



UNIVERSIDAD NACIONAL DEL SUR

TESIS DE DOCTOR EN QUÍMICA

*“DETERMINACIÓN DE LA ADULTERACIÓN DE QUESOS RALLADOS
MEDIANTE METODOLOGÍAS ANALÍTICAS DE SCREENING Y ANÁLISIS
MULTIVARIANTE”*

Licenciado en Química Lucas Gabriel Visconti

BAHÍA BLANCA

ARGENTINA

2023

PREFACIO

Esta tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Química de la Universidad Nacional del Sur y no ha sido presentada para la obtención de otro título en esta Universidad u otras. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Departamento de Química durante el período comprendido entre el 3 de Julio del 2018 y el 12 de julio del 2023, bajo la dirección de la Dra. María Susana Rodríguez, Profesora Titular, y la codirección de la Dra. Carolina Vanesa Di Aníbal, Profesora adjunta e Investigadora Adjunta. Ambos cargos docentes lo desarrollan en el Área III, Química Analítica, del Departamento de Química de la Universidad Nacional del Sur, mientras que el cargo de investigador de la Dra. Di Aníbal pertenece al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

Lic. En Química Lucas Gabriel Visconti

	<p>UNIVERSIDAD NACIONAL DEL SUR</p> <p>Secretaría General de Posgrado y Educación Continua</p>
<p>La presente Tesis ha sido aprobada el 25/08/2023 , mereciendo la calificación de 10 (diez)</p>	

AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a mis directores de tesis, la Dra. María Susana Rodríguez y la Dra. Carolina Di Aníbal, por sus enseñanzas, consejos y confianza brindada. Por su cariño y palabras de apoyo. Porque juntos pudimos sobrellevar cada circunstancia, y sin su dedicación esta tesis doctoral no hubiera sido posible.

Agradezco de corazón a los docentes del Departamento de Química, en particular a las Dras. Liliana Albertengo y Adriana Debbaudt, por su invaluable apoyo y orientación durante mis primeros años de doctorado.

Agradezco sinceramente a aquellos que han colaborado y realizado aportes a lo largo de mi formación, en particular al Dr. Claudio Delrieux y al Ing. Steven Martínez Vargas. Su apoyo y contribuciones han sido invaluable para mi desarrollo académico y profesional. También quiero expresar mi agradecimiento a los Dres. Paulo Henrique Gonçalves Dias Diniz y David Douglas de Sousa Fernández. Su contribución y asesoramiento en mi investigación han sido esenciales para enriquecer los resultados y las conclusiones de esta tesis.

Agradezco de corazón al Dr. Marcelo Fabián Pistonesi y a su grupo de investigación por recibirme cálidamente durante mi última etapa del doctorado. Su acogida y apoyo fueron fundamentales para el éxito de mi trabajo de investigación. A Fede y Marianito, quiero expresar mi profundo agradecimiento por su disposición y voluntad de ayudarme en todo momento.

Agradezco de todo corazón a todos aquellos que me brindaron palabras de apoyo para poder continuar en mi camino, en especial a mis amigos Nico, Flor y Damy. Su constante aliento y respaldo han sido un bálsamo en los momentos difíciles y han sido una fuente de motivación para seguir adelante.

A mis compañeros docentes y amigos que han compartido conmigo lindos momentos, quiero expresar mi sincero agradecimiento. A Nati, Anita, Claus, Gaby y Agus, gracias por ser parte de mi trayectoria y por hacer de nuestro entorno de trabajo un lugar lleno de camaradería y alegría. Su compañía y su amistad han enriquecido mi experiencia como docente y han hecho de cada día algo especial.

A la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC), y al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), por haberme otorgado becas para formación doctoral.

Por último, pero más importante, quiero decirle gracias a mi pareja Cyn. Durante este arduo proceso, has sido mi mayor apoyo, estando a mi lado tanto en los momentos felices como en los desafiantes. Su confianza inquebrantable en mí ha sido fundamental para superar los obstáculos y perseverar hasta el final. Su amor, paciencia y aliento constante han sido mi mayor fortaleza. No puedo expresar lo agradecido que estoy por tenerte en mi vida. Te amo y te estaré agradecido por siempre.

Finalmente, gracias Dios por estar siempre en mi corazón y guiar cada paso de mi vida.

RESUMEN

El valor nutricional, la característica sensorial y la seguridad higiénica de los alimentos son campos de gran repercusión social y elevado interés científico por los cuales, tanto las industrias alimentarias como los organismos nacionales e internacionales de regulación, juegan un papel fundamental a la hora de garantizar la calidad y seguridad de los alimentos que llegan a los consumidores. El control de calidad de los alimentos y el cumplimiento de regulaciones establecidas hacen necesario el desarrollo de métodos analíticos para detectar, entre otras cosas, la presencia de aditivos alimentarios en concentraciones superiores a las establecidas, como también el agregado de sustancias no autorizadas, para poder asegurar la inocuidad de los alimentos.

El objetivo principal de esta tesis es proponer herramientas analíticas para determinar la adulteración de alimentos de alto valor comercial como quesos rallados de pasta dura, susceptibles de prácticas fraudulentas para obtener beneficios comerciales. Para ello, se realizaron los siguientes trabajos.

- Se desarrolló una nueva metodología basada en el uso de la espectroscopía NIR junto con clasificación multivariante, que permite identificar adulteraciones.

- Se diseñó y fabricó, en colaboración con el Laboratorio de Ciencias de las Imágenes (LCI) del Instituto de Investigaciones en Ingeniería Eléctrica de la UNS, un novedoso dispositivo portátil de toma de imágenes basado en LEDs y sensores específicos, que permite efectuar determinaciones *in-situ*.

- Se implementó una metodología original basada en el análisis de imágenes digitales, junto con técnicas de aprendizaje automático, y empleando el dispositivo portátil diseñado.

- Se llevó a cabo una nueva metodología basada en la espectroscopía NIR y regresión multivariante, que permite cuantificar adulteraciones.

Los métodos analíticos propuestos son fiables, rápidos, simples y económicos, y representan una excelente alternativa para su aplicación por organismos de regulación y laboratorios de control de calidad de alimentos, con potencial de poder aplicarse *in-situ* debido a su ventaja de portabilidad.

ABSTRACT

Nutritional value, sensory characteristics and hygienic safety of foods are fields of significant social impact and high scientific interest. In this context, both food industries and national and international regulatory authorities play a fundamental role in ensuring the quality and safety of foods destined to consumers. Food quality control and the compliance with established regulations require the development of analytical methods to detect the presence of food additives in concentrations higher than those established as well as the addition of unauthorized substances, in order to ensure food safety.

The main objective of this Thesis is to propose analytical tools to determine the adulteration of high-commercial value foods such as grated-hard cheeses, which are susceptible to fraud for commercial benefits. For this purpose, the following work was performed.

- A new methodology based on the use of NIR spectroscopy together with multivariate classification was developed to identify adulterations.

- A novel portable imaging device based on LEDs and specific sensors was designed and manufactured to perform in-situ determinations, in collaboration with the Imaging Sciences Laboratory (LCI) of the Electrical Engineering Research Institute of the National South University (UNS).

- A novel methodology based on digital-image analysis together with machine learning techniques was performed by using the designed device.

- A new methodology based on NIR spectroscopy and multivariate regression was used to quantify adulterations.

The proposed methods are reliable, fast, simple and economic, and represent an excellent alternative for regulatory agencies and food quality control laboratories, with potential to be applied in any stage of the analytical problem because they have the advantage of being portable.

INDICE GENERAL

ESTRUCTURA.....	11
CAPÍTULO I: ALCANCES Y OBJETIVOS	13
1.1. Alcances	13
1.2. Objetivos de esta tesis	16
1.3. Referencias.....	17
CAPÍTULO II: ASPECTOS TEÓRICOS.....	19
2.1. Alimentos.....	19
2.2. Quesos	19
2.2.1. Definición y proceso de elaboración	19
2.2.2. Clasificación de quesos	24
2.2.3. Quesos de pasta dura	24
2.2.3.1. Quesos rallados: aditivos antiaglutinantes permitidos y agregados no permitidos	25
2.3. Adulteraciones alimentarias	29
2.4. Métodos de screening.....	31
2.5. Técnicas analíticas	33
2.5.1. Espectroscopía de infrarrojo cercano (NIR).....	34
2.5.2. Imágenes digitales	39
2.6. Análisis de datos multivariante.....	47
2.6.1. Pre-procesamiento de espectros NIR.....	47
2.6.1.1. Métodos de corrección de scattering (dispersión).....	47
2.6.1.2. Derivadas espectrales	50
2.6.2. Preparación de conjuntos de muestras (Algoritmo Kennard-Stone).....	51
2.6.3. Análisis cuantitativo: Regresión multivariante	52
2.6.3.1. Mínimos Cuadrados Parciales (PLS).....	52
2.6.3.2. Parámetros de calidad en regresión multivariante	53

2.6.4. Análisis cualitativo: Clasificación multivariante	55
2.6.4.1. Técnicas de reconocimiento de patrones no supervisadas.....	58
I. Análisis por Componentes Principales (PCA)	59
2.6.4.2. Técnicas de reconocimiento de patrones supervisadas.....	60
I. <i>K</i> -Vecinos más Cercanos (<i>K</i> -NN)	61
II. Análisis Discriminante de Mínimos Cuadrados Parciales (PLS-DA).62	
III. Árboles de decisión (DTs) y árboles aleatorios (RTs)	64
IV. Máquinas de Soporte Vectorial (SVM).....	66
V. Regresión logística (LR)	67
VI. Modelado Suave e Independiente por Analogías de Clases (SIMCA).	
.....	68
Técnicas de Clasificación aplicadas al modelado de una clase (One-	
class)	69
I. Modelado Suave e Independiente por Analogías por Clases Impulsado	
por Datos (DD- SIMCA).....	69
II. Mínimos Cuadrados Parciales de Una Clase (OC-PLS)	70
2.6.4.3. Parámetros de calidad en clasificación multivariante.....	71
2.6.5. Validación cruzada	74
2.7. Referencias.....	76
CAPÍTULO III: DETERMINACIÓN DE LA ADULTERACIÓN DE QUESOS RALLADOS	
MEDIANTE ESPECTROSCOPÍA NIR Y ANÁLISIS MULTIVARIANTE	85
3.1. Objetivo	85
3.2. Materiales y métodos.....	85
3.2.1. Muestras y reactivos.....	85
3.2.2. Instrumentación.....	86
3.2.3. Análisis composicional.	87
3.2.4. Análisis multivariante.....	87
3.3. Resultados y discusión	88

3.3.1. Espectros NIR.	88
3.3.2. Análisis exploratorio.	89
3.3.3. Clasificación Supervisada.....	90
3.3.4. Análisis composicional de muestras sospechosas de estar adulteradas. ...	91
3.3.5. Nuevos modelos de clasificación.....	93
3.3.6. Predicción de muestras.....	95
3.4. Conclusiones	97
3.5. Referencias.....	98
CAPÍTULO IV: DETERMINACIÓN DE LA ADULTERACIÓN DE QUESOS RALLADOS MEDIANTE IMÁGENES DIGITALES Y ANÁLISIS MULTIVARIANTE.....	100
4.1. Objetivos.....	100
4.2. Materiales y métodos.....	100
4.2.1. Muestras y reactivos.....	100
4.2.2. Instrumentación y software.....	101
4.2.3. Análisis multivariante.....	103
4.2.3.1. Modelos de clasificación.....	103
4.2.3.2. Procesamiento de los datos	103
4.3. Resultados y discusión	105
4.3.1. Adquisición de imágenes.....	105
4.3.2. Análisis multivariante.....	106
4.3.2.1. Análisis exploratorio	106
4.3.2.2. Resultados de modelos de clasificación	107
4.3.2.3. Ventajas de la técnica propuesta.....	109
4.5. Conclusiones	110
4.5. Referencias.....	111
CAPÍTULO V: CUANTIFICACIÓN DE ADULTERACIONES Y AUTENTICACIÓN DE QUESOS RALLADOS DE PASTA DURA MEDIANTE ESPECTROSCOPIA NIR Y ANÁLISIS MULTIVARIANTE	113

5.1. Objetivos.....	113
5.2. Antecedentes.....	113
5.3. Materiales y métodos.....	114
5.3.1. Muestras y reactivos.....	114
5.3.2. Instrumentación.....	115
5.3.3. Análisis multivariante.....	115
5.3.3.1. Regresión multivariante.....	115
5.3.3.2. Clasificación multivariante.....	116
5.4. Resultados y discusión.....	117
5.4.1. Espectros NIR.....	117
5.4.2. Análisis multivariante.....	118
5.4.3. Análisis exploratorio.....	118
5.4.4. Cuantificación de adulterantes.....	120
5.4.5. Modelos de autenticación.....	124
5.5. Conclusiones.....	128
5.6. Referencias.....	129
CAPÍTULO VI: CONCLUSIONES GENERALES.....	131
APENDICE.....	135
A.1. Artículos científicos presentados en revistas internacionales.....	135
A.2. Artículos científicos actualmente en redacción.....	136
A.3. Contribuciones en reuniones científico-tecnológicas.....	136

ESTRUCTURA

La presente tesis está estructurada en 6 capítulos.

Capítulo I: Alcances y objetivos. En este capítulo se presenta el marco que se utilizará para estudiar y aplicar métodos de *screening* junto con análisis multivariante de datos destinados a determinar alimentos adulterados. Seguidamente, el capítulo define los objetivos de la tesis.

Capítulo II: Aspectos teóricos. Este capítulo está dividido en tres partes. La primera describe las diferentes muestras, la problemática vigente y los analitos en estudio. La segunda parte describe los aspectos teóricos haciendo énfasis en las técnicas analíticas utilizadas (espectroscópicas e imágenes digitales). Finalmente, la tercera parte contiene una descripción de las herramientas quimiométricas utilizadas: pre-procesamiento de espectros, modelos multivariados de clasificación (supervisada y no supervisada), modelos de regresión y parámetros de evaluación de calidad de los modelos multivariantes.

Capítulo III: Determinación de la adulteración de quesos rallados mediante espectroscopía NIR y análisis multivariante.

Capítulo IV: Determinación de la adulteración de quesos rallados mediante imágenes digitales y análisis multivariante.

Capítulo V: Cuantificación de adulteraciones y autenticación de quesos rallados de pasta dura mediante espectroscopía NIR y análisis multivariante.

Capítulo VI: Conclusiones generales. Este capítulo contiene todas las conclusiones finales de la tesis.

Apéndice: Esta sección contiene la lista de publicaciones científico-tecnológicas y las presentaciones en congresos nacionales e internacionales durante el período de la tesis.



C

APÍTULO 1

Alcances y objetivos

CAPÍTULO I: ALCANCES Y OBJETIVOS

1.1. Alcances

La producción de alimentos altamente industrializada y la globalización del mercado de alimentos a los que nos enfrentamos en la actualidad, hacen que sea una tarea desafiante rastrear y monitorear los contaminantes de los alimentos desde su origen hasta la mesa del consumidor. Las diferentes partes interesadas, como son los productores, la industria, los organismos reguladores y los consumidores, están cada vez más preocupadas en cuanto a la seguridad alimentaria. Existe una gran variedad de peligros asociados a los alimentos, y tanto los ingredientes como los aditivos alimentarios presentan un dilema muy complejo porque, en general, carecen de características que permitan distinguirlos de otros ingredientes similares o de posibles adulterantes. En este sentido, la industria alimentaria es la responsable del producto final, que debe responder a las características nutricionales e higiénicas adecuadas (Sørensen, Khakimov, y Engelsen, 2018). El desafío analítico incluye la detección de fraudes relacionados con la contaminación, el agregado de sustancias no descritas en el producto, o adulteraciones de ingredientes y materias primas, con el fin de asegurar alimentos inocuos.

La seguridad alimentaria es uno de los puntos que la Organización de Naciones Unidas (ONU) incorporó en su agenda 2030 para el Desarrollo Sostenible. En el objetivo 2 se expresa textualmente *“Poner fin al hambre, lograr la seguridad alimentaria y la mejora de la nutrición y promover la agricultura sostenible”* (ONU, 2020). En este sentido, es de interés evaluar la adulteración de alimentos dado que esta práctica influye en la inocuidad y calidad nutricional de los mismos. Según el Código Alimentario Argentino (CAA) *“un alimento adulterado es aquel que ha sido privado, en forma parcial o total, de sus elementos útiles o característicos, reemplazándolos o no por inertes o extraños, que ha sido adicionado de aditivos no autorizados o sometido a tratamientos de cualquier naturaleza para disimular u ocultar alteraciones, deficiente calidad de materias primas o defectos de elaboración”* (CAA, Cap. I, 2023). Debido a que la adulteración de alimentos puede tener un efecto adverso en la salud de la población, es de vital importancia prevenirla y detectarla adecuadamente. Esto obliga a los diferentes países a sancionar leyes y crear instituciones y comisiones reguladoras, como por ejemplo el Código Alimentario Argentino (CAA, Argentina), la Administración de Drogas y Alimentos (FDA, Estados Unidos), la Autoridad Europea de Seguridad Alimentaria (EFSA, Unión Europea) y el Codex Alimentarius, entre otros. Estos

organismos regulan, entre otros aspectos, el agregado o sustitución de sustancias prohibidas en alimentos con fines netamente económicos.

En la actualidad, es difícil estimar la frecuencia con la que ocurre un fraude alimentario o su impacto económico exacto, puesto que el mismo está diseñado para evitar su detección. Estimaciones externas realizadas por expertos han encontrado que los fraudes afectan a la industria alimentaria mundial, a un costo de alrededor de \$10 a \$15 mil millones de dólares al año, aunque algunas estimaciones más recientes sitúan el costo en \$40 mil millones al año (FDA, 2023). Un caso de trascendencia global fue la adulteración de quesos de pasta dura con agregados de aditivos alimentarios, como celulosa microcristalina como agente antiaglutinante, en cantidades superiores a las permitidas por la regulación de Estados Unidos, siendo los responsables del fraude condenados a tres años de prisión en libertad condicional y a abonar una abultada suma económica (Cheese reporter, 2018). Por otro lado, existe la posibilidad de incorporar agregados no permitidos en estos alimentos como son la harina y sémola de trigo y el polvo de madera. La razón por la cual podrían ocurrir estas prácticas fraudulentas es que representa, no sólo, una ganancia económica, sino también que podrían reemplazar el uso de aditivos antiaglutinantes. Los métodos oficiales de análisis para la determinación de los aditivos permitidos en quesos de pasta dura (celulosa microcristalina y dióxido de silicio), como así también de agregados no permitidos, incluyen metodologías de laboratorio tediosas con la generación de residuos (AOAC, 1995). Por consiguiente, la implementación de métodos de screening constituye una buena estrategia alternativa para poder incorporarlos al análisis de rutina, con el fin de controlar la calidad de alimentos de consumo masivo como son los quesos.

En química analítica, los métodos de screening han constituido una herramienta efectiva y de gran importancia a lo largo de los años en muchas áreas y laboratorios de rutina. Se han utilizado ampliamente en ciencias medioambientales, farmacéuticas, médicas, clínicas y en análisis de alimentos, pudiendo ser adaptados a problemáticas específicas (Muñoz-Olivas, 2004). Desde un punto de vista práctico, es bastante habitual realizar un análisis inicial para determinar si los analitos de interés están presentes o ausentes en una muestra a un nivel determinado y, en algunos casos, sí están presentes poder cuantificarlos. Los métodos de screening tienen la ventaja de reducir el tiempo y el costo del análisis, además de involucrar instrumentos analíticos más sencillos y menos costosos que los empleados en los métodos oficiales de análisis en general. Estos métodos ofrecen una solución en situaciones que requieren una respuesta rápida por parte del productor, consumidor o

analista de alimentos. Concretamente, existe una necesidad urgente de métodos analíticos fiables, exactos, precisos, económicos, simples, rápidos y, en la medida de lo posible, portátiles, que permitan un screening previo al análisis de confirmación de adulterantes en alimentos. De este modo, se podrá hacer frente más fácilmente a los efectos sociales y económicos derivados del consumo de alimentos contaminados o adulterados. Tsagkaris y colaboradores exponen, al menos, 270 artículos científicos donde se utilizaron métodos de detección y confirmación de contaminantes y adulterantes en matrices alimentarias bajo la utilización de métodos que incluían el uso de teléfonos móviles, fluorescencia y quimioluminiscencia, entre otros (Tsagkaris, et. al, 2019).

Los métodos de screening pueden estar asociados a distintas técnicas espectroscópicas que ofrecen la posibilidad de adquirir una gran cantidad de datos de una muestra en pocos segundos. En consecuencia, en química analítica juega un rol importante la utilización del análisis multivariante para poder interpretar estos datos. Dentro del análisis multivariante, las técnicas de clasificación no supervisadas y supervisadas son utilizadas para agrupar y clasificar objetos. Pueden ser empleadas, por ejemplo, para identificar las variaciones en la composición de un alimento adulterado o para identificar cambios en la cadena de producción de la industria. En términos generales, se basan en la construcción de una regla de clasificación que permita clasificar futuras muestras a correspondientes clases predefinidas en base a los atributos de las muestras bajo estudio (Oliveri, Malegori, y Casale, 2018).

Los métodos de screening junto con el análisis multivariante de datos representan una muy útil herramienta para el aseguramiento de la inocuidad de alimentos, y por este motivo han cobrado un gran interés en todo el campo científico.

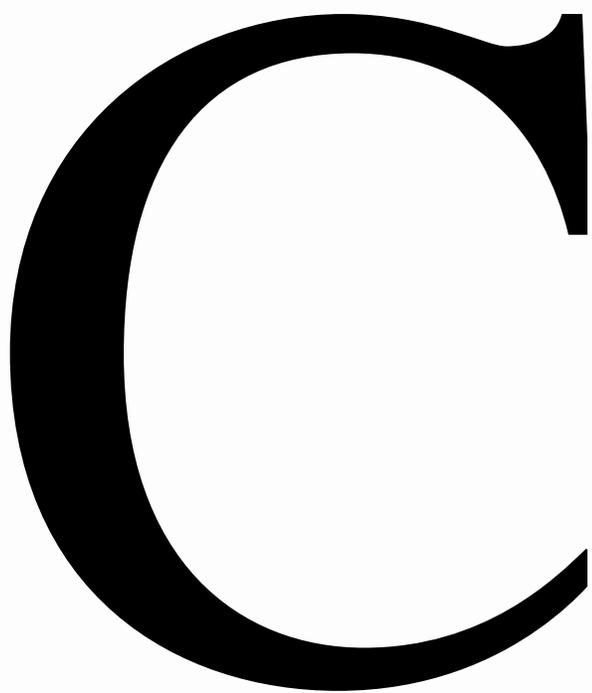
1.2. Objetivos de esta tesis

El objetivo principal de esta tesis es proponer herramientas analíticas que sean fiables, rápidas, simples y económicas para determinar la adulteración de quesos rallados de pasta dura por el uso de aditivos en cantidades superiores a las reguladas o por la adición de agregados no permitidos con el fin de obtener beneficios comerciales. Para ello se desarrollan metodologías analíticas de screening, basadas en técnicas espectroscópicas e imágenes digitales, junto con análisis multivariante de datos con el fin de proponer alternativas que puedan garantizar la correcta identidad de los quesos y proteger al consumidor de engaños por fraudes alimentarios. Para el desarrollo del mencionado objetivo se llevan a cabo diferentes estrategias:

1. Aplicación de técnicas espectroscópicas e imágenes digitales que abarcan el estudio de la zona del infrarrojo cercano y el visible del espectro electromagnético (espectros NIR e histogramas de color).
2. Utilización de herramientas quimiométricas como pre-procesamiento de datos, análisis exploratorio de datos, técnicas de reconocimiento de patrones (supervisadas y no supervisadas) y modelos de regresión multivariantes. Dentro de las técnicas de clasificación supervisadas, se evalúan diferentes enfoques: discriminación que incluye técnicas clásicas y aprendizaje supervisado (*machine learning*) o autenticación que incluye modelado de una clase (*one-class*).
3. Desarrollo de un novedoso dispositivo construido en el laboratorio para la captura de imágenes digitales.
4. Validación de las metodologías analíticas desarrolladas.
5. Postulación de las metodologías desarrolladas, tanto como métodos de screening como cuantitativos para determinar la adulteración de quesos rallados de pasta dura.

1.3. Referencias

- AOAC, Association of Official Analytical Chemists. (1995). Fibra dietética total, soluble e insoluble en alimentos, método n°. 991.43. *Métodos oficiales de análisis de AOAC Internacional (16ª ed.)*. Gaithersburg, MD. Estados Unidos: AOAC Internacional.
- CAA, Código Alimentario Argentino. (2010). Capítulo I: Disposiciones generales, artículo 6, disponible en http://www.anmat.gov.ar/alimentos/codigoa/CAPITULO_I.pdf (Último acceso julio 2023).
- Cheese reporter. (2018). Anti-Caking Suppliers Clear Up Inaccuracies Regarding Appropriate Levels Of Cellulose. <http://www.cheesereporter.com/cellulose.htm> (Último acceso julio 2023).
- FDA, Food and Drug administration. (2023). <https://www.fda.gov/food/compliance-enforcement-food/economically-motivated-adulteration-food-fraud> (Último acceso julio 2023).
- Muñoz-Olivas R. (2004). Screening analysis: an overview of methods applied to environmental, clinical and food analyses. *Trends in Analytical Chemistry*, 23, 203-216.
- Oliveri P., Malegori C., y Casale M. (2018). Multivariate classification techniques. *In the Reference module in chemistry, molecular Sciences and chemical engineering*, 481-486. Amsterdam, Netherlands: Elsevier.
- ONU, Organización de las Naciones Unidas. (2020). <https://www.un.org/sustainabledevelopment/es/sustainable-development-goals/> (Último acceso julio 2023).
- Sørensen K. M., Khakimov B., y Engelsen S. B. (2016). The use of rapid spectroscopic screening methods to detect adulteration of food raw materials and ingredients. *Current Opinion in Food Science*, 10, 45–51.
- Tsagkaris A.S., Nelis J.L.D., Ross G.M.S., Jafari S., Guercetti J., Kopper K., Zhao Y., Rafferty K., Salvador J.P., Migliorelli D., Salentijn G.IJ., Campbell K., Marco M.P., Elliot C.T., Nielen M.W.F., Pulkrabova J., y Hajslova J. (2019). Critical assessment of recent trends related to screening and confirmatory analytical methods for selected food contaminants and allergens, *TrAC Trends in Analytical Chemistry*, 121, 115688.

A large, bold, black letter 'C' is positioned on the left side of the page. It is partially enclosed by a thick, blue, curved shape that resembles a stylized 'C' or a decorative element. The background is white with a light gray shadow effect behind the blue shape.

APÍTULO 2

Aspectos teóricos

CAPÍTULO II: ASPECTOS TEÓRICOS

2.1. Alimentos

Los alimentos son productos naturales o elaborados que, debido a la calidad tanto nutricional como de inocuidad y de aceptación sensorial, contribuyen al mantenimiento de las funciones normales del organismo, es decir al mantenimiento de la salud. Desde el punto de vista químico, los alimentos en general son una matriz compleja y heterogénea formada por diversos componentes. La evaluación de la calidad, características sensoriales, valor nutricional y estabilidad de productos alimentarios es el principal interés de los químicos analíticos que trabajan tanto en industrias de procesamiento de alimentos como en laboratorios académicos y gubernamentales de control de calidad. El rol principal del análisis de alimentos está enfocado en asegurar una correcta calidad, tanto nutricional como higiénica de los mismos. El desafío constante comprende la implementación de normas reguladoras y legislaciones actualizadas, asociadas a la utilización de metodologías analíticas para determinar adulteraciones de forma rápida y que sean adaptables a laboratorios de rutina.

A medida que la cadena de suministros de alimentos se ha ido transformando progresivamente, siendo cada vez más compleja y global, nuevos y desafiantes riesgos han surgido en materia de seguridad alimentaria, sobre todo en relación a la incorporación de aditivos y de ingredientes que no corresponden a la formulación de un alimento genuino (Cozzolino, 2015). El uso de aditivos alimentarios exige el cumplimiento, de acuerdo con la norma general del Codex (Codex Alimentarius, Codex Stan 192-1995, 2021), de los siguientes principios: inocuidad, justificación de uso, Buenas Prácticas de Manufactura (BPM) y especificaciones de identidad y pureza. La utilización de sustancias no reguladas, capaces de cumplir la función de un aditivo alimentario, conlleva a la contaminación y adulteración de los alimentos.

2.2. Quesos

2.2.1. Definición y proceso de elaboración

El queso es uno de los productos lácteos más consumidos a escala mundial y es un alimento muy apreciado por su sabor. Se define, según el C.A.A en su art. 605, como *“el producto fresco o maduro que se obtiene por separación del suero de la leche o de la leche reconstituida (entera, parcial o totalmente descremada) coagulada por acción del cuajo y/o enzimas específicas, completadas o no por bacterias específicas o por ácidos orgánicos*

permitidos a este fin, con o sin el agregado de sustancias colorantes permitidas, especias o condimentos u otros productos alimenticios” (CAA, Cap. VIII, 2023). Este alimento está formado por caseína coagulada que retiene grasas, otras proteínas, sales y componentes del suero. Además, contiene compuestos aromáticos del suero producidos en procesos fermentativos microbianos, que son específicos de cada una de las numerosas clases de quesos que existen para consumo.

Las etapas de elaboración del queso, a partir de la leche, se encuentran resumidas en la Figura 2.1:



Figura 2.1. Esquema del proceso de elaboración del queso.

A► La **coagulación** es el proceso en el cual tienen lugar las modificaciones físico-químicas de las proteínas (caseína), que bajo la acción de enzimas proteolíticas llevan a la formación de un entramado o “esqueleto” proteico que hace que la leche pase de un estado líquido a un estado sólido, formándose un coágulo o gel. La coagulación enzimática se produce al añadir a la leche, a una temperatura entre 30-32°C, una sustancia o enzima coagulante denominada cuajo o renina, capaz de modificar el extremo de la caseína kappa (κ), que es la que tiene la mayor carga negativa y está en la parte exterior de la micela, posibilitando que se unan las proteínas y coagulen de la solución micelar. Esta coagulación puede ser complementada por la presencia de ácido láctico (coagulación ácida) producido por los microorganismos de la leche cruda o los del cultivo iniciador que se añaden a la leche pasteurizada al fermentar la lactosa. La disminución del pH contribuye a la formación de coágulo.

B► Una vez obtenida la cuajada se procede al **desuerado** que consiste en la separación del lactosuero que es la fase acuosa del coágulo. Tras el corte de éste se obtiene el queso sin madurar o queso fresco. Las principales etapas de desuerado incluyen: a) corte de la cuajada: se corta el coágulo o cuajada para dividirla, en diferentes tamaños según el tipo de queso, que se llaman “granos de cuajada”; b) agitación: se agita, de forma manual o mecánica, para que los granos de cuajada se desueren fácilmente por todas sus caras y no se unan entre sí. También la agitación contribuye para un buen reparto del calor; c) calentamiento: proceso opcional en el cual la elevación de la temperatura de cuajada provoca que el lactosuero salga de los granos de la cuajada; d) lavado de los granos de cuajada: se realiza un lavado para reducir la lactosa y así controlar el crecimiento de los microorganismos y e) separación del lactosuero de los granos de cuajada: en algunos quesos, antes de la separación del lactosuero, se aplica un pre-prensado en la cuba ejerciendo una ligera presión.

