

Del 2 al 5 de mayo de 2023

**CENTRO UNIVERSITARIO  
SANTA ANA  
ALMENDRALEJO**



Joaquín Sorolla Bastida. Comiendo uvas, 1898. Acuñera sobre papel. Museo Sorolla, n° inv. 00427

**XLV JORNADAS  
DE VITICULTURA Y ENOLOGÍA  
TIERRA DE BARROS  
V CONGRESO AGROALIMENTARIO  
DE EXTREMADURA**

XLV JORNADAS DE VITICULTURA Y ENOLOGÍA  
DE LA TIERRA DE BARROS  
V CONGRESO AGROALIMENTARIO DE EXTREMADURA

**Edita:**

Centro Universitario Santa Ana  
C/ IX Marqués de la Encomienda, nº 2  
Almendralejo  
Tel. 924 661 689  
<http://www.univsantana.com>

**Colabora:** Cajalmendralejo

**Ilustración de portada:**

Joaquín Sorolla Bastida. "Comiendo uvas". 1898. Acuarela sobre papel.  
Museo Sorolla. n: inv. 00427. © Fundación Museo Sorolla

**Diseño original:**

Tecnigraf S.A.

**Maquetación:** María Sabater

ISBN: 84-7930-113-9

D.L.: BA-000169-2024

Imprime: Impresal

# Cuantificación inteligente de adulteraciones en cacao en polvo mediante espectroscopía NIR

CALLE, J. L. <sup>1</sup>

NUR MILLATINA, N. R. <sup>2</sup>

BAREA-SEPULVEDA, M. <sup>1</sup>

FERREIRO-GONZÁLEZ, M. <sup>1</sup>

SETYANINGSIH, W. <sup>2</sup>

PALMA, M. <sup>1</sup>

1 Departamento de Química Analítica, Facultad de Ciencias,  
Universidad de Cádiz.

2 Departamento de Productos Alimentarios y Agricultura,  
Facultad tecnológica de Agricultura,  
Universidad de Gadjah Mada, Indonesia.

## RESUMEN

La adulteración del cacao en polvo es un problema común en la industria alimentaria, el cual ha sido agravado debido a la creciente y alta demanda de este producto. Por ello, el presente estudio tiene como objetivo desarrollar una metodología que utilice la espectroscopia de infrarrojo cercano (NIRs) junto con algoritmos de *machine learning* (ML) para cuantificar las adulteraciones en el cacao en polvo de manera fiable y automatizada. Para ello, las muestras de cacao puras se

adulteraron a distintos porcentajes en un rango de 0.5-40% utilizando varios sustitutivos. En el análisis exploratorio se observó una clara tendencia de agrupación de las muestras en función del tipo de adulterante, así como del porcentaje empleado. El análisis supervisado permitió crear modelos de ML con rendimientos muy satisfactorios, tanto para la cuantificación como para la identificación del adulterante. Los mejores resultados se obtuvieron para el modelo de bosque aleatorio (RF) y las máquinas de vectores de soporte (SVM) los cuales identificaban correctamente el 100% de los adulterantes. Respecto a la cuantificación, el mejor resultado se obtuvo para la regresión de vectores de soporte (SVR), con un coeficiente de determinación ( $R^2$ ) superior a 0.99 y una raíz de error cuadrático medio (RMSE) inferior a 1. Estos resultados indican que los modelos son muy precisos, y por ello se ha desarrollado una aplicación web en la que pueden ser utilizados, facilitando así la caracterización de las muestras a cualquier usuario.

**Palabras claves:** Cacao en polvo, NIRs, Machine Learning (ML), Adulteración.

#### SUMMARY

The adulteration of cocoa powder is a common problem in the food industry, which has been aggravated due to the increasing and high demand for this product. Therefore, the present study aims to develop a methodology using near infrared spectroscopy (NIRs) coupled with machine learning (ML) algorithms to quantify adulterations in cocoa powder in a reliable and automated way. For this purpose, pure cocoa samples were adulterated at different percentages in the range of 0.5-40% using various substituents. In the exploratory analysis, it was observed that the samples tended to cluster based on both the type and percentage of adulterant used. The supervised analysis allowed the creation of ML models with very satisfactory performances for both quantification and adulterant detection. The best results were obtained for the random forest (RF) model and the support vector machines (SVM) which correctly identified 100% of the adulterants. In the quantification, the best result was obtained for support vector regression (SVR) model with a coefficient of determination ( $R^2$ ) higher than 0.99 and a root mean square error (RMSE) lower than 1. The results indicate that the models are highly accurate. As a result, a web application has been developed to make these models accessible to users, which facilitates the characterization of the samples.

