

# XLIV JORNADAS DE VITICULTURA Y ENOLOGÍA TIERRA DE BARROS

IV Congreso Agroalimentario de Extremadura

CENTRO UNIVERSITARIO SANTA ANA ALMENDRALEJO



Del 3 al 6 de Mayo 2022

XLIV JORNADAS DE VITICULTURA Y ENOLOGÍA  
DE LA TIERRA DE BARROS  
IV 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:**

© ALBERTO CATILLO

**Diseño original:**

Tecnigraf S.A.

**Maquetación:** Virginia Pedrero

ISBN: 978-84-7930-112-0

D.L.:

Imprime: Impresal

# Detección automatizada de adulteraciones en zumos de fruta mediante HS-GC-IMS

CALLE, J.L.P.

FERREIRO-GONZÁLEZ, M.

BAREA-SEPÚLVEDA, M.

PALMA, M.

Departamento de Química Analítica, Facultad de Ciencias, Universidad de Cádiz,  
Campus de Excelencia Internacional Agroalimentario (ceiA3), IVAGRO, 11510, Puerto  
Real, Cádiz, España.

## RESUMEN

Los zumos de fruta son una de las bebidas más consumidas mundialmente y su producción está sujeta a regulaciones europeas. Sin embargo, este alimento es frecuentemente adulterado mediante la adición de agua, azúcares u otros zumos de frutas menos costosos. Esta última es una de las más comunes, debido a la mayor dificultad para ser detectada. Por ello, el presente estudio tiene como objetivo desarrollar una metodología basada en la técnica de Espacio de Cabeza-Cromatografía de Gases-Espectrometría de Movilidad Iónica (HS-GC-IMS) en combinación con algoritmos de aprendizaje automatizado o *machine learning* (ML) para detectar y cuantificar de forma fiable y automatizada las adulteraciones en zumos. Para ello, se evaluaron tres tipos de zumos de frutas 100% exprimidos (piña, naranja y manzana)

y se adulteraron con zumo de uva en diferentes porcentajes (5%, 10%, 15%, 20%, 30%, 40% y 50%). Los resultados del análisis exploratorio revelaron una clara tendencia de agrupación de las muestras según el tipo de zumo analizado. El análisis supervisado, basado en el desarrollo de modelos de ML para la detección de la adulteración, obtuvo un rendimiento satisfactorio siendo el mejor resultado para las máquinas de vectores de soporte (SVM) con una *accuracy* del 95.45% en el conjunto de validación. En cuanto a la cuantificación del porcentaje del adulterante los mejores resultados se obtuvieron con la regresión de vectores de soporte (SVR) mostrando una  $R^2$  de 0.987 y un RMSE de 1.831 para el conjunto de validación. Además, se ha desarrollado una sencilla aplicación web en la que se pueden utilizar los modelos entrenados, por lo que cualquier usuario puede aplicar el método desarrollado en el laboratorio.

**Palabras claves:** Zumos de fruta, HS-GC-IMS, Machine Learning (ML), Adulteración.

#### ABSTRACT

Fruit juices are one of the most widely consumed beverages worldwide and their production is subject to European regulations. However, this food is often adulterated by adding water, sugars or other less expensive fruit juices. The latter is very common due to the greater difficulty to be detected. Therefore, the present study aimed to develop a methodology based on Headspace-Headspace-Gas Chromatography-Ion Mobility Spectrometry (HS-GC-IMS) in combination with machine learning (ML) algorithms to reliably detect fruit juice adulterations. For this purpose, three types of 100% squeezed fruit juices (pineapple, orange and apple) were evaluated and adulterated with grape juice in different ratios (5%, 10%, 15%, 20%, 30%, 40% and 50%). The exploratory analysis revealed a clustering trend to classify the samples according to the type of juice analyzed. The supervised analysis, based on the development of machine learning models for the detection of adulteration, achieved successful performance, obtaining the best result for the support vector machines with an accuracy of 95.45% in the test set. For adulterant quantification the best results were obtained with support vector regression showing an  $R^2$  of 0.987 and a RMSE of 1.831 for the test set. In addition, a simple web application has been developed where the trained models can be used, so any researcher can apply the method to detect this type of adulteration.

**Key words:** Fruits juices, HS-GC-IMS, Machine Learning (ML), Adulteration.

## 1. INTRODUCCIÓN

La adulteración de alimentos se trata de una práctica fraudulenta, llevada a cabo por motivos fundamentalmente económicos, que consiste en una pérdida de la calidad intencionalmente, ya sea mediante la adición de una sustancia ajena o la eliminación de algunas de las especies químicas que este contiene [1]. Esta práctica ilegal se trata de uno de los problemas más importantes que afronta hoy en día la industria agroalimentaria, y entre las adulteraciones alimenticias más comunes se encuentra los zumos de fruta [2].