Una vez eliminado el suero se procede al moldeado y volteos de los quesos que consisten en la colocación de los granos de cuajada en moldes especiales para cada tipo de queso. Esta operación asegura la forma del queso y que continúe el desuere. Las perforaciones de los moldes permiten una salida más rápida del suero, y los volteos que se realizan, mientras el queso está en el molde, permiten destapar los orificios de salida que se hallan obstruidos por la cuajada y aseguran la homogeneidad del producto (Figura 2.2).



Figura 2.2. Proceso de moldeado y volteo de quesos.

La etapa del salado se realiza para: 1) favorecer la eliminación del lactosuero del interior del queso, 2) dar sabor, 3) formar corteza, y 4) dar protección contra ciertos microorganismos. Normalmente se realiza por inmersión en salmueras, aunque hay tecnologías en las que los granos de cuajada se salan en la cuba.

Por último, en el oreado se trata de secar la superficie del queso para que se forme una corteza firme y seca. El oreado se lleva a cabo en cámaras especiales con una temperatura y humedad ambiental determinadas (13-15 °C y 75 % de humedad relativa), dependiendo del tipo de queso. Normalmente la humedad de esta cámara es menor que la que luego tendrá en la maduración, pues se trata de secar la superficie del queso para que se forme la corteza, en los quesos en los que ésta es firme y seca.

C► La **maduración** corresponde a la etapa en la que la masa del queso cambiará sus características, transformándose de una masa blanquecina, algo insípida, más o menos consistente, a una masa con nuevas características en su estructura, aspecto, composición, consistencia y color, al mismo tiempo que el olor y el gusto adquieren nuevos matices y se

acentúan (Kongo y Malcata, 2016; Kapoor y Metzger, 2008; Upadhyay y McSweeney, 2007). Todo esto tiene lugar porque la masa del queso está poblada de microorganismos propios de la leche y los añadidos para su coagulación que, con sus enzimas, llevarán a cabo transformaciones físico-químicas y bioquímicas (Figura 2.3). Estos cambios serán diferentes para cada tipo de queso, dependiendo del tipo de leche, del tipo de coagulación, de los trabajos realizados en la fase de desuerado y de las condiciones de la maduración.

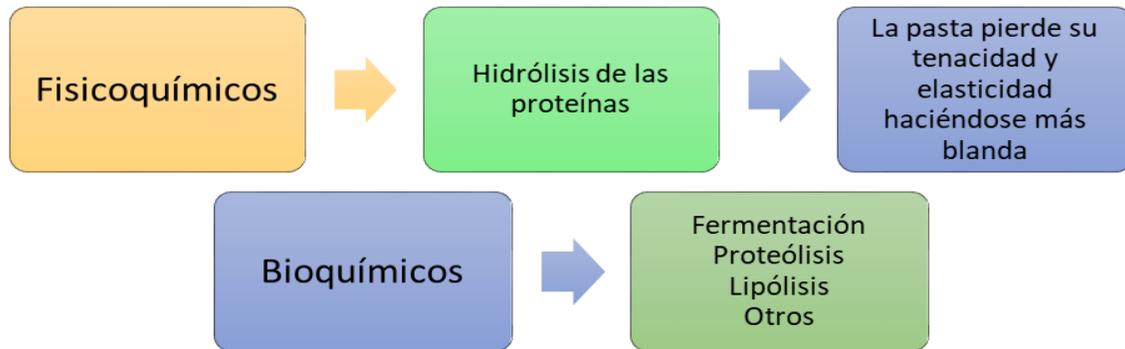


Figura 2.3. Procesos asociados a la fabricación de quesos.

La maduración se puede dar en ambientes naturales o en cámaras de maduración. En ambientes naturales se maduran quesos de pasta compacta y quesos donde se necesita el desarrollo de mohos. En las cámaras de maduración existe una temperatura determinada (9-15 °C) y una humedad relativa dada (85-99%) dependiendo de que sean quesos de pasta prensada y corteza seca o de pasta blanda y corteza húmeda o con mohos. La duración de esta fase también dependerá del tipo de queso, desde unos días hasta varios meses y definirá la edad del queso (Gastalver Robles, 2015) (Figura 2.4).



Figura 2.4. Cámaras de maduración de quesos (9-11°C y 85% humedad).

2.2.2. Clasificación de quesos

El Código Alimentario Argentino establece, en el artículo 605, la siguiente clasificación para quesos (CAA, Capítulo VIII, 2023):

A► De acuerdo con el contenido de materia grasa del extracto seco, en porcentaje, los quesos se clasifican en:

- Extra graso o Doble crema: cuando contengan no menos del 60%.
- Grasos: cuando contengan entre 45,0 y 59,9%.
- Semigrasos: cuando contengan entre 25,0 y 44,9%.
- Magros: cuando contengan entre 10,0 y 24,9 %.
- Descremados: cuando contengan menos de 10,0%.

B► De acuerdo con el contenido de humedad, en porcentaje, los quesos se clasifican en:

- Quesos de baja humedad: (conocidos como de pasta dura): humedad hasta 35,9%.
- Quesos de mediana humedad: (conocidos como de pasta semidura): humedad entre 36,0 y 45,9%.
- Quesos de alta humedad (conocidos como de pasta blanda o macíos): humedad entre 46,0 y 54,9%.
- Quesos de muy alta humedad (conocidos como de pasta muy blanda o mole): humedad no menor a 55,0%.

Además de la clasificación del CAA también se pueden ordenar según su forma de preparación, región de origen y tipo de leche empleado, entre otras cosas. Teniendo en cuenta las diversas clasificaciones se puede afirmar que existe en el mercado una amplia variedad de quesos, destacándose entre ellos los de pasta dura que se caracterizan por tener muy baja humedad (entre 49 y 56%), alta consistencia de su pasta, color amarillento, sabor fuerte y aromático que se intensifica con su maduración a temperatura y humedad controladas.

2.2.3. Quesos de pasta dura

El CAA, en el capítulo VIII (artículo 635) Alimentos Lácteos, define al **queso de pasta dura** como “*con el nombre de Queso Parmesano, Queso Parmesão, Queso Reggiano, Queso Reggianito y Queso Sbrinz se entienden los quesos madurados que se obtienen por coagulación de la leche por medio del cuajo y/u otras enzimas coagulantes apropiadas, complementada por la acción de bacterias lácticas específicas*”. Los Quesos Parmesano, Parmesão, Reggiano, Reggianito y Sbrinz son quesos de baja humedad y semigrasos o

grasos, de acuerdo con lo establecido en el artículo 605 inciso 2) del presente Código. Deberán tener un contenido mínimo de 32 g. de materia grasa /100 g. del extracto seco. En la elaboración de estos quesos se utilizarán los denominados Ingredientes obligatorios que incluye leche entera o estandarizada en su contenido de materia grasa, cultivos de bacterias lácticas específicas, cuajo y/u otras enzimas coagulantes apropiadas y cloruro de sodio, y los Ingredientes opcionales como crema, concentrado de proteínas lácteas y cloruro de calcio. Además, podrán ser utilizados los aditivos previstos en el artículo 605 inciso 3.c) del presente Código, para los quesos de baja humedad (CAA, Cap. VIII, 2023).

Los quesos de pasta dura maduran en cámaras cuya temperatura y humedad son debidamente controladas. Antes de su comercialización son pintados y rotulados adecuadamente. Su forma de consumo más generalizada es como queso rallado, que industrialmente provienen de la disgregación mecánica de los quesos de pasta dura.

En el **Artículo 640** del CAA se define a los **quesos rallados** como “*con el nombre de Queso Rallado o Quesos Rallados, según corresponda, se entiende el producto obtenido por desmenuzado o rallado de la masa de una o hasta cuatro variedades de quesos de baja y/o mediana humedad aptos para el consumo humano. El producto podrá ser parcialmente deshidratado o no*” (CAA, Cap. VIII, 2023).

2.2.3.1. Quesos rallados: aditivos antiaglutinantes permitidos y agregados no permitidos

El CAA establece que, en la elaboración de quesos rallados, se admitirá el uso de los aditivos autorizados para los quesos utilizados como materia prima, en una concentración en el producto final que deberá responder a los máximos autorizados en el artículo 640 (CAA, Cap. VIII, 2023). Se admite, además, el uso de otros aditivos: antiaglutinantes y conservadores. Entre los primeros se incluyen a la celulosa microcristalina en cantidades de acuerdo a las Buenas Prácticas de Manufactura (BPM) y al dióxido de silicio a una concentración máxima 5 g/kg. Entre los segundos se encuentran ácido sórbico, sólo o en combinación con sorbato de sodio, potasio o calcio y natamicina (CAA, Cap. VIII, 2023). Por otro lado, la norma europea establece la utilización de celulosa microcristalina en cantidades *quantum satis* en quesos de pasta dura (EFSA E.U., 2011) y en Estados Unidos, la FDA también permite su agregado en quesos rallados, como antiaglutinante, en cantidades adecuadas y seguras (FDA, 2018). Expertos en materia de quesos aseguran

que valores entre un 1-4% de celulosa microcristalina son aceptables para lograr el efecto funcional como aditivo.

La siguiente tabla muestra un resumen de los aditivos antiaglutinantes permitidos en quesos rallados según diferentes organismos de regulación.

Tabla 2.1. Límites permitidos para aditivos antiaglutinantes en quesos rallados.

Aditivo	CAA	EFSA	FDA
Celulosa microcristalina	Quantum satis*	Quantum satis*	Quantum satis*
Dióxido de Silicio	0,5 g/Kg	1 g/Kg	2 g/Kg

*cantidad necesaria para que sea alcanzado el efecto deseado.

El CAA, en su capítulo XVIII (aditivos alimentarios) establece los parámetros de calidad que deberán tener los aditivos antiaglutinantes permitidos en quesos rallados como son la celulosa microcristalina y el dióxido de silicio (CAA, Cap. VIII, 2023). La celulosa microcristalina es un polvo fino, blanco, cristalino, insoluble en: agua, alcohol, éter, ácidos diluidos; parcialmente solubles en soluciones alcalinas diluidas. Para el caso del dióxido de silicio se presenta como un producto en polvo blanco espumoso o granulado; insoluble en agua y en alcohol y en otros solventes orgánicos; soluble en solución de álcalis a 80-100°C, en ácido fosfórico caliente y en ácido fluorhídrico.

La celulosa es el biopolímero más abundante en la superficie terrestre. Se forma por la unión de moléculas de β -D-glucosa mediante enlaces β -1,4-O-glucosídico (Figura 2.5). Contiene zonas amorfas y cristalinas. Las cadenas en la zona amorfa son susceptibles a bioconversión (celulasas).

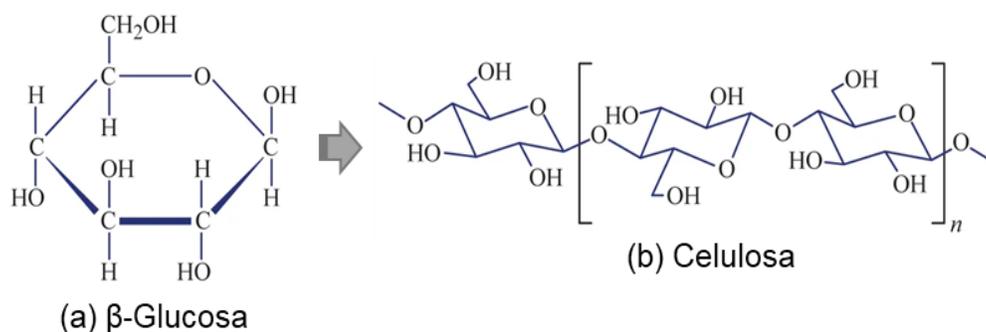


Figura 2.5. (a) Estructura de la β -glucosa y (b) celulosa.

Las principales fuentes de celulosa son la madera y, en menor medida, el algodón. Para aislar la celulosa, la pulpa de madera se hierva bajo presión con soda cáustica, una mezcla de hidróxido sódico y sulfito de sodio (proceso Kraft) o con ácido sulfúrico y sales (proceso sulfúrico). También se utilizan, aunque en menor medida, procesos de obtención que incluyen solventes orgánicos y procesos oxidativos (Céliz, Forneris, Severini y Filippin, 2018). La celulosa en los alimentos se puede determinar como parte de la fibra insoluble, pero el análisis no es preciso y requiere métodos largos y laboriosos como la cromatografía líquida (HPLC) con hidrólisis enzimática (AOAC, 1995; McCleary et al., 2012). Además, se pueden utilizar métodos colorimétricos para identificar la celulosa en alimentos (Jablonský et al., 2015).

La celulosa microcristalina (CMC) se obtiene por hidrólisis ácida de la celulosa inicial hasta llegar al mínimo grado de polimerización correspondiente a la despolimerización de cristales individuales, llegando a una celulosa de alta cristalinidad que se caracteriza por ser un polvo blanco y fino, inodoro, inerte, insoluble en agua, en solventes orgánicos y ácidos diluidos, y parcialmente soluble en álcalis diluidos. La CMC posee propiedades únicas, se utiliza como excipiente en medicamentos, suplementos dietarios y como agente estabilizante, espesante y antiaglutinante en alimentos (Loelovich, 2022). La CMC debe cumplir con los requisitos de calidad para ser utilizada como aditivo alimentario y debe obtenerse de proveedores autorizados a tal fin. No existen métodos oficiales para determinar la CMC. Los análisis de composición de alimentos se centran en determinar los macronutrientes (como carbohidratos, proteínas y grasas) y los micronutrientes (como vitaminas y minerales). La CMC, al ser un polímero de glucosa, generalmente se incluye dentro del análisis de carbohidratos totales o fibra dietética.

En términos generales, los métodos utilizados para analizar la celulosa y la fibra dietética en los alimentos suelen basarse en el uso de enzimas específicas, como las enzimas de la familia de las celulasas, que ayudan a descomponer los componentes fibrosos de los alimentos en sus constituyentes básicos. Estos métodos pueden incluir técnicas como la hidrólisis ácida, la hidrólisis enzimática y la cromatografía de exclusión molecular.

Por otro lado, el dióxido de silicio o sílice es un aditivo muy utilizado en la industria alimentaria por sus propiedades antiaglomerantes, ya que evita la formación de grumos y que los ingredientes de los alimentos formen aglomerados, precipitados, etc. El dióxido de silicio tiene un alto poder absorbente, por lo que absorbe el exceso de humedad y evita que

los componentes de los alimentos se adhieran entre sí cuando están expuestos a condiciones de humedad elevada.

El dióxido de silicio generalmente se determina mediante espectrometría de emisión óptica de plasma acoplado inductivamente (ICP-OES) (Ka et al., 2017). Además del costo del análisis, estas técnicas no son adecuadas para realizar controles de rutina.

Los carbohidratos constituyen uno de los macronutrientes orgánicos presentes en los alimentos, junto a las proteínas y a las grasas. Se los puede clasificar como carbohidratos digeribles y no digeribles. Entre los primeros, se encuentran el almidón, almidón despolimerizado, lactosa, glucosa, fructosa y sorbitol. En el grupo de carbohidratos no digeribles se ubican la rafinosa, estaquiosa, almidón resistente, pectina, celulosa, hemicelulosa, entre otros (Voragen, 1998). Por otro lado, los carbohidratos se pueden añadir a los alimentos debido a sus múltiples funciones tecnológicas: antiaglomerantes, emulsificantes, humectantes, estabilizantes y espesantes, entre otras.

La norma general del Codex Alimentarius, norma de grupo para el queso no madurado, incluido el queso fresco, establece como ingredientes autorizados solo en quesos de alta humedad o de pasta blanda a las *“harinas y almidones de arroz, maíz y papa. Estas sustancias pueden utilizarse con los mismos fines que los antiaglutinantes para el tratamiento de la superficie de productos cortados, rebanados y desmenuzados únicamente”*. Por consiguiente, el almidón no puede ser utilizado como antiaglutinante en quesos de pasta dura o madurados (Codex Alimentarius, Codex Stan 221-2001, 2021).

Los agregados no permitidos estudiados en esta tesis incluyen: harina y sémola de trigo y polvo de madera.

En términos generales, la harina es el producto obtenido de la molienda del *Triticum vulgare* conocido como trigo harinero o trigo pan; mientras que la sémola se obtiene de la molienda del *Triticum durum* comúnmente llamado trigo duro. Ambos productos tienen al almidón como componente principal, que en los quesos rallados puede cumplir la función antiaglutinante.

Por otro lado, el polvo de madera tiene como componente principal la celulosa, que lo convierte en un potencial adulterante debido a sus propiedades antiaglomerantes y su gran disponibilidad a muy bajo precio.

2.3. Adulteraciones alimentarias

Los alimentos aptos para el consumo tienen tres dimensiones: inocuidad, calidad y autenticidad, lo que es similar al concepto de “alimentos con integridad” (Comisión del Codex Alimentarius, 2018). El fraude y la adulteración de alimentos son prácticas que contradicen los principios de integridad de los alimentos. La siguiente Figura 2.6 representa una síntesis de las definiciones más importantes.

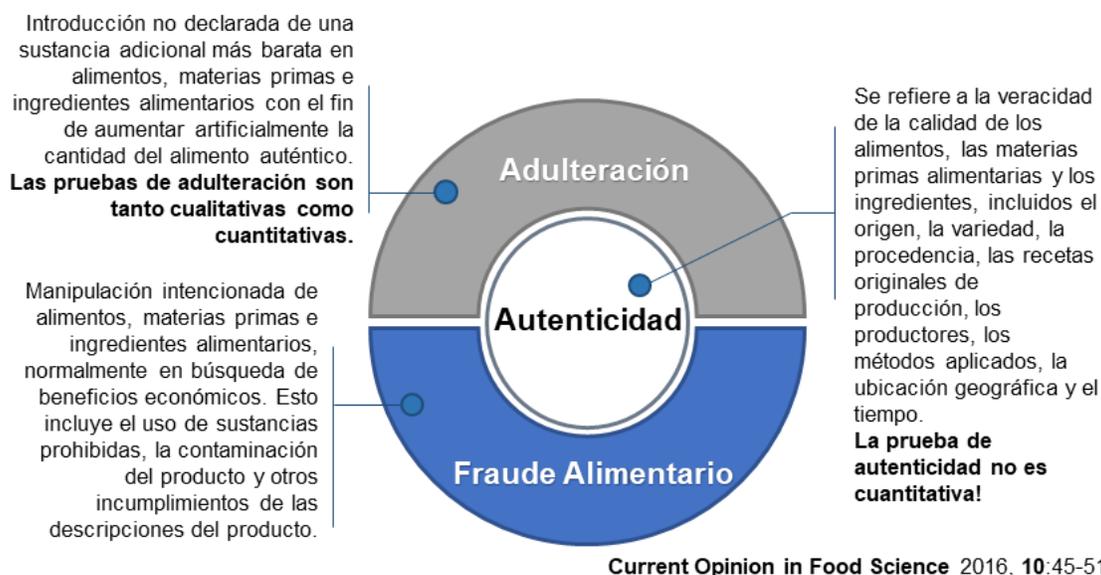


Figura 2.6. Definiciones de adulteración, autenticidad y fraude alimentario. (Sorensen, Khakimov y Engelsen, 2016).

Las adulteraciones alimentarias no constituyen un problema contemporáneo, son tan antiguas como los sistemas de producción y procesamiento de alimentos. Sin embargo, en la actualidad, se han transformado en prácticas más sofisticadas, siguiendo o anticipándose a los avances en el campo de las ciencias analíticas. La adulteración económicamente motivada (de sus siglas en inglés, EMA) se define como la modificación fraudulenta intencionada de un producto acabado o ingrediente para obtener un beneficio comercial a través de las siguientes prácticas: mejoras no aprobadas, dilución con un ingrediente de menor valor, ocultamiento de daños o contaminación, etiquetado incorrecto de un producto o ingrediente, sustitución de un ingrediente por uno de menor valor o la no divulgación de la información requerida sobre el producto (Everstine, Splink y Kennedy, 2013; Stadler, 2019). Los principales motivadores de estas adulteraciones son el bajo riesgo de detección

y la poca rigurosidad de las sanciones por parte de los organismos de control. La Figura 2.7 ilustra los motivadores del fraude bajo EMA en alimentos.



Figura 2.7. Motivadores de adulteraciones en alimentos.

Por consiguiente, en materia de seguridad alimentaria se requiere de una vigilancia continua para asegurar la calidad de los alimentos que llegan al consumidor, siendo de primordial importancia la investigación y desarrollo de técnicas analíticas rápidas y simples para detectar adulteraciones en alimentos y en materias primas utilizadas en su elaboración. En este contexto, los métodos de screening rápidos, eficientes y que puedan ser adaptados a dispositivos manuales, representan una herramienta muy útil para ser utilizada en puntos de vulnerabilidad a lo largo de la compleja red de suministros de materias primas y alimentos (Ellis et al., 2012).

Un ejemplo de adulteración alimentaria en quesos rallados es el agregado de aditivos permitidos, pero en cantidades superiores a las reglamentarias según Buenas Prácticas de Manufactura (BPM), como así también el empleo de sustancias que, aún sin ser peligrosas para la salud, se agregan con el fin de abaratar costos. En los últimos años, esta adulteración ha cobrado gran importancia desde que se detectó en Estados Unidos la adulteración de queso Parmesano rallado, cuyos rótulos denominaban 100% queso Parmesano, con cantidades de celulosa en polvo superiores a las consideradas por las BPM (Washingtonpost, 2016; Cheese reporter, 2016). Por otro lado, en nuestro país, específicamente en la ciudad de Bahía Blanca (Laboratorio de Bromatología de la Municipalidad de Bahía Blanca), se detectaron en el año 2012 quesos rallados adulterados con harina de trigo debido a la denuncia de un consumidor que encontró las características organolépticas modificadas. Si bien el agregado de celulosa en cantidades superiores a las reglamentadas y el uso de harinas no son prácticas que ponen en riesgo la salud de la

mayoría de los consumidores, pueden ser potencialmente peligrosos en personas con alergias alimentarias. Por consiguiente, el uso de estas sustancias en quesos rallados constituye una adulteración alimentaria y un claro engaño al consumidor.

En bibliografía se pueden encontrar estudios sobre la adulteración del queso con almidones (Barreto, Cruz-Tirado, Siche y Quevedo, 2018), aceites vegetales (Juarez-Barrientos et al., 2018) o bien la adulteración de la leche utilizada en la fabricación de queso (Cuibus et al., 2014; Hurley, Coleman, Irlanda y Williams, 2005).

El uso de aditivos antiaglutinantes permitidos celulosa y dióxido de silicio, utilizados en quesos rallados en cantidades superiores a las reglamentadas y de agregados que se podrían añadir para aumentar el volumen y peso de lotes comerciales, como son harinas y sémolas de trigo y el polvo de madera (aserrín), constituyen las adulteraciones estudiadas en esta tesis.

2.4. Métodos de screening

La mayoría de las problemáticas analíticas requieren la cuantificación de uno o más analitos presentes en una determinada muestra (análisis cuantitativo). También existe el enfoque semi-cuantitativo o cualitativo, como por ejemplo cuando se quiere autenticar una sustancia/producto o verificar si una sustancia está presente por encima o debajo de una concentración preestablecida. En estos casos, el uso de métodos cualitativos que proveen una respuesta binaria (positivo/negativo) son adecuados para tal fin. Los métodos cualitativos han sido utilizados ampliamente en situaciones que requieren una rápida respuesta, y representan una alternativa muy útil a los métodos cuantitativos, los cuales generalmente proporcionan una gran cantidad de información que no siempre es necesaria, con la consecuente inversión de un mayor costo y tiempo de análisis. En los últimos años, el desarrollo de métodos cualitativos ha crecido ampliamente y se han aplicado en campos como medicina, biología y química (López, Callao y Ruisánchez, 2015).

El análisis cualitativo ha sido definido por diferentes entidades internacionales. La Unión Internacional de Química Pura y Aplicada (IUPAC) lo define como el análisis por *“el cual se identifican o clasifican las sustancias en base a sus propiedades físicas o químicas”* (IUPAC, 1995). Por otro lado, la AOAC y la FDA han ampliado la definición a *“método en el cual se identifican o clasifican las sustancias en base sus características químicas físicas o biológicas”* (FDA 2017; AOAC, 2007). Su respuesta es ausencia/presencia de analitos detectados directa o indirectamente por un determinado análisis. De estas definiciones se

puede inferir que el análisis cualitativo se caracteriza por respuestas binarias (positivo/negativo), pero también se relaciona con propiedades categóricas de la muestra en vez de ausencia/presencia.

En este contexto, los métodos de screening se utilizan para detectar la presencia de una sustancia o clase de sustancias a una determinada concentración de interés. Estos métodos ofrecen la posibilidad de procesar un elevado número de muestras y son utilizados para cribar un elevado número de muestras de resultados de no conformidad (por ejemplo, muestras que no cumplen con la legislación de un determinado aditivo). Los métodos cualitativos multivariantes se caracterizan por proporcionar una respuesta binaria de múltiples variables analizadas, como por ejemplo un espectro NIR. En el caso particular de la espectroscopía NIR, es una técnica que ofrece una muy rica información molecular, pero suele ser poco específica en matrices complejas como son los alimentos, por lo que debe ir siempre acompañada de un adecuado tratamiento quimiométrico para extraer de los datos la información útil necesaria para abordar el problema analítico estudiado. La Figura 2.8 muestra un esquema de la aplicación de métodos cualitativos multivariantes: Se emplean técnicas analíticas como NIR o imágenes digitales para extraer información del sistema en estudio, esta información multivariante se procesa mediante tratamiento quimiométrico, con el fin de obtener respuestas como ausencia/presencia de adulterantes en alimentos (respuestas binarias) basadas en un criterio de decisión, como por ejemplo, concentraciones de aditivos permitidos en alimentos (niveles regulatorios).

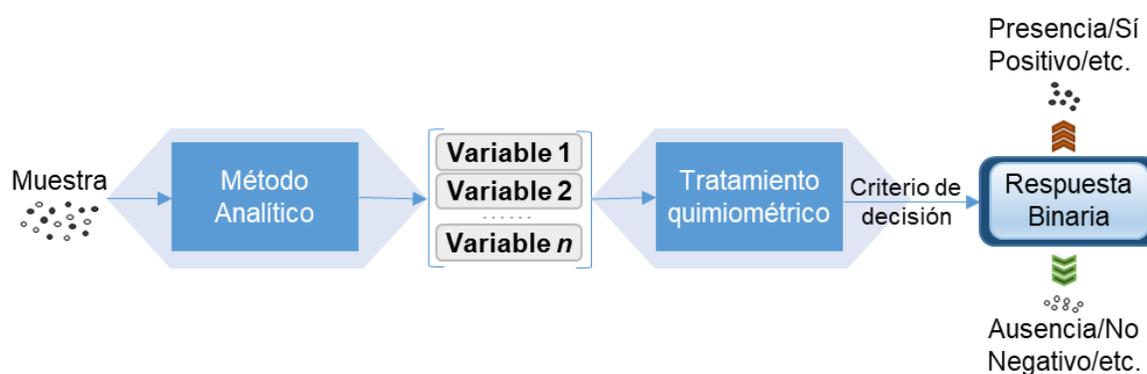


Figura 2.8. Esquema de análisis cualitativo.

La necesidad de una rápida respuesta en laboratorios de control de calidad es una demanda cada vez más creciente. En el ámbito de la química, los métodos de screening han sido una herramienta muy útil y eficaz a lo largo de los años en muchas disciplinas y laboratorios de rutina (Muñoz-Olivas, 2004). En resumen, los métodos de screening se

basan en extraer, aislar e identificar un compuesto o grupo de componentes de una muestra con mínima manipulación de la misma. En general, estos métodos presentan las siguientes características:

- Poseen un mayor énfasis cualitativo.
- Usualmente no incluyen pre-tratamiento de muestra.
- Son simples y rápidos.
- La respuesta obtenida requiere a veces la confirmación mediante un método alternativo.
- Son muy útiles en situaciones que requieren respuestas inmediatas.
- Evitan la necesidad de procesar un elevado número de muestras.
- Promueven la toma rápida de decisiones.
- Obtienen medidas globales del analito.
- Minimizan operaciones preliminares de métodos convencionales que usualmente son tediosos, largos y fuentes de errores sistemáticos, etc.
- Minimizan el uso de instrumentos de elevado costo y mantenimiento, ya que sólo las muestras positivas requerirían de los mismos (Muñoz-Olivas, 2004).

La implementación de metodologías analíticas de screening es de gran utilidad en el campo de los alimentos y, más aún, en lo que respecta a seguridad alimentaria, ya que resulta una herramienta muy útil y práctica para determinar, a *priori*, el estado de una muestra. Por ejemplo, una muestra a analizar puede ser positiva (adulterada) o negativa (no adulterada), y en caso de detectar muestras positivas, éstas deben ser analizadas mediante un análisis confirmatorio, que generalmente involucra instrumentación de mayor sensibilidad y más costosa. Por lo tanto, el ahorro en tiempo y costo de análisis al analizar un menor número de muestras es considerable.

2.5. Técnicas analíticas

Durante los últimos años, ha aumentado el interés por el desarrollo de metodologías de control de parámetros analíticos en tiempo real. Específicamente, en el campo de la seguridad alimentaria, el objetivo primordial ha sido llevar a cabo el análisis de productos de una manera simple, rápida, cualitativa o cuantitativamente, con la posible implementación de análisis *in-situ* en cadenas de producción o laboratorios de control de calidad de rutina. Para tal fin, esta tesis tiene su foco en el estudio de dos técnicas analíticas con potencial de portabilidad para análisis *in-situ*: la Espectroscopía de Infrarrojo Cercano

(Near Infrared Spectroscopy, NIR) y el análisis de imágenes digitales. Estas técnicas tienen la ventaja de proveer una respuesta analítica rápida sin necesidad de destruir la muestra, lo que las hace ideales para analizar muestras en cualquier estadio de la cadena de producción. No obstante, estas técnicas no son altamente específicas cuando se trabaja con matrices complejas como los alimentos. Por ello, es importante el tratamiento de datos mediante el análisis multivariante de la señal para obtener información útil y relevante.

2.5.1. Espectroscopía de infrarrojo cercano (NIR)

La región del Infrarrojo está comprendida entre $700\text{-}10^6$ nm y, tanto por razones instrumentales como por las características de interacción de la radiación con la materia, esta región se divide en tres zonas: Infrarrojo cercano (NIR), Infrarrojo medio (MIR) e Infrarrojo lejano (FIR) (Tabla 2.2).

Tabla 2.2. Zonas de la región Infrarroja del espectro electromagnético.

