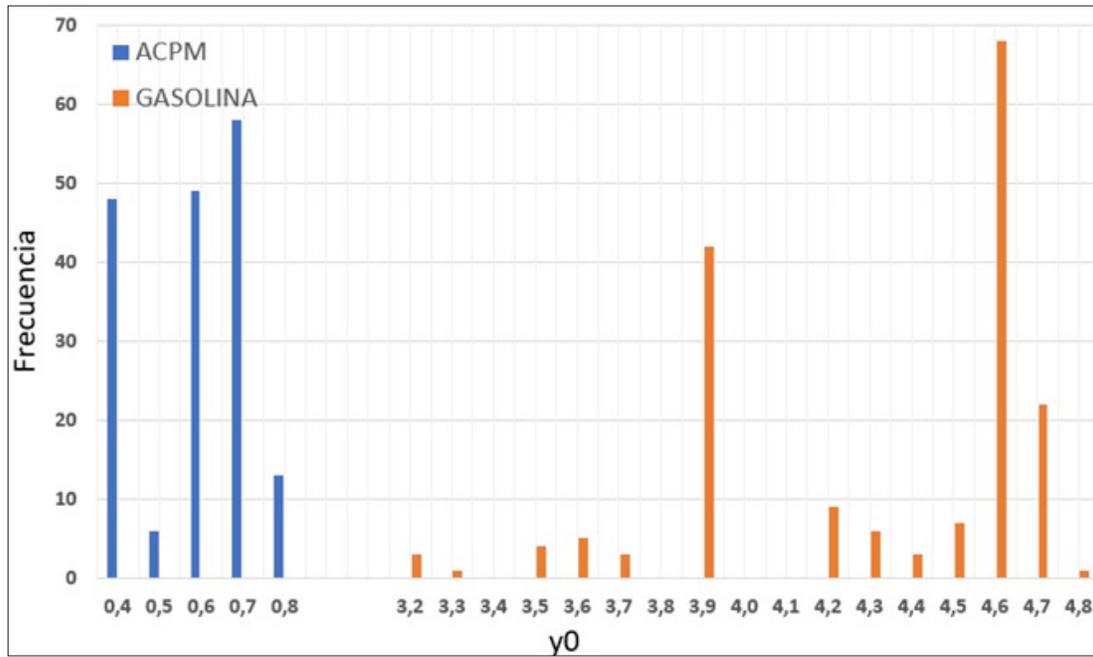
FIGURA 10. GRÁFICA DE DISPERSIÓN DE A Y TEMPERATURA.



Fuente: los autores.

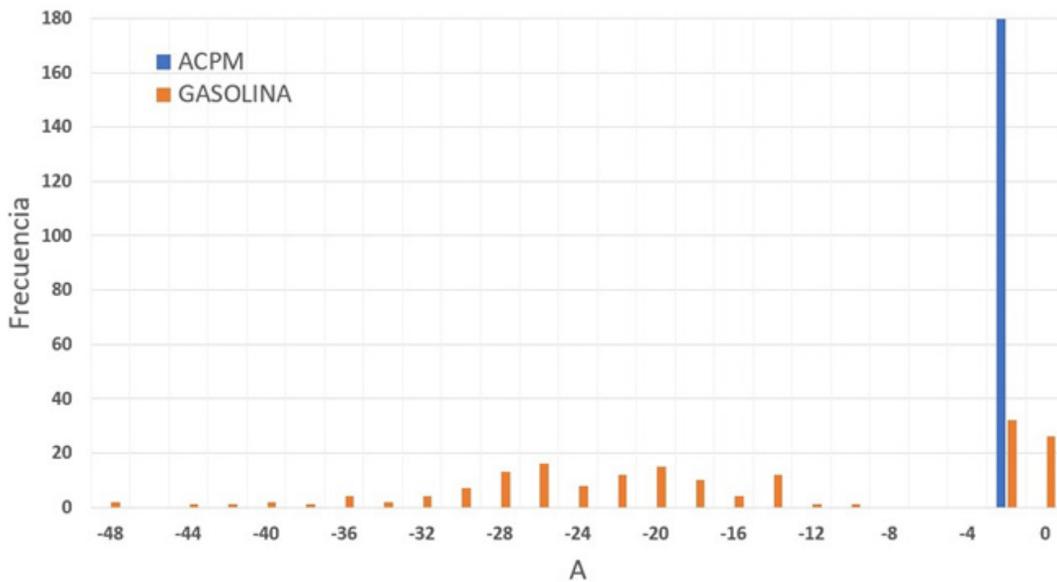
FIGURA 11. GRÁFICA DE A Y PRESIÓN.