En España existe una estricta regulación para garantizar la calidad y el origen en la producción de zumos, contemplada en el Real Decreto 781/2013 [3] de 11 de octubre, por el que se establecen normas relativas a la elaboración, composición, etiquetado, presentación y publicidad de los zumos de frutas y otros productos similares destinados a la alimentación humana. La Directiva 2001/112/CE del Consejo, de 20 de diciembre de 2001, relativa a los zumos de frutas y otros productos similares destinados a la alimentación humana, fue incorporada al ordenamiento jurídico nacional mediante el Real Decreto 1050/2003, de 1 de agosto, por el que se aprueba la Reglamentación técnico-sanitaria de zumos de frutas y de otros productos similares, destinados a la alimentación humana. Esta Directiva 2001/112/CE, de 20 de diciembre, ha sido objeto de sucesivas modificaciones, una de ellas por la Directiva 2009/106/CE de la Comisión, de 14 de agosto de 2009, que se dictó teniendo en cuenta la evolución de las normas internacionales pertinentes, en particular, la Norma General del Codex para Zumos (jugos y regulada en el ámbito europeo por la directiva 2012/12 UE [4]. Además, en estas regulaciones se describen varias categorías dentro de los zumos, así como productos similares. De esta forma, se encuentran: I) Zumo de fruta, que proviene 100% del exprimido de frutas sanas y maduras donde la adicción de azúcares está prohibida; II) Zumo concentrado, es el producto obtenido a partir del zumo por eliminación del agua; III) Zumo de fruta a partir de concentrado, es el producto obtenido al reconstituir un zumo concentrado con la adicción de agua; y IV) Néctar de fruta, son zumos, puros o concentrados de frutas, a los que se añade agua y, los únicos donde la adicción de azúcares y/o miel está permitido. Aquello que no esté englobado en estas categorías se considerarán bebidas de frutas.

A pesar de las normativas, estos productos siempre han sido objeto de adulteración en el mercado, de hecho, ha sido reportado como uno de los 7 alimentos más comunes de adulteración entre 1980 y 2010 [5]. Entre las adulteraciones más frecuentemente empleadas se encuentra la dilución con agua, la adición de edulcorantes artificiales o la adición de zumos de fruta menos costosos [6]. Esta última es muy popular debido a la mayor dificultad para ser detectada [7]. Es importante tener en cuenta que adicionar otros zumos de fruta y no informar al consumidor supone un riesgo adicional pudiendo producirle reacciones alérgicas [8]. Es por ello que, además de ser un fraude alimentario para obtener beneficios económicos, esta práctica podría afectar a la salud del consumidor [9]. Además, las formas de adulterar cada vez son más sutiles. Por todo esto, se hace necesario disponer de técnicas y metodologías analíticas adecuadas que permitan caracterizar y garantizar la calidad de los productos en función de su origen y tipo.

Las técnicas que se han empleado para la caracterización de zumos son muy diversas y estas incluyen, análisis fisicoquímicos [10,11], técnicas isotópicas y elementales [12-14], basadas en ADN [15,16] de separación [7,17-20] y espectroscópicas [6,8,21-23]. De todas ellas, actualmente la cromatografía tanto líquida como gaseosa es la metodología más utilizada [24]. Sin embargo, estas técnicas requieren el uso de disolventes, preparación de la muestra, largos tiempos de análisis y personal altamente cualificado. Por este motivo, se necesita disponer de técnicas analíticas rápidas entre las que destacan las denominadas narices electrónicas (e-Nose).

Técnicas como la cromatografía de gases-espectrometría de movilidad iónica (GC-IMS) se utilizan cada vez más en el fraude alimentario y pueden emplearse como nariz electrónica. La GC-IMS suele ir acoplada a técnicas de espacio de cabeza, (HS-GC-IMS) y el análisis está relacionado principalmente con los compuestos orgánicos volátiles presentes en las muestras. Además, tienen límites de detección muy bajos, en su mayoría en el rango de ppb [25]. Otra ventaja importante es la capacidad de analizar en tiempo real, lo que resulta muy útil para el control rutinario del fraude alimentario. Aunque estas técnicas se han utilizado en muchos campos diferentes [25-27] han demostrado ser muy prometedoras en la caracterización y el control de calidad de alimentos como la miel [28,29], el café [30] o el aceite [31-33]. Sin embargo, según la literatura existente, nunca se ha aplicado al análisis de muestras de zumos.

Por otra parte, la mayoría de los métodos analíticos se basan en la identificación de uno o varios compuestos individuales que son usados como biomarcadores para detectar el adulterante o la autenticidad del zumo, por lo que son fáciles de eludir y, por tanto, cada vez menos exitosas. El hecho de centrarse en unos compuestos diana, a pesar de que las técnicas analíticas proporcionan una gran cantidad de información, se debe fundamentalmente a la complejidad en el tratamiento de datos y en la interpretación de los mismos para extraer la información más relevante. Sin embargo, en la actualidad, debido al auge de metodologías basadas en el aprendizaje automatizado o *machine learning* (ML), las técnicas de perfiles globales se están volviendo cada vez más populares para desarrollar métodos de cribado rápido [9].

En base a todo lo anterior, el objetivo de este estudio es desarrollar una metodología para la detección y cuantificación de la adulteración de diferentes zumos de fruta (naranja, piña y manzana) con otro de menor coste (uva) utilizando HS-GC-IMS en combinación con distintos algoritmos de ML. Entre dichos algoritmos se encuentran los modelos de bosque aleatorio (RF), las máquinas de vector de soporte (SVM), el análisis discriminante lineal (LDA) y la regresión de mínimos cuadrados parciales (PLS).

