

Jornadas de Automática

Sistema eNose BME688 para clasificar aceites de oliva por calidad

Paula Molina^a, Diego M. Martínez^{a,b,*}, Silvia Satorres^{a,b}, Sergio Illana Rico^{a,b}, Erbert Osco^c, Ildefonso Ruano Ruano^a

^aGrupo de Robótica, Automática y Visión por Computador, Departamento de Ingeniería Electrónica y Automática, Universidad de Jaén, Campus Las Lagunillas s/n, Jaén 23071, España.

^bEl Instituto Universitario de Investigación en Olivar y Aceites de Oliva (INUO), Universidad de Jaén, Campus Las Lagunillas s/n, Jaén 23071, España.

^cDepartamento de Informática y Sistemas, Universidad Nacional Jorge Basadre Grohmann, Tacna 23003, Perú.

To cite this article: Molina, P., Martínez, D.M., Satorres, S., Illana, S., Osco, E., Ruano, I. 2026. BME688 eNose system for classifying olive oils by quality. *Jornadas de Automática*, 47.
<https://doi.org/10.17979/ja-cea.2026.47.13715>

Resumen

Este trabajo presenta el desarrollo y evaluación de una nariz electrónica basada en sensores BME688 para la clasificación de aceites de oliva virgen y virgen extra. El sistema combina una cámara de medida modular, perfiles térmicos programables, extracción de características temporales y técnicas de aprendizaje automático. A partir de las respuestas de tres sensores y diez pasos térmicos por sensor se obtuvieron 180 características por muestra. Se evaluaron distintos clasificadores lineales y no lineales mediante validación cruzada estratificada, así como estrategias de selección de características. El mejor rendimiento se obtuvo con una máquina de soporte vectorial (SVM) lineal combinado con SelectKBest y filtrado de variables altamente correlacionadas, reduciendo el conjunto de entrada a 26 características y alcanzando una precisión balanceada de $0,822 \pm 0,058$. Los resultados muestran el potencial del sistema para discriminar ambas categorías, aunque la validación entre repeticiones evidencia la necesidad de mejorar la robustez frente a variabilidad experimental y deriva instrumental.

Palabras clave: Procesado postcosecha y alimentario, Sistemas de clasificación y evaluación de calidad, Biosensores en agricultura, Reconocimiento de patrones e IA en agricultura, Aprendizaje automático, Validación de modelos

BME688 eNose system for classifying olive oils by quality

Abstract

This work presents the development and evaluation of an electronic nose based on BME688 sensors for the classification of virgin and extra virgin olive oils. The system combines a modular measurement chamber, programmable thermal profiles, temporal feature extraction, and machine learning techniques. From the responses of three sensors and ten thermal steps per sensor, 180 features were extracted for each sample. Linear and non-linear classifiers were evaluated using stratified cross-validation, together with different feature selection strategies. The best performance was obtained with a linear Support Vector Machine (SVM) combined with SelectKBest and filtering of highly correlated variables, reducing the input space to 26 features and achieving a balanced accuracy of $0,822 \pm 0,058$. The results show the potential of the proposed system to discriminate between both categories, although validation across independent repetitions highlights the need to improve robustness against experimental variability and sensor drift.

Keywords: Post-harvesting and food processing, Grading systems and Quality assessment, Biosensors in agriculture, Pattern recognition and AI in agriculture, Machine Learning, Model Validation

*Autor para correspondencia: dmgila@ujaen.es

1. Introducción

La producción de aceite de oliva virgen (AOV) es una actividad económica en la que España es líder mundial, representando el 46.1% de la producción mundial, con la provincia de Jaén aportando el 50% de esta producción. Esto hace que el AOV sea un producto de elevado valor económico cuya calidad puede verse afectada por procesos de oxidación, almacenamiento inadecuado o adulteración, provocando cambios en el perfil aromático y la aparición de compuestos volátiles asociados a rancidez y defectos sensoriales que a su vez pueden mermar su valor nutricional (Lechhab et al., 2022). Los métodos convencionales empleados para evaluar la calidad de aceites son llevados a cabo en laboratorios acreditados y ofrecen resultados fiables, pero presentan limitaciones relacionadas con el coste, el tiempo de análisis y la necesidad de equipamiento especializado y personal entrenado (Ghniimi et al., 2024). Por ello, existe un interés creciente en el desarrollo de metodologías instrumentales rápidas, portátiles y no destructivas que permitan realizar evaluaciones objetivas de calidad directamente en entornos productivos o de control industrial (Ortega et al., 2016).