Adicionalmente, se generaron gráficas que muestran la densidad de frecuencia en los atributos yo , A y Ro . Para tal fin se usó la opción *Distributions* contenida en el *software* ORANGE; así se pudo observar las diferencias de magnitud entre los combustibles para un mismo atributo. En las Figuras 12, 13 y 14 se muestran los resultados obtenidos tanto para gasolina como para ACPM.



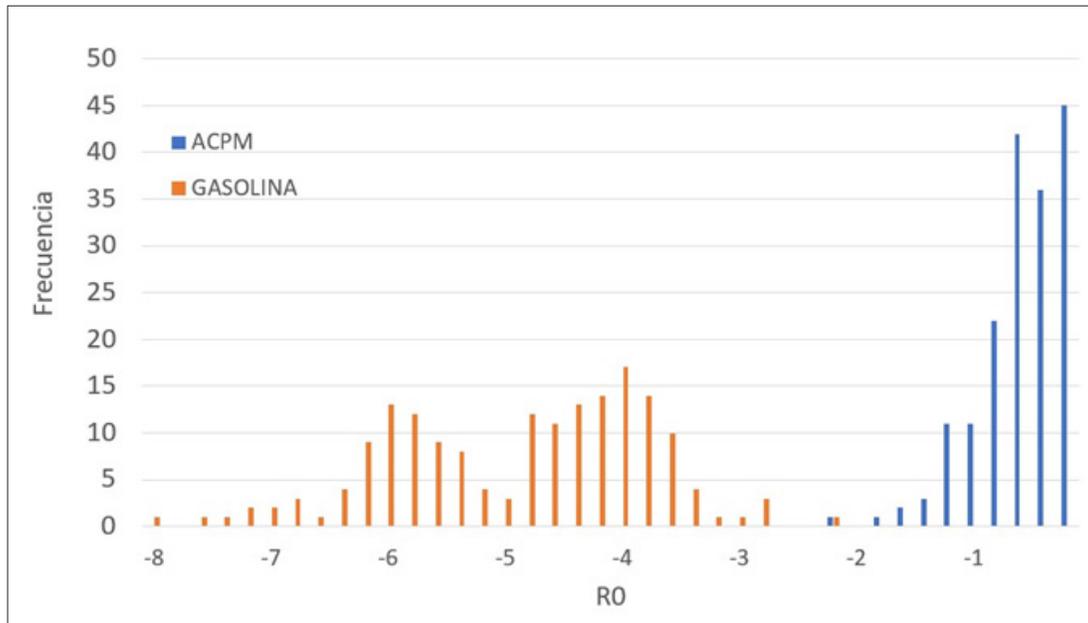
Fuente: los autores.

FIGURA 12. DENSIDAD DE LA VARIABLE YO.



Fuente: los autores.

FIGURA 13. DENSIDAD DE LA VARIABLE A.



Fuente: los autores.

FIGURA 14. DENSIDAD DE LA VARIABLE RO.

Finalmente, se verificó el nivel de acierto de los clasificadores, para lo cual se utilizó el método de validación cruzada en la matriz de confusión con un número de divisiones de 10.

De la respuesta de la matriz de confusión para el método KNN (Figura 15) se determinó que la tasa de verdaderos positivos (TPR) fue de 100%, la tasa de falsos positivos (FPR) fue de 0%, la tasa de verdaderos negativos (TNR) fue de 100%, la tasa de falsos negativos (FNR) fue de 0%. En consecuencia, la exactitud (E) obtenida fue de 1, como figura en la ecuación 3.

$$E = \frac{TPR + TNR}{TPR + TNR + FPR + FNR} = 1 \quad (E3)$$

		Predicted		Σ
		ACPM	Gasolina	
Actual	ACPM	100.0 %	0.0 %	174
	Gasolina	0.0 %	100.0 %	174
Σ		174	174	348

Fuente: los autores.