## 2. MATERIALES Y MÉTODOS

### 2.1. Muestras

Se seleccionaron cuatro tipos de zumos de frutas 100% (manzana, piña, naranja y uva) de diferentes proveedores de España. De cada tipo de zumo, se eligieron un mínimo de tres marcas distintas y de cada una de ellas diversos lotes con el fin de aumentar la heterogeneidad dentro de cada tipo de zumo. De esta forma, las muestras se etiquetaron de acuerdo con el siguiente código de carácter: "A\_BB\_C" donde "A" hace referencia al tipo de zumo: "M" (manzana), "P" (piña), "N" (naranja) y, "U" (Uva), "BB" indica la marca y "C" el número de lote (1, 2 o 3). Además, cada una de ellas se analizó por duplicado quedando etiquetado como "R1" o "R2". Así, la etiqueta correspondiente al segundo lote de un zumo de manzana de la marca "HC" sería "M\_HC\_2\_R1" para la primera réplica y "M\_HC\_2\_R2" para la segunda. El número de muestras total ascendió a 76 con un mínimo de 18 muestras por cada tipo de zumo.

## 2.2. Adulteración

El zumo de uva se seleccionó como adulterante dado que su bajo coste, en comparación con el resto de los zumos, lo hace ideal para este tipo de prácticas fraudulentas. Con el objetivo de cubrir la mayor heterogeneidad posible en el proceso de adulteración, se crearon dos matrices madres de cada tipo de zumo. Dicha matriz madre consistió en mezclar de forma equilibrada (mismo volumen) una muestra de cada una de las marcas de cada tipo de zumo. Posteriormente, para simular las adulteraciones se mezcló la matriz madre de cada tipo de zumo de fruta puro (piña, manzana y naranja) con el correspondiente volumen de la matriz madre del adulterante (uva). Así, de cada tipo de zumo se obtuvieron 2 muestras diferentes que fueron adulteradas a distintos porcentajes (en volumen) comprendidos en el rango de 5-50% (5%, 10%, 15%, 20%, 30%, 40% y 50%). Además, se analizaron las muestras madres puras (0% de adulteración). Finalmente, el número de muestras analizadas fue de 96 correspondiente a 3 tipos de zumos x 2 Matrices Madres x 8 Porcentajes de adulteración (de 0 a 50%) x 2 réplicas. Las muestras fueron etiquetadas con el siguiente código de carácter: "AB\_C", donde "A" hace referencia al tipo de matriz madre utilizada ("P" para piña, "N" para naranja y "M" para manzana), "B" indica la matriz madre empleada (1 o 2) y "C" el porcentaje de adulteración (0, 5, 10, 15, 20, 30, 40 y 50). Al final de dicho código se añade "R1" para indicar la primera réplica y "R2" para la segunda. Por tanto, la primera matriz madre de piña adulterada al 10% quedaría etiquetada como "P1\_10\_R1" para la primera réplica y "P1\_10\_R2" para la segunda.

## 2.3. Espacio de Cabeza-Cromatografía de Gases-Espectrometría de Movilidad Iónica (HS- GC-IMS)

El análisis de las muestras se realizó mediante el equipo HS-GC-IMS concretamente el modelo Flavour Spec (GAS, Dortmund, Alemania). Este equipo cuenta con un sistema de creación de espacio de cabeza (HS) seguido de una columna de cromatografía de gases (GC) y un detector de espectrometría de movilidad iónica (IMS). Las condiciones utilizadas para el análisis de las muestras quedan recogidas en la Tabla 1, siendo el tiempo total de análisis de 20 minutos por muestra. Una vez analizadas todas las muestras, se obtuvo el espectro de suma de movilidad iónica (IMSS) que consiste en la suma de todas las intensidades en función de los tiempos de retención



procedentes del IMS, quedando anulada la información procedente de la separación cromatográfica [30,33]. El IMSS fue reducido a aquellas zonas donde aparecían señales características de los compuestos volátiles de las muestras analizadas. De esta forma, se obtuvieron un total de 882 tiempos de deriva que van desde el 1.020 hasta el 1.90 (relativo al RIP).

## 2.4. Análisis de datos

Los IMSS en bruto se almacenaron en matrices  $D_{n \times m}$ , donde  $n$  denota el número de muestras y  $m$  el número de variables. Por lo tanto, para el análisis de los datos se utilizó una matriz  $D_{172 \times 882}$  (172 muestras y 882 tiempos de derivas). Todos los análisis se realizaron con el software RStudio versión 4.0.2 (RStudio Team 2021, Boston, MA, USA). Se utilizó el paquete *mclust* [34] para realizar el *model-based clustering* y el paquete *ggplot2* [35] para las representaciones gráficas. El entrenamiento de los modelos se realizó con el paquete *caret* [36] y el desarrollo de la aplicación se realizó mediante *shiny* [37].