En este contexto, las narices electrónicas (*electronic noses*, eNose) han emergido como una alternativa prometedora para la caracterización de aceites mediante el análisis de compuestos orgánicos volátiles. Estos sistemas combinan matrices de sensores químicos con técnicas de reconocimiento de patrones y aprendizaje automático para generar huellas aromáticas características asociadas a diferentes estados de calidad o tipos de aceite (Devika et al., 2026). Diversos estudios han demostrado la viabilidad de sensores MOS para aplicaciones relacionadas con autenticidad, rancidez, adulteración y clasificación de aceites de oliva (Fernandez et al., 2025; Amini et al., 2026; Martínez et al., 2026; Soto et al., 2024; Gila et al., 2020). Asimismo, algoritmos de aprendizaje automático han mostrado un elevado potencial para discriminar perfiles aromáticos complejos y mejorar el rendimiento predictivo de estos sistemas (Xu et al., 2021).

En este trabajo se presenta el desarrollo y evaluación de una nariz electrónica basada en sensores BME688 para la clasificación de aceites de oliva virgen y virgen extra mediante el análisis de su perfil aromático. El sistema propuesto combina perfiles térmicos programables, extracción de características temporales y técnicas de aprendizaje automático con el objetivo de evaluar la capacidad discriminante del dispositivo e identificar las variables más relevantes para la clasificación.

2. Materiales y métodos

2.1. Diseño del sistema experimental

El sistema experimental es una nariz electrónica modular diseñada para analizar compuestos volátiles de aceite de oliva. Su arquitectura vertical consta de tres módulos: ventilación inferior, cámara central de medida y electrónica de control superior (Figura 1).

La cámara de medida (volumen aproximado de 4 L) aloja la muestra e integra en su parte superior una matriz de tres sensores de gas BME688. Estos sensores operan bajo perfiles térmicos programables para registrar la respuesta frente a los compuestos liberados en condiciones controladas.

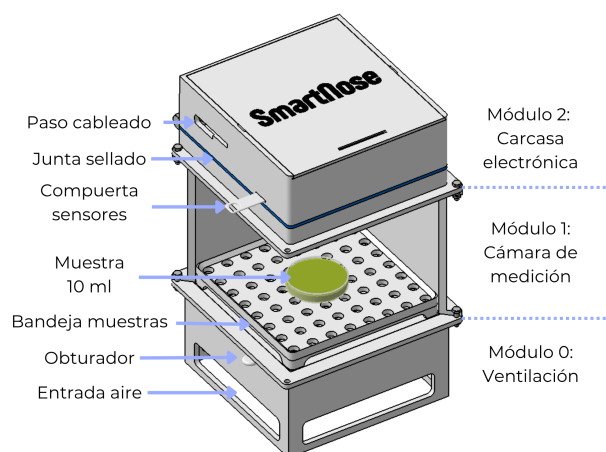


Figura 1: Representación tridimensional del prototipo. Se muestra la arquitectura modular del dispositivo, señalando características operativas clave. El detalle seccionado de la cámara de medición ilustra la colocación de las muestras durante los ensayos.

El control recae sobre un microcontrolador ESP32-S3 Nano que gestiona la adquisición sensorial vía Interfaz Serie para Periféricos (SPI) y la ventilación activa mediante Modulación por Ancho de Pulso (PWM). En cuanto a la alimentación, la etapa de potencia y ventilación utiliza una fuente externa de 12 V, mientras que la unidad de control recibe 5 V vía USB desde el ordenador (empleado también para la configuración y almacenamiento de datos). Los sensores operan a 3.3 V suministrados por la propia unidad de control (Figura 2).

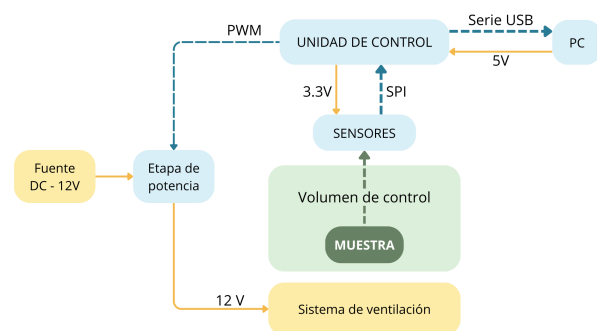


Figura 2: Diagrama de bloques del dispositivo experimental. Se muestran las conexiones de alimentación y comunicación entre la unidad de control, la matriz de sensores BME688, el sistema de ventilación activa y el ordenador de adquisición.