**Keywords:** Cocoa powder, NIRs, Machine Learning (ML), Adulteration.

## 1. INTRODUCCIÓN

En la industria alimentaria existe un gran interés en el desarrollo de metodologías que permitan detectar adulteraciones en distintos alimentos. Estas prácticas fraudulentas, además de afectar a la calidad del producto, suponen un riesgo extra a la salud del consumidor, ya que algunos aditivos o sustancias agregadas pueden ser tóxicos o alérgenos [1].

El cacao en polvo es el producto que se obtiene a partir de la mezcla de granos de cacao tostados y molidos, sin la adición de azúcares u otros ingredientes [2]. Este producto es muy popular en la industria alimentaria y se utiliza en la elaboración de una gran variedad de productos como chocolates, bebidas y postres. Además, es muy valorado por sus beneficios nutricionales, que incluyen su alto contenido en fibra, antioxidantes y minerales [3]. Dada la gran versatilidad, en Europa se ha establecido una estricta regulación recogida en la directiva 2000/36/EC para garantizar la calidad en la producción del cacao en polvo [4]. Sin embargo, este alimento se encuentra entre los 20 productos más adulterados desde 1980 hasta 2010 [5]. Entre los adulterantes comúnmente utilizados se encuentra el polvo de algarroba, arrurruz o achicoria, así como las cáscaras de cacao, cacahuete o castaña, debido al bajo coste respecto al cacao en polvo [6]. Además, las harinas de trigo, soja y mijo también se utilizan como adulterantes, las cuales son imprescindibles ser detectadas pues producen problemas relacionadas con alérgenos. Cabe destacar que, las harinas de cáscara de cacao, algarroba y mijo no alteran el color, el aroma, ni el sabor del cacao en polvo, lo que supone una mayor dificultad para ser detectada [7].

Entre las metodologías más ampliamente extendidas para la detección de adulteraciones en alimentos se encuentra la cromatografía de gases (GC), la espectroscopía de infrarrojo cercano (NIR) y la cromatografía líquida (LC) [7]. Sin embargo, los análisis tanto con cromatografía líquida como de gases requieren largos tiempos de preparación de la muestra, optimización del método y, materiales y reactivos de alto coste. Por el contrario, la espectroscopia NIR es rápida, fiable, menos costosa y no utiliza reactivos ni genera residuos, por lo que se encuadraría dentro de la “Química Verde” [8]. En el campo de la detección de adulteraciones en el cacao en polvo, la LC

ha reportado excelentes resultados tanto para determinar la autenticidad de este producto [9] como para detectar adulteraciones [10]. Análogamente la GC también se ha empleado para caracterizar y garantizar su autenticidad [11,12]. En el caso de la espectroscopía NIR los estudios de adulteraciones del cacao en polvo son escasos, y se ha empleado con éxito para detectar adulterantes como la cáscara de cacao [13], la harina de algarroba [14] así como para cuantificar determinados parámetros relacionados con la calidad del cacao en polvo [15].

Respecto al análisis de datos, la mayoría de los estudios se centran en la identificación de determinados compuestos dianas que garantizan la autenticidad del producto o confirman la presencia de una adulteración. Esta tarea en muchas ocasiones es compleja e impide la automatización del proceso. Además, necesita de un analista experimentado que determine la calidad del producto, lo que añade cierta subjetividad al proceso [16]. Sin embargo, los espectros NIR pueden actuar como una “huella dactilar” de la muestra y, en combinación con algoritmos de *machine learning* (ML) permitir su identificación [17]. De hecho, el uso del espectro NIR y los algoritmos de ML se han empleado con éxito para detectar adulteraciones en distintos alimentos como zumos [18], miel [19,20] o aceites [21,22]. Cabe destacar que, en la mayoría de los estudios se utilizan modelos de ML paramétricos como el análisis discriminante lineal (LDA) para la clasificación y, los mínimos cuadrados parciales (PLS) para la regresión. Sin embargo, modelos no-paramétricos más complejos como las máquinas de vectores de soporte (SVM) o los bosques aleatorios (RF) son menos utilizados, pese a que en muchas ocasiones ofrecen mejores resultados [23,24].