Región	Longitud de onda (nm)	Origen Absorbancia
NIR	700-2500	Sobretonos y bandas de combinación vibraciones moleculares fundamentales
MIR	2500-50000	Vibraciones moleculares fundamentales
FIR	$50000\text{-}10^6$	Rotaciones moleculares

La absorción en el infrarrojo cercano se debe a los armónicos y bandas de combinación de las vibraciones fundamentales de las moléculas. Una molécula absorbe radiación infrarroja cuando experimenta un cambio neto en el momento dipolar como consecuencia de su movimiento de vibración o rotación. Si esto ocurre, el campo eléctrico asociado a la radiación puede interactuar con el campo eléctrico originado por la fluctuación del momento dipolar de la molécula. Si la frecuencia de radiación iguala a la frecuencia de vibración natural de la molécula, ocurre una transferencia neta de energía que da lugar a un cambio en la amplitud de la vibración molecular y, como consecuencia, se absorbe radiación (Skoog, Holler y Nieman, 2001). En el caso de la espectroscopía de infrarrojo cercano (NIR), es una herramienta poderosa para identificar compuestos orgánicos e inorgánicos puros, porque a excepción de las moléculas homonucleares (O_2 , N_2 , Cl_2), todas las especies químicas absorben radiación infrarroja. Cada compuesto molecular tiene un espectro de absorción infrarrojo único. Por lo tanto, una coincidencia exacta entre el espectro de un compuesto de estructura conocida y el espectro del analito, identifica

inequívocamente el analito (Skoog, et. al, 2014). La energía NIR con longitudes de onda específicas puede ser parcialmente absorbida por enlaces moleculares específicos formados por hidrógeno, por ejemplo, C-H, O-H, S-H y N-H para convertirse en energía mecánica de sobretonos moleculares y vibraciones (Pasquini y Braz, 2003). La cantidad de esta energía absorbida se ve afectada por la composición química y las propiedades físicas de una muestra examinada.

Al incidir sobre una muestra, la radiación infrarroja puede sufrir diferentes fenómenos: absorción, transmisión y reflexión. Así, de la intensidad de radiación incidente (P_0), una parte es transmitida (P_T), otra parte es reflejada (P_R) mientras que otra parte es absorbida por la sustancia (P_A). En líneas generales, las mediciones en el NIR pueden realizarse teniendo en cuenta la radiación transmitida/absorbida (medidas por transmisión), la radiación reflejada (reflectancia difusa, reflectancia total interna) o ambas a la vez (transflectancia). En esta tesis, las medidas realizadas se llevaron a cabo en modo de reflectancia difusa. El instrumento FT-NIR posee un compartimento para medir muestras sólidas (modo reflectancia) que consta de una esfera integradora que recolecta el espectro de diferentes partes de la muestra (número de scans), resultando en la obtención de espectros más representativos. En el modo transmitancia la luz atraviesa la muestra y llega al detector, se pueden medir gases, líquidos, semilíquidos y sólidos; en el modo reflectancia el haz de luz es reflejado por la muestra y recogido por el detector, se pueden medir semisólidos y sólidos y en modo transflectancia el haz de luz atraviesa la muestra, se refleja en un reflector que está en contacto con la misma y finalmente llega al detector, se aplica a muestras líquidas y semilíquidas. Las diferentes formas de interacción de la radiación infrarroja con una muestra se pueden ver en la Figura 2.9 (Peguero Gutiérrez, 2010).

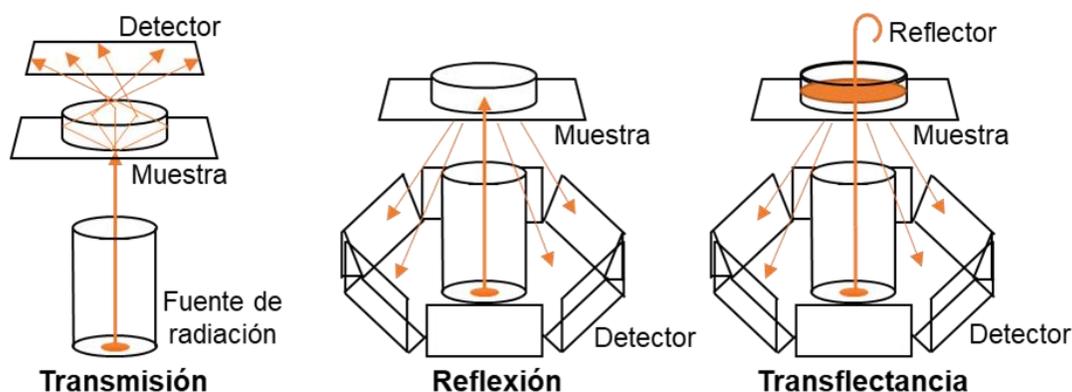


Figura 2.9. Diferentes tipos de medida en la región del infrarrojo cercano. **(a)** Transmisión, **(b)** Reflectancia difusa, **(c)** Transflectancia.

Los parámetros instrumentales bajo los cuales se obtuvieron los espectros NIR durante la presente tesis son: número de scans, resolución, ganancia (gain), tiempo de integración, apertura y corrección de fondo (background). La Tabla 2.3 ofrece una explicación de cada parámetro con su correspondiente rango de variabilidad. El valor óptimo de cada parámetro dependerá de la muestra y objetivo de estudio.

Tabla 2.3. Parámetros instrumentales para la adquisición de espectros NIR.

Parámetro	Descripción	Rango de variabilidad
Número de escaneos	Se refiere al número de veces que el espectrofotómetro toma mediciones en un rango de longitudes de onda determinado. Realizar múltiples escaneos y promediarlos puede mejorar la precisión de la medición y reducir el ruido.	10-100
Resolución	Capacidad del instrumento para distinguir pequeñas diferencias en la absorbancia o la intensidad de la luz en diferentes longitudes de onda. Una resolución más alta permite una mayor precisión en las mediciones, ya que puede detectar cambios más sutiles en la absorción.	2-8 cm ⁻¹
Ganancia	La ganancia se refiere al ajuste de la amplificación de la señal. Permite aumentar o disminuir la amplitud de la señal detectada. La ganancia se utiliza para optimizar la relación señal-ruido y ajustar la sensibilidad del instrumento a diferentes muestras.	1-10, donde 10 es la ganancia más alta
Tiempo de integración	Tiempo que el detector necesita para recopilar datos en cada punto de longitud de onda. Un tiempo de integración más largo puede mejorar la relación señal-ruido, pero también puede aumentar el tiempo total requerido para adquirir un espectro completo.	Microsegundos a segundos, o modo automático
Rango espectral	Rango de longitudes de onda	700-2500 nm
Apertura	Refiere al tamaño del área de la muestra que se expone a la luz en el espectrofotómetro. Una apertura más grande puede ser adecuada para muestras sólidas o heterogéneas, mientras que una apertura más pequeña puede ser preferible para muestras líquidas o concentradas.	4-10 mm
Corrección de fondo	Algunos espectrofotómetros NIR permiten realizar una corrección de fondo, teniendo en cuenta propiedades ópticas del módulo de muestreo y el instrumento.	<i>Background</i> interno o externo

En el caso de muestras sólidas, los espectros NIR pueden estar significativamente influenciados por variaciones sistemáticas no deseadas causadas principalmente por dispersión de la radiación, debido a la variación del tamaño de partícula y diferencias en el camino óptico efectivo. Estas variaciones no deseadas provocan desplazamientos en la línea de base y efectos de no linealidad (Rinnan, Van Den Berg y Engelsen, 2009). Por otro lado, las técnicas espectroscópicas como NIR generan una gran cantidad de datos relativos a las muestras analizadas en breves intervalos de tiempo. En NIR, las bandas de absorción no están bien definidas, apareciendo en forma de bandas de poca intensidad, anchas y solapadas, lo que afecta la especificidad del sistema. Por consiguiente, el uso de herramientas matemáticas para el tratamiento de los datos se necesita cuando se trabaja con espectros NIR, con el fin de extraer la información analítica útil y relevante para tener una correcta interpretación del sistema en estudio. En este contexto, la quimiometría ofrece soluciones para procesar la información proveniente de espectros NIR, proporcionando una amplia variedad de algoritmos en función de la naturaleza del problema en estudio (Wei Hong y Seng Chia, 2020).

La espectroscopía NIR es una técnica ampliamente utilizada para análisis de screening en el campo de los alimentos y materias primas, con resultados exitosos en la determinación de calidad de productos, adulteraciones alimentarias, control y monitoreo de procesos. El interés creciente de la espectroscopía NIR en el sector industrial es consecuencia directa de las ventajas que esta técnica analítica ofrece para el control de calidad del producto terminado. A continuación, se describen las principales ventajas de la espectroscopía NIR:

- Es una técnica no destructiva ni invasiva de la muestra. Permite el análisis tanto *in-line* como *on-line*.
- Se pueden analizar muestras sólidas, líquidas y gaseosas.
- La preparación de la muestra es prácticamente nula, el análisis es rápido y con costos reducidos.
- No se utilizan solventes y no se generan residuos.
- Se obtiene tanto información química como física de la muestra.
- A partir de un único conjunto de espectros se pueden establecer modelos multivariantes para realizar la determinación de múltiples parámetros, clasificar muestras en función de sus atributos, realizar controles de calidad, etc.

En base a las características descritas anteriormente, NIR es una técnica que se ha aplicado en numerosas problemáticas de adulteraciones en alimentos. La Tabla 2.4 enumera algunas aplicaciones de uso de NIR junto con análisis multivariado de datos para determinar adulteraciones y realizar el control de calidad de productos lácteos.

Tabla 2.4. NIR en detección de fraudes alimentarios y control de calidad de productos lácteos.

Muestra	Aplicación	Tratamiento Quimiométrico*	Referencia
Manteca	Adulteración de manteca con aceite	PCA, PLS-DA	Madeiras, Lima, Gonçalves, Godoy, Barbin, 2023
Leche en polvo	Adulteración de leche en polvo con leche de menor calidad	K-NN	Yuan, Chen, Huang, Chen, y Pan, 2023
Leche	Adulteración de leche de cabra y vaca con suero	DD-SIMCA	Galvan et al., 2022
Queso	Adulteración en el proceso de cocción/estiramiento del queso	SIMCA	Ozturk, Ali Dogan, Menevseoglu, y Ayvaz, 2022
Leche de burra	Adulteración de leche de burra con leche de vaca	PLS-DA	Di Donato, Biancolillo, Ferretti, D'Archivio, y Marini, 2023
Yogurt y queso de cabra	Adulteración de yogurt y queso de cabra con leche de vaca	PCA, PLS-DA	Teixeira et al., 2021
Mantequilla, productos lácteos para untar y queso Cheddar	Predicción de ácidos grasos trans naturales, inducidos industrialmente y totales en mantequilla, productos lácteos para untar y queso Cheddar	PLS	Zhao, Renwick, Anna, O'Donnell, y Downey, 2015
Aceite de mantequilla	Adulteración de aceite de mantequilla con otras grasas más económicas o caducadas	DD-SIMCA, SVM	Da Silva Medeiros Lima, Cruz-Tirado, Freitas Lima, Teixeira Godoy, Fernandes Barbin, 2023

*PCA: Análisis por Componentes Principales, PLS-DA: Análisis Discriminante de Mínimos Cuadrados Parciales, K-NN: K-Vecinos más Cercanos, DD-SIMCA: Modelado Suave e Independiente por Analogía de Clases Impulsado por Datos, SIMCA: Modelado Suave e Independiente por Analogía de Clases, PLS: Mínimos Cuadrados Parciales, SVM: Máquinas de Soporte Vectorial.

La rapidez de NIR permite realizar medidas frecuentes desde la recolección de materias primas, producción y control del proceso hasta la calidad del producto final. El desarrollo de nuevas herramientas quimiométricas para el tratamiento de datos junto con la miniaturización de componentes instrumentales que facilitan la portabilidad de los equipos hace que esta técnica esté en plena fase de auge y expansión.

2.5.2. Imágenes digitales

Durante los últimos años, las imágenes digitales se han utilizado en el ámbito de la química analítica como una herramienta potencial emergente para el análisis tanto cuantitativo como cualitativo de alimentos (Herrero-Latorre, Barciela-García, García-Martín, y Peña-Crecente, 2019; Beltrame et al., 2021). A modo de ejemplo, existen trabajos de control de calidad en bebidas, salsas, autenticación de té, mieles, y determinación de adulteraciones en café, entre otros (Diniz et. al, 2012; Botelho, De Assis, y Sena, 2014; Domínguez, Diniz, Di Nezio, Araújo, y Centurión, 2014; Souto et. al, 2015).

El uso de imágenes digitales reemplaza al ojo humano, eliminando la naturaleza subjetiva del análisis, el cual es sustancialmente influenciado por las condiciones del medio ambiente, generando inconsistencias. Las imágenes pueden obtenerse mediante varios dispositivos como cámaras digitales, cámaras web, scanners y teléfonos celulares. El análisis de las imágenes involucra diferentes pasos: inicialmente el proceso de captura, luego, y en caso de ser necesario, acondicionamiento de la imagen obtenida y, finalmente, mediante el empleo de algoritmos computacionales, la extracción de la información relevante de las características que son de interés. Este análisis involucra el empleo de modelos de color, con la finalidad de especificarlos de una forma normalizada y aceptada genéricamente. Los sistemas de color más utilizados son: **Rojo-Verde-Azul** (Red-Green-Blue, RGB), **Tono-Saturación-Brillo** (Hue-Saturation-Brightness, HSB) y **escala de grises** (Grayscale, EG), que se definen en un espacio de coordenadas, donde cada color es representando por un solo punto. Con el fin de obtener una señal analítica característica del sistema en estudio, las imágenes digitales son descompuestas en histogramas de color que describen la distribución estadística de píxeles (número de píxeles) en función de la componente de color registrado. Así, se extrae gran cantidad de información mediante el empleo de algoritmos informáticos avanzados en tiempos muy cortos, de forma objetiva y a un bajo costo (Diniz, 2020). Los histogramas de color obtenidos de las imágenes digitales constituyen la señal analítica característica del sistema en estudio. En general, los sistemas

de detección de imágenes están formados por una fuente de iluminación, un dispositivo de toma de imágenes (cámara) y un software de tratamiento de datos (Figura 2.10).

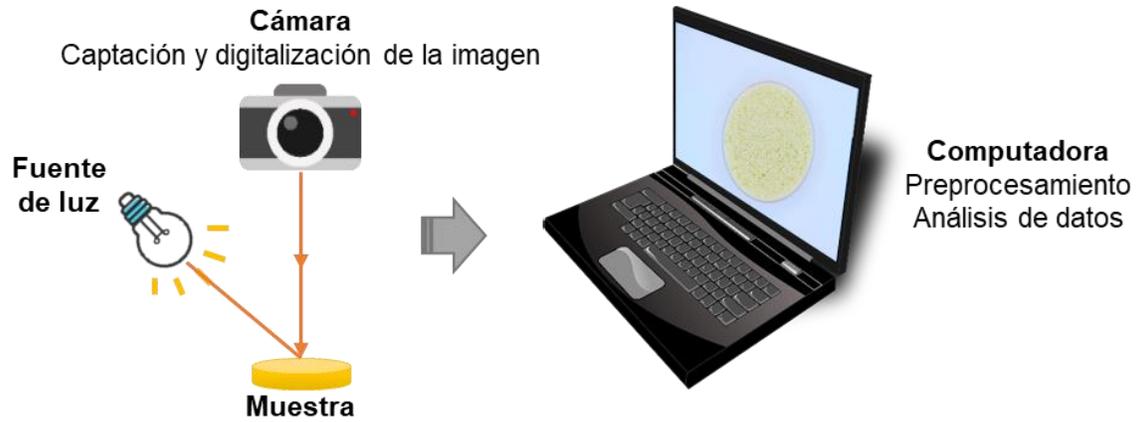


Figura 2.10. Sistema de detección y procesamiento de imágenes.

Características de una imagen digital

Una imagen puede definirse como una función bidimensional $f(x, y)$, donde “ x ” e “ y ” son coordenadas espaciales (planas) y la amplitud de “ f ” en cualquier par de coordenadas (x, y) se llama intensidad de la imagen en ese punto (x, y). Los valores de intensidad de “ f ” son cantidades finitas y discretas (Figura 2.11). Una imagen digital está compuesta de un número finito de elementos, cada uno de los cuales tiene una ubicación y un valor particular. Estos elementos se denominan elementos de imagen o píxeles (González, y Woods, 2008).

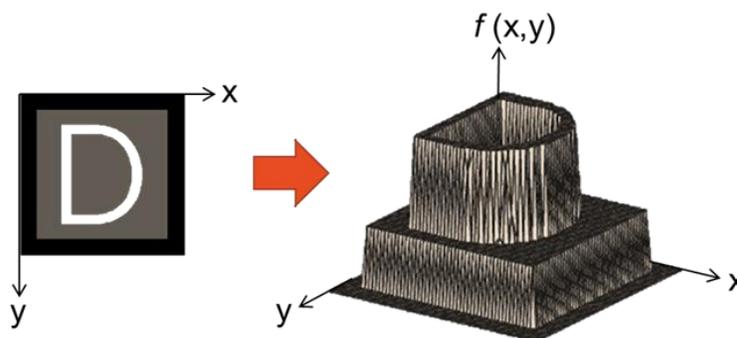


Figura 2.11. Representación gráfica de una imagen.

La imagen capturada mediante dispositivos como cámaras digitales, celulares, etc. es almacenada con un formato determinado como “.jpg” o “.tiff”. Luego se selecciona un área de interés, se procesa mediante un software científico y, de esta manera, se extrae la

información relevante y las características de la imagen. De acuerdo a la información almacenada en píxeles, se pueden obtener cuatro características distintas de la imagen: color, tamaño, forma y textura. El color es la intensidad de los píxeles, el tamaño refleja el número de píxeles, la forma describe el límite físico de los productos capturados con la imagen y la textura es normalmente la dependencia entre cada pixel y los píxeles vecinos o la variación de la intensidad de los píxeles (Zheng C., Sun, y Zheng L., 2006).

La característica del color en el análisis de las imágenes involucra modelos de color normalizados y aceptados genéricamente. Un modelo de color es un modelo matemático que permite representar los colores en forma numérica. En esta tesis se estudiarán los modelos RGB, HSB y EG.

Modelo de color RGB

El ojo humano tiene dos tipos de células sensibles a la luz o foto receptores, los bastones y los conos. Aunque tanto los bastones como los conos responden a un rango de longitudes de onda de luz, la percepción del color, así como la visión a niveles de luz elevados, depende principalmente de los conos. La retina del ojo humano contiene tres tipos de conos, donde cada uno es especialmente sensible a la luz en diferentes zonas del espectro visible. El cono S, sensible a longitudes de onda cercanas a 480 nm, correspondiente al color azul; el cono M, sensible a longitudes de onda cercanas a 550 nm, correspondiente al color verde; y el cono L, sensible a longitudes de onda cercanas a 680 nm, correspondiente al color rojo (Péres, 2007) (Figura 2.12).

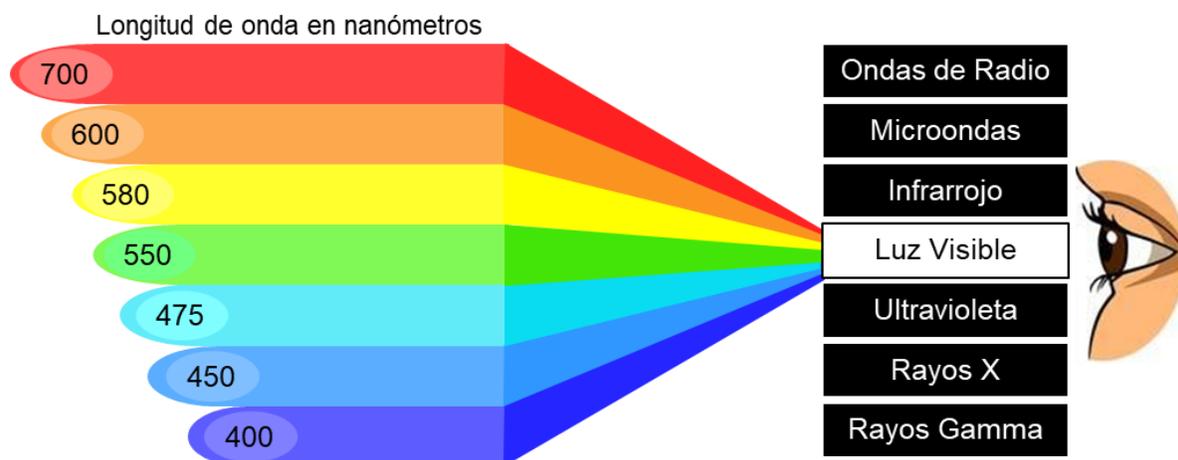


Figura 2.12. Región visible del espectro electromagnético.

RGB es el modelo de color más común en dispositivos de captura de imágenes ya que el sistema visual humano trabaja de forma similar. RGB utiliza los componentes (o canales): rojo, verde y azul, que son los tres colores utilizados para definir completamente un color (Capitán-Vallvey, López-Ruiz, Martínez-Olmos, Erenas, y Palma, 2015). En una imagen de 8 bits por canal, cada componente puede tomar 256 valores diferentes de intensidad y se expresa como un número entre 0 y 255 (Figura 2.13).

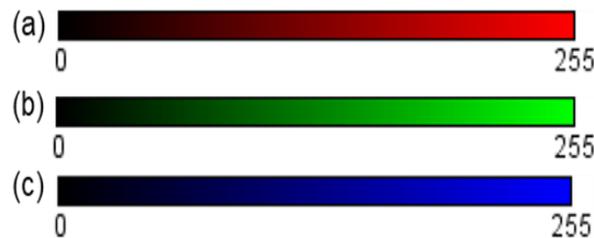


Figura 2.13. Escala de valores de los canales (a) R, (b) G, y (c) B.

El canal 0 corresponde a la ausencia total de un color y el canal 255 corresponde a la máxima intensidad de color. Cada píxel está formado por una combinación de estos colores. La combinación de los canales RGB crea los casi 16,8 millones de colores diferentes (Botelho, De Assi, y Sena, 2014). Estos colores juntos determinan un espacio de colores. Se puede imaginar este espacio como un cubo alineado al sistema de coordenadas de un espacio tridimensional, en el cual la intensidad de color rojo se representa a lo largo del eje "X", la intensidad de color verde se representa a lo largo del eje "Y", y la intensidad de color azul se representa a lo largo del eje "Z" (Figura 2.14). En esta representación, cada color tiene una posición única.

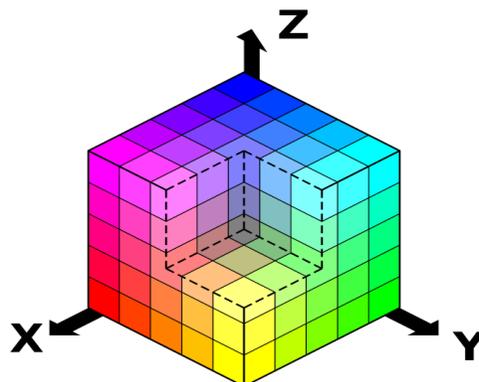
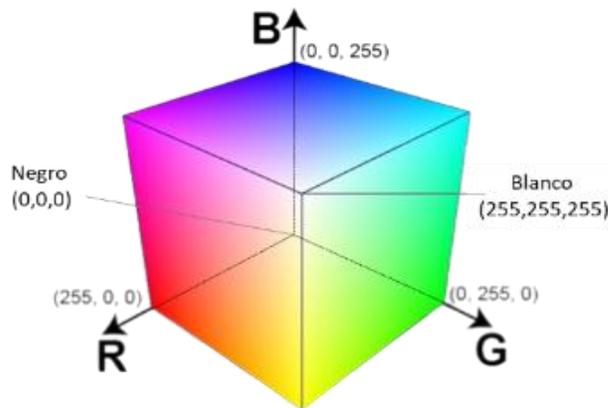


Figura 2.14. Modelo RGB representado en un espacio tridimensional.

Cuando los canales rojo, azul y verde se combinan en su máxima intensidad (255), el ojo percibe el color resultante como blanco. Cuando la intensidad de cada componente es 0, indica que hay ausencia de los tres colores y el ojo percibe el color negro. Cuando las intensidades de cada componente son iguales, se obtiene un tono de gris neutro. En la Figura 2.15 se muestran los valores de intensidad de cada canal de color.



Color	Coordenadas
Negro	RGB (0,0,0)
Rojo	RGB (255,0,0)
Amarillo	RGB (255,255,0)
Verde	RGB (0,255,0)
Cian	RGB (0,255,255)
Azul	RGB (0,0,255)
Magenta	RGB (255,0,255)
Blanco	RGB (255,255,255)

Figura 2.15. Representación de la intensidad en cada canal.

Modelo de color HSB o HSV

El modelo de color HSB utiliza el matiz (H, del inglés "Hue"), la saturación (S, del inglés "Saturation") y el brillo (B, del inglés "Brightness") como componentes para definir los colores. HSB también se denomina HSV (con los componentes matiz, saturación y valor). El matiz se define comúnmente como la longitud de onda dominante de la radiancia espectral de color, la saturación define la pureza del color y el brillo, que suele ser también conocido como valor, representa la intensidad de la luz presente en el color descrito. Los valores de intensidad de HSB pueden obtenerse fácilmente del espacio de color RGB mediante ecuaciones específicas (Capitán-Vallvey, López-Ruiz, Martínez-Olmos, Erenas, y Palma, 2015).

El matiz se expresa en grados para representar la ubicación del espectro de colores estándar. El color rojo corresponde a 0 grados, el verde 120 grados, el azul 240 grados, y nuevamente el rojo a 360 grados (Figura 2.16). Esta escala gradual es ajustada a la escala de 256 valores de intensidad usada en el modelo de color RGB.

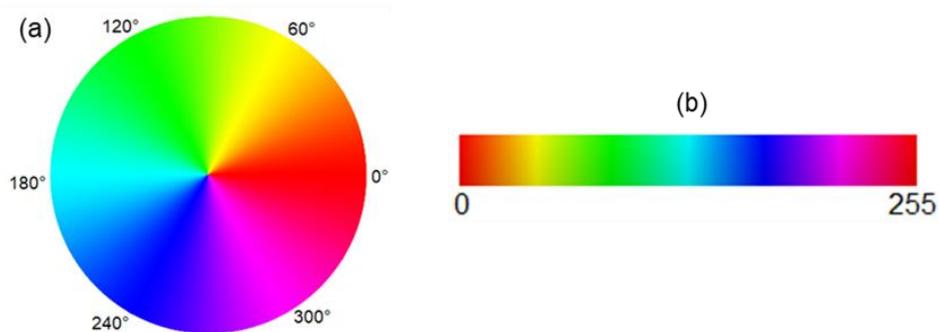


Figura 2.16. Representación **(a)** angular y **(b)** lineal de H.

La saturación varía de 0 a 100 y se expresa en porcentajes (cuanto mayor es el valor, más puro es el color). Esta escala porcentual es ajustada a la escala de 256 valores de intensidad utilizada en el modelo de color RGB de manera tal que el 0% de saturación corresponde al valor de intensidad 0, y el 100% de saturación corresponde al valor 255 (Figura 2.17).



Figura 2.17. Representación lineal de S.

El brillo o valor determina la cantidad de blanco que contiene el color. Como ocurre con la saturación, los valores de brillo varían de 0 a 100 y representan porcentajes (cuanto mayor es el valor, más brillante es el color). Nuevamente, esta escala es ajustada a la escala usada en el modelo de color RGB, donde el 0% de brillo (ausencia de blanco) corresponde al valor de intensidad 0, y el 100% de brillo (color blanco) corresponde al valor 255 (Figura 2.18).



Figura 2.18. Representación lineal de V.

Este modelo de color puede representarse en un sistema cilíndrico que contiene los tres componentes de color (Figura 2.19).

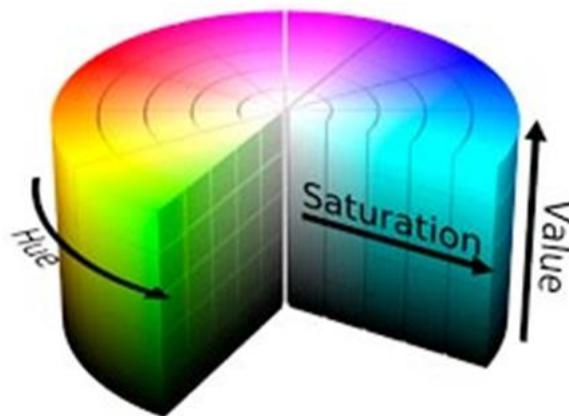


Figura 2.19. Representación modelo de color HSV.

De la misma manera que el modelo de color RGB, cada pixel tendrá un valor de intensidad determinado de cada componente del HSB, es decir que tendrá una intensidad determinada de matiz, una intensidad determinada de saturación y una intensidad determinada de brillo, las cuales pueden tomar valores entre 0 y 255. Para fines prácticos se utilizarán las siglas HSV en vez de HSB.

Modelo de color de escala de grises (EG)

El modelo de color de escala de grises define el color con un único componente, la luminosidad, que se mide con valores comprendidos entre el 0 y el 255. Cada uno de los colores de la escala de grises tiene un valor de intensidad equivalente de los componentes rojo, verde y azul del modelo de color RGB. El valor 0 corresponde a la ausencia de todos los colores y el valor 255 corresponde a la máxima intensidad de los tres colores (Figura 2.20).



Figura 2.20. Escala de grises.

Esta escala puede observarse en la diagonal que va desde el punto (0,0,0) hasta el punto (255,255,255) en el modelo tridimensional de color RGB que muestra la Figura 2.15.

Histogramas de color

De acuerdo a lo expresado anteriormente, cada pixel cuenta con una componente de la escala de grises, una componente R (rojo del RGB), una componente G (verde del RGB), una componente B (azul del RGB), una componente H (matiz del HSV), una componente S (saturación del HSV) y una componente V (valor del HSV).

Una vez seleccionada un área de interés de la imagen, la cual tiene una determinada cantidad de píxeles, la imagen se descompone en histogramas de color que describen la distribución estadística de píxeles en función de la intensidad de cada componente de color registrada (variables). De esta manera se obtienen siete histogramas diferentes, los cuales, para fines prácticos, se representan en un mismo gráfico (Figura 2.21).

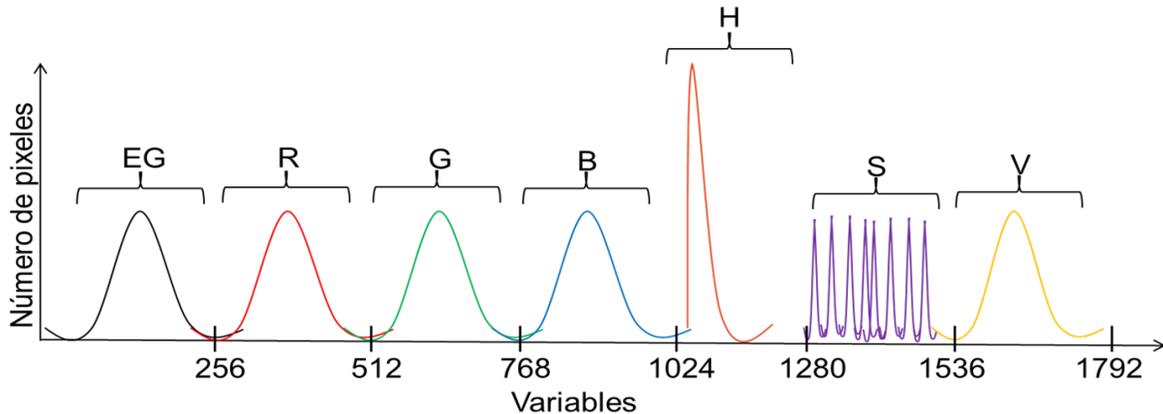


Figura 2.21. Histogramas de todos los modelos de color.