El módulo inferior incorpora un ventilador activo y compuertas mecánicas para alternar entre las fases de exposición y purgado, garantizando condiciones reproducibles entre medidas consecutivas. Esta disposición modular facilita significativamente el montaje, la limpieza y la sustitución de componentes durante campañas experimentales extensas.

2.2. Muestras de aceite de oliva empleadas en el estudio

Para el estudio se emplearon 125 muestras de aceite de oliva (78 virgen extra y 47 virgen), previamente caracterizadas mediante análisis sensorial de laboratorio y categorizadas según la normativa reglamentaria.

La información de referencia (*ground truth*) de cada muestra incluyó su código, categoría, fecha de recepción y, cuando

estaban disponibles, las medianas sensoriales de frutado y defecto (Figura 3). Estos datos se utilizaron exclusivamente como variable objetivo, y nunca como entrada para los modelos de clasificación.

En cada ensayo se distribuyó un volumen estandarizado de aceite en placas de Petri para maximizar la superficie de contacto con la atmósfera interna de la cámara. Las medidas se realizaron a temperatura ambiente y sin precalentamiento externo, favoreciendo así la liberación natural de compuestos volátiles y garantizando condiciones experimentales homogéneas.

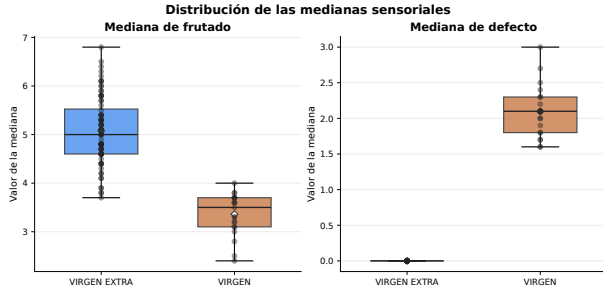


Figura 3: Distribución de las medianas sensoriales de frutado y defecto para las muestras incluidas en el estudio, diferenciando entre aceites de oliva virgen extra y virgen.

2.3. Procedimiento de medida

El procedimiento de medida se estructuró en dos fases consecutivas de 10 minutos cada una. En la fase inicial de ventilación, con la cámara abierta y el ventilador activo, se renovó la atmósfera interna con aire ambiente para estabilizar los sensores. El último valor registrado en esta etapa se empleó posteriormente como referencia de línea base.

A continuación, durante la fase de medición, la cámara permaneció cerrada. De este modo, los sensores BME688 quedaron expuestos a los compuestos volátiles liberados por la muestra de aceite, registrándose la evolución temporal de sus respuestas.

Cada sensor se operó mediante un perfil térmico específico compuesto por ciclos de 26 segundos (Figura 4, 5 y 6), lo que permite obtener respuestas diferenciadas bajo distintas temperaturas. Dado que en cada ciclo se tomaron diez lecturas asociadas a los pasos térmicos de cada sensor, para cada instante de análisis se generaron un total de 30 señales de gas (tres sensores por diez pasos).

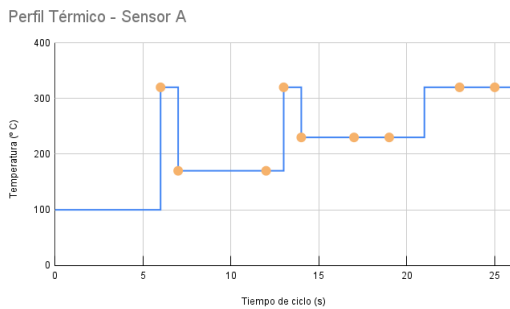


Figura 4: Perfil térmico del sensor BME688 definido como [A].

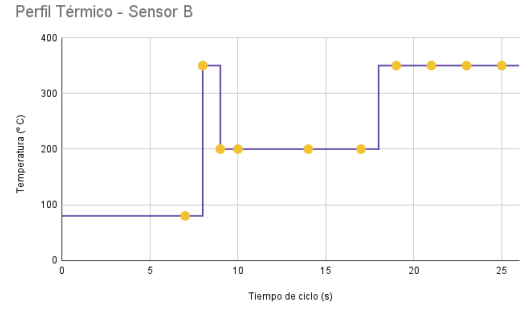


Figura 5: Perfil térmico del sensor BME688 definido como [B].