En base a todo lo anterior, el objetivo del presente estudio es desarrollar una metodología basada en el uso de la espectroscopia NIR y ML para identificar y cuantificar el porcentaje de distintos adulterantes en el cacao en polvo. De esta forma, la metodología propuesta es inteligente y automatizada, donde las adulteraciones se cuantifican haciendo uso de una aplicación web que contiene los modelos de ML creados.

## 2. MATERIALES Y MÉTODOS

### 2.1. Muestras

Se obtuvieron dos cacao puros en polvo de la variedad forastero, uno de ellos de PT Perke-bunan Nusantara XII (Banyuwangi, Indonesia) y, el otro procedía de Kampung Coklat Senara (Gerung, Indonesia). En ambas muestras se garantizaba la autenticidad de las mismas, y estas fueron mezcladas en la misma proporción con el fin de abarcar la mayor heterogeneidad posible. En el caso de los adulterantes, se utilizaron los siguientes: cáscara de cacao, soja en polvo, algarroba en polvo, harina de mijo de cola de zorra y harina de trigo integral. Todos ellos se compraron en mercados locales de Indonesia y, de cada uno de ellos se adquirieron 3 marcas distintas que fueron mezcladas en la misma cantidad. Nuevamente, este proceso se realizó con el fin de garantizar la máxima heterogeneidad posible.

Los niveles de adulteración se situaron en el rango de 0.5-40%, pues niveles superiores alterarían las características sensoriales del cacao en polvo y, porcentajes inferiores no serían económicamente rentables. Por tanto, los ratios de adulteración fueron los siguientes: 0.5, 1, 1.5, 2, 2.5, 5, 10, 15, 20, 25, 30, 35 y 40%. Este proceso de adulteración se realizó por duplicado, por lo que el total de muestras generadas fue de 130 (5 adulterantes x 13 puntos de adulteración x 2 duplicados). Además, también se analizó por duplicado el cacao en polvo sin adulterar, por lo que el número final de muestras fue de 132.

Las muestras analizadas se nombraron con el siguiente código: mezcla de cacao en polvo sin adulterar (C), adulterado con cáscara de cacao (S), adulterado con soja en polvo (B), adulterado con algarroba en polvo (R), adulterado con harina de mijo cola de zorra (F) y adulterado con harina de trigo integral (W). Seguido a la correspondiente letra se indica el porcentaje de adulteración. De esta forma, una muestra adulterada con cáscara de cacao al 10% se codificó como S10\_1 para el primer duplicado y S10\_2 para el segundo duplicado.

### 2.3. Espectroscopía de infrarrojo cercano (NIRs)

Para el análisis de las muestras se utilizó el espectrómetro FOSS XDS Rapid Content™ (FOSS Analytical, Hilleroed, Dinamarca). Este equipo cuenta

con un analizador de haz de luz único, y los espectros se adquirieron de 400 a 2500 nm con una resolución de 0,5 nm. Para cada muestra se escanearon dos espectros y, el resultado final fue la media de ambos. El tiempo total de análisis por muestra fue de 30s.

## 2.4. Análisis de datos

Todo el análisis de datos se realizó mediante el software RStudio v.4.2.2 (Rstudio Team 2022, Boston, MA, USA). Los datos se almacenaron en matrices  $D_{n \times p}$ , donde  $n$  denota el número de muestras y  $p$  el número de predictores. Los principales paquetes de RStudio utilizados fueron: *ggplot2* [25] para las representaciones gráficas; *prospectr* [26] usado para el pretratamiento de los datos consistente en aplicar la primera derivada y el filtro de Savitzky -Golay; *factoextra* [27] para el análisis de componentes principales; *caret* [28] para crear los diferentes modelos de *machine learning* tanto de clasificación como de regresión y *Shiny* [29] para el desarrollo de la aplicación.

Los diferentes modelos de ML de clasificación utilizados para la identificación del tipo de adulterante fueron: el análisis discriminante lineal (LDA), las máquinas de vectores de soporte (SVM) y el bosque aleatorio (RF). De forma análoga, los modelos de ML de regresión usados para la cuantificación del adulterante fueron: la regresión de mínimos cuadrados parciales (PLS), la regresión de vectores de soporte (SVR) y el bosque aleatorio (RF).