FIGURA 15. MATRIZ DE CONFUSIÓN MÉTODO KNN.

En relación con la respuesta de la matriz de confusión con el método de Naive Bayes (Figura 16), se pudo observar que la tasa de verdaderos positivos (TPR) fue de 100 %, la tasa de falsos positivos (FPR) fue de 0 %, la tasa de verdaderos negativos (TNR) fue de 100 %, la tasa de falsos negativos (FNR) fue de 0 %. Por lo tanto, la exactitud fue de 1 (cálculo semejante a la ecuación 3).

		Predicted		Σ
		ACPM	Gasolina	
Actual	ACPM	100.0 %	0.0 %	174
	Gasolina	0.0 %	100.0 %	174
Σ		174	174	348

Fuente: los autores.

FIGURA 16. MATRIZ DE CONFUSIÓN MÉTODO NAIVE BAYES.

DISCUSIÓN

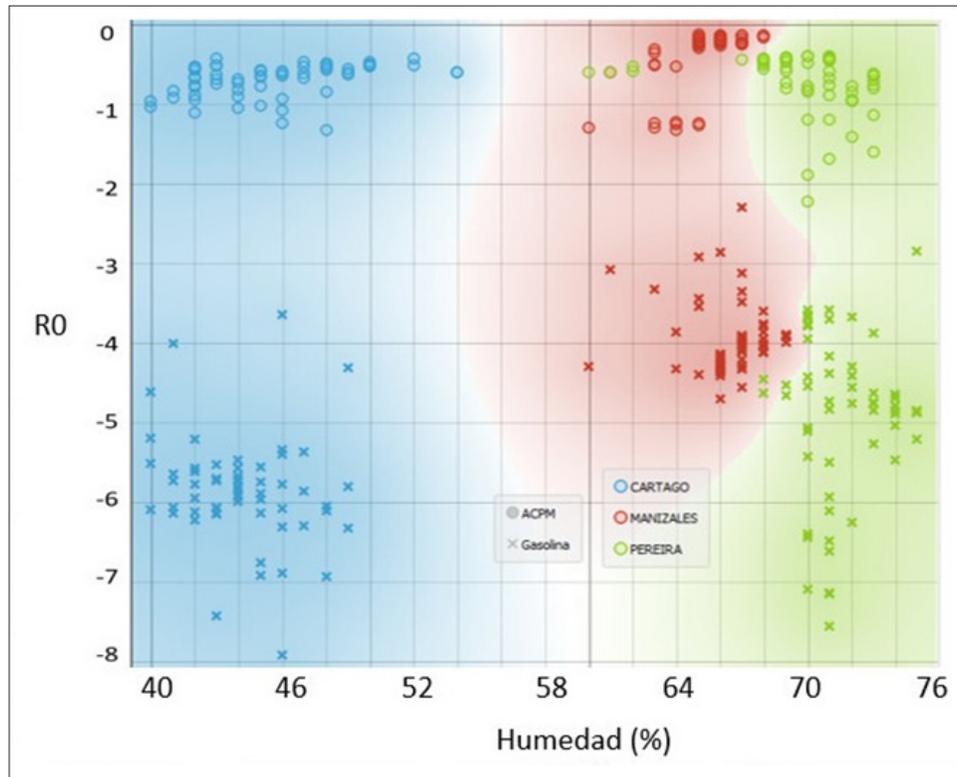
Las investigaciones realizadas en el campo de los hidrocarburos [10], [11], [29], [30] demuestran la clasificación de GLP a través de NE con objetivos en seguridad, caracterización y calidad de los compuestos derivados del petróleo. Sin embargo, es importante resaltar que en estas las condiciones adversas no se consideraron, mientras que el arreglo experimental diseñado e implementado en esta investigación se centraba

en determinar la premisa de, si bajo las condiciones ambientales variables (condiciones adversas de temperatura ambiente, presión atmosférica y humedad relativa) el equipo, basado en un sistema FDMM, podría hacer la diferenciación de GLP en una aplicación concreta.