## 3. RESULTADOS Y DISCUSIÓN

### 3.1. Análisis exploratorio

El objetivo de esta primera parte del estudio fue observar si existe una tendencia de agrupación de las muestras en función del tipo de zumos, es decir, una diferencia espectral basada en el perfil aromático de cada tipo de muestra. Para ello, se utilizaron todas las muestras de zumo no adulteradas y se comprobó mediante una técnica no supervisada conocida como *model-based clustering*. Esta técnica permite encontrar patrones en los datos usando modelos de mezcla gaussiana. En este tipo de algoritmo, el número de grupos ( $k$ ) debe ser previamente determinado por el analista. En este caso, se tomó el valor de  $k=4$  debido precisamente a la existencia de 4 tipos de zumos y el resultado queda recogido en la Figura 1. En esta se representa las observaciones en función de las dos primeras componentes principales (PC1 y PC2) obtenidas del análisis de componentes principales (PCA), además la forma y el color se da en función del grupo obtenido por el *model-based clustering* cuyos centroides se resaltan con un mayor tamaño.

Como se puede observar todas las muestras de zumo de uvas se encuentran en el mismo grupo (clúster verde oscuro, símbolo "+"). Además, si se

observa la información proporcionada por la PCA, estas muestras poseen altos valores en la PC1 y esto sirve para diferenciarlas del resto. También, se observa que las muestras de zumo de uva de la marca "HC" son más distintas a la marca "CA" y "GR" que se encuentran a una menor distancia entre ellas. Es muy importante, diferenciar el zumo de uva del resto, dado que este es el utilizado como adulterante. Por tanto, se ve claramente una tendencia de agrupación en función del tipo de zumo y en menor medida de la marca. Por otro lado, las muestras de zumo de piña forman un único clúster (coloreado de verde claro, símbolo "."), y se pueden diferenciar en función de las dos primeras PCs. Centrándose en las muestras de zumo de manzana se observan que todas ellas están dentro del clúster rojo (símbolo "■"). Sin embargo, en este mismo clúster se encuentran algunas muestras de zumo de naranja, concretamente de la marca "ZS". Por último, el resto de muestras de zumo de naranja se encuentran en el clúster naranja (símbolo "▲").

Por tanto, existe una diferencia en el perfil aromático de los distintos tipos de zumos y, el IMSS está claramente influenciado, en primer lugar, por el tipo de zumo de fruta y, en segundo lugar, por la marca utilizada. En general, el *model-based clustering* permitió distinguir casi perfectamente el tipo de zumo analizado independientemente de la marca o la muestra utilizada.

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

Una vez que se ha comprobado la tendencia de agrupación de las muestras en función del tipo de fruta, se procedió a estudiar la capacidad de la técnica HS-GC-IMS en combinación con modelos supervisados para obtener un método que permita la correcta identificación del tipo de zumo, así como la detección de posibles adulterantes. Es por ello, que en este caso además de utilizar los IMSS de las muestras procedente de los distintos tipos de zumos de fruta (naranja, piña, manzana y uva) usadas en el análisis exploratorio, se obtuvieron los IMSS de las muestras madres de cada tipo de zumo adulteradas a distintos porcentajes (5%, 10%, 15%, 20%, 30%, 40% y 50%). Por lo tanto, la matriz de datos obtenida para este análisis es  $D_{172 \times 882}$ . Para los métodos supervisados, se establecieron un total de 4 grupos a priori en función del tipo de zumo empleado ("Piña", "Manzana" y "Naranja") y de la presencia de adulteración ("Adulterado"). En este último, se incluyen las muestras adulteradas procedentes de todos los zumos a los distintos porcentajes, así como las muestras de zumo de uvas pura ya que es el adul-

terante utilizado. Para resolver este problema se han evaluado distintos algoritmos entre ellos, técnicas paramétricas como el análisis discriminante lineal (LDA) y no paramétrica como bosque aleatorio (RF) y las máquinas de vectores de soporte (SVM). Todos los algoritmos se han ajustado con un conjunto de entrenamiento constituido por el 75% de las muestras y, el 25% restante constituyen el conjunto de validación que es usado para evaluar el rendimiento del modelo. Además, se aseguró que ambos conjuntos estuvieran balanceados y, formados por muestras de todos los tipos de zumos, así como de todos los porcentajes de adulteración. En la Tabla 2 se recoge un resumen de la *accuracy* obtenida para cada uno de los modelos entrenados. Como se puede observar el mejor rendimiento se obtiene para el modelo SVM con una *accuracy* del 100% en el conjunto de entrenamiento y del 95.45% en el conjunto de validación. Por tanto, este modelo únicamente clasificó incorrectamente 2 observaciones como adulteradas, cuando realmente pertenecía a un zumo de manzana puro. El segundo mejor modelo fue el LDA que mostró una *accuracy* del 100% y 93.18% para el conjunto de entrenamiento y validación respectivamente. En este último se clasificaron incorrectamente 3 observaciones, dos de ellas se trataban de un zumo de naranja y manzana puro predicho como adulterado, y la otra era una muestra de piña adulterada al 20% que se predijo como manzana pura. Por último, el modelo RF obtuvo un rendimiento más bajo siendo del 90.91% para el conjunto de validación.