## 3. RESULTADOS Y DISCUSIÓN

### 3.1. Análisis exploratorio

Antes de realizar los análisis hay que tener en cuenta que los espectros NIR se caracterizan por tener bandas de absorción superpuesta y dispersión de la luz y, por ello, no aportan una información fácilmente interpretable. Por tanto, para mejorar las características espectrales y compensar los cambios de la línea base y las diferencias de dispersión de la luz, los espectros originales deben someterse a pretratamientos antes de ser utilizados [30].

En este caso el pretratamiento consistió en aplicar la primera derivada y un filtro de suavizado de Savitzky-Golay (de grado polinómico 3 y tamaño de

ventana 11). De esta forma, se eliminaron los problemas relacionados con el desplazamiento de la línea base, así como la superposición de picos y, además se redujo el ruido aleatorio.

Una vez realizado dicho pretratamiento, el objetivo de esta primera parte del estudio fue observar si existe una tendencia de agrupación de las muestras en función del tipo de adulterante utilizado (cáscara de cacao, algarroba, soja, harina de trigo integral o de mijo) así como del nivel de adulteración (0-40%). Por este motivo, se realizó un análisis de componentes principales (PCA) utilizando la matriz de datos completa ( $D_{133 \times 4200}$ ) consistente en las 133 muestras y, las 4200 variables (números de ondas). En la Figura 1, se representa las puntuaciones obtenidas por las muestras para las dos primeras componentes principales (PC1 y PC2). Además, se han establecido 6 grupos a priori diferenciados por el color y la forma. Estos grupos se han establecido en función del tipo de adulterante, siendo los siguientes: I) "Algarroba" que contiene todas las muestras adulteradas con polvo de algarroba (coloreado de rosa, símbolo "."); II) "Cáscara" que contiene las muestras adulteradas con cáscara de cacao (coloreado de amarillo, símbolo "▲"); III) "Mijo" contiene las muestras adulteradas con la harina de mijo de cola de zorra (coloreadas de verde, símbolo "■"); IV) "No\_adulterada" está formado por las dos muestras de cacao en polvo sin adulterar (coloreados de azul oscuro, símbolo "+"); V) "Soja" contiene las muestras adulteradas con la harina de soja (coloreados de azul claro, símbolo "⊠"); VI) "trigo" formado por las muestras adulteradas con harina de trigo integral (coloreadas de rosa, símbolo "\*").

Como se puede observar en la Figura 1, debido a la gran cantidad de muestras y con el objetivo de intentar evitar la superposición de los nombres, no todas las muestras han podido ser etiquetadas. La PC1 explica el 59% de la variabilidad de los datos, y es principalmente la encargada de separar a las muestras en función de porcentaje de adulteración. De esta forma, se observa que mayores valores en la PC1 se asocian con porcentajes de adulteraciones más altos. Por otra parte, la PC2 que explica el 36.2% de la variabilidad de los datos, es la responsable de separar las muestras adulteradas con polvo de algarroba y cascara de cacao del resto. De esta forma, valores negativos en la PC2 se asocian con los adulterantes "mijo", "soja" y "trigo" mientras que valores positivos pertenecen a los grupos "cáscara" y "algarroba".

Centrándose en la combinación de PC1 y PC2, se observa que dentro de cada grupo el porcentaje de adulteración sigue una relación lineal. Todo esto indica que el espectro FT-IR está influenciado tanto por el tipo de adulterante como por su porcentaje.

### 3.2. Análisis supervisado para la identificación del adulterante

Una vez que se ha comprobado la tendencia de agrupación de las muestras en función del porcentaje y tipo de adulterante, se procedió a entrenar los modelos de *machine learning* para la identificación del tipo de adulterante. Por tanto, se establecieron un total de 5 grupos a priori en función del tipo de adulterante utilizado (“Algarroba”, “Cáscara”, “Mijo”, “Trigo” y “Soja”) y, los modelos de ML probados fueron: LDA, RF y SVM. Hay que destacar que, previo a la creación de los modelos, el conjunto de datos ( $D_{130 \times 4100}$ ) se dividió en dos subconjuntos.

El primero de ellos es el subconjunto de entrenamiento constituido por el 75% de las muestras y se usó para la creación de los modelos. El segundo de ellos es el conjunto de validación constituido por el 25% restante de las muestras y se utilizó para poner a prueba los modelos. Cabe destacar que, este subconjunto de validación no participa en ningún momento en la creación de los modelos, por lo que el error obtenido será insesgado y bastante próximo al que se obtendría en una situación real. En la Tabla 1, se muestra un resumen de la *accuracy* (porcentaje de clasificación correcta) obtenida para los distintos modelos.