En esta investigación se presentó la clasificación de GLP a través de una NE, para lo cual se usaron dos técnicas de aprendizaje de máquina, KNN y Naive Bayes. La exactitud para ambas técnicas fue de 1 como resultado de emplear una rica base de datos multimodal; por lo tanto, las metodologías empleadas son idóneas para la clasificación de ACPM y gasolina.

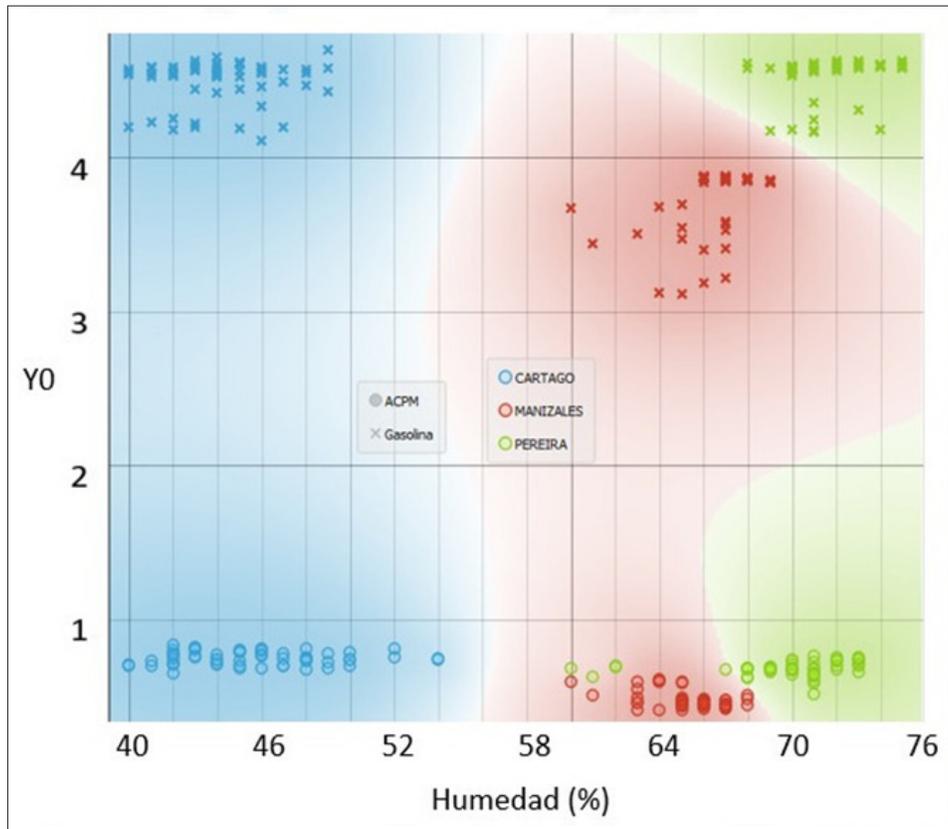
El análisis de los datos, mediante las gráficas de dispersión, permite reconocer la sensibilidad que tienen los parámetros de la gasolina — RO , YO y A — frente a las variaciones de humedad (Figuras 17, 18 y 19). Asimismo, en estas figuras se observa una menor sensibilidad de los parámetros del ACPM frente a dichas variaciones. Las gráficas de dispersión y de densidad muestran que los atributos — RO , YO y A — no se traslapan y que son altamente diferenciables, lo que demuestra la razón por la cual con un solo sensor de detección de GLP fue suficiente para discriminar entre gasolina y ACPM.

En la selección de la técnica más apropiada para la clasificación se consideraron el bajo consumo de recursos computacionales y la respuesta de la matriz de confusión. De lo anterior, se determinó que la técnica que consume una menor cantidad de recursos es el Naive Bayes, pues solo debe operar 14 valores correspondientes a 2 datos obtenidos de la desviación estándar y el promedio multiplicado por los 7 atributos, frente a la técnica KNN la cual debe comparar 2436 valores que comprenden 348 datos multiplicado por los 7 atributos.