Como ya se ha mencionado, cada componente de color puede tomar 256 valores de intensidad, de esta manera obtenemos un total de 1792 variables, donde cada rango representa a un componente de color diferente (Tabla 2.5).

Tabla 2.5. Números de variables correspondientes a los distintos modelos de color.

Variable	Modelo de Color	Abreviatura
1-256	Escala de Grises	EG
257-512	Rojo del RGB	R
513-768	Verde del RGB	G
769-1024	Azul del RGB	B
1025-1280	Matriz del HSV	H
1281-1536	Saturación del HSV	S
1537-1792	Valor (brillo) del HSV	V

Las imágenes digitales presentan la ventaja principal de que son adquiridas con dispositivos muy asequibles (cámaras de fotos, teléfonos celulares, scanners, etc.) y, al igual que NIR, no destruyen la muestra, no generan residuos, son simples y fáciles de utilizar, generan información analítica en un muy corto período de tiempo y pueden ser empleadas para análisis *in-situ* debido a sus condiciones de portabilidad.

La adulteración de quesos rallados generalmente no es detectada por el ojo del consumidor, por lo que el empleo de técnicas de screening simples, rápidas y económicas representan una potencial y útil alternativa para realizar el control de calidad de alimentos con elevado valor comercial, como son los quesos de pasta dura, que son alimentos muy susceptibles de prácticas fraudulentas. Además, cabe destacar que las técnicas analíticas empleadas en esta tesis están en concordancia con los principios de la química verde, al no generar ningún tipo de residuo durante el análisis de muestra.

2.6. Análisis de datos multivariante

2.6.1. Pre-procesamiento de espectros NIR

El pre-procesado de espectros es el paso más importante antes de aplicar un modelo multivariante, ya sea un análisis exploratorio de datos o un modelo de calibración o clasificación. Consiste en correcciones basadas en tratamientos matemáticos con la finalidad de aumentar la calidad de la señal instrumental, reduciendo al máximo la información no asociada a la propiedad estudiada, como pueden ser desplazamientos de la línea base, ruido espectral, etc. Se denomina ruido al componente espectral no relacionado con ninguna propiedad física o química de la muestra, como puede ser el ruido instrumental o de iluminación. Además, otras fuentes de variación como la temperatura y humedad pueden aportar interferencias a la señal analítica. En la literatura se han propuesto varios algoritmos matemáticos utilizados comúnmente en NIR, estos pueden dividirse en dos grupos: métodos de corrección de dispersión “*scattering*” y derivadas espectrales (Rinnan, Van Den Berg y Engelsen, 2009).

2.6.1.1. Métodos de corrección de scattering (dispersión)

Corrección multiplicativa de dispersión (Multiplicative Scatter Correction, MSC)

MSC es una técnica utilizada para corregir la dispersión de la luz causada por factores como la reflectancia y la dispersión múltiple. La dispersión de la radiación puede provocar distorsiones en los espectros adquiridos, lo que dificulta la interpretación y comparación de las muestras. El método busca compensar estas variaciones sistemáticas para mejorar la calidad de la información obtenida. Se basa en la suposición que la dispersión es una fuente de variabilidad que afecta a todos los espectros de manera proporcional. El objetivo es eliminar esta variabilidad sistemática mediante la división de cada espectro por una referencia, generalmente la media de un conjunto de datos, produciendo una corrección de efectos de escala y compensación (línea de base) (Figura 2.22).

El primer paso consiste en establecer un modelo lineal con los coeficientes de corrección teniendo en cuenta el espectro promedio:

$$X_{org} = b_0 + b_{ref} \cdot X_{ref} + e$$

Donde X_{org} es el espectro original, X_{ref} es el espectro promedio y b son los coeficientes de regresión. El espectro corregido (X_{corr}) se obtiene como:

$$X_{corr} = \frac{X_{org} \cdot b_0}{b_{ref}}$$

El resultado de la corrección MSC es un conjunto de espectros que han sido normalizados en relación con la dispersión, lo que permite una mejor comparación entre muestras y una mayor capacidad para detectar diferencias sutiles (Fearn, Riccioli, Garrido-Varo, y Guerrero-Ginel, 2009).

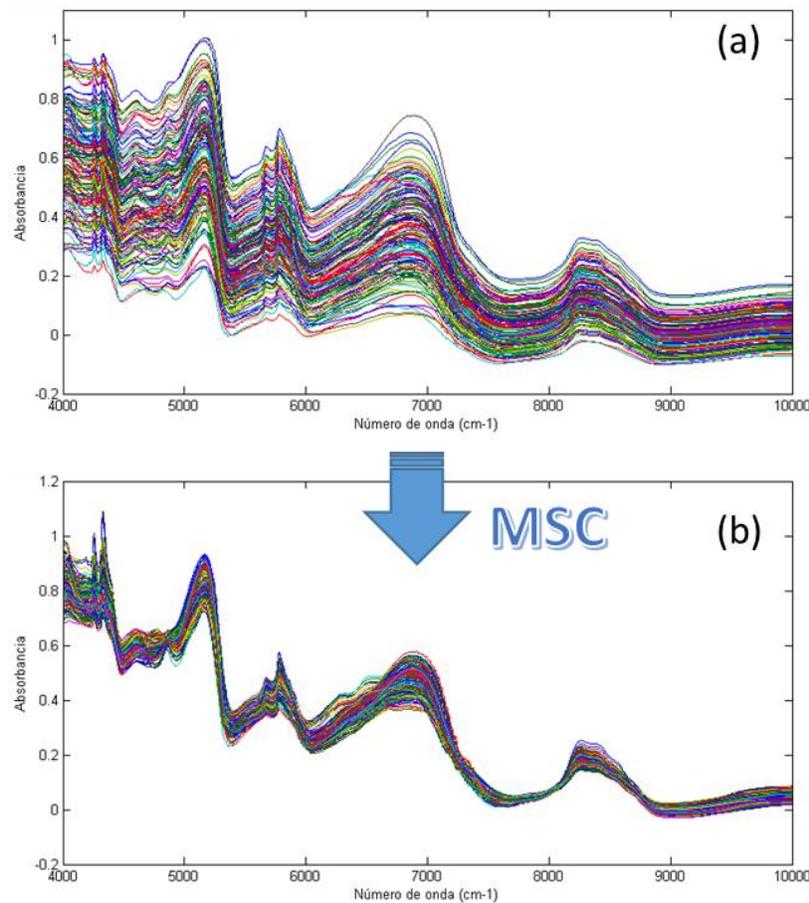


Figura 2.22. (a) Espectros NIR sin pre-procesar, (b) Espectros pre-procesados.

Variable Normal Estándar (Standard Normal Variate, SNV)

SNV es una técnica utilizada para eliminar o reducir las variaciones sistemáticas de las desviaciones de línea de base de los espectros sin eliminar la información relacionada con las diferencias químicas o físicas entre muestras. La línea de base puede variar debido a la dispersión de la luz y la desviación instrumental. La corrección se basa en dividir cada punto espectral por la desviación estándar de los puntos espectrales de una muestra y luego restar la media de esos puntos. El objetivo es transformar los datos originales en una distribución de variable normal estándar, es decir, con una media de cero y una desviación estándar de uno (Figura 2.23). La transformación SNV se define como:

$$X_{corr} = X_{org} - \frac{a_0}{a_1}$$

Donde a_0 y a_1 son el promedio y la desviación estándar del espectro corregido y X_{org} es el espectro original, X_{corr} es el espectro corregido (Fearn, Riccioli, Garrido-Varo, y Guerrero-Ginel, 2009).

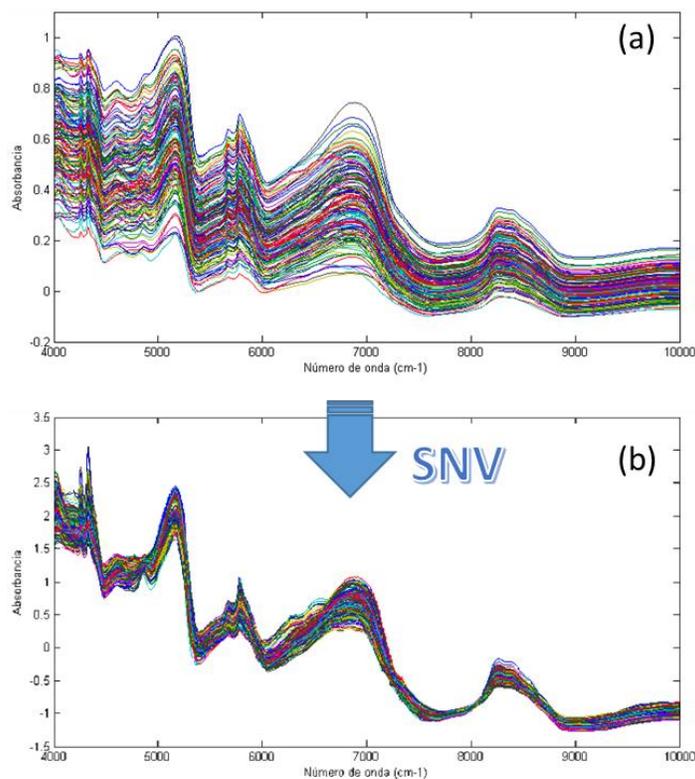


Figura 2.23. (a) Espectros NIR sin pre-procesar, **(b)** Espectros pre-procesados.

Modelado explícito de línea de base: Detrending (eliminación de tendencia)

Detrending es muy utilizado en casos de desplazamiento constante de la línea de base ya sea lineal o curvo. Este método consiste en aproximar la línea de base del espectro con una función que puede ser lineal o polinómica, y que se resta posteriormente para obtener el espectro corregido. Cuando se emplea un polinomio de grado 0 (una constante) se denomina “*offset correction*” y consiste en restar a cada espectro la absorbancia medida a una determinada longitud de onda (zona sin absorción) (Figura 2.24).

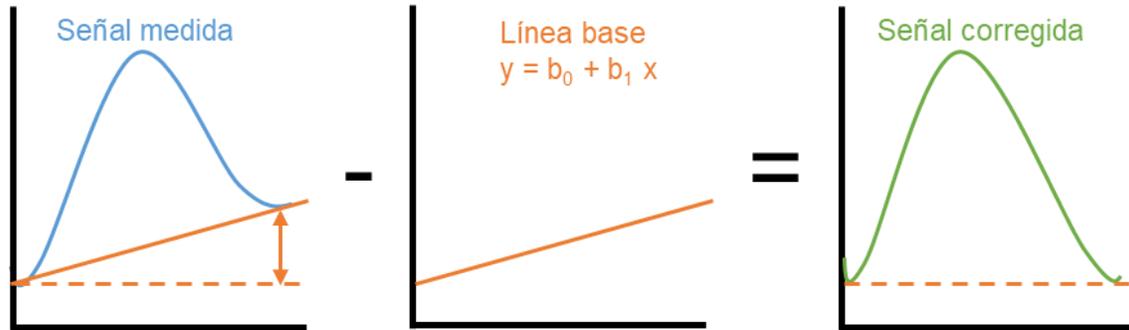


Figura 2.24. Esquema de corrección de la línea de base utilizando un modelo explícito lineal.

2.6.1.2. Derivadas espectrales

El método de derivadas consigue diferenciar mejor los picos solapados y elimina desplazamientos lineales y cuadráticos de la línea de base. La derivada reduce la correlación entre variables y el efecto de la dispersión debido al tamaño de las partículas, problema muy frecuente en espectros NIR. La primera derivada, se obtiene de restar el punto vecino inmediato de una variable, de manera de eliminar la señal común entre las dos variables y deja solo la parte de la señal que es diferente. La primera derivada elimina desplazamientos en el espectro y resta importancia a las señales de frecuencia más baja. La segunda derivada se calcula repitiendo el proceso, lo que acentúa las características de frecuencia más alta.

Las derivadas presentan el inconveniente de magnificar el ruido en la señal, por lo que se utiliza en combinación con métodos de suavizado de la señal, como el de Savitzky-Golay. No suelen utilizarse superiores a orden 2 (segunda derivada), ya que decrece la relación señal-ruido (S/N). El algoritmo Savitzky-Golay requiere la selección de un tamaño de ventana (ancho del filtro) y del orden del polinomio necesarios para el suavizado (Figura 2.25) (Griffiths, 2017).

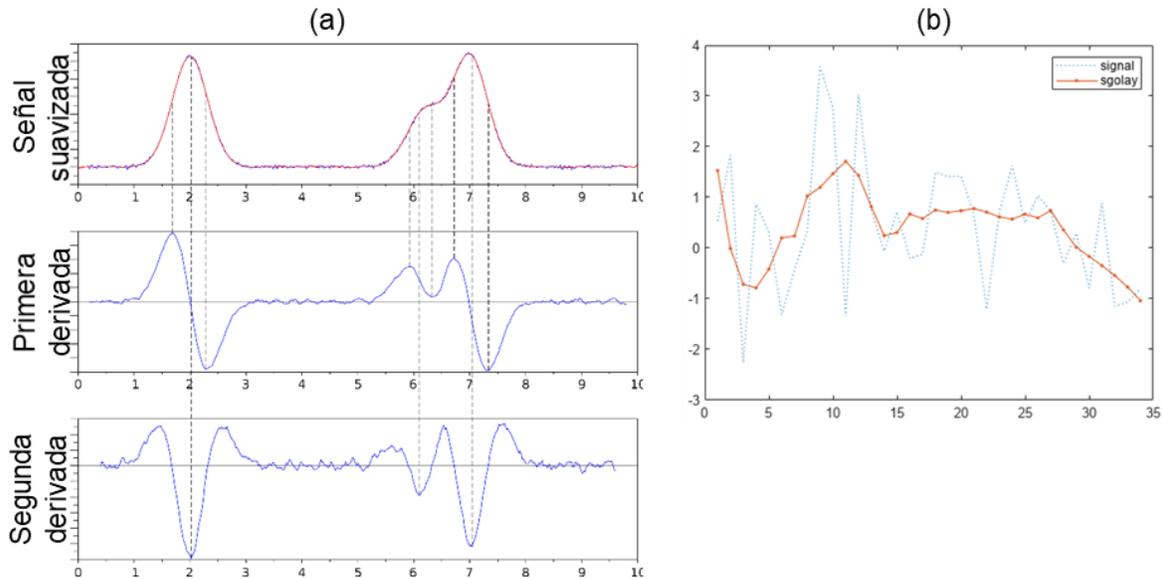


Figura 2.25. (a) Algoritmo de Savitzky-Golay aplicado sobre señales con ruido aleatorio: suavizado (arriba), primera derivada (medio) y segunda derivada (abajo). **(b)** Señal suavizada con una ventana de 11 puntos y un polinomio de tercer grado.

2.6.2. Preparación de conjuntos de muestras (Algoritmo Kennard-Stone)

Previo a la construcción de modelos multivariante, de regresión y clasificación, es necesario establecer el procedimiento para trabajar con el conjunto de muestras. Por tal motivo Kennard y Stone propusieron un método para dividir al conjunto de muestras en entrenamiento y prueba. Se trata de un método secuencial que selecciona un conjunto de “N” puntos en el espacio multidimensional definido por las variables (conjunto candidato). El método comienza al encontrar las dos muestras (o puntos candidatos) que están más alejadas dentro del espacio experimental usando una distancia geométrica, que es normalmente la distancia Euclídea, aunque es también posible utilizar otras como la Mahalanobica. De todos los puntos candidatos restantes, se selecciona aquel que esté más alejado de los dos primeros previamente seleccionados y se añade al conjunto. El proceso se repite hasta que se haya agregado al conjunto el número requerido de muestras. En la práctica, esto produce una red de puntos seleccionados muy uniformemente distribuidos sobre el conjunto de datos (Figura 2.26) (Kennard y Stone, 1969).

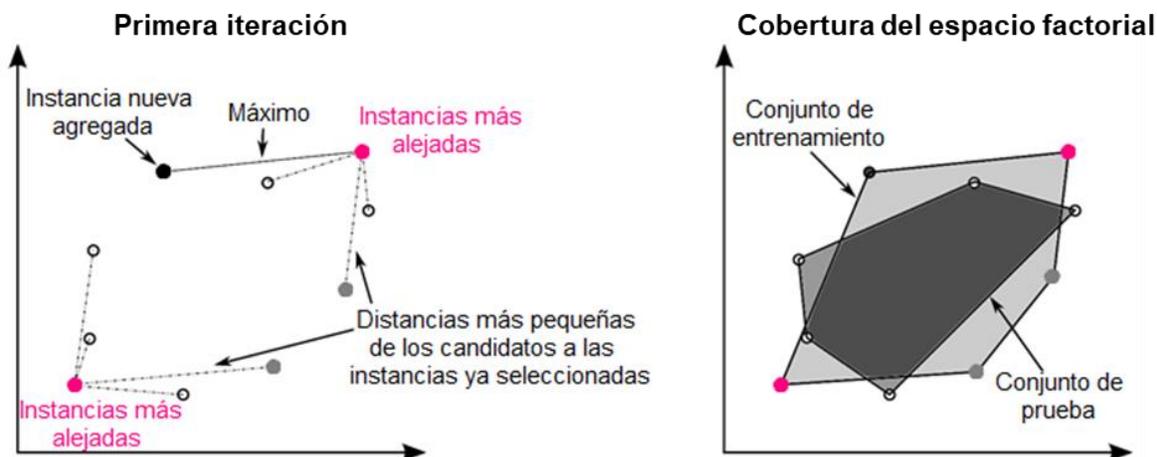


Figura 2.26. Esquema representativo del algoritmo de Kennard-Stone.

Con las muestras de entrenamiento seleccionadas, se construye una regla de clasificación, se valida este modelo mediante validación cruzada y finalmente se evalúa la capacidad predictiva del modelo con las muestras de predicción.

2.6.3. Análisis cuantitativo: Regresión multivariante

2.6.3.1. Mínimos Cuadrados Parciales (PLS)

PLS (Partial Least Squares) es una de las técnicas de regresión multivariante más conocidas y populares en el ámbito de la química analítica. Es una herramienta muy útil cuando se trabaja con datos espectroscópicos y se tiene un conocimiento parcial del sistema a cuantificar. Por ejemplo, si se desea cuantificar adulterantes en un alimento (muestras) mediante el empleo de técnicas espectroscópicas (señal analítica), se construye un modelo PLS a través de una serie de medidas de las muestras, con el fin de establecer una relación entre las características típicas de este conjunto de datos (X) y la concentración conocida de los analitos en estudio (Y). Similarmente, PLS apunta a modelar la covarianza de X e Y. La idea es proporcionar un modelo de regresión multivariante que permita la predicción de la concentración de futuras muestras que contengan características similares a las muestras utilizadas para entrenar el modelo. El algoritmo PLS extrae componentes (variables latentes, LVs) que contienen la información relevante de ambos conjuntos de datos: la señal analítica en estudio (variables independientes) y la concentración de diferentes compuestos en una muestra (variable dependiente). Cada variable latente contiene la información relevante del sistema de manera decreciente, por ejemplo LV2 no contiene información proporcionada por LV1 y así sucesivamente. Un aspecto importante de PLS es que considera los errores provenientes tanto en las concentraciones como en la

señal analítica y también es adecuado cuando existe multicolinealidad en los datos, es decir, variables que están altamente correlacionadas entre sí (Mark y Workman, 2018).

El cálculo matemático de PLS se mostrará más adelante, cuando se aborde una extensión de este algoritmo en un contexto de clasificación (PLS-DA), en donde cambia el bloque de concentraciones por clases o categorías a las cuales pertenecen las muestras.

2.6.3.2. Parámetros de calidad en regresión multivariante

En regresión multivariante, se utilizan diferentes cifras de mérito y métricas para evaluar la calidad de los modelos obtenidos. A continuación, se detallan algunas de las cifras de mérito junto con la métrica de elipses de confianza:

Error cuadrático medio de validación cruzada (RMSECV): Se utiliza para evaluar la calidad de un modelo de regresión multivariante durante el proceso de validación cruzada en datos de entrenamiento. En general, cuanto más bajo sea el valor de RMSECV, mejor será el modelo, ya que indica una menor discrepancia entre los valores predichos y los valores reales. Hay que destacar que un valor de RMSECV muy cercano a cero no es correcto, ya que, si bien indicaría un ajuste perfecto del modelo a los datos de entrenamiento, existe una alta probabilidad de que el modelo esté sobreajustando, lo que significa que el modelo se adapta demasiado a los datos de entrenamiento. El sobreajuste podría tener consecuencias negativas en la predicción de futuras muestras. Finalmente, un valor de RMSECV alto indica un ajuste deficiente del modelo y puede ser indicativo de que el modelo no puede capturar la relación subyacente entre la señal analítica y la concentración de las muestras. La fórmula general del error cuadrático medio es la siguiente:

$$MSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}$$

Donde n es el número de muestras del conjunto de entrenamiento, x_i es la concentración predicha por el modelo PLS e y_i es la concentración real.

Error cuadrático medio de predicción (RMSEP): es un parámetro utilizado para evaluar la calidad de un modelo de regresión en la predicción de nuevos datos, a través de la medida de la discrepancia entre los valores predichos por el modelo y los valores reales en un conjunto de datos de prueba. El RMSEP es similar al RMSECV utilizado en la validación cruzada, pero se aplica a un conjunto de datos independiente que no se utilizó

durante el entrenamiento del modelo. La evaluación en la obtención de valores de RMSEP altos, bajos o cercanos a cero son las mismas que las descritas para el caso anterior (RMSECV).

Relación entre rendimiento y desviación (RPD): es una métrica utilizada para evaluar la capacidad predictiva de un modelo de regresión en comparación con la variabilidad intrínseca de los datos. El RPD es un parámetro comúnmente utilizado en la espectroscopía y la calibración de instrumentos analíticos. El RPD es la relación entre la desviación estándar (SD) y el error estándar de validación cruzada (SECV). Cuanto mayor sea el valor del RPD, mejor será el desempeño del modelo en términos de predicción en comparación con la variabilidad de los datos. La fórmula para el cálculo del RPD es la siguiente:

$$RPD = \frac{SD}{SECV}$$

Donde *SD* es la desviación estándar de los valores de referencia de la propiedad que se está modelando y *SECV* es el error estándar de validación cruzada.

- RPD < 1: Se consideran valores no satisfactorios. El modelo no puede superar la variabilidad intrínseca de los datos y tiene una capacidad predictiva limitada.
- RPD entre 1 y 2: Se considera una pobre capacidad predictiva del modelo, ya que éste puede predecir la variable de respuesta, pero su capacidad predictiva es limitada en relación con la variabilidad de los datos.
- RPD entre 2 y 3: Se considera una capacidad predictiva moderada del modelo, el cual podría proporcionar resultados satisfactorios, pero no son ideales.
- RPD > 3: Son los valores deseables a obtener. En este caso se considera un buen desempeño del modelo, ya que posee una buena capacidad predictiva y puede superar la variabilidad intrínseca de los datos, proporcionando predicciones confiables.

El valor **R-cuadrado de validación cruzada (R²CV):** es una métrica utilizada para evaluar la calidad de un modelo de regresión durante el proceso de validación cruzada. Mide la proporción de la variabilidad de la variable de respuesta (concentración) que puede ser explicada por el modelo. Un R²CV cercano a 1 indica que el modelo tiene un buen ajuste a los datos y puede explicar una gran proporción de la variabilidad de la variable respuesta. Valores altos de R²CV son deseables y generalmente indican un buen rendimiento del modelo. Valores bajos de R²CV indican que el modelo no puede capturar bien la relación

entre las variables de entrada y la variable de respuesta. La fórmula para el **R²CV** es la siguiente:

$$R_{CV}^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Donde SS_{res} es la suma de los cuadrados de los residuos obtenidos durante la validación cruzada y SS_{tot} es la suma total de los cuadrados de las diferencias entre los valores reales de la respuesta y su media.

El **R-cuadrado de predicción (R²pred)**: es un parámetro utilizado para evaluar la calidad de un modelo de regresión en la predicción de nuevos datos. Mide la proporción de la variabilidad de la variable de respuesta (concentración) que puede ser explicada por el modelo en los datos de prueba. La evaluación en la obtención de valores de R²pred altos, bajos o cercanos a cero son las mismas que las descritas para el caso anterior (R²CV). La fórmula para el **R²pred** es la siguiente:

$$R_{pred}^2 = 1 - \frac{PRESS}{SST}$$

Donde $PRESS$ es la suma de los errores de predicción al cuadrado para cada observación dejando fuera una por una y SST es la suma total de los cuadrados de las diferencias entre los valores reales de la respuesta y su media.

Elipses de confianza: Son una representación gráfica utilizada en estadística multivariante para mostrar la incertidumbre asociada con la estimación de los parámetros de un modelo. Estas elipses se basan en la matriz de covarianza y representan regiones elípticas que contienen una determinada probabilidad de contener los verdaderos valores de los parámetros. En general, cuanto más pequeña sea la elipse de confianza, mayor será la precisión de la estimación y menor será la incertidumbre asociada.

En resumen, las cifras o figuras de mérito en regresión multivariante sirven para evaluar la calidad de los modelos obtenidos y para determinar su capacidad predictiva. Cada cifra tiene un significado particular y debe ser interpretada en el contexto del problema específico que se está abordando (Massart, Vandeginste, Buydens, Lewi, y Smeyers-Verbeke, 1997).

2.6.4. Análisis cualitativo: Clasificación multivariante

El análisis multivariante es una herramienta muy útil para extraer de datos experimentales, la mayor cantidad posible de información, interpretarlos y tener conocimiento del sistema químico.

Dentro del conjunto de herramientas multivariantes, existen las **técnicas de reconocimiento de patrones**, que son muy utilizadas desde hace décadas para solventar diversos problemas analíticos. Su filosofía está basada en la identificación de relaciones y/o vínculos entre objetos previamente caracterizados a través de medidas (físicas, químicas o biológicas), con el fin de agruparlos y clasificarlos en una clase o categoría definida según la similitud o diferencias que existan entre ellos (Marini, 2013). La clase o categoría indica la colección de objetos que comparten características similares cuya definición depende del problema en estudio, de manera que el mismo set de muestras puede ser agrupado de distintas maneras, de acuerdo con la finalidad del estudio. Existe una amplia variedad de técnicas de clasificación multivariantes de diferente naturaleza, que pueden emplearse para identificar un alimento adulterado y/o contaminado, evaluar su calidad, autentificarlo, evaluar e identificar cambios causados por procesos industriales, entre otras aplicaciones (Oliveri y Downey, 2012).

Las técnicas de reconocimiento de patrones se pueden dividir en dos grupos: **no supervisadas y supervisadas**. Las primeras apuntan a localizar grupos o tendencias en los datos sin tener información previa del sistema en estudio. Dentro de este grupo se encuentra el análisis exploratorio de datos, el cual es una herramienta muy útil para mejorar el entendimiento y accesibilidad de los resultados a través de una exploración inicial de los datos. Este análisis permite detectar agrupamientos y/o tendencias naturales entre las muestras, tener una idea de la varianza de los datos y poder detectar valores anómalos (*outliers*). Por otro lado, las técnicas de reconocimiento de patrones supervisadas utilizan información previa de las clases a las cuales pertenecen las muestras en estudio. Su fundamento se basa en construir una regla de clasificación a partir de datos de entrenamiento (muestras cuya verdadera clase es conocida) con el fin de clasificar muestras desconocidas dentro de las clases previamente designadas (en nuestro caso muestras sin adulterar y muestras adulteradas con los diferentes adulterantes). El criterio en la elección de una regla de clasificación constituye la base de la diferenciación entre los algoritmos de clasificación propuestos en literatura, los cuales son de diferente naturaleza, ofreciendo una amplia variedad de opciones para resolver diversos problemas de clasificación (Brereton, 2009; Marini, 2013).

Las técnicas de clasificación se pueden dividir en diferentes categorías: técnicas lineales y no lineales, paramétricas y no paramétricas, discriminantes o de modelado de clase, entre otras. La primera agrupación se refiere a la forma matemática de establecer los límites entre

las distintas clases: **técnicas lineales y no lineales**. Las lineales construyen modelos a través del mejor límite lineal para discriminar entre las clases, mientras que las segundas se basan en establecer la mejor curva (límite no lineal cuadrático o polinómico) para separar las clases. Por otro lado, las técnicas pueden ser **paramétricas**, en las cuales se asume una distribución normal de los datos, o **no paramétricas**, en las cuales no se hace una estimación explícita de la distribución que siguen los datos. Finalmente, las **técnicas discriminantes** establecen límites separativos entre las diferentes clases, con el fin de asignar nuevos datos a una de las clases existentes. La clasificación se basa en la estimación de las distribuciones de probabilidad maximizando las diferencias entre las clases, dividiendo el espacio multidimensional en tantas regiones como número de clases en estudio, de manera que una muestra siempre será clasificada a una de las clases disponibles (Figura 2.27a). Las técnicas de **modelado de clase** están basadas en modelar cada clase independientemente una de otra, focalizándose en capturar la similitud entre las muestras de una misma clase en vez de capturar las diferencias entre las distintas clases. Desde el punto de vista algorítmico, cada clase es modelada individual e independientemente de las otras, y desde el geométrico equivale a identificar un volumen en el espacio multidimensional que encierra la clase, de manera que si una muestra cae dentro del mismo es aceptada por esa clase, mientras que si cae fuera es rechazada por ese modelo de clase (Figura 2.27b). En este caso existen varias posibilidades de clasificación: muestras que son aceptadas por una clase, muestras que son aceptadas por más de una clase y muestras que no son aceptadas por ninguna de las clases disponibles.

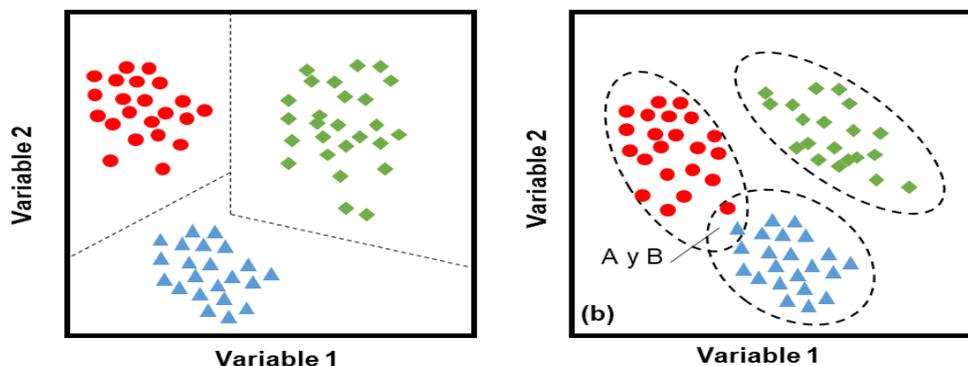


Figura 2.27. (a) Representación de técnicas discriminantes y **(b)** de modelado de clases en un espacio bidimensional.

Ejemplos de técnicas de clasificación discriminantes son: Análisis Discriminante Lineal (Linear Discriminant Analysis, LDA), Análisis Discriminante por Mínimos Cuadrados

Parciales (Partial Least Squares Discriminant Analysis, PLS-DA), Análisis Discriminante Cuadrático (Quadratic Discriminant Analysis, QDA), *K*-Vecinos más Cercanos (*K*-Nearest Neighbours, *K*-NN), Máquinas de Soporte Vectorial (Support Vector Machines, SVM). Asimismo, ejemplos de técnicas de modelado de clases incluyen: Modelado Suave e Independiente por Analogía de Clases (Soft Independent Modelling of Class Analogy, SIMCA) y Modelado de Clases Desiguales (Unequal Class-Modelling, UNEQ). La elección de la técnica de clasificación debe ser hecha en base a la característica de los datos y el objetivo del análisis.