Como se puede observar, tanto el modelo SVM como el RF, clasificaron correctamente el 100% de las muestras en el conjunto de validación. En el modelo LDA, el rendimiento fue algo inferior, clasificándose incorrectamente la muestra de trigo “W10\_2” como soja, lo que condujo a una *accuracy* del 96.88%. Por tanto, todos los modelos son adecuados para predecir el tipo de adulterante, no obstante, se preferiría el uso de RF o SVM.

### 3.2. Análisis supervisado para la cuantificación del adulterante

Una vez obtenido un modelo capaz de identificar el tipo de adulterante, la siguiente tarea es determinar el porcentaje de este. Para ello, se entrenaron distintos algoritmos de *machine learning* de regresión usando nuevamente la misma matriz de datos ( $D_{130 \times 4200}$ ) y, dividiéndolo en dos subconjuntos:

I) El conjunto de entrenamiento, el cual contenía el 75% de las muestras  
II) El conjunto de validación, el cual se utiliza como una comprobación externa y, contiene el 25% de las muestras restante. Además, se aseguró de que la partición fuera de una forma equilibrada. Los modelos de ML probados fueron: PLS, RF y SVR y, en la Tabla 2 se muestra un resumen del rendimiento obtenido por los mismos.

Las métricas utilizadas para evaluar el rendimiento consistieron en el coeficiente de determinación ( $R^2$ ) y la raíz del error cuadrático medio (RMSE). Como se puede observar en la Tabla 2, todos los modelos obtuvieron muy buenos resultados con errores (RMSE) inferiores a 1.5 y  $R^2$  superiores a 0.99 en todos los casos. No obstante, el mejor modelo para la cuantificación del porcentaje de adulteración fue el SVR con un  $R^2$  de 0.996 y un RMSE de 0.922 en el conjunto de validación.

### 3.4. Desarrollo de la aplicación

Con el objetivo de facilitar el análisis de las muestras para el resto de los usuarios, se creó una aplicación que incluye los modelos SVM y SVR para la identificación y para la cuantificación del adulterante, respectivamente. El enlace para acceder a la aplicación es el siguiente:

[https://agr291.shinyapps.io/App\\_Cacao/](https://agr291.shinyapps.io/App_Cacao/)

El funcionamiento de esta es bastante sencillo y, un ejemplo del mismo puede verse en el siguiente video:

<https://drive.google.com/file/d/118r4RmSH6TgC-I3T-zrUnoqlg8FNPIh/view?usp=sharing>

Como se puede observar, el usuario únicamente tendría que subir el archivo Excel generado directamente del análisis de la muestra mediante NIR. Al hacer click en el botón “submit” se realiza las predicciones directamente. Además, se muestran los gráficos del análisis exploratorio de una forma interactiva, donde se puede hacer zoom, seleccionar grupos, etc. De esta forma, se automatiza el análisis de datos y se consigue obtener los resultados en cuestión de segundos, por lo que esta metodología supone una gran ventaja para las agencias reguladoras.

Destacar que los modelos subidos a la página web se irán actualizando con el análisis de más muestras. De esta forma, cada vez se incorporará más información a los mismo y abarcará mayor variabilidad.

#### 4. CONCLUSIONES

Se ha demostrado el potencial de la técnica NIRs en combinación con las herramientas de *machine learning* para detectar y cuantificar distintos adulterantes en muestra de cacao en polvo. El análisis exploratorio mostró una tendencia de agrupación en función del tipo de adulterante, así como de porcentaje presente. Respecto al análisis supervisado, la mayoría de los modelos de clasificación identificaron correctamente el 100% del tipo de adulterante utilizado. En el caso de la cuantificación los mejores resultados se obtuvieron para el modelo SVR con un RMSE de 0.922 y una  $R^2$  de 0.996 para el conjunto de validación. Esto permite que la metodología sea capaz, no solo de detectar el adulterante usado, sino también de cuantificarlo. Además, posee unas ventajas muy importantes para las agencias reguladoras como la posibilidad del análisis *in-situ* y, la obtención de los resultados en un par de minutos. Destacar que es también una metodología limpia, y la creación de la página web facilita aún más su uso para el resto de usuarios. Por último, señalar que la web permite ver los gráficos de una forma interactiva y ayuda aún más el entendimiento de los mismo. Además, tanto estos gráficos como los modelos se irán actualizando con el análisis de más muestras. De esta forma, la metodología se puede considerar “inteligente” pues va “aprendiendo” mientras se le va “enseñando”.