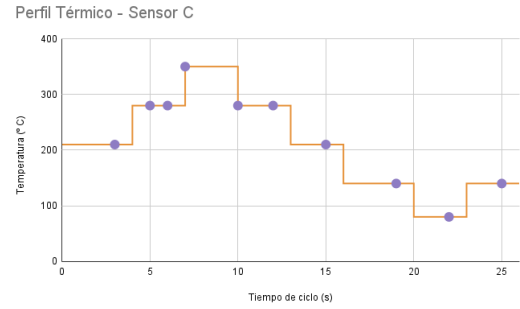


Figura 6: Perfil térmico del sensor BME688 definido como [C].

2.4. Extracción de características

Para cada muestra analizada se registró un conjunto de señales procedentes de los tres sensores BME688 integrados en el sistema eNose, identificados como $s \in \{A, B, C\}$. Cada sensor fue operado utilizando un perfil térmico compuesto por diez pasos de temperatura, denotados como $p \in \{1, \dots, 10\}$. En consecuencia, cada ensayo generó señales asociadas a todas las combinaciones sensor-paso, representadas como $x_{s,p}(t)$.

Cada ensayo experimental consta de dos etapas diferenciadas: una fase inicial de ventilación, destinada a estabilizar la línea base de los sensores, y una fase de medición, durante la cual los sensores quedan expuestos a los compuestos volátiles liberados por la muestra de aceite. Con el objetivo de reducir diferencias de línea base entre ensayos, cada señal de medición fue referenciada respecto al último valor registrado durante la fase de ventilación. Para cada combinación sensor-paso, la referencia se definió como:

$$x_{s,p}^{\text{ref}} = x_{s,p}(t_{\text{vent, end}}) \quad (1)$$

donde $t_{\text{vent, end}}$ representa el instante final de la fase de ventilación.

Posteriormente, la señal correspondiente a la fase de medición se transformó mediante:

$$\tilde{x}_{s,p}(t_i) = x_{s,p}(t_i) - x_{s,p}^{\text{ref}}, \quad t_i \in \mathcal{T}_{\text{med}} \quad (2)$$

donde $\mathcal{T}_{\text{med}}$ denota el conjunto de instantes pertenecientes a la fase de medición. Esta normalización, definida en la Ecuación (2), permite expresar la respuesta del sensor como una variación relativa al estado inmediatamente anterior a la exposición de la muestra, mejorando así la comparabilidad entre ensayos.

A partir de cada señal referenciada $\tilde{x}_{s,p}(t_i)$ se calcularon cinco descriptores estadísticos y dinámicos. El valor medio de la respuesta se obtuvo como:

$$\mu_{s,p} = \frac{1}{N} \sum_{i=1}^N \tilde{x}_{s,p}(t_i) \quad (3)$$

donde N representa el número total de muestras temporales durante la fase de medición.

El valor máximo de la respuesta se definió como:

$$M_{s,p} = \max_{i=1,\dots,N} \tilde{x}_{s,p}(t_i) \quad (4)$$

El valor final de la señal se calculó como:

$$F_{s,p} = \tilde{x}_{s,p}(t_N) \quad (5)$$

donde t_N corresponde al último instante registrado durante la fase de medición.

Con el fin de caracterizar la dinámica temporal de la respuesta, se estimó la pendiente de la señal mediante un ajuste lineal por mínimos cuadrados:

$$\tilde{x}_{s,p}(t_i) \approx a_{s,p}t_i + b_{s,p} \quad (6)$$

donde el coeficiente $a_{s,p}$ fue empleado como descriptor de pendiente:

$$P_{s,p} = a_{s,p} \quad (7)$$

Adicionalmente, el área bajo la curva (*Area Under the Curve*, AUC) se calculó mediante integración trapezoidal:

$$AUC_{s,p} = \sum_{i=1}^{N-1} \frac{\tilde{x}_{s,p}(t_i) + \tilde{x}_{s,p}(t_{i+1})}{2} (t_{i+1} - t_i) \quad (8)$$

De este modo, para cada combinación sensor-paso se obtuvo el siguiente conjunto de descriptores derivados de la fase de medición:

$$[\mu_{s,p}, M_{s,p}, P_{s,p}, AUC_{s,p}, F_{s,p}] \quad (9)$$

Además, se conservó el valor de referencia $x_{s,p}^{\text{ref}}$, definido en la Ecuación (1), correspondiente al final de la fase de ventilación, con el objetivo de mantener información asociada a la línea base previa a la exposición de la muestra. En consecuencia, cada combinación sensor-paso quedó finalmente representada por seis variables:

$$\mathbf{z}_{s,p} = [x_{s,p}^{\text{ref}}, \mu_{s,p}, M_{s,p}, P_{s,p}, AUC_{s,p}, F_{s,p}] \quad (10)$$