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

Una vez obtenido un modelo capaz de detectar la presencia de adulteración, la siguiente tarea es determinar la cantidad de adulterante utilizada. Para ello, se llevó a cabo una regresión global, con el fin de evaluar la posibilidad de detectar el porcentaje de adulterante independientemente del tipo de zumo. El conjunto de datos utilizado consistió en 96 muestras, es decir, 32 de cada tipo de zumo, donde se incluyen los distintos porcentajes de adulteración (5%, 10%, 15%, 20%, 30%, 40% y 50%), así como las muestras sin adulterar (0%). Los datos se dividieron de forma balanceada en un conjunto de entrenamiento y validación, constituido por el 80% y el 20% de las muestras, respectivamente. El conjunto de validación contenía un total de 24 muestras, una por cada nivel de adulteración y tipo de zumo. Las distintas técnicas de regresión evaluadas fueron: la regresión por mínimos cuadrados parciales (PLS), la regresión basada en máquinas de vector de soporte (SVR) y la regresión por bosque aleatorio.

Antes de aplicar dichos algoritmos, fue necesario realizar un tratamiento previo de selección de variables. El objetivo es reducir el número de variables y seleccionar sólo aquellas que estén relacionadas con la respuesta (porcentaje de adulteración). De este modo, se eliminaron los tiempos de deriva que generan ruido en los modelos de regresión, consiguiendo mejorar el rendimiento de los mismos. Para ello, se utilizó el algoritmo Boruta sobre el conjunto de entrenamiento, el cual seleccionó 53 variables. El resto de ellas (828 tiempos de deriva) se eliminaron tanto del conjunto de entrenamiento como del de validación. Una vez realizado esto, se entrenaron los modelos de regresión y un resumen del rendimiento de los mismo se encuentra en la Tabla 3. Como se puede observar el mejor resultado se obtuvo para SVR, con un error cuadrático medio (RMSE) de 1.374 y una  $R^2$  de 0.996 para el conjunto de entrenamiento mientras que para el conjunto de validación el RMSE fue de 1.831 y la  $R^2$  de 0.987. En base a los resultados, este modelo permite cuantificar el porcentaje de adulteración de una forma bastante precisa. Sin embargo, tanto RF como PLS obtienen un RMSE demasiado alto (superior a 6) por lo que no son modelos tan exactos para la cuantificación del adulterante.

### 3.4. Desarrollo de la aplicación

Los modelos de ML en combinación con el IMSS han obtenido un excelente rendimiento para la detección de adulteraciones en zumo de frutas, así como para la cuantificación del mismo. Por este motivo, se ha creado una sencilla aplicación que contiene los modelos basados en SVM lo que permite compartir el algoritmo y facilitar la identificación de las adulteraciones al resto de usuarios. Esto es muy importante ya que generalmente las muestras deben ser analizadas por GC-MS y la presencia de adulterante se determina por la identificación individual de determinados compuestos dianas, con esta simple aplicación se ahorraría tiempo y costes, y se puede acceder en el siguiente enlace:

[https://jose Luispecalle.shinyapps.io/Almendralejo\\_APP/](https://jose Luispecalle.shinyapps.io/Almendralejo_APP/)

Esta aplicación es un prototipo y realiza automáticamente la predicción mediante el modelo SVM para la detección del adulterante y SVR para la cuantificación del mismo. Por tanto, para su uso, sólo es necesario cargar el archivo Excel con el espectro (IMSS) obtenido del análisis de las muestras por HS-GC-IMS. Además, se ha introducido un archivo de ejemplo (des-

cargar clicando en el botón “Download”) que puede utilizarse para realizar la predicción.

Aunque es una aplicación muy sencilla, permite compartir los algoritmos creados y facilitar el análisis de las muestras de zumos a otros usuarios. Cabe destacar que tanto la aplicación como el modelo pueden actualizarse con el análisis de más muestra ya que son modelos de aprendizaje, lo que a su vez puede llevar a una mejora en la predicción al abarcar mayor heterogeneidad. De este modo, la base de datos se irá ampliando progresivamente y abarcará más situaciones de adulteración, satisfaciendo de este modo las actuales demandas de la industria agroalimentaria.

#### 4. CONCLUSIONES

Se ha demostrado el potencial de HS-GC-IMS en combinación con algoritmos de ML para la detección y cuantificación de adulteraciones con zumo de uva en otros tipos de zumos (naranja, manzana y piña). Los modelos basados en SVM han obtenido los mejores resultados, con una *accuracy* del 95.45% en el conjunto de validación para la detección del adulterante, y una  $R^2$  de 0.987 y un RMSE de 1.831 para la cuantificación del mismo. Además, el *model-based clustering* reveló que el tipo de zumo de fruta analizado afecta en gran medida al IMSS, y en menor medida la marca utilizada. Cabe destacar que la técnica HS-GC-IMS ofrece varias ventajas frente a los métodos tradicionales, como el corto tiempo de análisis, el bajo coste, la simplicidad y el hecho de ser respetuosa con el medio ambiente. Además, se trata de una técnica de fácil portabilidad, permitiendo el análisis *in situ* en la industria alimentaria, lo que supone una ventaja muy importante, ya que permite una menor manipulación de la muestra y facilita la toma de decisiones lo antes posible. Destacar que este es el primer estudio en el análisis de zumos utilizando HS-GC-IMS y proporciona un punto de partida en este campo de investigación.