#### 5. AGRADECIMIENTOS

José Luis Pérez Calle agradece al Ministerio de Ciencia e Innovación de España un contrato de doctorado bajo el programa FPU (FPU20/03377). Los autores agradecen al Instituto de Investigación Vitivinícola y Agroalimentario (IVAGRO) por facilitar las instalaciones necesarias para llevar a cabo esta investigación y al proyecto Singular AgroMIS. ceiA3 Instrumentos Estratégico hacia un tejido productivo Agroalimentario Moderno, Innovador y Sostenible: motor del territorio rural andaluz. Programa Operativo FEDER 2014-2020 de Andalucía – PAI-TAN-AT2019-AGROMIS-EC.

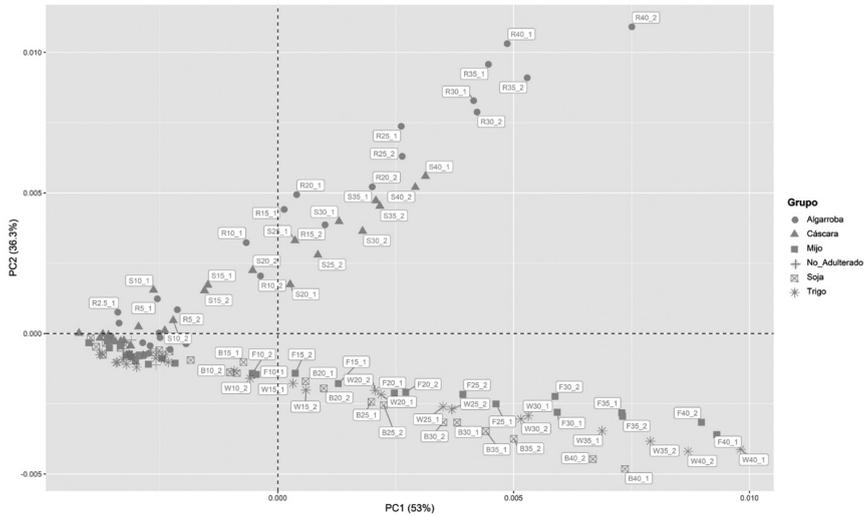
## 6. REFERENCIAS

1. Faith Ndlovu, P.; Samukelo Magwaza, L.; Zeray Tesfay, S.; Ramaesele Mphahlele, R. "Destructive and Rapid Non-Invasive Methods Used to Detect Adulteration of Dried Powdered Horticultural Products: A Review". *Food Research International* 2022, 157, 111198, doi:10.1016/j.foodres.2022.111198.
2. Beg, M.S.; Ahmad, S.; Jan, K.; Bashir, K. "Status, Supply Chain and Processing of Cocoa—A Review". *Trends Food Sci Technol* 2017, 66, 108–116, doi:10.1016/j.tifs.2017.06.007.
3. Całkosiński, A.; Kostrzewska, P.; Majewski, M.; Mandera, A.; Pawlikowska, A.; Szuster, E.M. "Impact of Cocoa on the Human Health". *Journal of Education, Health and Sport* 2019, 9, 145–154, doi:10.5281/zenodo.2718725.
4. European Parliament: Luxembourg, 2012; pp. 9–26. Directive 2000/36/EC of the European Parliament and of the Council of 23 June 2000 Relating to Cocoa and Chocolate Products Intended for Human Consumption.
5. Moore, J.C.; Spink, J.; Lipp, M. "Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010". *J Food Sci* 2012, 77, R118–R126, doi:10.1111/j.1750-3841.2012.02657.x.
6. Jack L *Risk Modelling of Food Fraud Motivation: "NSF Fraud Protection Model" Intelligent Risk Model Scoping Project FS 246004: Final Report*; Food Standards Agency: London, 2015.
7. Quelal-Vásconez, M.A.; Pérez-Esteve, É.; Arnau-Bonachera, A.; Barat, J.M.; Talens, P. "Rapid Fraud Detection of Cocoa Powder with Carob Flour Using near Infrared Spectroscopy". *Food Control* 2018, 92, 183–189, doi:10.1016/j.foodcont.2018.05.001.
8. Ma, H. Liang; Wang, J. Wen; Chen, Y. Jun; Cheng, J. Le; Lai, Z. Tian. "Rapid Authentication of Starch Adulterations in Ultrafine Granular Powder of Shanyao by Near-Infrared Spectroscopy Coupled with Chemometric Methods". *Food Chem* 2017, 215, 108–115, doi:10.1016/J.FOODCHEM.2016.07.156.