Considerando que cada ensayo incluye tres sensores y diez pasos de temperatura por sensor, el vector completo de características de cada muestra se definió como:

$$\mathbf{z} = [\mathbf{z}_{A,1}, \dots, \mathbf{z}_{A,10}, \mathbf{z}_{B,1}, \dots, \mathbf{z}_{B,10}, \mathbf{z}_{C,1}, \dots, \mathbf{z}_{C,10}] \quad (11)$$

lo que conduce a una dimensionalidad total de:

$$3 \times 10 \times 6 = 180 \quad (12)$$

características por muestra.

2.5. Clasificadores empleados y métricas de rendimiento

El problema se formuló como una tarea de clasificación binaria entre aceites de oliva virgen y virgen extra. Para cada muestra i , el vector de características extraído del sistema eNose se denotó como $\mathbf{x}_i \in \mathbb{R}^d$, donde d representa el número de características empleadas tras la estrategia de selección correspondiente. La etiqueta de clase asociada a cada muestra se definió como:

$$y_i \in \{0, 1\} \quad (13)$$

donde una clase representa los aceites vírgenes y la otra los aceites vírgenes extra.

Se evaluaron clasificadores lineales y no lineales. Entre los modelos lineales se incluyeron regresión logística y máquinas de vectores soporte (*Support Vector Machines*, SVM) con kernel lineal. En la regresión logística, la probabilidad posterior de pertenencia a la clase positiva se modela como:

$$P(y_i = 1 | \mathbf{x}_i) = \frac{1}{1 + \exp[-(\mathbf{w}^T \mathbf{x}_i + b)]} \quad (14)$$

donde $\mathbf{w}$ es el vector de pesos del modelo y b es el término independiente. La clasificación se realiza comparando la probabilidad obtenida en la Ecuación (14) con un umbral de decisión.

En el caso del SVM lineal, la función de decisión viene dada por:

$$f(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i + b \quad (15)$$

y la clase predicha se obtiene a partir del signo de $f(\mathbf{x}_i)$. Este modelo busca maximizar el margen de separación entre las dos clases en el espacio de características.

Como modelos no lineales se evaluaron un SVM con kernel radial, Random Forest y Extra Trees. En el SVM con kernel radial, la similitud entre dos muestras $\mathbf{x}_i$ y $\mathbf{x}_j$ se define mediante:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (16)$$

donde γ controla la anchura del kernel radial.

Los clasificadores Random Forest y Extra Trees se basan en conjuntos (*ensembles*) de árboles de decisión. En ambos casos, la predicción final se obtiene mediante agregación de las predicciones individuales de cada árbol:

$$\hat{y}_i = \text{mode} \{h_1(\mathbf{x}_i), h_2(\mathbf{x}_i), \dots, h_T(\mathbf{x}_i)\} \quad (17)$$

donde h_t representa el árbol t -ésimo y T es el número total de árboles del conjunto.

Antes del entrenamiento, las características fueron imputadas y estandarizadas cuando el clasificador lo requería. La estandarización de cada característica x_j se realizó como:

$$x'_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j} \quad (18)$$

donde μ_j y σ_j representan la media y la desviación estándar de la característica j , calculadas exclusivamente sobre el conjunto de entrenamiento. Todas las etapas de preprocesado, selección de características y clasificación se integraron en una misma tubería de aprendizaje (*pipeline*), de forma que, en cada partición de validación cruzada, los parámetros de preprocesado fueron estimados únicamente a partir de

los datos de entrenamiento, evitando así fugas de información (*data leakage*).

La evaluación se realizó mediante validación cruzada estratificada de cinco particiones. Sea $K = 5$ el número de particiones. En cada iteración k , el modelo se entrenó con $K - 1$ particiones y se evaluó sobre la partición restante. La estratificación aseguró que la proporción de muestras de cada clase fuese aproximadamente constante en todos los folds.

El rendimiento final se expresó como la media y la desviación estándar de cada métrica a lo largo de las cinco iteraciones:

$$\bar{m} = \frac{1}{K} \sum_{k=1}^K m_k \quad (19)$$

$$\sigma_m = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (m_k - \bar{m})^2} \quad (20)$$

donde m_k representa el valor de una métrica determinada en el fold k .