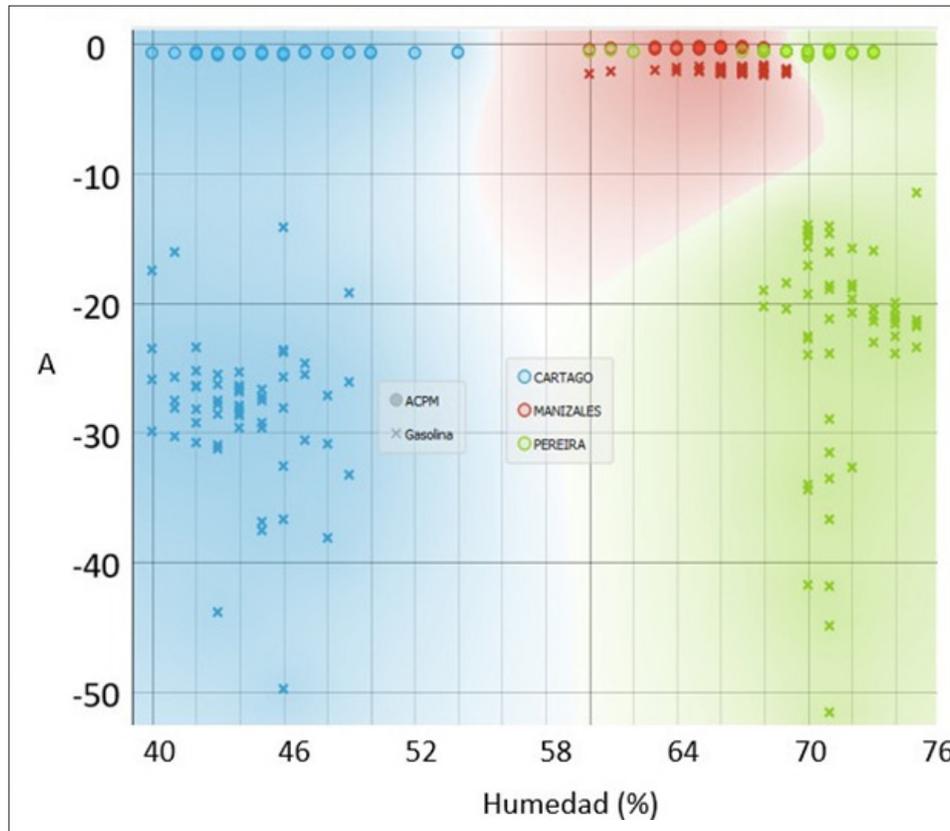
Fuente: los autores.

FIGURA 17. GRÁFICA DE RO Y HUMEDAD.



Fuente: los autores.

FIGURA 18. GRÁFICA DE YO Y HUMEDAD.



Fuente: los autores.

FIGURA 19. GRÁFICA DE A Y HUMEDAD.

CONCLUSIONES

La presente investigación responde a necesidades tanto en el área de FDMM —que abordan un gran número de modalidades por considerar condiciones dinámicas del entorno—, como en el desarrollo de una solución ingenieril a un problema latente. La tecnología implementada parte de una patente, de nombre “Equipo portátil para evitar la contaminación vehicular en las estaciones de servicio”, con resolución número 23016, con la cual se evidenció que en la actualidad ningún tipo de instrumento presenta semejanzas en su aplicación y funcionamiento.

El conjunto de la NE Desarrollada y las técnicas de clasificación empleadas permitieron la diferenciación de los GLP, ACPM y gasolina, en la totalidad de los ensayos realizados para tres ciudades con condiciones ambientales diferentes y no controladas. De esta manera, tanto el aparato como los algoritmos utilizados son una alternativa

tecnológica para ser implementados como parte de un sistema que pueda disminuir la probabilidad de contaminación vehicular en las estaciones de servicio.

Adicionalmente, se identificó que, entre las dos técnicas de clasificación utilizadas, la que presenta menor gasto computacional frente a una posible implementación es el Naive Bayes.