La Figura 2.28 muestra un resumen de las técnicas de clasificación de reconocimiento de patrones enunciadas anteriormente. Las técnicas discriminantes focalizadas en clasificación dirigida son adecuadas cuando se conoce *a priori* los posibles adulterantes que puedan estar presentes en las muestras, con la ventaja que proporcionan información detallada sobre la problemática abordada. Por otro lado, las técnicas de modelado de clase focalizadas en clasificación no dirigida o rigurosa de una clase son útiles para realizar el control de calidad cuando no se tiene conocimiento de los posibles adulterantes que puedan estar presentes, con la ventaja de necesitar menos muestras para construir los modelos ya que solo la clase de interés es modelada. Cabe destacar que la elección del mejor clasificador depende del enfoque que se le dé al problema analítico estudiado (De Araújo Gomes, Azcarate, Spánik, Khvalbota, Goicochea, 2023).

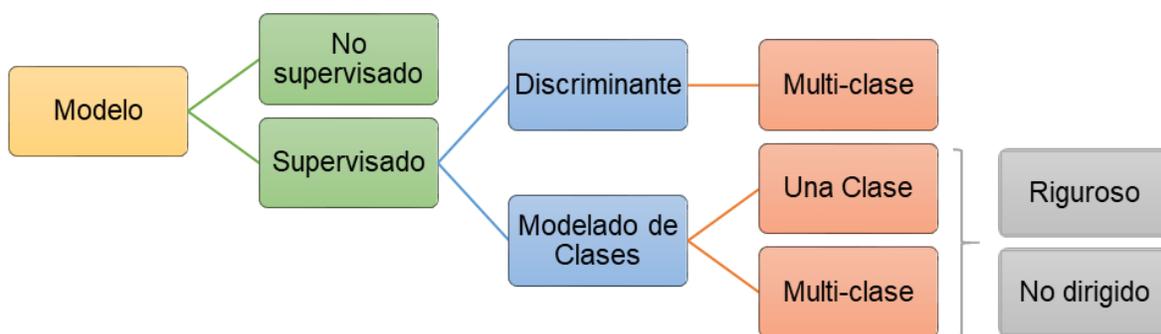


Figura 2.28. Diagrama esquemático de diferentes tipos de técnicas de reconocimiento de patrones.

2.6.4.1. Técnicas de reconocimiento de patrones no supervisadas

Como se mencionó anteriormente las técnicas de reconocimiento de patrón no supervisadas se basan en la agrupación de datos en diferentes grupos o *clusters*, sin tener la información previa del sistema (clases a las que pertenecen las muestras). Estas técnicas

son útiles en situaciones donde se busca patrones y estructuras en los datos que permitan agruparlos de manera natural, con el fin de determinar la estructura existente en el conjunto de datos sin tener información *a priori*.

I. Análisis por Componentes Principales (PCA)

Una de las técnicas de análisis no supervisado más utilizada en el ámbito de la química es el Análisis por Componentes Principales (Principal Components Analysis, PCA). PCA puede ser utilizado con diferentes propósitos: visualización de datos, análisis exploratorio, reducción de datos y como algoritmo en técnicas de clasificación y calibración, entre otros. PCA define las fuentes de variación significativas del conjunto de datos a través de componentes o factores, de modo que la información relevante quede representada con el menor número de dimensiones posibles conocidas como componentes principales (PCs). Las componentes principales (PCs) se obtienen como combinación lineal de las variables originales según el criterio de máxima varianza, de manera que se reduce la dimensionalidad de un conjunto de datos. La primera componente principal explica el mayor porcentaje de varianza y cada componente subsiguiente describe una cantidad decreciente de varianza no contemplada por las componentes anteriores (son ortogonales entre sí). Matemáticamente, PCA descompone la matriz de datos original según la siguiente ecuación:

$$X = TP + E$$

Donde X es la matriz de datos originales que contiene “n” filas (muestras) y “p” columnas (variables, por ejemplo, lecturas de absorbancia); T es la matriz de *scores*, del español “puntuaciones”, con “n” filas y “d” columnas (donde d es el número de componentes principales), P es la matriz de *loadings*, del español “cargas”, con “p” filas y “d” columnas y E es la matriz de los residuales con “n” filas y “p” columnas (Figura 2.29).

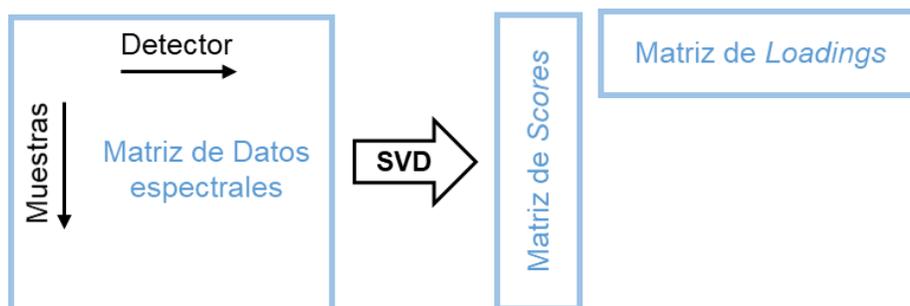


Figura 2.29. Matriz de *scores* y *loadings* obtenidas a partir de la matriz de datos originales.

Cada columna de la matriz de puntuaciones es una variable transformada que se obtiene multiplicando el peso o influencia que cada variable tiene en el modelo PCA (matriz de *loadings*) por cada variable original. La matriz de *scores* contiene información sobre las muestras que se describen en términos de sus proyecciones sobre las componentes principales (PCs), mientras que la matriz de *loadings* contiene información sobre las variables, e indica cuales son las más importantes para describir la variación en los datos originales (Vandeginste et al., 1998; Saccenti y Camacho, 2020).

PCA es muy útil para obtener una idea inicial sobre el comportamiento inherente de los datos analizados, en un espacio definido por unas pocas componentes principales antes de aplicar una técnica de clasificación supervisada. Por otro lado, mediante PCA se puede detectar y eliminar “*outliers*” (valores atípicos), es decir, aquellos que en principio son de naturaleza dudosa (Brereton, 2009).

2.6.4.2. Técnicas de reconocimiento de patrones supervisadas

Las técnicas de clasificación supervisadas requieren de un conjunto de muestras de entrenamiento cuya clase *a priori* es conocida, para luego poder establecer una regla de clasificación con el fin de asignar muestras desconocidas a las clases definidas previamente.

En esta tesis se utilizaron técnicas supervisadas de diferente naturaleza, con objetivos diferenciados: dos trabajos se centran en clasificación dirigida entre las diferentes clases establecidas, empleando técnicas discriminantes y de modelado de clase. Por otro lado, un tercer trabajo se basa en clasificación no dirigida mediante modelado de clases (técnicas *one-class*).

La siguiente **Tabla 2.6** enuncia las técnicas de clasificación utilizadas en esta tesis:

Discriminantes	Modelado de Clases
<i>K</i> -Vecinos Más Cercanos (<i>K</i> -Nearest Neighbours, K-NN)	Modelado Suave e Independiente por Analogía de Clases (Soft Independent Modelling of Class Analogy, SIMCA)
Análisis Discriminante de Mínimos Cuadrados Parciales (Partial Least Squares-Discriminant Analysis, PLS-DA)	
Máquinas de Soporte Vectorial (Support Vector Machine, SVM).	Mínimos Cuadrados Parciales de Una Clase (One Class - Partial Least Squares, OC-PLS)
Árboles de decisión (Decision Trees, DTs)	Modelado Suave e Independiente por Analogía de Clases Impulsado por Datos (Data-Driven-SIMCA, DD-SIMCA)
Árboles random (Random Trees, RTs)	
Regresión logística (Logistic Regression, RL)	

I. *K*-Vecinos más Cercanos (*K*-NN)

K-NN (*K*-Nearest Neighbours) es una de las técnicas más simples de clasificación supervisada no paramétrica. Esta técnica se basa en calcular la distancia entre cada muestra y el resto de muestras existentes, cuya filosofía radica en asignar una muestra a la clase cuyas “*K*” muestras vecinas pertenecen, las cuales están posicionadas a la distancia más corta en el espacio multidimensional. El primer paso consiste en determinar el número de vecinos a utilizar, el cual se puede optimizar seleccionando el valor de “*K*” que proporcione el menor error de clasificación. A continuación, para clasificar una muestra desconocida, *K*-NN selecciona los vecinos más cercanos a esta muestra y se aplica la regla de mayoría: la muestra desconocida se clasifica a la clase a la cual pertenecen la mayoría de las muestras vecinas, basada en la mínima distancia. Esta situación se representa en la Figura 2.30, donde se utiliza $K=3$.

Existen varias formas de determinar la distancia entre muestras, siendo las más utilizadas la distancia Euclídea y la Manhattan. La primera, se calcula como la raíz cuadrada de la suma de las diferencias al cuadrado entre las coordenadas de dos puntos. En cambio, la segunda, se calcula como la suma de las diferencias absolutas entre las coordenadas de dos puntos.

Distancia Euclidiana Distancia Manhattan

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2} \qquad \sum_{i=1}^k |x_i - y_i|$$

Donde “ x_i ” e “ y_i ” es la ubicación de los puntos en las coordenadas.

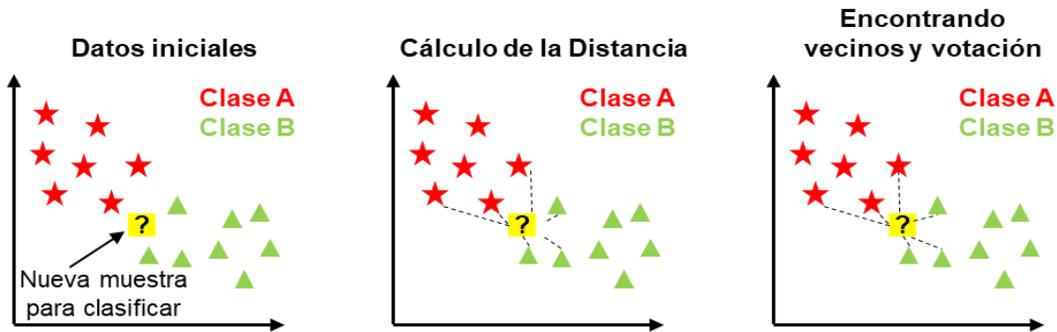


Figura 2.30. Utilizando un $K=3$, la muestra incógnita se asignaría a la clase B (triángulo verde) por cantidad de vecinos cercanos.

K -NN presenta la ventaja de ser conceptual y matemáticamente simple y puede clasificar correctamente objetos, aunque las clases no sean lineales o cuadráticamente separables, ya que establece fronteras irregulares entre las clases (Otto M., 2017). En contrapartida, presenta varias desventajas donde se destaca la sensibilidad a datos desbalanceados. En tal caso, si el conjunto de entrenamiento tiene una distribución desigual entre clases, los resultados pueden sesgar a las clases mayoritarias recibiendo menos atención las minoritarias y teniendo un impacto en las decisiones de clasificación.

II. Análisis Discriminante de Mínimos Cuadrados Parciales (PLS-DA).

PLS-DA (Partial Least Squares-Discriminant Analysis) es una extensión de la técnica de regresión de PLS, que se aplica cuando se tienen variables de respuestas categóricas (clasificación) en lugar de variables continuas (regresión). PLS-DA establece un modelo basado en factores o variables latentes (LVs), que son análogas a las componentes principales (PCs) en PCA. Matemáticamente, PLD-DA funciona de la siguiente manera:

Para la construcción del modelo de clasificación, el primer paso consiste en utilizar un número adecuado de factores (LVs). La matriz de datos originales X y el bloque Y de clases es descompuesta de acuerdo a las siguientes ecuaciones:

$$X = TP^T + E$$

$$Y = UQ + F$$

Donde “T” y “U” son las matrices *scores*, y “P y Q” son las correspondientes matrices *loadings*. El producto TP se relaciona con los datos experimentales, y el error está contenido en las correspondientes matrices residuales “E y F”.

A continuación, se construye una relación interna que vincula “T y U”, y una vez calculado el modelo PLS se obtiene un vector “b” con los coeficientes de regresión del modelo. La predicción para una muestra “x” puede ser obtenida de acuerdo a:

$$y = x^T b$$

Donde “y” es la predicción de los valores dependientes para muestras desconocidas, en donde las variables independientes “x” fueron medidas (Marini, 2013). PLS-DA se focaliza en encontrar una relación lineal entre un conjunto de variables independientes (señal analítica) y variables de respuesta categórica (clases). En este sentido, se crea un modelo donde se maximiza la correlación entre X (datos originales o respuestas instrumentales) e Y (clases a las que pertenecen las muestras) mediante LVs, donde la identificación de pertenencia a las clases se realiza a través de una matriz de variables ficticias (bloque Y) que consiste en 1 y 0, donde 0 es no pertenencia a la clase y 1 corresponde a pertenencia a la clase, por lo que es necesario establecer un valor umbral entre ambos (Tabla 2.7). Estos valores umbrales se basan en el teorema de Bayes, el cual considera una distribución normal de los valores predichos de las muestras tras aplicar el modelo, y que se estima usando la media y desviación estándar de los valores predichos para cada clase (Brereton, 2009; Brereton y Lloyd, 2014). Estos límites se calculan de manera de obtener la menor tasa de resultados falsos (falsos positivos y negativos). El modelo PLS-DA construido se utiliza para predecir nuevas muestras, en donde a cada clase pertenecen las muestras cuyos valores de Y estén por encima del valor delimitado (umbral) correspondiente a esa clase.

Tabla 2.7. Codificación de la matriz Y en PLS-DA.

		Bloque Y				
		1	2	3	4	5
Clase 1		1	0	0	0	0
Clase 2		0	1	0	0	0
Clase 3	Datos espectrales	0	0	1	0	0
Clase 4		0	0	0	1	0
Clase 5		0	0	0	0	1

III. Árboles de decisión (DTs) y árboles aleatorios (RTs)

Los DTs (Decision Trees) son modelos de aprendizaje automático no paramétricos que permiten la construcción de modelos de clasificación y predicción en función de las características estudiadas del sistema. El árbol de decisión es una estructura que está formada por ramas y nodos de distintos tipos:

- Nodos internos: representan cada una de las características a considerar para tomar una decisión
- Ramas: representan la decisión en función de una determinada condición (por ejemplo, probabilidad de ocurrencia)
- Nodos finales: representan el resultado de la decisión.

El DT se construye en base a una serie de decisiones lógicas similares a un diagrama de flujo, con nodos de decisión que indican una decisión sobre una característica. Estos se dividen en ramas que indican las elecciones de la decisión. El árbol termina con nodos finales que denotan el resultado de seguir una combinación de decisiones. La división o partición se va realizando de manera recursiva buscando un mínimo local de acuerdo a medidas como la entropía de los datos (Figura 2.31).

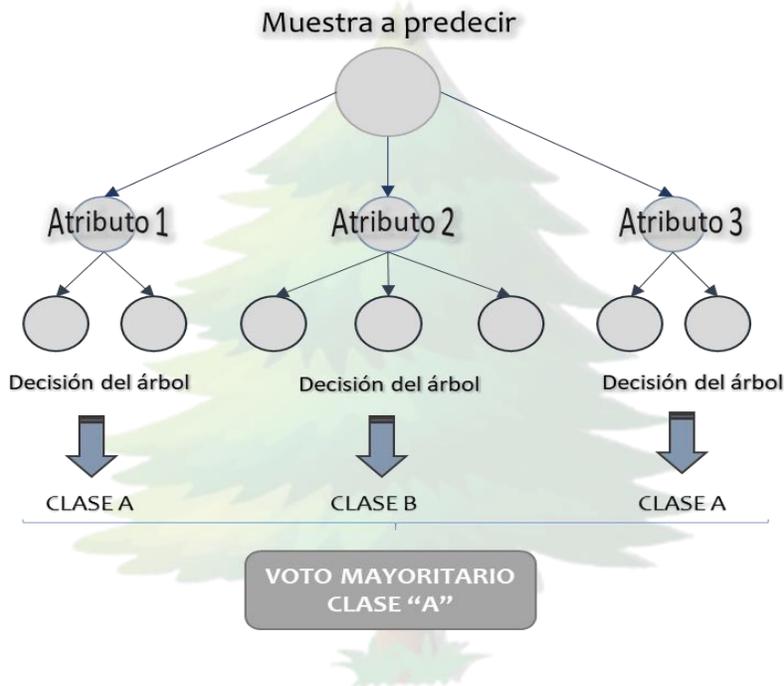


Figura 2.31. Funcionamiento de un árbol de decisión.

Por otro lado, los árboles aleatorios, también conocidos como bosques aleatorios o del inglés, *random forests*, son una extensión de los árboles de decisión. En lugar de construir un único árbol, los árboles aleatorios generan múltiples árboles de decisión de forma aleatoria. Cada árbol se construye utilizando una muestra aleatoria con reemplazo del conjunto de datos de entrenamiento, lo que se conoce como "*bootstrapping*". La aleatorización adicional se produce durante la construcción de cada nodo del árbol. En lugar de considerar todas las variables para la división, se selecciona un subconjunto aleatorio de ellas, con lo que se evita que los árboles estén altamente correlacionados y permite que cada árbol aprenda de diferentes características relevantes. Una vez que se construyen los árboles aleatorios, la clasificación de una muestra se realiza mediante la votación de todos los árboles individuales. Cada árbol emite una predicción y la clase más votada se selecciona como predicción final del bosque aleatorio (Figura 2.32) (Breiman, 1996; Goel y Abhilasha, 2017).

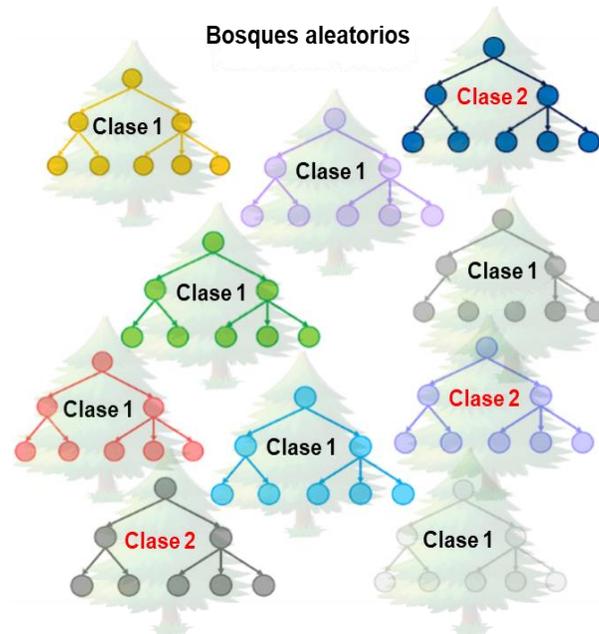


Figura 2.32. Bosques aleatorios.

En resumen, los árboles de decisión se basan en la construcción de árboles generados a través de la división de los datos en función de los atributos, mientras que los árboles aleatorios agregan aleatorización en la construcción de múltiples árboles y en la selección de atributos para mejorar la precisión y reducir la correlación entre los modelos individuales (Geurts, Ernst, y Wehenkel, 2006; Saravanan y Sujatha, 2018).

IV. Máquinas de Soporte Vectorial (SVM)

El objetivo de las SVM (Support Vector Machines) es construir un hiperplano que maximice la separación entre las diferentes clases en estudio. Un hiperplano es una superficie que divide el espacio multidimensional en tantas regiones como clases en estudio. Idealmente se busca encontrar el hiperplano que tenga la mayor distancia posible a las muestras más cercanas de cada clase, lo que se conoce como el margen máximo. Sin embargo, en muchos casos los datos no son linealmente separables, es decir, no se pueden separar perfectamente con un solo hiperplano. Para resolver este problema, SVM utiliza una técnica llamada "*kernel trick*". El truco del kernel permite transformar los datos de entrada a un espacio de mayor dimensión donde sí sean linealmente separables. Esto se logra mediante la aplicación de una función kernel que mapea los datos originales a un nuevo espacio. SVM es especialmente útil cuando se trabaja con conjuntos de datos complejos y no linealmente separables.

Los vectores soporte son puntos de datos que se encuentran cercanos a este hiperplano e influyen su posición y orientación y son utilizados para maximizar el margen del clasificador, logrando el hiperplano separador de clases óptimo (Figura 2.33). Este tipo de modelos se pueden aplicar a datos de cualquier tipo mediante funciones no lineales (como polinómicas, hiperbólicas o gaussianas). Estas funciones representan los datos en un espacio con un mayor número de dimensiones que el espacio de datos original, en el cual los datos pueden ser representados como linealmente separables (Otto, 2017; Kotsiantis, 2007; Trentanni Hansen et al., 2019; Wu et al., 2008).

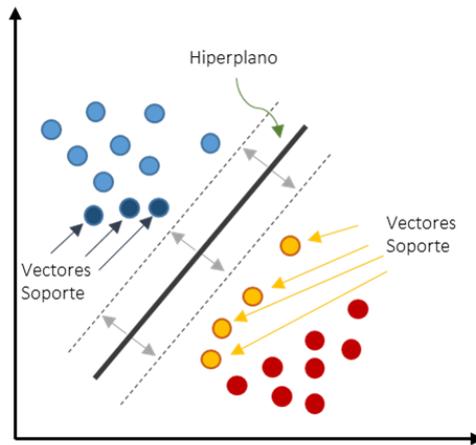


Figura 2.33. Algoritmo de Máquinas de Soporte Vectorial.

V. Regresión logística (LR)

Aunque su nombre incluye la palabra "regresión", la LR (Logistic Regression) se utiliza para problemas de clasificación, no de regresión. El nombre se debe a que utiliza una función logística (también conocida como sigmoide) para realizar la transformación de una combinación lineal de características en una probabilidad. La idea básica, detrás de la regresión logística, es modelar la relación entre las características o variables independientes y la probabilidad de pertenecer a una clase específica. A diferencia de la regresión lineal, que produce una salida continua, la regresión logística produce una salida discreta en forma de una probabilidad entre 0 y 1, utilizando una sigmoidea para transformar una combinación lineal de características en probabilidades (Figura 2.34). Durante el entrenamiento, se ajustan los coeficientes para maximizar la probabilidad de los datos de entrenamiento y durante la predicción se utiliza un umbral para tomar la decisión final de clasificación. La regresión logística también se puede extender a problemas de clasificación multiclase utilizando técnicas como la regresión logística multinomial o la regresión logística uno vs el resto (*one-vs-rest*).

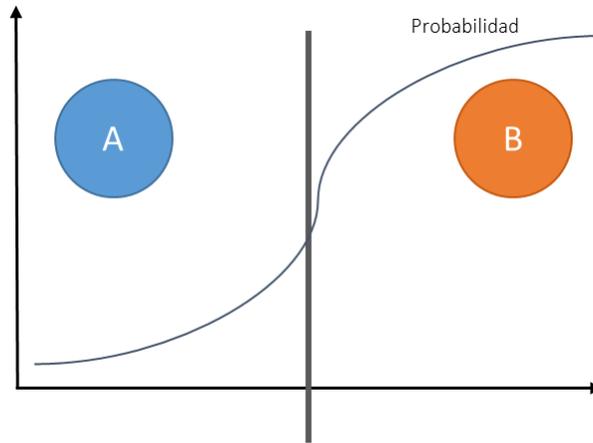


Figura 2.34. Función sigmoidea al aplicar RL para clasificación binaria.

VI. Modelado Suave e Independiente por Analogías de Clases (SIMCA).

A diferencia de los métodos discriminantes, la pertenencia de muestras a una determinada clase puede también determinarse mediante la descripción de las clases individuales, a través de un modelado separado e independiente de las clases restantes. SIMCA (Soft Independent Modelling of Class Analogy) se basa en el uso de PCA para modelar cada clase de forma independiente. El primer paso consiste en generar un modelo para cada una de las clases, los cuales quedan descritos con un determinado número de componentes principales (PCs), como se puede observar en la Figura 2.35, donde cada clase puede ser modelada por un número diferente de PCs. Es decir, el modelo captura la variabilidad intrínseca de cada clase y se utiliza como referencia para la clasificación de futuras muestras. Para asignar una muestra desconocida a una determinada clase, se realiza una comparación estadística mediante un test F: la varianza residual de cada muestra con la varianza residual total de cada una de las clases. Además, se verifica si la muestra desconocida cae dentro de los límites establecidos para dicha clase. SIMCA ofrece tres posibilidades de resultados de clasificación: muestras pertenecientes a una determinada clase, muestras pertenecientes a más de una clase y muestras no pertenecientes a ninguna de las clases existentes (Otto, 2017).

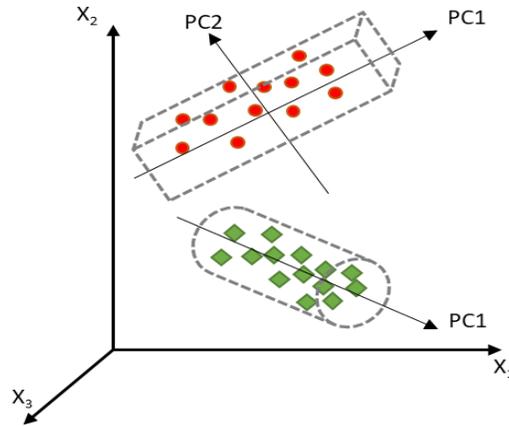


Figura 2.35. Modelo de SIMCA para dos clases con diferentes números de PCs.

Técnicas de Clasificación aplicadas al modelado de una clase (One-class)

En el análisis multivariante, las técnicas de clasificación "one-class" se focalizan en modelar la clase de interés, independientemente de que existan otras clases. En este sentido, estas técnicas apuntan a capturar las características descriptivas de una clase objetivo y en detectar si nuevas muestras se asemejan a esta clase o no. Por ejemplo, al abordar una problemática de adulteración de alimentos, podría interesar modelar la clase de muestras sin adulterar o auténticas, aunque existan adulterantes contenidos en otras clases adulteradas. La principal diferencia entre técnicas de una clase y análisis discriminante es que en las primeras se construye un modelo solo con las muestras en la clase de interés y se centran en encontrar similitudes entre las mismas, mientras que las discriminantes se basan en la distinción entre múltiples clases. En el contexto de la autenticidad, estas técnicas pueden aplicarse para detectar elementos que difieran significativamente de la norma o del comportamiento esperado (Khan, y Madden, 2014).

En esta tesis se utilizaron los modelos DD-SIMCA y OC-PLS con fines de clasificación *one-class* para verificar la autenticidad de los quesos rallados de pasta dura.

I. Modelado Suave e Independiente por Analogías por Clases Impulsado por Datos (DD- SIMCA).

La técnica DD-SIMCA (Data-driven Soft Independent Modelling of Class Analogy) construye una regla de decisión que delinea los objetos de la clase de interés a través de la exploración de la correspondiente matriz de datos. El primer paso es aplicar PCA, para capturar las características más relevantes y establecer el área de aceptación basada en cálculos de distancias (*scores* vs ortogonales) que se denomina gráfico de aceptación,

definido por un determinado valor alfa (α) (nivel de significancia), el cual especifica errores tipo I (tasa de falsos negativos). Cada muestra del set de entrenamiento se caracteriza por su posición en el gráfico de aceptación y tiene el estatus de muestra regular (perteneciente a la clase de interés) o muestra extrema (localizada fuera del área de aceptación y se considera muestra extraña o anómala). A continuación se establece un segundo valor límite gamma (γ) utilizado para construir una frontera de anómalos. Este valor establece la probabilidad de que al menos un valor regular del conjunto de datos sea erróneamente considerado como un valor anómalo. Los valores de gamma oscilan entre 0,05 y 0,01.

Finalmente, para evaluar muestras no contempladas durante el modelado de clase, se calculan los errores tipo II (tasa de falsos positivos), para evaluar si estas muestras pertenecen o no a la clase de interés (Oxana, Radionova, Oliveri, y Alexey, 2016).

La Figura 2.36 es una representación gráfica del modelo DD-SIMCA y su correspondiente región de aceptación.

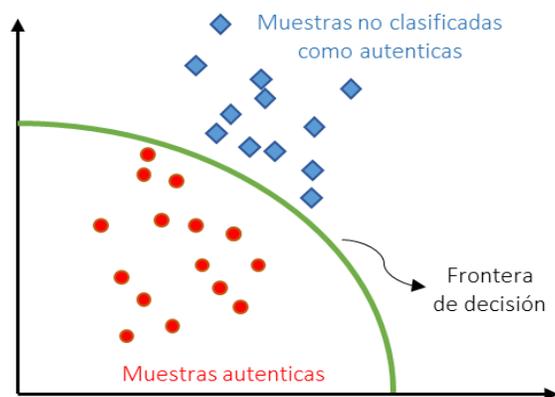


Figura 2.36. Representación gráfica DD-SIMCA.

II. Mínimos Cuadrados Parciales de Una Clase (OC-PLS)

En OC-PLS (One Class - Partial Least Squares), se ajusta un modelo PLS a una metodología de una clase a través de la correlación de los espectros o señal analítica con un vector compuesto de valores igual a 1. Esta técnica considera simultáneamente la varianza explicada y las correlaciones de la clase objetivo. La asignación de muestras se realiza a través de dos parámetros: la distancia de Hotelling T^2 basadas en distancias *score* (SDs) y los valores absolutos de los residuales centrados (ACR). Se asume que los valores ACR siguen una distribución normal con una media de 0. Los valores SDs son una medida de la distancia de un objeto al centro de la clase en un espacio descrito por variables latentes (LVs), mientras que los valores ACR se consideran como medida de dispersión de

la proyección en el vector de coeficientes de regresión construidos en el modelo OC-PLS. Valores elevados de SDs o ACR indican que un objeto está alejado del comportamiento característico de la clase de interés. Existen cuatro posibilidades de asignación: muestras regulares que tienen valores bajos de SDs y ACR, puntos satisfactorios de influencia que poseen altos valores de SDs y bajos de ACR, anómalos de clase que poseen bajos valores de SDs y elevados de ACR y puntos no satisfactorios de influencia que poseen valores altos de de SDs y ACR. Los puntos satisfactorios y no satisfactorios de influencia, así como los anormales de clase pueden ser detectados como diferentes tipos de valores anómalos globales (Xu et al., 2014).

2.6.4.3. Parámetros de calidad en clasificación multivariante

La validación de un modelo se considera el proceso para asegurar que un procedimiento analítico sea fiable y pueda satisfacer las expectativas de una aplicación en particular. En resumen, significa que el modelo se puede utilizar con confianza. El desarrollo y la validación de un método están estrechamente relacionados, ya que los parámetros de calidad usualmente se evalúan como parte del desarrollo del método. Al abordar un problema analítico, se debe tener en cuenta ciertas consideraciones como se esquematizan en la Figura 2.37.



Figura 2.37. Validación y Evaluación de la habilidad de las metodologías desarrolladas (López, Callao y Ruisánchez, 2015).