Las métricas empleadas fueron exactitud global (*accuracy*), exactitud balanceada (*balanced accuracy*), F1 macro y área bajo la curva ROC (ROC-AUC). A partir de la matriz de confusión binaria, se definen los verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN).

La exactitud global se calculó como:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

Dado que el número de muestras por clase no es idéntico, también se empleó la exactitud balanceada, definida como la media entre la sensibilidad de la clase positiva y la especificidad de la clase negativa:

$$Balanced Accuracy = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (22)$$

El valor F1 de cada clase se calculó a partir de la precisión y la sensibilidad (*recall*):

$$Precision = \frac{TP}{TP + FP} \quad (23)$$

$$Recall = \frac{TP}{TP + FN} \quad (24)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (25)$$

Para evitar favorecer a la clase mayoritaria, se utilizó el promedio macro, definido como la media aritmética del valor F1 obtenido para cada clase:

$$F1_{macro} = \frac{1}{C} \sum_{c=1}^C F1_c \quad (26)$$

donde $C = 2$ es el número de clases.

Finalmente, se calculó el área bajo la curva ROC (ROC-AUC), que resume la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos para distintos umbrales de

decisión. La tasa de verdaderos positivos (*True Positive Rate*, TPR) se definió como:

$$TPR = \frac{TP}{TP + FN} \quad (27)$$

mientras que la tasa de falsos positivos (*False Positive Rate*, FPR) se definió como:

$$FPR = \frac{FP}{FP + TN} \quad (28)$$

El área bajo la curva ROC se expresó como:

$$AUC = \int_0^1 TPR(FPR) dFPR \quad (29)$$

Esta métrica proporciona una medida de la capacidad discriminante del clasificador independientemente de un umbral de decisión específico.

3. Resultados y discusión

La Tabla 1 resume la clasificación binaria (aceite virgen vs. virgen extra) empleando las características extraídas de la eNose. Utilizando el conjunto completo de 180 variables, los modelos lineales mostraron un comportamiento superior; por ejemplo, la regresión logística y el SVM lineal alcanzaron una *balanced accuracy* de $\sim 0,74$, frente al $\sim 0,68$ de los modelos no lineales (basados en árboles y SVM-RBF). Esto sugiere que la información discriminante entre ambas categorías se captura adecuadamente mediante combinaciones lineales.

Aplicar técnicas de selección de características (SelectK-Best, $k = 80$) mejoró significativamente el rendimiento de los modelos lineales, evidenciando que el conjunto original contenía información redundante o ruidosa. El mejor resultado global se logró combinando SelectKBest con un filtrado posterior de variables altamente correlacionadas ($|r| > 0,99$), seguido de un SVM lineal. Esta estrategia redujo el modelo a un subconjunto muy compacto de solo 26 características, alcanzando una *balanced accuracy* de $0,822 \pm 0,058$ y un Macro-F1 de $0,802 \pm 0,055$. Esta reducción favorece la interpretabilidad y disminuye el riesgo de sobreajuste.

El análisis de estabilidad durante la validación cruzada (Tabla 2) identificó un núcleo robusto de 15 características seleccionadas en todos los *folds* y 22 presentes en al menos cuatro. Físicamente, esta información discriminante se concentró casi en su totalidad en los sensores C (9 variables) y A (6 variables). Los descriptores más frecuentes fueron las pendientes y los valores finales, demostrando que la dinámica temporal de la respuesta del sensor contiene información más relevante que su simple magnitud media.

Asimismo, regiones específicas del perfil térmico del BME688 (como el paso inicial 1, y los intermedios 5 y 7) resultaron recurrentes. Esto indica que dichas zonas térmicas son especialmente sensibles a los volátiles de interés, siendo la velocidad de cambio de la señal en estos pasos un descriptor altamente informativo.

En conjunto, el sistema demuestra una capacidad prometedora para discriminar calidades de aceite apoyándose en tan solo 26 variables. Aunque será necesario realizar validaciones independientes para confirmar la robustez del enfoque frente a

Tabla 1: Rendimiento en validación cruzada de los clasificadores evaluados. Los valores se expresan como media $\pm$ desviación estándar en cinco particiones estratificadas.