Es importante resaltar que en este artículo no se adelantó una evaluación que permitiera evidenciar la incidencia de cada atributo de forma individual respecto del comportamiento de la NE o sobre las técnicas de clasificación. En razón a esto, se considera relevante desarrollar estudios adicionales orientados a identificar el número de atributos necesarios o suficientes para la correcta operación del clasificador.

Finalmente, existen algunos elementos dentro del prototipo que hacen parte del experimento y que no se trataron en este manuscrito, dado que mucha de la información hace parte del contenido protegido a través de la patente indicada en acápites iniciales.

REFERENCIAS

- [1] Universidad Tecnológica de Pereira, “Vicerrectoría de Investigaciones, Innovación y Extensión: Proyectos”, Desarrollo del prototipo que aplica sobre la patente titulada: equipo portátil para evitar la contaminación vehicular en las estaciones de servicio, 2018, [https://www2.utp.edu.co/vicerrectoria/investigaciones/investigaciones/ DetallesProyecto/2222](https://www2.utp.edu.co/vicerrectoria/investigaciones/investigaciones/DetallesProyecto/2222)
- [2] T. Meng, X. Jing, Z. Yan y W. Pedrycz, “A survey on machine learning for data fusion”, *Inf. Fusion*, vol. 57, pp. 115-129, 2020, doi: <https://doi.org/10.1016/j.inffus.2019.12.001>.
- [3] R. Bokade *et al.*, “A cross-disciplinary comparison of multimodal data fusion approaches and applications: Accelerating learning through trans-disciplinary information sharing”, *Expert Syst. Appl.*, vol. 165, p. 113885, 2021, doi: <https://doi.org/10.1016/j.eswa.2020.113885>
- [4] H. Kaur, D. Koundal y V. Kadyan, “Image Fusion Techniques: A Survey”, *Arch. Comput. Methods Eng.*, vol. 28, n.º 7, pp. 4425-4447, 2021, doi: [10.1007/s11831-021-09540-7](https://doi.org/10.1007/s11831-021-09540-7)
- [5] J. Ma, Y. Ma y C. Li, “Infrared and visible image fusion methods and applications: A survey”, *Inf. Fusion.*, vol. 45, pp. 153-178, 2019, doi: <https://doi.org/10.1016/j.inffus.2018.02.004>
- [6] I. Moreno, R. Caballero, R. Galán, F. Matía y A. Jiménez, “La nariz electrónica: estado del arte”, *Rev. Iberoam. Automática e Informática Ind. RIAI*, vol. 6, n.º 3, pp. 76-91, jul. 2009, doi: [10.1016/S1697-7912\(09\)70267-5](https://doi.org/10.1016/S1697-7912(09)70267-5)