## 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.

## 5. REFERENCIAS

1. Spink, J.; Moyer, D.C. Defining the Public Health Threat of Food Fraud. *J. Food Sci.*, 2011, 76, doi:10.1111/j.1750-3841.2011.02417.x.
2. Valand, R.; Tanna, S.; Lawson, G.; Bengtström, L. A review of Fourier Transform Infrared (FTIR) spectroscopy used in food adulteration and authenticity investigations. *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* 2020, 37, 19–38.
3. La Presidencia, M. DE *Disposición 10611 del BOE núm. 245 de 2013*; 2013.
4. European Parliament Directive 2012/12/EU of the European Parliament and of the Council of 19 April 2012 amending Council Directive 2001/112/EC relating to fruit juices and certain similar products intended for human consumption. *Off. J. Eur. Union*, 2012, 1–11.
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. Ammari, F.; Redjal, L.; Rutledge, D.N. Detection of orange juice frauds using front-face fluorescence spectroscopy and Independent Components Analysis. *Food Chem.*, 2015, 168, 211–217, doi:10.1016/j.foodchem.2014.06.110.

7. Róžańska, A.; Dymerski, T.; Namieśnik, J. Novel analytical method for detection of orange juice adulteration based on ultra-fast gas chromatography. *Monatshefte fur Chemie*, 2018, 149, 1615–1621, doi:10.1007/S00706-018-2233-8/TABLES/3.
8. Boggia, R.; Casolino, M.C.; Hysenaj, V.; Oliveri, P.; Zunin, P. A screening method based on UV-Visible spectroscopy and multivariate analysis to assess addition of filler juices and water to pomegranate juices. In *Proceedings of the Food Chemistry*; Elsevier, 2013; Vol. 140, pp. 735–741.
9. Dasenaki, M.E.; Thomaidis, N.S. Quality and authenticity control of fruit juices-a review. *Molecules* 2019, 24, 1014.
10. Lorente, J.; Vegara, S.; Martí, N.; Ibarz, A.; Coll, L.; Hernández, J.; Valero, M.; Saura, D. Chemical guide parameters for Spanish lemon (*Citrus limon* (L.) Burm.) juices. *Food Chem.*, 2014, 162, 186–191, doi:10.1016/j.foodchem.2014.04.042.
11. Dzugan, M.; Wesołowska, M.; Zagała, G.; Puchalski, C. The comparison of the physicochemical parameters and antioxidant activity of homemade and commercial pomegranate juices. *Acta Sci. Pol. Technol. Aliment.*, 2018, 17, 59–68, doi:10.17306/J.AFS.2018.0529.
12. Bononi, M.; Quaglia, G.; Tateo, F. Preliminary LC-IRMS Characterization of Italian Pure Lemon Juices and Evaluation of Commercial Juices Distributed in the Italian Market. *Food Anal. Methods*, 2016, 9, 2824–2831, doi:10.1007/s12161-016-0479-5.
13. Nuncio-Jáuregui, N.; Calín-Sánchez, Á.; Hernández, F.; Carbonell-Barrachina, Á.A. Pomegranate juice adulteration by addition of grape or peach juices. *J. Sci. Food Agric.*, 2014, 94, 646–655, doi:10.1002/jsfa.6300.
14. Cristea, G.; Dehelean, A.; Voica, C.; Feher, I.; Puscas, R.; Magdas, D.A. Isotopic and Elemental Analysis of Apple and Orange Juice by Isotope Ratio Mass Spectrometry (IRMS) and Inductively Coupled Plasma-Mass Spectrometry (ICP-MS). *Anal. Lett.*, 2021, 54, 212–226, doi:10.1080/00032719.2020.1743717.