Tipo	Clasificador	Características	Accuracy	Bal. Acc.	Macro-F1	ROC-AUC
Linear	Linear SVM	SelectKBest ($k = 80$)	0.808 ± 0.059	0.814 ± 0.059	0.797 ± 0.061	0.854 ± 0.117
Linear	Linear SVM	SelectKBest ($k = 80$) + corr. filter ($ r > 0.99$)	0.808 ± 0.052	0.822 ± 0.058	0.802 ± 0.055	0.847 ± 0.107
Linear	Linear SVM	All features	0.752 ± 0.033	0.741 ± 0.034	0.733 ± 0.028	0.818 ± 0.062
Linear	Logistic Regression	SelectKBest ($k = 80$)	0.760 ± 0.075	0.767 ± 0.079	0.747 ± 0.074	0.844 ± 0.100
Linear	Logistic Regression	SelectKBest ($k = 80$) + corr. filter ($ r > 0.99$)	0.768 ± 0.052	0.777 ± 0.056	0.759 ± 0.050	0.822 ± 0.101
Linear	Logistic Regression	All features	0.752 ± 0.044	0.748 ± 0.052	0.735 ± 0.041	0.827 ± 0.073
Non-linear	Extra Trees	SelectKBest ($k = 80$)	0.688 ± 0.104	0.656 ± 0.093	0.657 ± 0.097	0.745 ± 0.129
Non-linear	Extra Trees	SelectKBest ($k = 80$) + corr. filter ($ r > 0.99$)	0.720 ± 0.094	0.676 ± 0.086	0.681 ± 0.090	0.749 ± 0.108
Non-linear	Extra Trees	All features	0.720 ± 0.063	0.675 ± 0.056	0.681 ± 0.061	0.784 ± 0.128
Non-linear	RBF SVM	SelectKBest ($k = 80$)	0.632 ± 0.052	0.616 ± 0.060	0.608 ± 0.052	0.708 ± 0.077
Non-linear	RBF SVM	SelectKBest ($k = 80$) + corr. filter ($ r > 0.99$)	0.664 ± 0.078	0.640 ± 0.095	0.633 ± 0.091	0.726 ± 0.065
Non-linear	RBF SVM	All features	0.688 ± 0.072	0.677 ± 0.048	0.670 ± 0.061	0.786 ± 0.064
Non-linear	Random Forest	SelectKBest ($k = 80$)	0.680 ± 0.098	0.646 ± 0.094	0.644 ± 0.098	0.740 ± 0.123
Non-linear	Random Forest	SelectKBest ($k = 80$) + corr. filter ($ r > 0.99$)	0.712 ± 0.072	0.671 ± 0.062	0.672 ± 0.072	0.737 ± 0.118
Non-linear	Random Forest	All features	0.728 ± 0.077	0.693 ± 0.072	0.697 ± 0.077	0.770 ± 0.098

la variabilidad experimental y la deriva instrumental, la combinación de capacidad predictiva y reducción de dimensionalidad proporciona una base sólida para futuras investigaciones.

Tabla 2: Características seleccionadas de forma más estable por el procedimiento SelectKBest ($k = 80$) seguido de filtrado por correlación ($|r| > 0.99$). Se muestran las variables seleccionadas en al menos cuatro de los cinco folds de validación cruzada.

Característica	Sensor	Paso	Descriptor	Folds
SA.T01_maximo	A	1	Maximo	5/5
SA.T01_media	A	1	Media	5/5
SA.T01_pendiente	A	1	Pendiente	5/5
SA.T01_valor_final	A	1	Valor final	5/5
SA.T02_pendiente	A	2	Pendiente	5/5
SA.T05_maximo	A	5	Maximo	5/5
SC.T01_media	C	1	Media	5/5
SC.T01_pendiente	C	1	Pendiente	5/5
SC.T01_valor_final	C	1	Valor final	5/5
SC.T02_valor_final	C	2	Valor final	5/5
SC.T05_pendiente	C	5	Pendiente	5/5
SC.T05_valor_final	C	5	Valor final	5/5
SC.T07_pendiente	C	7	Pendiente	5/5
SC.T07_valor_final	C	7	Valor final	5/5
SC.T08_pendiente	C	8	Pendiente	5/5
SA.T02_media	A	2	Media	4/5
SB.T05_pendiente	B	5	Pendiente	4/5
SB.T07_pendiente	B	7	Pendiente	4/5
SC.T02_maximo	C	2	Maximo	4/5
SC.T02_media	C	2	Media	4/5
SC.T02_pendiente	C	2	Pendiente	4/5
SC.T10_valor_final	C	10	Valor final	4/5

4. Conclusiones

En este trabajo se ha desarrollado y evaluado una nariz electrónica basada en sensores BME688 para la clasificación de aceites de oliva virgen y virgen extra a partir de su perfil volátil. Los resultados obtenidos en validación cruzada muestran que el sistema contiene información discriminante útil, especialmente cuando se emplean modelos lineales combinados con selección de características. El mejor rendimiento se obtuvo con un SVM lineal tras aplicar SelectKBest y filtrado de variables altamente correlacionadas, reduciendo el espacio de entrada de 180 a 26 características y alcanzando una *balanced accuracy* de $0,822 \pm 0,058$ y un Macro-F1 de $0,802 \pm 0,055$.