La solución del problema analítico debe considerarse como un proceso cíclico e interactivo de verificación y evaluación del método, que no se detiene hasta que el método

se considere capaz de satisfacer los requisitos. El proceso comienza con el estudio del problema, lo que se sabe sobre él y cuáles son los requisitos analíticos, de manera de elegir el método que mejor responda, y si se ajusta al propósito, éste debe ser validado mediante la evaluación de los parámetros de calidad del modelo estudiado.

La habilidad de las metodologías desarrolladas se evalúa en términos de los parámetros evaluadores de la calidad de los modelos de clasificación construidos en cuyo caso se pueden obtener asignaciones correctas y errores de clasificación. Las asignaciones correctas comprenden los **aciertos positivos (true positives, TP)** que son las muestras clasificadas correctamente a su clase y los **aciertos negativos (true negatives, TN)** que son muestras pertenecientes a otra clase e identificadas como externas correctamente. En cambio, dentro de los errores de clasificación tenemos los **falsos negativos (False negative, FN)** que son muestras de una clase que no son reconocidas como tal y **falsos positivos (False positive, FP)** que son muestras de otra clase incorrectamente identificadas como pertenecientes a la clase. Los parámetros de calidad son un conjunto de atributos medibles que definen la calidad de un método analítico, que, dependiendo de la naturaleza del problema, del método propuesto y del nivel de validación, pueden ser los siguientes: sensibilidad, especificidad, precisión, eficiencia del modelo, coeficiente de correlación de Mathews, curvas ROC (Receiver Operating Characteristic Curves), área bajo la curva (AUC) (Oliveri y Downey, 2012).

La **sensibilidad** es la fracción de todas las muestras objetivo que se clasifican correctamente como muestras objetivo y se representa matemáticamente como:

$$\text{Sensibilidad} = 100 \times \frac{(TP)}{(TP + FN)}$$

TP: Verdadero Positivo; FN: Falso negativo.

La **especificidad** representa la proporción de todas las muestras no objetivo que se clasificaron correctamente como muestras no objetivo y puede representarse matemáticamente como:

$$\text{Especificidad: } 100 \times \frac{(TN)}{(TN + FP)}$$

TN: Verdadero negativo; FP: falso positivo.

La **precisión** indica la relación entre el número de muestras correctamente aceptadas y el número total de muestras aceptadas por una clase.

$$\text{Precisión: } 100 \times \frac{(TP)}{(TP + FP)}$$

FP: falso positivo; TP: Verdadero Positivo

La **eficiencia** corresponde a la capacidad global de clasificación y/o predicción y está directamente relacionada con los parámetros de sensibilidad y especificidad.

$$\text{Eficiencia: } \frac{(TN + TP)}{(TN + FP + TP + FN)}$$

TN: Verdadero negativo y FP: falso positivo. TP: Verdadero Positivo y FN: Falso negativo.

El **F1-Score** es una métrica comúnmente utilizada en problemas de clasificación para evaluar el rendimiento de un modelo. Es una medida que combina la precisión y el recall (recuperación) de un modelo en un solo valor.

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

El *Recall* (recuperación) es la proporción de predicciones correctas positivas respecto a todas las instancias que realmente pertenecen a la clase positiva (verdaderos positivos más falsos negativos).

Curvas ROC: Una clase de parámetros comúnmente utilizados para evaluar la calidad de un modelo de clasificación son las curvas de características operativas del receptor (ROC). Una curva ROC es una representación gráfica de la sensibilidad frente a la tasa de falsos positivos (Breretón, 2009).

Las curvas ROC se obtienen graficando la sensibilidad vs. (1 - especificidad) variando alguno de los parámetros de clasificación (límite establecido entre clases, nivel de confianza, etc. que dependerá de la técnica de clasificación utilizada). Como se muestra en la Figura 2.38, una línea diagonal desde el extremo inferior izquierdo hasta el extremo superior derecho representa modelos que proporcionan asignación de muestras al azar (i) (modelos pobres), mientras que curvas que se ubican por encima de esta línea y tienden hacia el extremo superior izquierdo indican modelos eficientes (iii), pudiendo haber casos

intermedios (ii). Estas curvas pueden ser utilizadas para comparar diferentes modelos y para identificar el óptimo compromiso entre sensibilidad y especificidad para un determinado modelo.

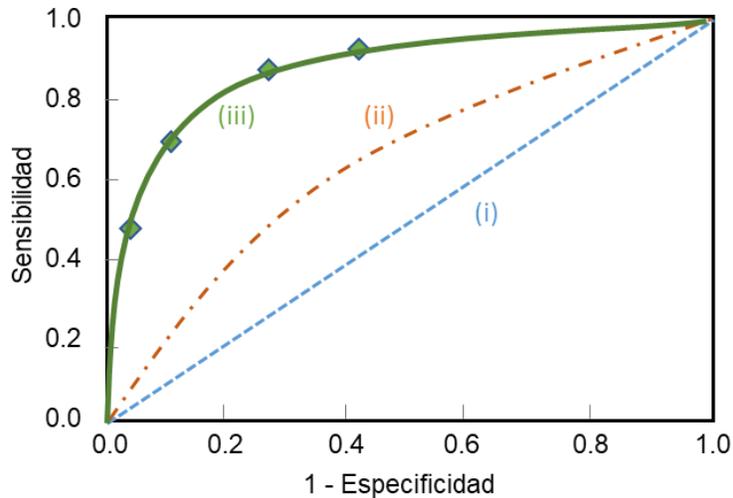


Figura 2.38. Curvas ROC.

2.6.5. Validación cruzada

Para construir un modelo de clasificación, lo ideal es dividir al conjunto de datos en conjunto de entrenamiento (70% muestras) y conjunto de prueba (30% muestras). En este caso, se establece el modelo de clasificación y se optimizan los parámetros propios de cada técnica de clasificación (número de factores, componentes principales o variables latentes, número de vecinos más cercanos, etc.) a través de la medida de la tasa del error de clasificación. Una vez construido y optimizado el modelo, se utiliza el conjunto de prueba, el cual contiene muestras desconocidas que no se utilizaron para establecer el modelo de clasificación, con el fin de predecir la clase a la que pertenecen dichas muestras desconocidas.

A veces no se cuenta con la cantidad suficiente de muestras para dividir a los datos como se mencionó anteriormente, ya sea porque el análisis es costoso, la obtención de datos resulta complicada experimentalmente o porque no es posible armar un conjunto de datos representativo del problema en cuestión. En estos casos, existen estrategias de validación cruzada que permiten dividir al conjunto de datos en entrenamiento y prueba sistemáticamente. Este proceso se denomina Validación cruzada o *Cross-validation*. La Figura 2.39 muestra diferentes formas de realizar la división de los datos al utilizar estrategias de validación cruzada.

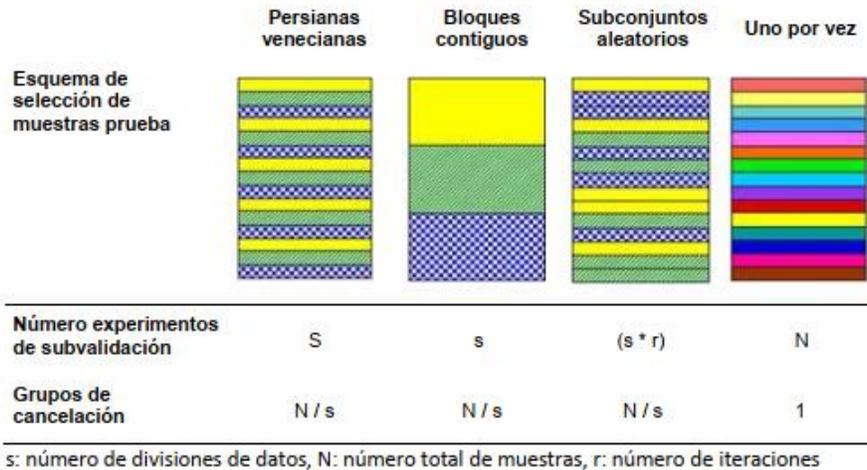


Figura 2.39. Esquemas de validación cruzada.

En términos generales la validación cruzada incluye los siguientes pasos:

1. Extraer un subconjunto de datos (muestras de prueba).
2. Utilizar las muestras restantes para construir el modelo de clasificación.
3. Aplicar el modelo construido para predecir las muestras extraídas en el paso 1.

El conjunto de calibración inicial se divide en diferentes grupos de cancelación. Cada vez que se calcula el modelo se deja un grupo de cancelación fuera y se calcula, un modelo con los demás datos. Una vez realizada la calibración se prueba el modelo con los datos del grupo de cancelación excluido. Se repite el proceso de manera sucesiva dejando cada vez un grupo de cancelación fuera del proceso de cálculo. Si el número de grupos de cancelación es igual al número de objetos en el conjunto de calibración, cada vez quedará un objeto fuera del cálculo del modelo y el proceso se denomina *full-cross-validation* o *leave-one-out*.

2.7. Referencias

- AOAC, Association of Official Analytical Chemists. (1995). Fibra dietética total, soluble e insoluble en alimentos, método n°. 991.43. *Métodos oficiales de análisis de AOAC Internacional (16ª ed.)*. Gaithersburg, MD. Estados Unidos: AOAC Internacional.
- AOAC, Association of Official Analytical Chemists. (2007). <https://www.aoac.org/wp-content/uploads/2019/09/ALACC-method-verification.pdf> S. Brunelle, R. Labudde, M. Nelson, P. Wehling, AOAC INTERNATIONAL. (Último acceso julio 2023).
- Barreto A., Cruz-Tirado J. P., Siche R., y Quevedo R. (2018). Determination of starch content in adulterated fresh cheese using hyperspectral imaging. *Food Bioscience*, 21, 14-19.
- Beltrame, K. K., Gonçalves, T. R., Gomes, S. T. M., Matsushita, M., Rutledge, D. N., Março, P. H., et al. (2021). Digital images and independent components analysis in the determination of bioactive compounds from grape juice. *LWT e Food Science and Technology*, 152, 112308.
- Botelho B.G., De Assis L.P., y Sena M.M. (2014). Development and analytical validation of a simple multivariate calibration method using digital scanner images for sunset yellow determination in soft beverages. *Food Chemistry*, 159, 175-180.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123-140.
- Brereton, R.G. (2009). Chemometrics for pattern recognition (1st ed.). *Chichester: John Wiley & Sons*. Centre for Chemometrics, School of Chemistry, Bristol University, Bristol, UK.
- Brereton R.G., y Lloyd R.G. (2014). Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometric* 28, pág. 213-225.
- CAA, Código Alimentario Argentino. (2023). Capítulo VIII, Artículos: 553 al 642 - Alimentos Lácteos. Disponible en <https://www.argentina.gob.ar/anmat/codigoalimentario> (Último acceso julio 2023).
- Capitán-Vallvey L.F., López-Ruiz N., Martínez-Olmos A., Erenas M.M., y Palma A.J. (2015). Recent developments in computer vision-based analytical chemistry: A tutorial review. *Analytica Chimica Acta*, 899, 23-56.
- Céliz F., Forneris I., Severini H., y Filippin A. (2018). Diseño de una planta de generación de carbón activado obtenido a partir de carozos de aceitunas mediante activación térmica. *Facultad de Ciencias Exactas Físicas y Naturales*. <https://rdu.unc.edu.ar/bitstream/handle/11086/6459/PI%20C%C3%A9liz-Forneris.pdf?isAllowed=y&sequence=1> (Último acceso julio 2023).

- Cheese reporter. (2016). Anti-Caking Suppliers Clear Up Inaccuracies Regarding Appropriate Levels of Cellulose, *Cheese Reporter*, 04-01-2016, Page 19. <https://npaper-wehaa.com/cheese-reporter/2016/04/?article=2731489&output=html> (Último acceso julio 2023).
- Codex Alimentarius. (2018). Documento de debate sobre la integridad y la autenticidad de los alimentos CX/FICS 18/24/7. Programa conjunto FAO/OMS sobre normas alimentarias. https://www.fao.org/fao-who-codexalimentarius/sh%20proxy/en/?Ink=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252FMeetings%252FCX-733-24%252FWorking%2BDocuments%252Ffc24_07s.pdf (Último acceso julio 2023).
- Codex Alimentarius. (2021). Norma general para los aditivos alimentarios: Codex Stan 192-1995 (Adopted in 1995, amended in 2021). *Codex Alimentarius Commission, joint FAO/WHO food standards program*. <https://www.fao.org/fao-who-codexalimentarius/codex-texts/dbs/gsfa/es/> (Último acceso julio 2023).
- Codex Alimentarius (2021). Norma de grupo para el queso no madurado, incluido el queso fresco. Adoptada en 2001. Enmendada en 2008, 2010, 2013, 2018, 2021. https://www.fao.org/fao-who-codexalimentarius/sh-proxy/en/?Ink=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252FStandards%252FCXS%2B221-2001%252FCXS_221s.pdf (Último julio 2023).
- Cozzolino D. (2015). The role of vibrational spectroscopy as a tool to assess economically motivated fraud and counterfeit issues in agricultural products and foods. *Analytical Methods*, 7, 9390-9400.
- Cuibus L., Maggio R., Muresan V., Diaconeasa Z., Fetea F., y Socaciu C. (2014). Preliminary discrimination of cheese adulteration by FT-IR spectroscopy. *Food Science and Technology*, 71, 142-147.
- Da Silva Medeiros M. L., Lima Y., Cruz-Tirado L. J. P., Freitas Lima A., Teixeira Godoy H., Fernandes Barbin D. (2023). Portable NIR spectrometer and chemometric tools for predicting quality attributes and adulteration levels in butteroil, *Food Control*, 144, 109349.
- De Araújo Gomes A., Azcarate S. M., Spanik I., Khvalbota L., Goicoechea H. C. (2023). Pattern recognition techniques in food quality and authenticity: A guide on how to process multivariate data in food analysis. *Trends in Analytical Chemistry*, 164, 117105.

- Di Donato F., Biancolillo A., Ferretti A., D'Archivio A.A., y Marini F. (2023). Near infrared spectroscopy coupled to chemometrics for the authentication of donkey milk. *Journal of Food Composition and Analysis*, 115, 105017.
- Diniz P.H.G.D. (2020). Chemometrics-assisted color histogram-based analytical systems. *Journal of Chemometrics*, 34, 3242.
- Diniz P.H.G.D., Dantas H.V., Melo K.D.T., Barbosa M.F., Harding D.P., Nascimento E.C.L., Pistonesi M.F., Band B.S.F., y Araújo M.C.U. (2012). Using a simple digital camera and SPA-LDA modeling to screen teas. *Analytical Methods*, 4, 2648-2652.
- Domínguez M.A., Diniz P.H.G.D., Di Nezio M.S., Araújo M.C.U., Centurión M.E. (2014). Geographical origin classification of Argentinean honeys using a digital image-based flow-batch system. *Microchemical Journal*, 112, 104-108.
- EFSA, E.U., Autoridad Europea de Seguridad Alimentaria. (2011). Commission regulation 1129/2011. Official Journal of European Union. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011R1129&from=ENL295/1-177> (Último acceso julio 2023).
- Ellis D. I., Brewster V.L., Dunn W. B., Allwood J. W., Golovanov A. P., y Goodacre R. (2012). *Chemical Society Reviews*, 41, 5706-5727.
- Everstine K, Spink J, y Kennedy S. (2013). Economically motivated adulteration (EMA) of food: common characteristics of EMA incidents. *Journal of Food Protection*, 76, 723-35.
- FDA, Food and Drug Administration (2017). Foods Program Guidelines for Chemical Methods e Guidelines for the Validation of Chemical Methods for the FDA Foods Program. <https://www.fda.gov/science-research/field-science-and-laboratories/method-validation-guidelines> (Último acceso julio 2023).
- FDA, Food and Drug Administration. (2018). Food additive status list. Code of Federal Regulations (CFR). Washington, DC, USA: Food and Drug Administration (FDA). <https://www.fda.gov/Food/IngredientsPackagingLabeling/FoodAdditivesIngredients/ucm091048.htm>
- Fearn T., Riccioli C., Garrido-Varo A., y Guerrero-Ginel J. E. (2009). On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems*, 96, 22-26.
- Galvan D., Aparecida L. C., Effting L., Melquiades F. L., Bona E., y Conte-Junior C. A. (2022). Low-cost spectroscopic devices with multivariate analysis applied to milk authenticity, *Microchemical Journal*, 181, 107746.
- Gastalver Robles M. C. (2015). Procesos básicos de elaboración de quesos. EDITORIAL ELEARNING S.L. ISBN: 978-84-16424-76-4.

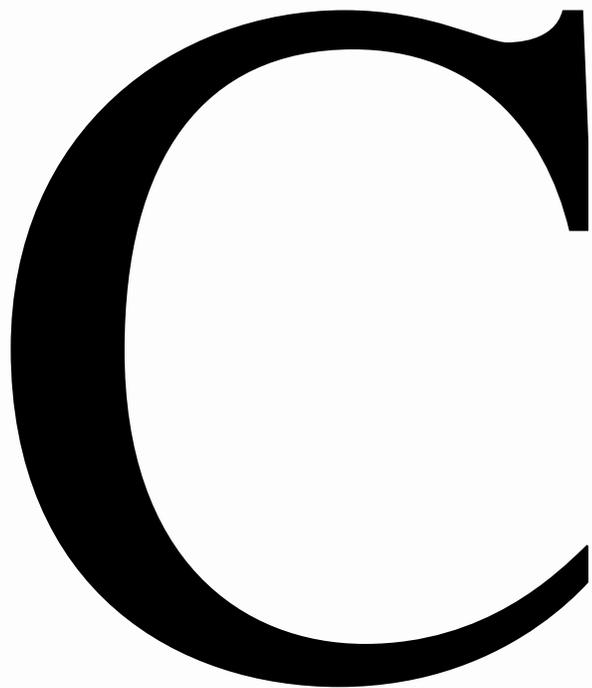
- Geurts P., Ernst D., y Wehenkel L. (2006). Extremely randomized trees. *Mach Learn*, 63, 3-42.
- Goel, E., y Abhilasha, E. (2017). Random forest: A review. *International Journal of Advanced Research in Computer Science and Software Engineering*, 7, 251-257.
- González R.C., y Woods R.E. (2008). Digital Image Processing. 3rd Edition, ed. Pearson Prentice Hall, New Jersey.
- Griffiths P.R. (2017). IR Spectroscopic Data Processing. *Encyclopedia of Spectroscopy and Spectrometry*, Third Edition, 428-436.
- Herrero-Latorre, C., Barciela-García, J., García-Martín, S., y Peña-Crecente, R. M. (2019). Detection and quantification of adulterations in aged wine using RGB digital images combined with multivariate chemometric techniques. *Food Chemistry*, 3, 100046.
- Hurley I. P., Coleman R. C., Ireland H. E., y Williams J. H. (2005). Use of sandwich IgG ELISA for the detection and quantification of adulteration of milk and soft cheese. *International Dairy Journal*, 16, 805-812.
- IUPAC. (1995). L.A. Currie, Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995), *Pure Appl. Chem.* 67, 1699-1723.
- Ioelovich M. J. (2022). Microcellulose Vs Nanocellulose. *A Review World Journal of Advanced Engineering Technology and Sciences*, 5, 01–15.
- Jablonský M., Dubinyova L., Varga S., Vizarova K., Sima J., y Katuscak S. (2015). Cellulose fibre identification through color vectors of stained fibre. *Bioresources*, 10, 5845-5862.
- Juarez-Barrientos, J. M., Tejeda-Paz, M., Ramírez-Rivera, E., Aguirre-Cruz, A., Rodríguez-Miranda, J., Martínez Sanchez, C., et al. (2018). Use of quantitative descriptive analysis (QDA) coupled with multivariate statistical methods to detection and discrimination of adulterated fresh cheeses. *Journal of Sensory Studies*, 34, 1-8.
- Ka M.-H., Lee K.-G., Lim H.-Y., Lee G., Yun S. S., Lim H. S., et al. (2017). Analytical method of silicon dioxide in health functional food products using ICP-OES. *Journal of Food Hygiene and Safety*, 32, 343-347.
- Kapoor R., y Metzger L. E. (2008). Process cheese: Scientific and technological aspects - a review. *Comprehensive Reviews in Food Science and Food Safety*, 7, 194-214.
- Kennard R.W., y Stone L.A. (1969). *Technometrics*, 11, 137-148.
- Khan S., y Madden M. (2014). One-class classification: taxonomy of study and review of techniques. *The Knowledge Engineering Review*, 29, 345–374.

- Kongo J. M., y Malcata F. X. (2016). Cheese: Types of cheeses - soft. In B. Caballero, P. M. Fringlas, & F. Toldra (Eds.), *Encyclopedia of food and health*, 768-773. Oxford, UK: Academic press.
- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatics*, 31, 249-268.
- Lopez M. I., Callao M. P., y Ruisanchez I. (2015). A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. *Analytica Chimica Acta*, 891, 62-72.
- Marini F. (2013). Chemometrics in food analysis. *Handling in science and technology*, 28, 1-498.
- Mark H., y Workman J. (2018). Jr. Chapter 23 - Calculating the Solution for Regression Techniques: Part 3-Partial Least Squares Regression Made Simple. *Chemometrics in Spectroscopy*, 2, 125-128.
- Massart D. L., Vandeginste B.G.M., Buydens S. D.J., Lewi P.J., y Smeyers-Verbeke J. (1997). Handbook of chemometrics and qualimetrics: Part A. Data Handling in science and technology.
- McCleary B. V., DeVries J. W., Rader J. I., Cohen G., Prosky L., Mugford D. C., et al. (2012). Determination of insoluble, soluble, and total dietary fiber (CODEX Definition) by enzymatic-gravimetric method and liquid chromatography: Collaborative study. *Journal of AOAC International*, 95, 824-844.
- Medeiros M.L.S., Lima A.F., Gonçalves M.C., Godoy H.T., Barbin D.F. (2023). Portable near-infrared (NIR) spectrometer and chemometrics for rapid identification of butter cheese adulteration. *Food Chemistry* 425, artículo 136461.
- Muñoz-Olivas R. (2004). Screening analysis: an overview of methods applied to environmental, clinical and food analyses. *Trends in Analytical Chemistry*, 23, 203-216.
- Oliveri P., y Downey G. (2012). Multivariate class modeling for the verification of food-authenticity Claims. *Trends in Analytical Chemistry*, 35, 74-86.
- Otto, M. (2017). Chemometrics, statistics and computer application in analytical chemistry (3rd ed.). Weinheim, Germany: Wiley-VCH.
- Otto, M. (2017). Pattern recognition and classification. In *Chemometrics, statistics and computer application in analytical chemistry*, 135-211. Weinheim, Germany: Wiley VCH.
- Oxana Y. R., Oliveri P., y Pomerantsev A.L., (2016). Rigorous and compliant approaches to one-class classification, *Chemometrics and Intelligent Laboratory Systems*, 159, 89-96.

- Ozturk M., Ali Dogan M., Menevseoglu A., y Ayvaz H. (2022). Infrared spectroscopy combined with chemometrics as a convenient method to detect adulterations in cooking/stretching process in commercial cheese. *International Dairy Journal*, 128, 105312.
- Pasquini C., y Braz J. (2003). Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *Chemical Society*, 14, 198-219.
- Peguero Gutiérrez A. (2010). Tesis Doctoral: La espectroscopía NIR en la determinación de propiedades físicas y composición química de intermedios de producción y productos acabados, Universidad Autónoma de Barcelona. <https://www.tesisenred.net/handle/10803/3316> (Último acceso julio 2023).
- Pérez M.R. (2007). *Focal Encyclopedia of Photography* 4th edition, ed. Focal Press (Elsevier), Oxford.
- Rinnan A., Van Den Berg F., y Engelsen S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28, 1201-1222.
- Saccenti E., y Camacho J. (2020). Multivariate exploratory data analysis using component models. In A. Cifuentes (Ed.), *Comprehensive foodomics*, 241-268. Amsterdam, Netherlands: Elsevier.
- Saravanan R., y Sujatha P. (2018). "A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification" *Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, 945-949.
- Skoog D.A., Holler F.J., y Nieman T.A. (2001). "Principios de análisis instrumental", 5ª ed., McGraw- Hill.
- Skoog, D. A., Donald M. West, F. James Holler y Stanley R. Crouch. (2014). Fundamentos de química analítica. Novena edición. ISBN: 978-607-519-937-6
- Sørensen K. M., Khakimov B., Engelsen S. B. (2016). The use of rapid spectroscopic screening methods to detect adulteration of food raw materials and ingredients. *Current Opinion in Food Science*, 10, 45-51.
- Souto U.T.C.P., Barbosa M.F., Dantas H.V., de Pontes A.S., Lyra W.S., Diniz P.H.G.D., de Araújo M.C.U., da Silva, E.C. (2015). *Food Analytical Methods*, 8, 1515-1521.
- Stadler, R. H. (2019). Introduction to the volume: Food adulteration & contamination. In L. Melton, F. Shahidi, & P. Varelis (Eds.), *Encyclopedia of food chemistry*, 317-319. Oxford, UK: Academic Press.

- Teixeira J. L. da Paixão, Dos Santos Caramês, E. T., Baptista D. P., Gigante M. L., Lima Pallone J. A. (2021). Rapid adulteration detection of yogurt and cheese made from goat milk by vibrational spectroscopy and chemometric tools, *Journal of Food Composition and Analysis*, 96, 103712.
- Trentanni Hansen G. J., Almonacid J., Albertengo L., Rodriguez M. S., Di Anibal C., & Delrieux C. (2019). NIR-based Sudan I to IV and Para-Red food adulterants screening. *Food Additives & Contaminants: Part A*, 36, 1163-1172.
- Upadhyay V. K., y McSweeney P. L. H. (2007). Acceleration of cheese ripening. In P. L. H. McSweeney (Ed.), *Cheese problems solved*, 1st ed., 419-447. Duxford, UK: Woodhead Publishing.
- Vandeginste B.G.M., Massart D.L., Buydens L.M.C., De Jong S., Lewi P.J., y Smeyers Verbeke J. (1998). Supervised pattern recognition, In: *Handbook of Chemometrics and Qualimetrics: Part B* Elsevier, Amsterdam.
- Voragen A. G. J. (1998). Technological aspects of functional food-related carbohydrates, *Trends in Food Science & Technology*, 9, 328-335.
- Xu L., Goodarzi M., Wei Shi, Chen-Bo Cai, Jian-Hui Jiang. (2016). A MATLAB toolbox for class modeling using one-class partial least squares (OCPLS) classifiers. *Chemometrics and Intelligent Laboratory Systems*, 139, 58–63.
- Washingtonpost. (2016). <https://www.washingtonpost.com/news/wonk/wp/2016/02/17/parmesan-cheese-is-not-what-it-seems/> (Ultimo acceso julio 2023).
- Wei Hong F., y Seng Chia K. (2020). A Review on Recent Near Infrared Spectroscopic Measurement Setups and Their Challenges, *Measurement*, 171, 108732.
- Wu X., Kumar V., Ross Quinlan J., Ghosh J., Yang Q., Motoda H., et al. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14, 1-37.
- Yuan L., Chen X., Huang Y., Chen J., y Pan. (2023). Spectral separation degree method for Vis-NIR spectroscopic discriminant analysis of milk powder adulteration, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 301, 122975.
- Zhao M., Renwick J. B., Anna M. F., O'Donnell C. P., y Downey G. (2015). Prediction of naturally-occurring, industrially-induced and total trans fatty acids in butter, dairy spreads and Cheddar cheese using vibrational spectroscopy and multivariate data analysis, *International Dairy Journal*, 51, 41-51.

Zheng C., Sun D.W., y Zheng L. (2006). Recent developments and applications of image features for food quality evaluation and inspection – a review. *Trends in Food Science and Technology*, 17, 642-655.

A large, bold, black letter 'C' is positioned on the left side of the page. It is partially enclosed by a thick, blue, curved graphic element that extends from the top left towards the center. The background is white with a light gray shadow effect behind the blue graphic.

APÍTULO 3

Determinación de la
adulteración de quesos
rallados mediante
espectroscopía NIR y análisis
multivariante

CAPÍTULO III: DETERMINACIÓN DE LA ADULTERACIÓN DE QUESOS RALLADOS MEDIANTE ESPECTROSCOPIA NIR Y ANÁLISIS MULTIVARIANTE

3.1. Objetivo

El objetivo general de esta parte del trabajo de tesis es proponer una herramienta sencilla, rápida y fiable para determinar adulteraciones en quesos rallados. Para ello se evalúa el potencial de la espectroscopía NIR, junto con análisis multivariante de datos. Se siguieron dos objetivos específicos, la determinación de (i) cantidades de aditivos, celulosa y dióxido de silicio, superiores a las permitidas, y (ii) agregados no permitidos como sémola de trigo, harina de trigo y polvo de madera. De esta manera, se proponen herramientas que puedan proteger al consumidor de fraudes alimentarios, que incluyen el etiquetado engañoso y el detrimento nutricional del alimento.

3.2. Materiales y métodos.

3.2.1. Muestras y reactivos.

Se adquirieron un total de 22 muestras de quesos rallados de puntos comerciales ubicados en diferentes provincias de Argentina (Buenos Aires, Córdoba, Santa Fe y La Pampa). Además, también se recolectaron 10 muestras bajo la denominación “*alimento a base de queso rallado y agregados como sémola, sólidos lácteos, harina de arroz, almidones modificados, fibra dietaria, entre otros*”. La celulosa microcristalina se adquirió de Biopack (Biopack Argentina) y el dióxido de silicio de Merck (Merck Argentina). La harina y la sémola de trigo se compraron en comercios locales. El polvo de madera se adquirió en aserraderos ubicados en la provincia de Misiones (Argentina) y se presentó como un polvo de color claro y molienda fina.

Para construir una regla de clasificación se necesita contar tanto con muestras no adulteradas (originales) como adulteradas. Por lo tanto, se añadieron a las muestras originales los posibles adulterantes a dos niveles de concentración, seguido de una homogeneización manual. La concentración de ambos aditivos (celulosa y dióxido de silicio) implicó niveles superiores a las cantidades máximas permitidas (Codex Alimentarius, FDA, Código Alimentario Argentino y UE) (Tabla 2.1). La harina y sémola de trigo y el polvo de madera no deberían estar presentes a ningún nivel de concentración, por lo que se eligieron niveles donde las características organolépticas del alimento no se vieran modificadas.

La siguiente tabla muestra la disposición de las clases definidas y los respectivos niveles de adulteración.

Tabla 3.1. Descripción de las muestras en estudio. Niveles de concentración entre paréntesis (% m/m).

Clases	Tipo de muestra	Codificación
Clase 1	Originales (sin adulterar)	1 – 22
Clase 2	Adulteradas con celulosa microcristalina	23 - 38 (7%), 39 - 60 (5%)
Clase 3	Adulteradas con dióxido de silicio	61 - 76 (6%), 77 - 98 (5%)
Clase 4	Adulteradas con harina de trigo	99 - 114 (6%), 115 - 136 (5%)
Clase 5	Adulteradas con sémola de trigo	137 - 152 (7%), 153 - 174 (5%)
Clase 6	Adulteradas con polvo de madera	175 - 196 (3%), 197 - 212 (4%)

3.2.2. Instrumentación.

Los espectros NIR se adquirieron en un equipo Nicolet iS50 FT-IR equipado con el accesorio del módulo NIR (Thermo Scientific, Madison, WI, EE. UU.) en modo reflectancia difusa (Figura 3.1). Las muestras se midieron en viales de vidrio con módulo de esfera integradora, por triplicado, y el vial de muestra se rellenó en cada medida. Se utilizó un material de referencia de teflón como “background”. El rango espectral de estudio comprende los 10.000 y 4000 cm^{-1} , una resolución de 6 cm^{-1} y un valor de ganancia (gain) de 1 a temperatura ambiente. Se adquirieron un total de 64 escaneos para cada espectro.