- [7] Z. Wang, P. Chen y X. Yang, “Application of electronic nose technology in detection of combustible gas”, *Proc. World Congr. Intell. Control Autom.*, pp. 6848-6852, 2010, doi: 10.1109/WCICA.2010.5554209
- [8] J. Palacín, E. Rubies, E. Clotet y D. Martínez, “Classification of Two Volatiles Using an eNose Composed by an Array of 16 Single-Type Miniature Micro-Machined Metal-Oxide Gas Sensors”, *Sensors 2022*, vol. 22, n.º 3, p. 1120, p. 1120, 2022, doi: 10.3390/S22031120
- [9] S. Kiani, S. Minaei y M. Ghasemi-Varnamkhasi, “Application of electronic nose systems for assessing quality of medicinal and aromatic plant products: A review”, *J. Appl. Res. Med. Aromat. Plants*, vol. 3, n.º 1, pp. 1-9, 2016, doi: 10.1016/J.JARMAP.2015.12.002
- [10] M. Ferreira-González, G. F. Barbero, M. Palma, J. Ayuso, J. A. Álvarez y C. G. Barroso, “Characterization and Differentiation of Petroleum-Derived Products by E-Nose Fingerprints.”, *Sensors (Basel)*., vol. 17, n.º 11, 2017, doi: 10.3390/s17112544
- [11] A. Oseev, M. Zubtsov y R. Lucklum, “Gasoline properties determination with phononic crystal cavity sensor”, *Sensors Actuators, B Chem.*, vol. 189, pp. 208-212, 2013, doi: 10.1016/j.snb.2013.03.072
- [12] H. A. Q. Vallejo, *Diseño y construcción de un sistema prototipo para la supervisión del suministro vehicular de gasolina o ACPM, en las estaciones de servicio*, Universidad Tecnológica de Pereira, 2014.
- [13] H. A. Q. Vallejo, *Equipo Portátil para Evitar la Contaminación Vehicular en las Estaciones de Servicio*, 2017.
- [14] X.-D. W. X.-D. Wang, H.-R. Z. H.-R. Zhang y C.-J. Z. C.-J. Zhang, “Signals recognition of electronic nose based on support vector machines”, *Int. Conf. Mach. Learn. Cybern.*, vol. 6, n.º. ago., pp. 18-21, 2005, doi: 10.1109/IPCon.2016.7831065
- [15] S. Bedoui, H. Samet, M. Samet y A. Kachouri, “Gases identification with Support Vector Machines technique (SVMs)”, in 2014 1st International Conference on Advanced Technologies for Signal and Image Processing, ATSSIP 2014, 2014, pp. 271-276, doi: 10.1109/ATSSIP.2014.6834620
- [16] V. A. Zarate, *Diseño de un manual de procedimientos y medidas de seguridad para una estación de servicio de combustible en zona rural colombiana*, Universidad Católica de Colombia, 2021.
- [17] Elmotor, “Qué debes hacer si en la gasolinera te ponen el combustible equivocado”, 25 de noviembre de 2017, 2017, <https://motor.elpais.com/conducir/que-pasa-combustible-equivocado/>
- [18] M. de C. S. A. E. S. P, *Manual de uso GLP* [online], p. 15, 2020, http://multigasdecolombia.com.co/archivos/manejo_seguro.pdf

- [19] N. A. Pérez, *Aplicabilidad del procesamiento de lenguaje natural al análisis de errores en logs de aplicativos utilizados en el área de BSS (Business Support System)*, Universidad de Valladolid, 2019.
- [20] T. C. Pearce, S. S. Schiffman, H. T. Nagle y J. W. Gardner, *Handbook of machine olfaction: electronic nose technology*, John Wiley & Sons, 2006.
- [21] N. I. Corporation, *Módulo LabVIEW*, 2019, <https://rb.gy/bi12n>
- [22] A. Industries, *BMP085 Library*, 2014. <https://github.com/adafruit/Adafruit-BMP085-Library>
- [23] A. Industries, *DHT sensor library*, 2015. <https://github.com/adafruit/DHT-sensor-library>
- [24] A. Industries, *Common sensor library*, 2017, https://github.com/adafruit/Adafruit_Sensor
- [25] Arduino.cc, *Arduino-Wire*, <https://www.arduino.cc/en/Reference/Wire>
- [26] A. D. Wilson y M. Baietto, “Advances in electronic-nose technologies developed for biomedical applications”, *Sensors*, vol. 11, n.º 1, pp. 1105-1176, 2011, doi: 10.3390/s110101105
- [27] F. An ISO9001 Company, *TGS 2610-for the detection of Combustible Gases*, [online], p. 13, 2000, <http://www.meditronik.com.pl/doc/bo-b9999/tgs2610.pdf>
- [28] S. R. Bosch, *Diseño y realización de una nariz electrónica para la discriminación de aceites*, Universitat Rovira i Virgili, 2001.
- [29] M. Ferreiro-González, J. Ayuso, J. A. Álvarez, M. Palma y C. G. Barroso, “Application of an HS-MS for the detection of ignitable liquids from fire debris”, *Talanta*, vol. 142, pp. 150-156, 2015, doi: 10.1016/j.talanta.2015.04.030
- [30] T. Sobański, A. Szczurek, K. Nitsch, B. W. Licznerski y W. Radwan, “Electronic nose applied to automotive fuel qualification”, *Sensors Actuators, B Chem.*, vol. 116, n.º 1-2, pp. 207-212, 2006, doi: 10.1016/j.snb.2005.11.087