9. Greño, M.; Plaza, M.; Luisa Marina, M.; Castro Puyana, M. "Untargeted HPLC-MS-Based Metabolomics Approach to Reveal Cocoa Powder Adulterations". *Food Chem* 2023, 402, 134209, doi:10.1016/J.FOODCHEM.2022.134209.
10. Yang, W. Li; Hu, M. Hua; Chen, S. Wei; Wang, Q.; Zhu, S.; Dai, J.; Li, X. Zhong. "Identification of Adulterated Cocoa Powder Using Chromatographic Fingerprints of Polysaccharides Coupled with Principal Component Analysis". *Food Anal Methods* 2015, 8, 2360–2367, doi:10.1007/S12161-015-0126-6/FIGURES/7.
11. Quelal-Vásconez, M.A.; Macchioni, R.; Livi, G.; Pérez-Esteve, É.; Lerma-García, M.J.; Talens, P.; Barat, J.M.; Petersen, M.A.; Bro, R. "Automatic and Non-Targeted Analysis of the Volatile Profile of Natural and Alkalized Cocoa Powders Using SBSE-GC-MS and Chemometrics". *Food Chem* 2022, 389, 133074, doi:10.1016/J.FOODCHEM.2022.133074.
12. Rodríguez-Carrasco, Y.; Gaspari, A.; Graziani, G.; Santini, A.; Ritieni, A. "Fast Analysis of Polyphenols and Alkaloids in Cocoa-Based Products by Ultra-High Performance Liquid Chromatography and Orbitrap High Resolution Mass Spectrometry (UHPLC-Q-Orbitrap-MS/MS)". *Food Research International* 2018, 111, 229–236, doi:10.1016/J.FOODRES.2018.05.032.
13. Quelal-Vásconez, M.A.; Lerma-García, M.J.; Pérez-Esteve, É.; Arnau-Bonachera, A.; Barat, J.M.; Talens, P. "Fast Detection of Cocoa Shell in Cocoa Powders by near Infrared Spectroscopy and Multivariate Analysis". *Food Control* 2019, 99, 68–72, doi:10.1016/J.FOODCONT.2018.12.028.
14. Quelal-Vásconez, M.A.; Pérez-Esteve, É.; Arnau-Bonachera, A.; Barat, J.M.; Talens, P. "Rapid Fraud Detection of Cocoa Powder with Carob Flour Using near Infrared Spectroscopy". *Food Control* 2018, 92, 183–189, doi:10.1016/J.FOODCONT.2018.05.001.
15. Krähmer, A.; Engel, A.; Kadow, D.; Ali, N.; Umaharan, P.; Kroh, L.W.; Schulz, H. "Fast and Neat – Determination of Biochemical Quality Parameters in Cocoa Using near Infrared Spectroscopy". *Food Chem* 2015, 181, 152–159, doi:10.1016/J.FOODCHEM.2015.02.084.
16. Ferreiro-González, M.; Ayuso, J.; Álvarez, J.A.; Palma, M.; Barroso, C.G. "Gasoline Analysis by Headspace Mass Spectrometry and near Infrared Spectroscopy". *Fuel* 2015, 153, 402–407, doi:10.1016/J.FUEL.2015.03.019.