Figura 3.1. Espectrofotómetro Nicolet iS50 FT-IR con iS50 módulo de NIR.

3.2.3. Análisis composicional.

El contenido total de carbohidratos (TCC) en 100 g de alimento se calculó por diferencia, aplicando la siguiente fórmula (Merrill y Watt, 1973):

TCC/100g de alimento= 100 - peso en gramos (proteína+grasas+agua+cenizas/100g de alimento).

El contenido de proteína bruta se obtuvo por el método Kjeldahl considerando un factor de multiplicación de 6,38 (AOAC, 2005a), los lípidos se determinaron por el método Soxhlet con hexano como solvente de extracción (AOAC, 2005b), el contenido de humedad se obtuvo por secado al vacío hasta alcanzar peso constante (AOAC, 2005c) y la ceniza se determinó por incineración a 525 °C en un horno mufla durante 6 h (AOAC, 2005d). Se aplicó la prueba de lugol para determinar cualitativamente la presencia de almidón por medio de una solución de KI/I₂. Se usó un microscopio Zeiss Primo Star equipado con una cámara de microscopía digital (modelo Carl Zeiss Brand AxioCam) con un aumento de 40X para determinar el origen del almidón.

3.2.4. Análisis multivariante.

Se evaluaron varios algoritmos de pre-procesamiento espectral para compensar las variaciones sistemáticas no deseadas propias de espectros NIR: Corrección de Dispersión Multiplicativa (MSC), *Detrending*, Variación Normal Estándar (SNV), primera y segunda derivada (Savitsky-Golay), Corrección de desplazamiento (*offset correction*) y combinaciones de pre-procesados (Rinnan, Van Den Berg y Engelsen, 2009).

Se aplicó el análisis de componentes principales (PCA) como análisis exploratorio (Bartolomé, 2010). Luego, se aplicaron técnicas de clasificación supervisada para construir un modelo que permita predecir futuras muestras desconocidas. Se utilizaron técnicas de clasificación discriminante como los *K*-vecinos más cercanos (*K*-NN) y el análisis discriminante por mínimos cuadrados parciales (PLS-DA) y también técnicas de modelado de clases como modelado independiente suave de analogía de clase (SIMCA) (Otto, 2017). Los parámetros, como el número de variables latentes (LV) en PLS-DA, el número de vecinos en *K*-NN y el número de PCs en cada clase en SIMCA, se optimizaron mediante una estrategia de validación cruzada con persianas venecianas con cuatro grupos de cancelación.

Se evaluaron los parámetros de calidad de clasificación en términos de sensibilidad, especificidad y precisión según se detalla en la sección 2.6.4.3. *Parámetros de calidad en clasificación multivariante*.

El análisis multivariante se realizó con el software MATLAB (Versión 2010), PLS-Toolbox (Eigenvector Research) y Milano Classification Toolbox (Milano 5.1).

3.3. Resultados y discusión

3.3.1. Espectros NIR.

La Figura 3.2 (a) muestra los espectros NIR de muestras de queso originales. Se pueden observar algunas diferencias espectrales entre los quesos comerciales, mayoritariamente en los rangos entre 4000 y 5000 cm^{-1} y entre 6000 y 7000 cm^{-1} . Además, los espectros contienen una variabilidad no deseada como desplazamientos de la línea de base y efectos de dispersión, por lo que fue necesario aplicar un pre-procesamiento previo para obtener la información útil y relevante. La Figura 3.2 (b) muestra los espectros de una muestra original aleatoria y de la misma muestra adulterada con celulosa microcristalina, dióxido de silicio, harina de trigo, sémola de trigo y polvo de madera. En este caso, no fue posible visualizar diferencias entre la muestra original y las adulteradas por inspección visual, por lo que se necesitaron herramientas quimiométricas para lograr la discriminación entre las diferentes muestras.

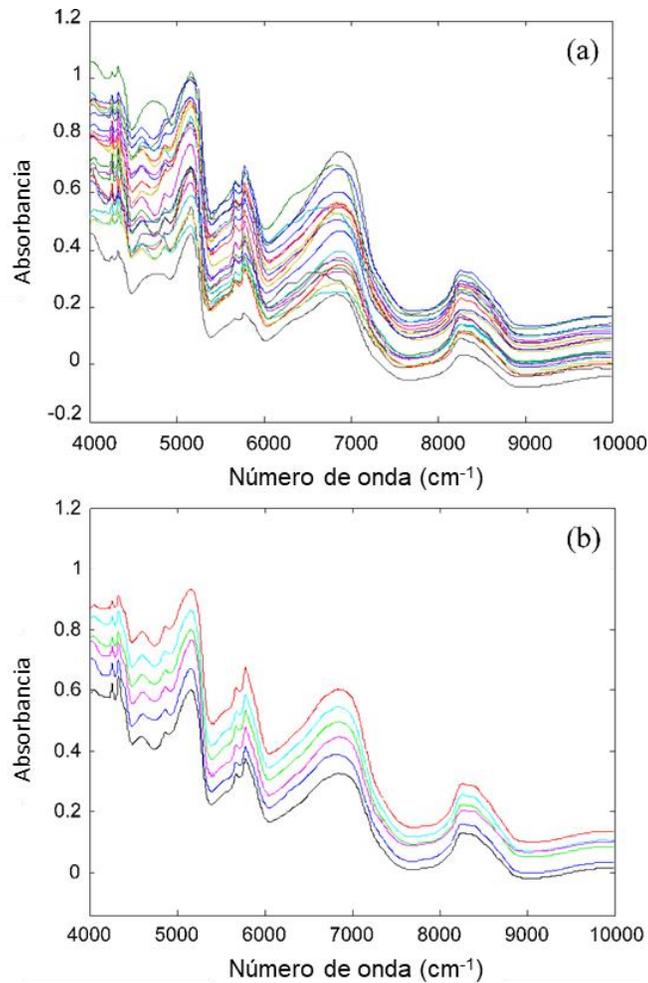


Figura 3.2. Espectros NIR de **(a)** todos los quesos rallados originales y **(b)** de una muestra original aleatoria (línea continua) y adulterada con celulosa microcristalina al 5% (línea marrón), dióxido de silicio al 5% (línea verde), harina de trigo al 5% (línea azul), sémola de trigo al 5% (línea roja) y polvo de madera al 4% (línea violeta).

3.3.2. Análisis exploratorio.

Se utilizó PCA para visualizar la posible tendencia de agrupaciones naturales en los datos y evaluar si existe discriminación entre las diferentes muestras no adulteradas y adulteradas. Además, también se utilizó PCA para detectar valores atípicos mediante la inspección del gráfico de Q residuales vs Hotelling T^2 . La Figura 3.3 muestra el gráfico de Scores 3D de los datos pre-procesados: primera derivada con algoritmo Savitzky-Golay (ventana de 11 puntos y polinomio de primer grado). Se explica una varianza total de 78,13% considerando las tres primeras PCs. Aunque existe superposición entre las diferentes clases, se puede apreciar que las muestras no adulteradas (clase 1 ▼) tienden a

agruparse a través de valores de PC2 en su mayoría positivos. De lo contrario, las muestras adulteradas con dióxido de silicio (clase 2 ■) se ubican predominantemente a lo largo de valores positivos de PC3 y presentan la tendencia a ser el grupo más separado entre las clases adulteradas. Cabe mencionar que no se encontraron valores atípicos considerando Hotelling T^2 y residuales Q.

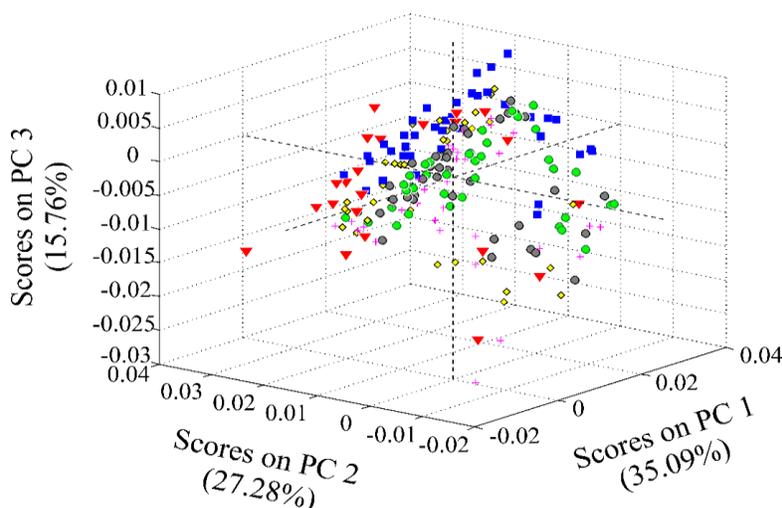


Figura 3.3. Gráfico de Scores PCA para datos pre-procesados con la primera derivada. Muestras no adulteradas (▼) y muestras adulteradas con: celulosa microcristalina (●), dióxido de silicio (■), harina de trigo (+), sémola de trigo (◆) and polvo de madera (●).

3.3.3. Clasificación Supervisada.

La Tabla 3.2 muestra los resultados de los modelos de clasificación K -NN, PLS-DA y SIMCA en términos de sensibilidad, especificidad y precisión. Se puede observar que no se obtuvieron resultados satisfactorios de clasificación con la técnica K -NN. Por otro lado, considerando los resultados globales, PLS-DA y SIMCA proporcionaron resultados globales satisfactorios (superiores al 87%) para las clases 2, 3 y 6. Además, con PLS-DA también se obtuvieron buenos resultados globales (alrededor del 90%) para las clases adulteradas con harina y sémola trigo (clases 4 y 5, respectivamente) y respecto a SIMCA, los resultados globales son menos satisfactorios (entre 42 y 87%). Finalmente, centrándonos en la clase 1 (muestras no adulteradas), los resultados son pobres en cuanto a sensibilidad para PLS-DA y precisión para SIMCA (por debajo del 58% en ambos casos). La baja sensibilidad sugiere que esta clase no puede reconocer correctamente sus propias muestras, mientras que la baja precisión significa que las muestras de clases adulteradas se asignaron erróneamente a la clase 1, específicamente muestras de las clases 4 y 5.

Tabla 3.2. Resultados obtenidos con diferentes técnicas de clasificación.

Técnica de Clasificación	Parámetros		C1	C2	C3	C4	C5	C6
K-NN	7 vecinos más cercanos	Sens.	18	34	63	21	18	29
		Esp.	98	83	95	75	80	86
		Prec.	30	30	73	15	17	31
PLS-DA	19 LVs	Sens.	54	97	100	92	83	100
		Esp.	99	98	98	98	97	100
		Prec.	75	94	94	88	83	100
SIMCA	10 PCs para C1 10 PCs para C2 12 PCs para C3 10 PCs para C4 10 PCs para C5 10 PCs para C6	Sens.	95	87	89	66	68	95
		Esp.	92	99	100	87	82	100
		Prec.	58	94	100	48	42	100

PLS-DA: Análisis discriminante por mínimos cuadrados parciales; K-NN: K-Vecinos más cercanos; SIMCA: Modelado Suave independiente por Analogía de Clases, LVs: Variables Latentes; PCs: Componentes Principales; Sens.: Sensibilidad; Esp.: Especificidad; Prec.: Precisión; C1: Clase 1; C2: Clase 2; C3: Clase 3; C4: Clase 4; C5: Clase 5; C6: Clase 6.

3.3.4. Análisis composicional de muestras sospechosas de estar adulteradas.

De los modelos de clasificación iniciales de la Tabla 3.2 se observó que muestras que deberían clasificarse a la clase 1 (muestras no adulteradas), fueron clasificadas a las clases 4 y 5 (adulteradas con harina y sémola de trigo, respectivamente). El principal carbohidrato presente tanto en la harina como en la sémola de trigo, es el almidón. Por lo tanto, a todas las muestras pertenecientes a la clase 1 se les aplicó una prueba colorimétrica cualitativa para la determinación de almidón. En este sentido, una muestra es positiva para almidón si se tiñe de violeta oscuro al agregar una solución de yodo en yoduro de potasio (solución de lugol) (Dar, 1923). Los resultados obtenidos sugieren que 5 de las 22 muestras comerciales de quesos rallados pertenecientes a la clase 1 fueron positivas (Figura 3.4.).

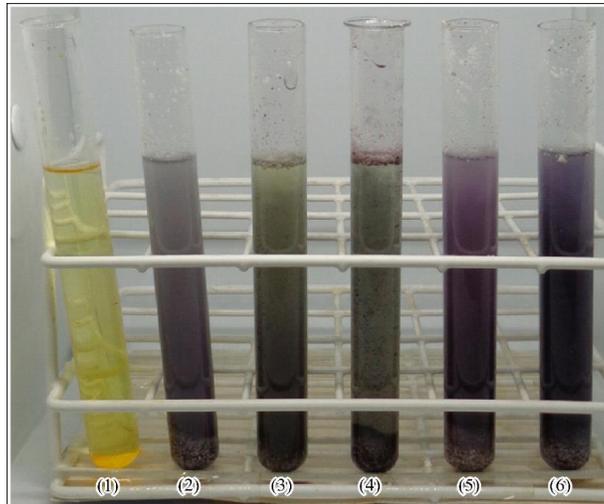


Figura 3.4. Prueba de Lugol para una muestra sin contenido de almidón **(1)** y para las 5 muestras con almidón **(2 a 6)**.

Posteriormente, para determinar el origen del almidón de estas muestras adulteradas, se evaluaron imágenes digitales microscópicas. El almidón de trigo tiene formas lenticulares y esféricas, mientras que el almidón de maíz tiene formas poliédricas irregulares compuestas por un hilo central en forma de estrella con tres ramas (BeMiller, 2018). La Figura 3.5 muestra las imágenes microscópicas obtenidas de las cinco muestras que fueron positivas al análisis del lugol. Se puede observar que tres de las cinco podrían contener almidón de trigo, mientras que las otras dos muestras sugieren la presencia de almidón de maíz.

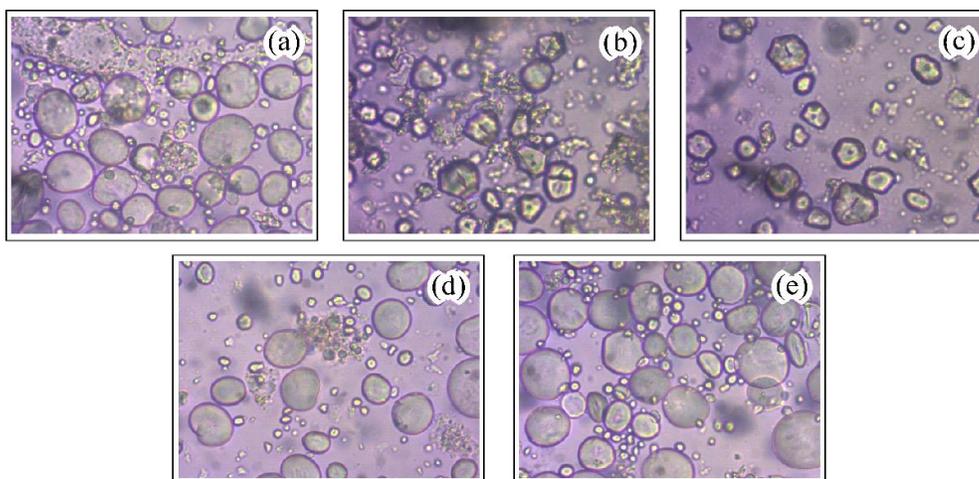


Figura 3.5. Micrografías de las 5 muestras positivas con Lugol: almidón de trigo **(a)**, **(d)**, **(e)** y almidón de maíz **(b)**, **(c)**.

Teniendo en cuenta que el almidón constituye la mayor proporción de carbohidratos en los cereales, se determinó el contenido total de carbohidratos como se indicó anteriormente (Tabla 3.3). Los resultados demuestran que las 5 muestras comerciales positivas a la prueba de lugol efectivamente contenían carbohidratos. Cabe mencionar que los rótulos de las muestras 3 y 4 declaran un contenido de carbohidratos del 10%, cuando en realidad las muestras contenían más del 20% y las muestras 1, 2 y 5 declaran un contenido nulo (0%) pero su contenido supera el 15%. Estas muestras no deben etiquetarse como 100% quesos rallados y están erróneamente etiquetadas, lo que representa un fraude al consumidor por esconder la adulteración.

Tabla 3.3. Análisis composicional de las 5 muestras positivas de test de Lugol (m/m %).

Muestra	Humedad	Proteínas	Grasas Totales	Cenizas	Carbohidratos*
1	13,7	14,1	16,3	4,1	51,8
2	16,0	25,7	32,7	7,8	17,8
3	17,2	24,7	26,9	8,0	23,2
4	14,7	17,7	14,0	9,6	44,0
5	19,1	22,4	10,5	11,5	36,5

*obtenidos por diferencia.

3.3.5. Nuevos modelos de clasificación.

Las muestras positivas al análisis del lugol se eliminaron del conjunto de datos. A partir del nuevo conjunto de datos, se construyeron nuevamente modelos de clasificación con PLS-DA y SIMCA. Estos valores de clasificación se muestran en Tabla 3.4 y se puede observar que se obtuvieron resultados globales muy satisfactorios para todas las clases considerando ambas técnicas de clasificación (en su mayoría por encima del 93%), siendo un poco mejores para PLS-DA (casi el 100%). Por lo tanto, se seleccionó PLS-DA como modelo de clasificación final. Considerando el problema analítico abordado en este estudio, la implicancia de los errores de clasificación (falsos positivos y falsos negativos) es diferente. Centrándonos en la clase 1, los resultados obtenidos en términos de especificidad son altamente positivos (100%), por lo que el riesgo de asignar muestras adulteradas como originales es nulo, lo que denota la capacidad del modelo de rechazar futuras muestras adulteradas y clasificarlas adecuadamente en las clases adulteradas. Por otro lado, la sensibilidad es muy satisfactoria (94%) y se relaciona con falsos negativos, es decir, muestras originales asignadas como adulteradas. Esto último implica un riesgo económico,

ya que estas muestras deben retirarse innecesariamente del mercado comercial. Respecto a todas las clases adulteradas, se obtuvo el máximo porcentaje de especificidad (100%), evidenciando una buena capacidad para rechazar muestras pertenecientes a otras clases. Finalmente, la precisión es excelente (100%) para las seis clases, lo que denota una alta probabilidad de que una decisión positiva sea correcta.

Tabla 3.4. Nuevo modelo de clasificación luego de eliminar muestras positivas en almidón de la clase 1.

Técnica de Clasificación	Parámetros óptimos		C1	C2	C3	C4	C5	C6
PLS-DA	19 LVs	Sens.	94	100	100	97	93	100
		Esp.	100	100	100	100	100	100
		Prec.	100	100	100	100	100	100
SIMCA	8 PCs para C1 8 PCs para C2 9 PCs para C3 8 PCs para C4 10 PCs para C5 8 PCs para C6	Sens.	100	97	93	100	80	97
		Esp.	98	100	100	99	100	100
		Prec.	85	100	100	94	100	100

PLS-DA: Análisis discriminante por mínimos cuadrados parciales; SIMCA: Modelado Suave independiente por Analogía de Clases. LVs: Variables Latentes; PCs: Componentes Principales; Sens.: Sensibilidad; Esp.: Especificidad; Prec.: Precisión; C1: Clase 1; C2: Clase 2; C3: Clase 3; C4: Clase 4; C5: Clase 5; C6: Clase 6.

Por otro lado, se realizó un análisis de las curvas ROC del modelo de clasificación final con PLS-DA. En la Figura 3.6 se muestran las curvas de la clase 1 (originales) y la clase 4 (adulteradas con harina de trigo). Las demás clases no se muestran ya que todas las clases adulteradas presentan un comportamiento similar. El punto de cruce entre la máxima sensibilidad (100%) y la máxima especificidad (100%) se calcula de tal forma que se maximiza el número de verdaderos positivos (sensibilidad) y verdaderos negativos (especificidad). Por lo tanto, se puede observar en ambas curvas una excelente capacidad de clasificación (100% de sensibilidad y especificidad), porque las curvas ROC parecen cuadrados en la esquina superior izquierda.

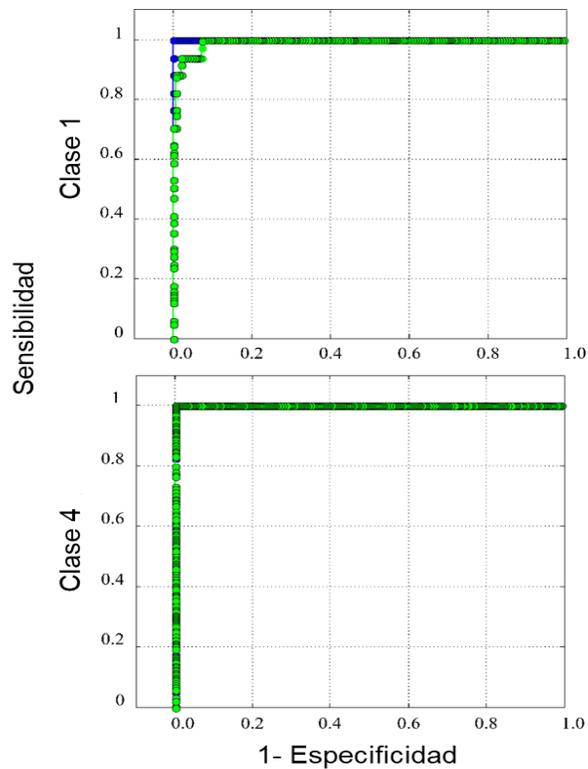


Figura 3.6. Curvas ROC para el modelo PLS-DA con datos pre-procesados con la primera derivada. Validación cruzada (línea verde) y estimada (línea azul).

3.3.6. Predicción de muestras

Las cinco muestras de quesos comerciales adulteradas (positivas al análisis del lugol) se utilizaron para probar la capacidad de predicción del modelo de clasificación final. Todas las muestras se clasificaron en clase 4 (muestras adulteradas con harina de trigo) y clase 5 (muestras adulteradas sémola de trigo). Estos resultados eran esperados, ya que la harina y sémola de trigo contienen principalmente carbohidratos donde el almidón es el componente principal. Por lo tanto, el modelo tiene la capacidad de predecir si una muestra de queso contiene almidones entre sus componentes, lo cual es altamente positivo porque se pueden detectar muestras adulteradas desconocidas, con el fin de proteger a los consumidores de fraudes alimentarios.

Hoy en día, se pueden encontrar productos similares a base de queso que se venden bajo el nombre de “*alimentos a base de queso rallado y agregados*”. Estos alimentos están apareciendo cada vez más en los puntos de venta comerciales con un precio considerablemente inferior al de los quesos rallados. Las etiquetas de estos alimentos

declaran almidones en su composición. Cabe mencionar que estos alimentos no cumplen con las mismas normas de calidad que los quesos rallados. Según el artículo 3 del CAA “*Todo proceso de elaboración que implícitamente no figure en el código será lícito si no introduce elementos extraños o indeseables, o no altera el valor nutritivo o aptitud bromatológica de los alimentos terminados que se trate*” (CAA, Cap. I, 2023). Por lo tanto, estas muestras se utilizaron sólo para evaluar y reforzar la capacidad predictiva del modelo de clasificación (Tabla 3.5). Los carbohidratos declarados en sus etiquetas corresponden a la sémola de trigo (cinco muestras), harina de arroz (una muestra) y a un almidón modificado, denominado fosfato de dialmidón de hidroxipropilo (tres muestras).

Tabla 3.5. Predicción de los “quesos a base de...”

Carbohidrato rotulado	Predicción					
	C1	C2	C3	C4	C5	C6
Sémola				X	X	
Sémola				X	X	
Sémola				X	X	
Sémola				X	X	
Almidón				X	X	
Harina de arroz				X	X	
Sémola				X	X	
Almidón				X	X	
Fibra dietaria		X				
Almidón modificado				X	X	

C1: Clase 1; C2: Clase 2; C3: Clase 3; C4: Clase 4; C5: Clase 5; C6: Clase 6

Se puede observar que la mayoría de las muestras se asignaron simultáneamente a las clases 4 y 5 (muestras adulteradas con harina de trigo y sémola, respectivamente). Estos resultados sugieren que NIR tiene la capacidad de detectar correctamente el contenido de almidón en estos alimentos. Finalmente, una muestra fue clasificada en clase 2 (muestras

adulteradas con celulosa) y tal asignación es coherente, ya que esta muestra declara contener fibra dietética y la celulosa es parte de la fibra dietética total.

3.4. Conclusiones

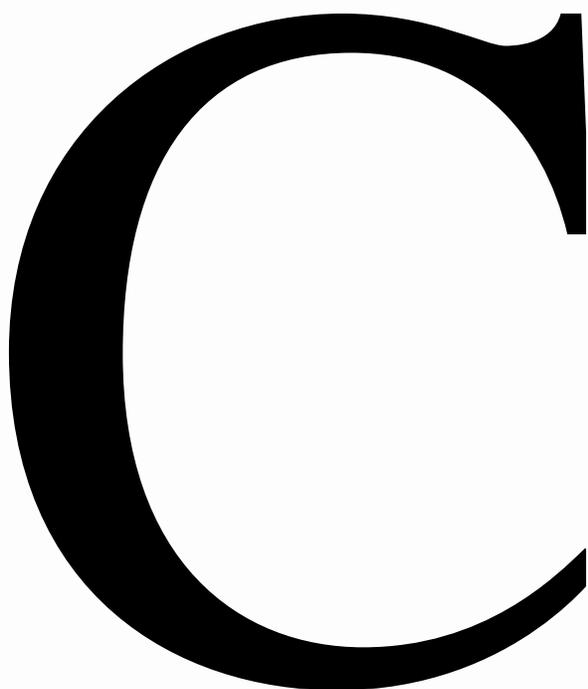
Este trabajo presenta una herramienta analítica de screening para determinar la adulteración de quesos rallados de pasta dura basada en la espectroscopía NIR y clasificación multivariante. Se evaluaron aditivos, como celulosa microcristalina y dióxido de silicio, en niveles superiores a los regulados y agregados no permitidos utilizados para aumentar volumen y peso, como harina de trigo, sémola de trigo y polvo de madera.

PLS-DA proporcionó excelentes resultados de clasificación en términos de sensibilidad, especificidad y precisión, ya que fue posible discriminar entre muestras originales y adulteradas. La capacidad predictiva del modelo se evaluó con dos tipos de alimentos diferentes: muestras de quesos rallados que contienen almidones no declarados y alimentos similares denominados "*alimentos a base de queso y agregados...*". Los resultados de predicción sugieren que NIR tiene la capacidad de detectar muestras adulteradas con almidones de diferentes orígenes, incluso cuando el almidón no se contempla como adulterante en el modelo.

La metodología representa una herramienta rápida, fiable y sencilla para realizar el control de calidad de quesos rallados. Esto es muy importante para asegurar la correcta identidad de muestras desconocidas, con el fin de evitar prácticas de venta fraudulentas como el etiquetado engañoso y el detrimento nutricional del queso, que claramente representan un fraude alimentario para los consumidores. Una clara ventaja de NIR, frente a las técnicas espectrométricas comunes, es que las medidas no son destructivas de la muestra, no existe la necesidad de procesos de extracción y no se generan residuos, siguiendo los principios de la química verde. Además, tiene una aplicabilidad potencial a lo largo de la industria láctea mediante la realización de análisis *in situ* con NIR portátiles.

3.5. Referencias

- AOAC. (2005a). *Official methods of analysis of AOAC International 920.123*. Gaithersburg, MD, USA: AOAC International.
- AOAC. (2005b). *Official methods of analysis of AOAC International 933.05*. Gaithersburg, MD, USA: AOAC International.
- AOAC. (2005c). *Official methods of analysis of AOAC International 926.08*. Gaithersburg, MD, USA: AOAC International.
- AOAC. (2005d). *Official methods of analysis of AOAC International 935.42*. Gaithersburg, MD, USA: AOAC International.
- Bartholomew D. J. (2010). Principal Components Analysis. *International Encyclopedia of Education*, 374-377.
- BeMiller J. N. (2018). Starches: Molecular and Granular Structures and Properties. *Carbohydrate Chemistry for Food Scientists* (3rd ed.). Whistler Center for Carbohydrate Research, Department of Food Science, Purdue University, West Lafayette, Indiana.
- CAA, Código Alimentario Argentino. (2010). Capítulo I: Disposiciones generales, artículo 6, disponible en http://www.anmat.gov.ar/alimentos/codigoa/CAPITULO_I.pdf (Último acceso julio 2023).
- Dhar, N. R. (1923). The Starch-Iodine Reaction. *The Journal of Physical Chemistry*, 28, 125–130.
- López, M. I., Callao, M. P., & Ruisanchez, I. (2015). A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. *Analytica Chimica Acta*, 891, 62-72.
- Merrill A. L., y Watt B. K. (1973). Energy Value of Foods-Basis and Derivation, USDA Handbook 74 (slightly revised 1973). https://www.ars.usda.gov/ARUserFiles/80400525/Data/SR13/sr13_doc.pdf (Último acceso julio 2023).
- Otto, M. (2017). Chemometrics, Statistics and Computer Application in Analytical Chemistry (3rd ed.). Wiley-VCH Verlag GmbH & Co.KGaA, Boschstr. Weinheim, Germany.
- Rinnan A., Van Den Berg F., y Engelsen S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28, 1201-1222.

A large, bold, black letter 'C' is positioned on the left side of the page. It is set against a white background that is partially enclosed by a blue and grey graphic element resembling a stylized 'C' or a bracket. The blue element is on the inside of the 'C', and the grey element is on the outside, creating a layered effect.

APÍTULO 4

Determinación de la
adulteración de quesos
rallados mediante imágenes
digitales y análisis
multivariante

CAPÍTULO IV: DETERMINACIÓN DE LA ADULTERACIÓN DE QUESOS RALLADOS MEDIANTE IMÁGENES DIGITALES Y ANÁLISIS MULTIVARIANTE

4.1. Objetivos

Este trabajo evalúa el potencial del análisis de imágenes digitales, junto con técnicas clásicas de clasificación y aprendizaje automático (machine learning), para determinar la adulteración de alimentos de alto valor comercial como son los quesos rallados de pasta dura. Se siguieron dos objetivos: (i) la determinación de aditivos permitidos, celulosa microcristalina y dióxido de silicio, en niveles superiores a los reglamentados y (ii) la determinación de agregados no permitidos como harina de trigo, sémola de trigo y polvo de madera, a tres niveles de adulteración. El objetivo general es proponer una herramienta analítica simple, rápida, de bajo costo y confiable para determinar los distintos adulterantes en quesos rallados, que tienen como consecuencia un impacto directo en la calidad nutricional de estos alimentos.