15. Liang, Y. lin; Ding, Y. jie; Liu, X.; Zhou, P. fei; Ding, M. xuan; Yin, J. jun; Song, Q. hou A duplex PCR-RFLP-CE for simultaneous detection of mandarin and grapefruit in orange juice. *Eur. Food Res. Technol.*, 2021, 247, 1-7, doi:10.1007/s00217-020-03602-z.
16. Pardo, M.A. Evaluation of a dual-probe real time PCR system for detection of mandarin in commercial orange juice. *Food Chem.* **2015**, 172, 377-384, doi:10.1016/j.foodchem.2014.09.096.
17. Li, S.; Hu, Y.; Liu, W.; Chen, Y.; Wang, F.; Lu, X.; Zheng, W. Untargeted volatile metabolomics using comprehensive two-dimensional gas chromatography-mass spectrometry – A solution for orange juice authentication. *Talanta*, 2020, 217, 121038, doi:10.1016/j.talanta.2020.121038.
18. Bocharova, O. V.; Reshta, S.P.; Bocharova, M.Y.; Eshtokin, V.I. Determination of apple juice authenticity using gas chromatography-mass spectrometry, volt-ampere and new conductometric methods. *J. Sci. Food Agric.*, 2018, 98, 4692-4699, doi:10.1002/JSFA.9003.
19. Abad-García, B.; Garmón-Lobato, S.; Sánchez-Ilárduya, M.B.; Berrueta, L.A.; Gallo, B.; Vicente, F.; Alonso-Salces, R.M. Polyphenolic contents in Citrus fruit juices: Authenticity assessment. *Eur. Food Res. Technol.*, 2014, 238, 803-818, doi:10.1007/s00217-014-2160-9.
20. Spinelli, F.R.; Dutra, S. V.; Carnieli, G.; Leonardelli, S.; Drehmer, A.P.; Vanderlinde, R. Detection of addition of apple juice in purple grape juice. *Food Control*, 2016, 69, 1-4, doi:10.1016/j.foodcont.2016.04.005.
21. Jahani, R.; Yazdanpanah, H.; van Ruth, S.M.; Kobarfard, F.; Alewijn, M.; Mahboubi, A.; Faizi, M.; Aliabadi, M.H.S.; Salamzadeh, J. Novel application of near-infrared spectroscopy and chemometrics approach for detection of lime juice adulteration. *Iran. J. Pharm. Res.*, 2020, 19, 34-44, doi:10.22037/ijpr.2019.112328.13686.
22. Ellis, D.I.; Ellis, J.; Muhamadali, H.; Xu, Y.; Horn, A.B.; Goodacre, R. Rapid, high-throughput, and quantitative determination of orange juice adulteration by Fourier-transform infrared spectroscopy. *Anal. Methods*, 2016, 8, 5581-5586, doi:10.1039/c6ay01480a.



23. Luis, J.; Calle, P.; Ferreiro-González, M.; Ruiz-Rodríguez, A.; Fernández, D.; Palma, M. Detection of Adulterations in Fruit Juices Using Machine Learning Methods over FT-IR Spectroscopic Data. *Agronomy*, 2022, 12, 683, doi:10.3390/AGRONOMY12030683.
24. Snyder, A.B.; Sweeney, C.F.; Rodriguez-Saona, L.E.; Giusti, M.M. Rapid authentication of concord juice concentration in a grape juice blend using Fourier-Transform infrared spectroscopy and chemometric analysis. *Food Chem.*, 2014, 147, 295–301, doi:10.1016/j.foodchem.2013.09.140.
25. Vautz, W.; Franzke, J.; Zampolli, S.; Elmi, I.; Liedtke, S. On the potential of ion mobility spectrometry coupled to GC pre-separation – A tutorial. *Anal. Chim. Acta*, 2018, 1024, 52–64, doi:10.1016/J.ACA.2018.02.052.
26. Calle, J.L.P.; Ferreiro-González, M.; Aliaño-González, M.J.; Barbero, G.F.; Palma, M. Characterization of Biodegraded Ignitable Liquids by Headspace-Ion Mobility Spectrometry. *Sensors*, 2020, 20, 6005, doi:10.3390/S20216005.
27. Aliaño-González, M.J.; Ferreiro-González, M.; Barbero, G.F.; Palma, M. Novel method based on ion mobility spectrometry sum spectrum for the characterization of ignitable liquids in fire debris. *Talanta*, 2019, 199, 189–194, doi:10.1016/J.TALANTA.2019.02.063.
28. Arroyo-Manzanares, N.; García-Nicolás, M.; Castell, A.; Campillo, N.; Viñas, P.; López-García, I.; Hernández-Córdoba, M. Untargeted headspace gas chromatography – Ion mobility spectrometry analysis for detection of adulterated honey. *Talanta*, 2019, 205, 120123, doi:10.1016/j.talanta.2019.120123.
29. Aliaño-González, M.J.; Ferreiro-González, M.; Espada-Bellido, E.; Palma, M.; F. Barbero, G. A Screening Method Based on Headspace-Ion Mobility Spectrometry to Identify Adulterated Honey. *Sensors*, 2019, 19, 1621, doi:10.3390/s19071621.
30. Piotr Konieczka, P.; Aliaño-González, M.J.; Ferreiro-González, M.; Barbero, G.F.; Palma, M. Characterization of Arabica and Robusta Coffees by Ion Mobility Sum Spectrum. *Sensors*, 2020, 20, 3123, doi:10.3390/s20113123.