17. Ríos-Reina, R.; Camiña, J.M.; Callejón, R.M.; Azcarate, S.M. "Spectralprint Techniques for Wine and Vinegar Characterization, Authentication and Quality Control: Advances and Projections". *TrAC–Trends in Analytical Chemistry* 2021, 134, 116121, doi:10.1016/j.trac.2020.116121.
18. Calle, J.L.P.; Barea-Sepúlveda, M.; Ruiz-Rodríguez, A.; Álvarez, J.Á.; Ferreiro-González, M.; Palma, M. "Rapid Detection and Quantification of Adulterants in Fruit Juices Using Machine Learning Tools and Spectroscopy Data". *Sensors* 2022, 22, 3852, doi:10.3390/S22103852.
19. Zhu, X.; Li, S.; Shan, Y.; Zhang, Z.; Li, G.; Su, D.; Liu, F. "Detection of Adulterants Such as Sweeteners Materials in Honey Using Near-Infrared Spectroscopy and Chemometrics". *J Food Eng* 2010, 101, 92–97, doi:10.1016/j.jfoodeng.2010.06.014.
20. Ferreiro-González, M.; Espada-Bellido, E.; Guillén-Cueto, L.; Palma, M.; Barroso, C.G.; Barbero, G.F. "Rapid Quantification of Honey Adulteration by Visible-near Infrared Spectroscopy Combined with Chemometrics". *Talanta* 2018, 188, 288–292, doi:10.1016/J.TALANTA.2018.05.095.
21. Du, Q.; Zhu, M.; Shi, T.; Luo, X.; Gan, B.; Tang, L.; Chen, Y. "Adulteration Detection of Corn Oil, Rapeseed Oil and Sunflower Oil in Camellia Oil by in Situ Diffuse Reflectance near-Infrared Spectroscopy and Chemometrics". *Food Control* 2021, 121, 107577, doi:10.1016/j.foodcont.2020.107577.
22. Vanstone, N.; Moore, A.; Martos, P.; Neethirajan, S. "Detection of the Adulteration of Extra Virgin Olive Oil by Near-Infrared Spectroscopy and Chemometric Techniques". *Food Quality and Safety* 2018, 2, 189–198, doi:10.1093/FQSAFE/FYY018.
23. Esteki, M.; Shahsavari, Z.; Simal-Gandara, J. "Gas Chromatographic Fingerprinting Coupled to Chemometrics for Food Authentication". *Food Reviews International* 2020, 36, 384–427, doi:10.1080/87559129.2019.1649691.
24. Qiu, S.; Wang, J.; Gao, L. "Discrimination and Characterization of Strawberry Juice Based on Electronic Nose and Tongue: Comparison of Different Juice Processing Approaches by LDA, PLSR, RF, and SVM". *J Agric Food Chem* 2014, 62, 6426–6434, doi:10.1021/JF501468B.
25. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag New York, 2016; ISBN 978-3-319-24277-4.

26. Stevens, A.; Ramirez-Lopez, L. *An Introduction to the Prospectr Package* 2020.
27. Kassambara, A.; Mundt, F. *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses* 2020.
28. Kuhn, M. *Caret: Classification and Regression Training* 2020.
29. Chang, W.; Cheng, J.; Allaire, J.J.; Xie, Y.; McPherson, J. Shiny: *Web Application Framework for R* 2020.
30. Finn, B.; Harvey, L.M.; McNeil, B.; McNeil, B. "Near-Infrared Spectroscopic Monitoring of Biomass, Glucose, Ethanol and Protein Content in a High Cell Density Baker's Yeast Fed-Batch Bioprocess". *Yeast* 2006, 23, 507-517, doi:10.1002/yea.1371.



**Figura 1.** Representación de las puntuaciones obtenidas utilizando el espectro NIR de todas las muestras ( $D_{133 \times 4200}$ ) en función de las dos primeras componentes principales (PC1 y PC2). Las muestras se han coloreado y simbolizado en función del tipo de adulterante empleado.

**Tabla 1.** Resultados obtenidos para cada modelo de clasificación aplicado en la identificación del tipo de adulterante utilizando los espectros NIRs de todas las muestras ( $D_{130 \times 4200}$ ).

Modelo	Hiperparámetros	Accuracy conjunto de entrenamiento	Accuracy conjunto de validación
lda	-	100%	100%
SVM	$C = 6$ $\gamma = 9.77 \times 10^{-4}$	100%	100%
RF	$mtry = 23$ $ntree = 500$	100%	97.67%

**Tabla 2.** Resultados obtenidos para cada modelo de regresión aplicado en la cuantificación del adulterante global utilizando los espectros NIRs de todas las muestras ( $D_{130 \times 4200}$ ).

Modelo	Hiperparámetros	Rendimiento conjunto de entrenamiento	Rendimiento conjunto de Vali
PLS	7 componentes principales	RMSE = 0.706 $R^2 = 0.997$	RMSE = 1.069 $R^2 = 0.995$
SVR	$C = 136$ $\gamma = 9.77 \times 10^{-4}$	RMSE = 0.929 $R^2 = 0.996$	RMSE = 0.922 $R^2 = 0.996$
RF	mtry = 3	RMSE = 0.812 $R^2 = 0.997$	RMSE = 1.304 $R^2 = 0.990$