El análisis de estabilidad mostró que las características más recurrentes se concentraron principalmente en los sensores A y C, destacando las pendientes y los valores finales de la respuesta como descriptores relevantes. Aunque los resultados son prometedores, será necesario ampliar el número de muestras, incorporar estrategias de normalización y validar el sistema en nuevas campañas experimentales antes de plantear su uso como herramienta robusta de clasificación.

Agradecimientos

Este trabajo ha sido financiado parcialmente por los proyectos del Plan Nacional con referencias PDC2022-133995-I00 y PID2023-150832OB-I00; y por la Junta de Andalucía a través del proyecto DGP-PIDI-2024-01419.

Referencias

- Amini, A., Anjileh, H. G., Amirfaridi, H., 1 2026. Fast and portable single-sensor electronic nose for accurate quality assessment of extra virgin olive oil using a temperature-modulated generic gas sensor. *Talanta* 296, 128490.
DOI: 10.1016/J.TALANTA.2025.128490
- Devika, K. K., Vijayalakshmi, K., Sameeksha, S., Srinivas, Y., Vivek, K., Nimbkar, S., Dalbhagat, C. G., Thivya, P., 1 2026. Advances of electronic nose technologies for rapid detection of edible oil quality. *Food Analytical Methods* 2026 19:2 19, 73–.
DOI: 10.1007/S12161-026-02996-Y
- Fernandez, L., Oller-Moreno, S., Fonollosa, J., Garrido-Delgado, R., Arce, L., Martín-Gómez, A., Marco, S., Pardo, A., 1 2025. Signal preprocessing in instrument-based electronic noses leads to parsimonious predictive models: Application to olive oil quality control. *Sensors* 2025, Vol. 25, Page 737 25, 737.
DOI: 10.3390/S25030737
- Ghnnimi, H., Ennouri, M., Chèné, C., Karoui, R., 12 2024. Non-destructive and rapid evaluation of the potentiality of faba bean lipoxygenase to promote lipid oxidation of rapeseed oil by using mid-infrared and near-infrared spectroscopies. *Food Analytical Methods* 2024 18:4 18, 517–531.
DOI: 10.1007/S12161-024-02735-1
- Gila, D. M. M., García, J. G., Bellincontro, A., Mencarelli, F., Ortega, J. G., 2 2020. Fast tool based on electronic nose to predict olive fruit quality after harvest. *Postharvest Biology and Technology* 160, 111058.
DOI: 10.1016/j.postharvbio.2019.111058
- Lechhab, T., Lechhab, W., Cacciola, F., Salmoun, F., 1 2022. Sets of internal and external factors influencing olive oil (*olea europaea* L.) composition: a review. *European Food Research and Technology* 2022 248:4 248, 1069–1088.
DOI: 10.1007/S00217-021-03947-Z
- Martínez, D. B., Braga, G. F. P., Gila, D. M. M., Martínez, S. S., 1 2026. Feasibility study of mos gas sensors for detecting mineral hydrocarbon contaminants in freshly harvested olives at different maturity stages. *Sensors* 26, 816.
DOI: 10.3390/s26030816
- Ortega, J. B., Gila, D. M. M., Puerto, D. A., García, J. G., Ortega, J. G., 11 2016. Novel technologies for monitoring the in-line quality of virgin olive oil during manufacturing and storage. *Journal of the Science of Food and Agriculture* 96, 4644–4662.
DOI: 10.1002/jsfa.7733
- Soto, J. P. N., Rico, S. I., Gila, D. M. M., Martínez, S. S., 4 2024. Influence of the degree of fruitiness on the quality assessment of virgin olive oils using electronic nose technology. *Sensors* 24.
DOI: 10.3390/S24082565
- Xu, A., Cai, T., Shen, D., Wang, A., 10 2021. Food odor recognition via multi-step classification.