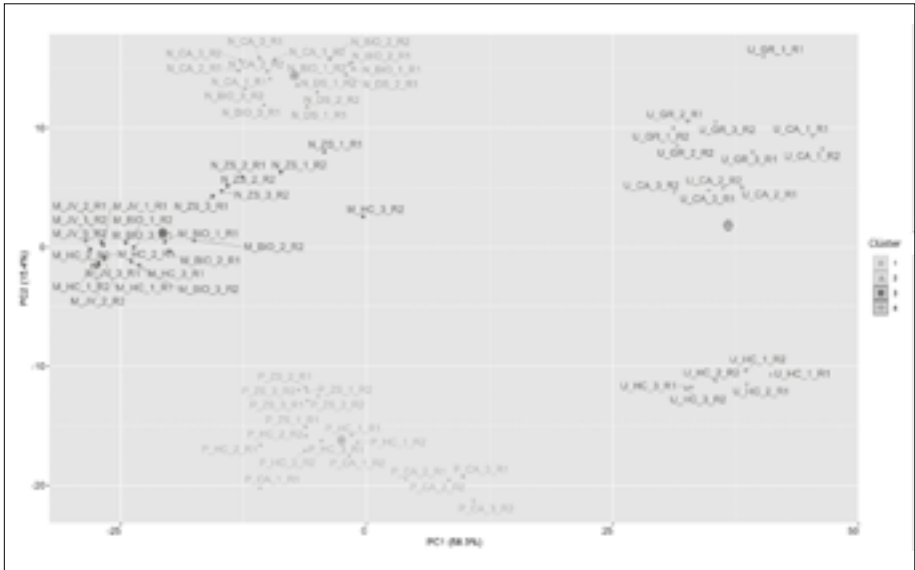
31. Gerhardt, N.; Schwolow, S.; Rohn, S.; Pérez-Cacho, P.R.; Galán-Soldevilla, H.; Arce, L.; Weller, P. Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: Comparison of different processing approaches by LDA, kNN, and SVM. *Food Chem.*, 2019, 278, 720–728, doi:10.1016/J.FOODCHEM.2018.11.095.
32. Contreras, M. del M.; Arroyo-Manzanares, N.; Arce, C.; Arce, L. HS-GC-IMS and chemometric data treatment for food authenticity assessment: Olive oil mapping and classification through two different devices as an example. *Food Control*, 2019, 98, 82–93, doi:10.1016/J.FOODCONT.2018.11.001.
33. Zhang, L.; Shuai, Q.; Li, P.; Zhang, Q.; Ma, F.; Zhang, W.; Ding, X. Ion mobility spectrometry fingerprints: A rapid detection technology for adulteration of sesame oil. *Food Chem.*, 2016, 192, 60–66, doi:10.1016/j.foodchem.2015.06.096.
34. Scrucca, L.; Fop, M.; Murphy, T.B.; Raftery, A.E. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J.*, 2016, 8, 289, doi:10.32614/rj-2016-021.
35. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag New York, 2016; ISBN 978-3-319-24277-4.
36. Kuhn, M. *caret: Classification and Regression Training* 2020.
37. Chang, W.; Cheng, J.; Allaire, J.J.; Xie, Y.; McPherson, J. *shiny: Web Application Framework for R* 2020.

## FIGURAS Y TABLAS

**Tabla 1.** Condiciones utilizadas para el análisis de las muestras mediante HS-GC-IMS.

	<b>Variable</b>	<b>Valor</b>
<b>HS</b>	Volumen de muestra ( $\mu\text{L}$ )	750
	Tiempo de incubación (min)	5
	Temperatura de incubación ( $^{\circ}\text{C}$ )	46.3
	Velocidad de agitación (rpm)	750
	Volumen de inyección ( $\mu\text{L}$ )	100
	Velocidad de llenado de jeringa ( $\mu\text{L}/\text{s}$ )	900
<b>GC-IMS</b>	Temperatura de jeringa ( $^{\circ}\text{C}$ )	51.3
	Tiempo de lavado (min)	5
	Velocidad de inyección ( $\mu\text{L}/\text{s}$ )	850
	EPC1 (mL/min)	250
	EPC2 (mL/min)	2 mL/min (t= 0 min), 10 mL/min (t=5 min), 25 mL/min (t=10 min)
	Temperatura del sistema ( $^{\circ}\text{C}$ )	45
	Temperatura de columna ( $^{\circ}\text{C}$ )	55
Temperatura del tubo de deriva ( $^{\circ}\text{C}$ )	80	
<b>Tiempo total de análisis (min)</b>		<b>20</b>

**Figura 1.** Resultado obtenido tras el análisis por *model-based clustering* con 4 clúster de los IMSS de todas las muestras de zumos puros ( $D_{882 \times 76}$ ). Las observaciones se representan usando las dos componentes principales y los colores y formas son dados en función del grupo resultante de aplicar *model-based clustering*.



**Tabla 2.** Resultados obtenidos para cada modelo de clasificación aplicado en la detección del adulterante utilizando los IMSS de todas las muestras analizadas ( $D_{172 \times 882}$ ).

Modelo	Hiperparámetros	Accuracy CONJUNTO DE ENTRENAMIENTO	Accuracy CONJUNTO de validación
LDA	-	100%	100%
SVM	$C = 2.83$ $\gamma = 0.022$	100%	100%
RF	$mtry = 23$ $ntree = 500$	100%	97.67%

**Tabla 3.** Resultados obtenidos para cada modelo de regresión aplicado en la cuantificación del adulterante global utilizando los 53 tiempos de derivas seleccionado de todas las muestras de zumo adulterado ( $D_{96 \times 53}$ ).

Modelo	HyperparÁmetROS	Rendimiento conjunto de entrenamiento	Rendimiento conjunto de validación
PLS	10 componentes principales	RMSE = 5.961 $R^2 = 0.809$	RMSE = 8.531 $R^2 = 0.711$
SVR	$C = 11.31$ $\gamma = 0.0883$	RMSE = 1.374 $R^2 = 0.996$	RMSE = 1.831 $R^2 = 0.987$
RF	$mtry = 3$	RMSE = 3.784 $R^2 = 0.962$	RMSE = 6.003 $R^2 = 0.813$