4.2. Materiales y métodos.

4.2.1. Muestras y reactivos.

Se adquirieron un total de 14 muestras de quesos rallados considerando diferentes marcas comerciales de regiones seleccionadas de Argentina (Buenos Aires, Córdoba, Santa Fe y La Pampa). La celulosa microcristalina se compró a Biopack (Biopack Argentina) y el dióxido de silicio a Merck (Merck Argentina). La harina de trigo y la sémola de trigo se obtuvieron en mercados comerciales. El polvo de madera se adquirió en aserraderos ubicados en la provincia de Misiones (Argentina) debido a que tiene características especiales como un color marrón claro y una molienda fina. Como sólo se disponía de 14 muestras comerciales de quesos rallados, esta cantidad no es suficiente para construir un modelo de clasificación robusto. Por lo tanto, se prepararon adicionalmente 30 muestras mezclando las muestras originales en una proporción del 50% cada una. Las muestras fueron seleccionadas al azar para obtener un conjunto representativo de las mismas. También se prepararon muestras adulteradas añadiendo los cinco agregados enumerados en la Tabla 4.1 a tres niveles de concentración, seguido de homogeneización manual.

Tabla 4.1 Descripción de las muestras estudiadas. Niveles de concentración entre paréntesis (% m/m).

Clases	Tipo de muestra	Codificación
Clase 1	Originales (sin adulterar)	1 – 44
Clase 2	Adulteradas con celulosa microcristalina	45-58 (7%), 59-72 (10%), 73-86 (12%)
Clase 3	Adulteradas con dióxido de silicio	87-100 (5%), 101-114 (6%), 115-128 (7%)
Clase 4	Adulteradas con harina de trigo	129-142 (7%), 143-156 (10%), 157-170 (15%)
Clase 5	Adulteradas con sémola de trigo	171-184 (7%), 185-198 (10%), 199-212 (15%)
Clase 6	Adulteradas con polvo de madera	213-226 (4%), 227-240 (5%), 241-254 (6%)

*Niveles de adulteración entre paréntesis.

Los aditivos estudiados, celulosa microcristalina y dióxido de silicio, se encuentran en niveles superiores a las cantidades máximas permitidas como se detallan en la Tabla 2.1 (Codex Alimentarius, FDA, Código Alimentario Argentino y UE). Es importante mencionar que las concentraciones seleccionadas de ambos aditivos no modifican las características organolépticas de los quesos rallados. Antes de la captura de la imagen digital, las muestras se colocaron en cajas de Petri de plástico, circulares, con un diámetro de 5,5 cm y una altura de 1,2 cm. La superficie se alisó con una varilla para obtener una imagen homogénea.

4.2.2. Instrumentación y software

Las imágenes digitales se obtuvieron con una cámara digital compacta Canon, línea PowerShot modelo SX730 HS (Nueva York, Canon USA, Inc.) con las siguientes características: resolución de 20,3 Mpx, zoom óptico de 40x, lentes de 1008 mm incluidas y una apertura de lente máxima de 3,5. La cámara cuenta con tres parámetros disponibles para el ajuste manual: apertura (F3,3 - F8,0), velocidad de obturación (1/3200-500) y sensibilidad ISO (Organización Internacional de Normalización; modo 80-3200). La apertura se define como el diámetro de la apertura del lente dividido por la distancia focal, y afecta la cantidad de luz que pasa a través de la lente. Los valores F más pequeños significan que entra más luz en la cámara. El tiempo que el obturador permanece abierto se denomina velocidad de obturación y también afecta la cantidad de luz que ingresa a la cámara. La

sensibilidad del sensor, conocida como ISO, se refiere a la amplificación de la señal. Los valores altos de ISO degradan la calidad de la imagen como resultado de la amplificación del ruido, la pérdida de contraste, el rango dinámico, la resolución y el color. Por lo tanto, se recomiendan valores bajos de ISO (Gilberto, 2022).

La cámara se colocó sobre un soporte en el centro de un novedoso dispositivo diseñado y construido en nuestro laboratorio para esta tesis doctoral. El mismo consta de una estructura metálica de aluminio de 25 cm de alto, 40 cm de ancho y 40 cm de largo, equipado con 6 fuentes de luz diferentes (fluorescente, LED fría, LED cálida, dicróica lámpara halógena y bombilla incandescente). La Figura 4.1 muestra el dispositivo con la disposición de las diferentes fuentes de luz. El dispositivo permite trabajar tanto con una sola lámpara como con combinaciones entre ellas. Las fuentes de luz seleccionadas involucran diferentes zonas de la región visible del espectro electromagnético, donde la radiación interactúa con la muestra de diferentes maneras. Este innovador dispositivo permite trabajar con muestras de diferente naturaleza, donde la fuente de luz representa un parámetro importante a optimizar. En la parte superior del dispositivo se utilizó un soporte metálico para sujetar la cámara en una posición fija, mientras que en la parte inferior se encontraba una plataforma móvil sobre la que se colocaba perfectamente la muestra en el centro del visor de la cámara. De esta forma, la distancia entre la cámara y la muestra se mantiene fija, asegurando la reproducibilidad, por lo que todas las imágenes se adquirieron en las mismas condiciones. Finalmente, el dispositivo se cubrió externamente con una cubierta de etil-vinil-acetato para reducir la interferencia del entorno.

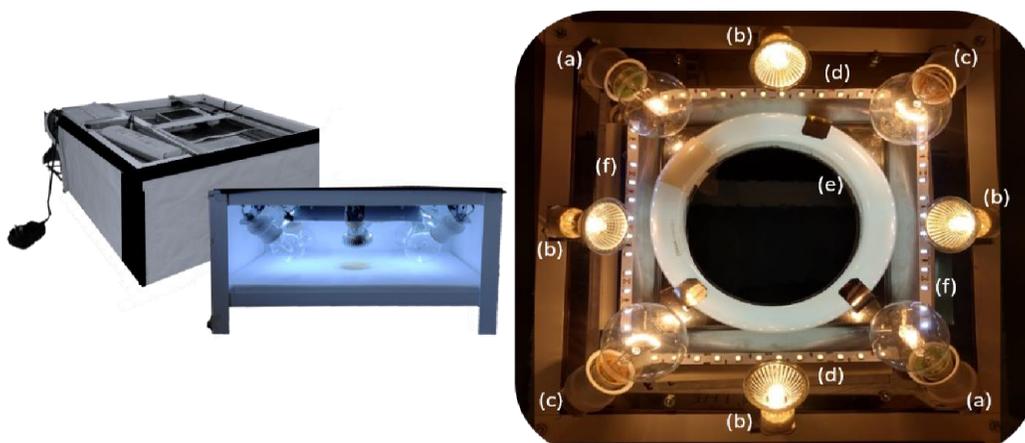


Figura 4.1. Dispositivo de captura de imágenes; Imagen externa e interna, fuentes de iluminación: (a) lámparas halógenas, (b) lámparas dicróicas, (c) lámpara incandescente, (d) Led Cálida, (e) lámpara fluorescente y (f) Led fría.

4.2.3. Análisis multivariante.

4.2.3.1. Modelos de clasificación

Se evaluaron varios algoritmos de clasificación. Por un lado, los modelos clásicos como Análisis Discriminante por Mínimos Cuadrados Parciales (PLS-DA) y los K -vecinos más cercanos (K -NN). Por el otro, modelos de aprendizaje automático supervisados (machine learning) que incluyen regresión logística (LR), máquinas de vectores de soporte (SVM), árboles de decisión (DT) y árboles aleatorios (RF).

4.2.3.2. Procesamiento de los datos

Inicialmente se realizó un análisis exploratorio de datos utilizando el análisis de componentes principales (PCA) para evaluar la variabilidad natural de las muestras. Los modelos discriminantes como PLS-DA y K -NN se construyeron seleccionando una región circular de interés en el centro de la imagen y, luego, los respectivos histogramas en rojo-verde-azul (RGB). Para los cálculos, se utilizó el histograma medio de cada muestra y se eliminaron los valores de frecuencia de píxeles iguales a cero para eliminar información no útil. Los parámetros como el número de variables latentes (LVs) en PLS-DA y el número de vecinos (K) en K -NN se optimizaron mediante una estrategia de validación cruzada utilizando persianas venecianas con cinco grupos de cancelación.

Teniendo en cuenta las técnicas de aprendizaje automático, cada imagen digital de 3888 - 3888 píxeles se dividió en nueve imágenes iguales de 1296 - 1296 píxeles. Esto permitió generar posteriormente un conjunto de datos balanceado (teniendo todo tipo de muestras en el entrenamiento, validación y prueba del modelo), nueve veces mayor que el original, que tiene más de una imagen por clase en cada partición del conjunto de datos. El hecho de tener una representatividad de los datos en todas las fases del entrenamiento del modelo permite obtener un modelo con un mayor nivel de generalización, lo que se traduce en una mejor clasificación. Para cada imagen, se obtuvieron los histogramas de los canales RGB. El histograma promedio de cada muestra se usó para los cálculos y los valores de frecuencia de píxeles iguales a cero se eliminaron para eliminar información no útil. Scikit-Learn, una biblioteca de amplia gama de algoritmos, se utilizó para el entrenamiento, validación y prueba del modelo (Pedregosa et al., 2011). El ajuste de los hiperparámetros se realizó utilizando GridSearchCV con validación cruzada k -fold con $k=5$. El entrenamiento se realizó con la partición estándar del 60%, 20%, 20% para entrenamiento, validación y prueba respectivamente, que se asignaron aleatoriamente para evitar el sobreajuste. Todos

los modelos de aprendizaje automático estudiados (DT, SVM, RF y LR) se construyeron utilizando la misma metodología descrita anteriormente (Fernández-Delgado, Cernadas, Barro y Amorim, 2014; Kotsiantis, 2007).

La Figura 4.2 muestra un diagrama de flujo ilustrativo que resume el procesamiento de imágenes previo a la construcción de los modelos de clasificación.

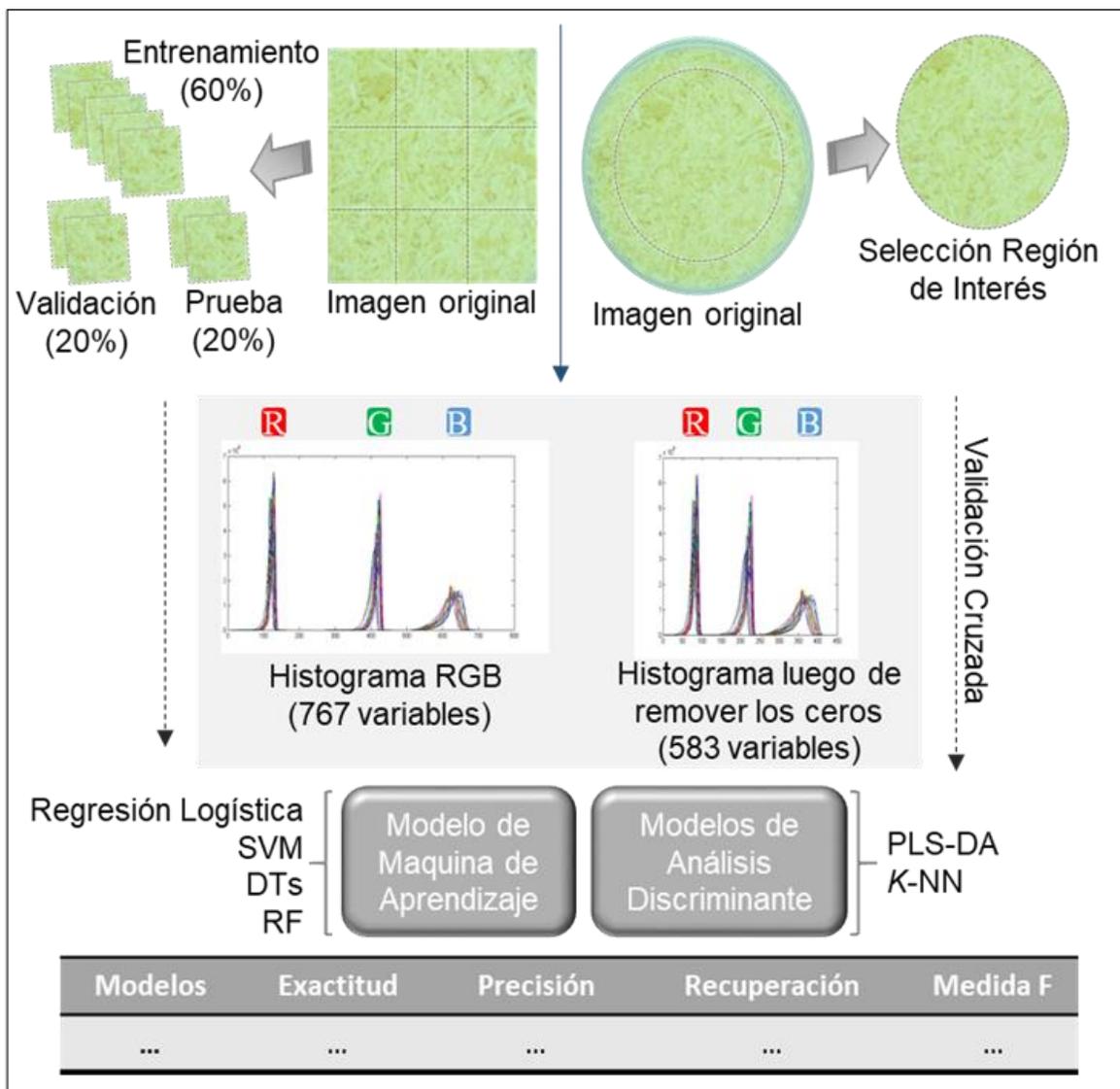


Figura 4.2. Diagrama de flujo ilustrativo que resume el tratamiento de imágenes previo a la construcción de los modelos quimiométricos.

4.3. Resultados y discusión

4.3.1. Adquisición de imágenes

Tanto los parámetros manuales de la cámara digital como la fuente de iluminación se deben optimizar previamente a obtener las imágenes digitales de la muestra. La Tabla 4.2 muestra todos los parámetros ajustables estudiados de la cámara digital y los valores finales seleccionados. Con respecto a la fuente de iluminación, la óptima fue la combinación de lámpara fluorescente con LED fría.

Tabla 4.2. Rango de variabilidad de los parámetros ajustables de la cámara digital y su valor óptimo.

Parámetros	Rango de variabilidad	Valor óptimo
Velocidad de obturación	1/3200" – 15"	1/125"
Apertura de diafragma	F 3.3 – F 8.0	F 8.0
ISO	80 – 3200	80

La Figura 4.3 (a) ilustra una muestra de queso rallado en todos los casos estudiados, original y adulterada con las cinco sustancias en el nivel de concentración más alto. Se puede observar que la percepción del ojo humano no puede distinguir una muestra original de una adulterada, incluso al nivel de concentración más alto. Esto ocurre porque los adulterantes en estudio presentan una granulometría pequeña que se mezcla perfectamente con el producto, por lo que la mezcla adulterada es imperceptible al ojo del consumidor.

Por otro lado, al evaluar los histogramas de color (Figura 4.3 b), es posible ver algunas diferencias entre los histogramas promedio de una muestra no adulterada y la misma muestra adulterada en la región RGB.

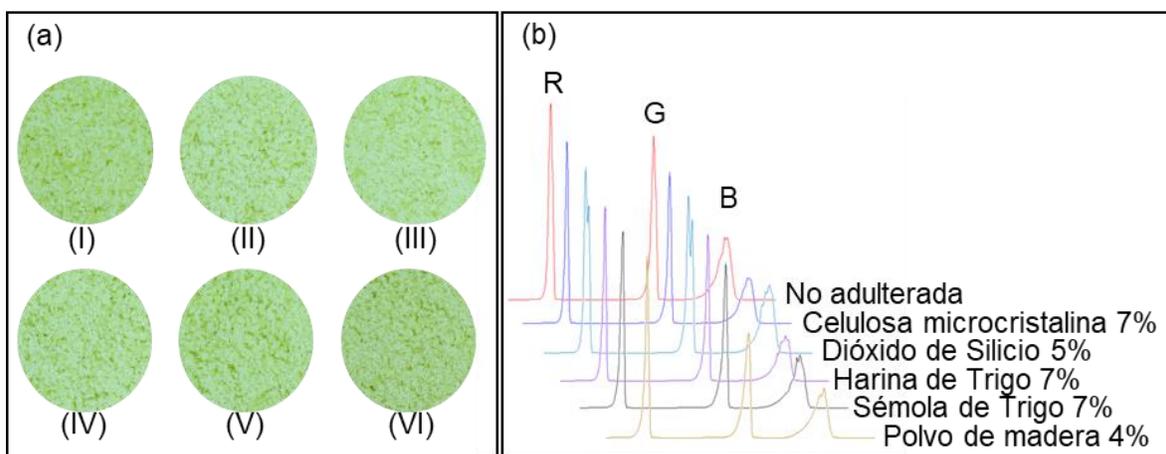


Figura 4.3. (a) Ilustraciones de la misma muestra de queso rallado con y sin la adición del adulterante estudiado: (I) no adulterada, adulterada con: (II) celulosa microcristalina, (III) dióxido de silicio, (IV) harina de trigo, (V) sémola de trigo, (VI) polvo de madera. (b) Histogramas de color promedios (en el canal RGB) de la muestra de queso rallado original y adulterado.

4.3.2. Análisis multivariante

4.3.2.1. Análisis exploratorio

La Figura 4.4 muestra el gráfico de Scores 2D para cada clase, junto con un mapa de calor. Los mapas de calor permiten una visualización rápida e intuitiva de los patrones de distribución, utilizando una escala de colores fríos y cálidos. Así, los elementos que presentan un comportamiento similar son representados por zonas cálidas, mientras que los colores fríos corresponden a elementos que se alejan del comportamiento común de los datos. Se explica una varianza total del 83,05% considerando las tres primeras componentes principales (PCs) y se muestran gráficas representadas por PC1 y PC3. Aunque existe una gran superposición entre las diferentes clases, se puede observar que las muestras no adulteradas (clase 1, círculos azules) tienden a agruparse a través de valores de PC3 en su mayoría positivos. Considerando las demás clases, las muestras adulteradas con celulosa microcristalina muestran un comportamiento hacia valores positivos de PC3 y negativos de PC1; mientras que las muestras adulteradas con polvo de madera muestran valores positivos de PC3 y PC1. El resto de las clases adulteradas muestran un comportamiento desigual. Considerando la inspección visual global de PCA, donde existe un gran solapamiento de las clases, se evaluarán técnicas de clasificación de distinta naturaleza.

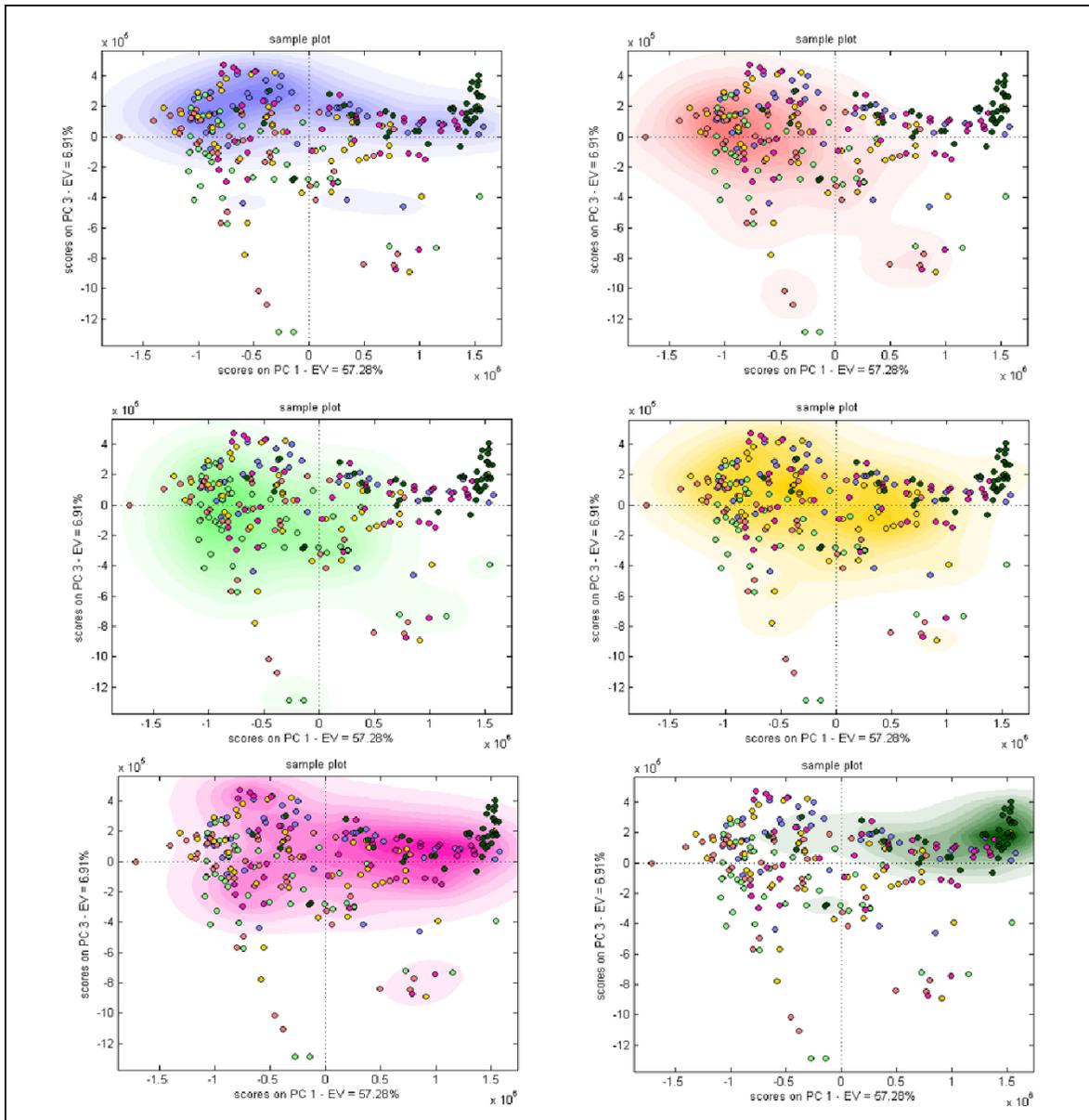


Figura 4.4. Gráfico de Scores PCA obtenidos mediante histogramas RGB: Muestras no adulteradas (*) y muestras adulteradas con: celulosa microcristalina (○), dióxido de silicio (●), harina de trigo (●), sémola de trigo (●) y polvo de madera (●).

4.3.2.2. Resultados de modelos de clasificación

Se evaluaron los parámetros de calidad de clasificación en términos de sensibilidad, precisión, exactitud y Valor-F1, según se detalla en la sección 2.6.4.3. *Parámetros de calidad en clasificación multivariante.*

Teniendo en cuenta los resultados generales, se puede observar que se obtuvieron resultados de clasificación bastante satisfactorios con técnicas de aprendizaje automático en comparación a las técnicas discriminantes (*K*-NN y PLS-DA). Específicamente, SVM reportó los mejores resultados globales. Estos resultados sugieren que las técnicas discriminantes no son adecuadas para abordar la adulteración de quesos rallados con análisis de imágenes digitales, donde existe una alta superposición entre las seis clases. Por el contrario, SVM representa una solución eficiente para separar clases no lineales al construir un límite lineal en una versión grande y transformada del espacio de características (Otto, 2017). A diferencia de los métodos de clasificación convencionales, como el análisis discriminante, no es necesario asumir ninguna distribución de clases con los clasificadores de vectores de soporte.

Tabla 4.3. Resultados de clasificación obtenidos con las distintas técnicas de clasificación estudiadas.

Modelos	Exactitud	Precisión	Sensibilidad	Medida F
SVM	81,7	81,6	81,6	81,5
RTs	73,2	73,1	73,2	73,1
DTs	61,8	62,3	61,8	62,0
LR	57,9	57,6	57,9	57,6
K-NN	51,9	52,8	52,0	52,4
PLS-DA	50,0	55,3	55,0	55,1

SVM: Máquinas de Soporte Vectorial; RTs: Árboles Aleatorios; DTs: Árboles de Decisión; RL: Regresión Logística; *K*-NN: *K* Vecinos más cercanos; PLS-DA: Análisis Discriminante por Mínimos Cuadrados Parciales.

Para explicar en detalle los resultados de clasificación de SVM, la Tabla 4.4 muestra los valores de los parámetros estadísticos obtenidos para cada clase individualmente. En el problema de adulteración en estudio, la implicancia de los errores de clasificación (falsos positivos y falsos negativos) es diferente. Centrándonos en la clase 1 (muestras no adulteradas) los resultados obtenidos en términos de precisión son muy positivos (83%), debido a que la probabilidad de clasificar muestras adulteradas como muestras originales es baja, lo que denota la capacidad del modelo para rechazar futuras muestras adulteradas y evitar posibles riesgos para la salud de la población. Además, la sensibilidad es muy satisfactoria (89%) y se relaciona con falsos negativos. Este resultado sugiere que existe una baja probabilidad de que las muestras originales puedan clasificarse como adulteradas,

evitando así una retirada innecesaria de los mercados y minimizando el riesgo económico. En cuanto a los resultados globales obtenidos de las clases adulteradas, se obtuvo una precisión superior al 81,4% y una sensibilidad superior al 80,2%, indicando un alto porcentaje de asignación correcta de muestras y una buena capacidad para rechazar muestras de otras clases. Finalmente, se obtuvo una sensibilidad aceptable para las muestras adulteradas, lo que señala una alta probabilidad de que una decisión positiva sea correcta.

Tabla 4.4. Resultados de Máquinas de Soporte Vectorial.

Técnica de Clasificación	Parámetros	C1	C2	C3	C4	C5	C6
SVM	Precisión	83	77	75	75	87	93
	Sensibilidad	89	67	77	77	80	100
	Medida F	86	71	76	76	83	97

SVM: Máquinas de Soporte Vectorial; C1: Clase 1 muestras originales; C2: Clase 2 adulteradas con celulosa microcristalina; C3: Clase 3 adulteradas con dióxido de silicio; C4: Clase 4 adulteradas con harina de trigo; C5: Clase 5 adulteradas con sémola de trigo; C6: Clase 6 adulteradas con polvo de madera.

4.3.2.3. Ventajas de la técnica propuesta

La detección de agentes antiaglomerantes permitidos en quesos rallados debe realizarse tanto para el aseguramiento de la calidad como para la prevención del fraude alimentario. Desafortunadamente, no existen métodos simples, confiables y oficiales para determinar agentes antiaglomerantes como celulosa microcristalina o dióxido de silicio en quesos rallados. El método AOAC para la determinación de celulosa en matrices alimentarias es un método gravimétrico que incluye digestión y separación, pero implica procedimientos largos y tediosos y suele dar resultados sobreestimados (AOAC 991.43, 1995; Vazquez-portalatin y Schoenfuss, 2020). Además, este método no está validado para medir celulosa microcristalina en quesos.

Por otro lado, los métodos más comunes para la determinación de dióxido de silicio en alimentos se basan en el uso de espectrometría de emisión atómica de plasma acoplado inductivamente (ICP-OES) y espectrometría de absorción atómica sin llama (AAS) (Ka et al., 2017). Estos métodos no son adecuados para análisis de rutina porque son costosos y de difícil implementación en una línea de producción para analizar matrices alimentarias complejas como son los quesos.

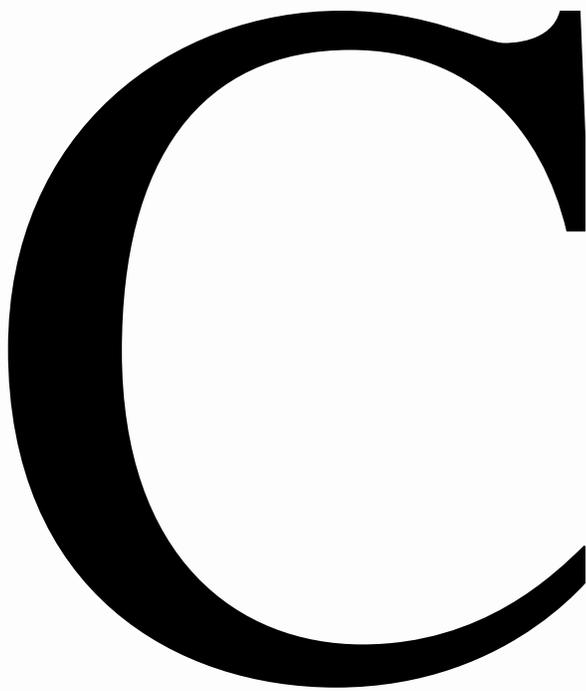
Finalmente, no existen métodos oficiales de determinación de los adulterantes estudiados en el presente trabajo. Por lo tanto, el desarrollo de una herramienta de detección fácil de usar, de bajo costo, portátil y confiable representa una ventaja para realizar análisis *in-situ* para el control de calidad de quesos de pasta dura, con el fin de proteger al consumidor de fraudes alimentarios.

4.5. Conclusiones

Este estudio demuestra la viabilidad de utilizar un dispositivo original de captura de imágenes digitales, junto con el análisis de clasificación multivariado, para identificar la adulteración de quesos rallados. Se evaluaron aditivos antiaglutinantes (celulosa y dióxido de silicio) utilizados en niveles superiores a los permitidos y agregados no permitidos utilizados para aumentar la masa del alimento (harina de trigo, sémola de trigo y polvo de madera). Los resultados de clasificación mostraron que se deben usar modelos no lineales para abordar el problema de adulteración en estudio. En este contexto, se obtuvieron resultados satisfactorios en términos de sensibilidad y precisión global con SVM considerando todas las clases (alrededor del 82%). La principal ventaja de la metodología propuesta es la rápida adquisición de la información analítica mediante un dispositivo de imágenes digitales muy accesible, sin utilizar procedimientos laboriosos, reduciendo tiempos y costos. Además, la imagen digital es una técnica no destructiva y no posee la necesidad de realizar extracciones de muestras, siguiendo los principios de la química verde al no generar residuos.

4.5. Referencias

- AOAC. (1995). Total, soluble, and insoluble dietary fibre in foods, method no. 991.43. *Official methods of analysis of AOAC International (16th ed.)*. Gaithersburg, MD. USA: AOAC International.
- Fernández-Delgado M., Cernadas E., Barro S., y Amorim D. (2014). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research*, 15, 3133-3181.
- Gilbert P. (2022). Photography. *Physics in the Arts*, 5, 77-99.
- Kotsiantis S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatics*, 31, 249-268.
- López M., I., Callao M., P., y Ruisánchez I. (2015). A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. *Analytica Chimica Acta*, 891, 62-72.
- Otto M. (2017). Pattern Recognition and Classification. *Chemometrics, statistics and computer application in analytical chemistry*, 135-211.
- Pedregosa F., G., Varoquaux A. Gramfort V., Michel B., Thirion O., Grisel M., Blondel P., Prettenhofer R., Weiss V., Dubourg J., Vanderplas A., Passos D., Cournapeau M., Brucher M., Perrot É., y Duchesnay. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Vazquez-Portalatin L., y Schoenfuss T. C. (2020). Modifications to established fiber methods may be required to quantify cellulose from flow aids in grated Parmesan cheese. *Dairy Foods: Chemistry and Materials Science*, 1, 1-5.

A large, bold, black letter 'C' is positioned on the left side of the page. It is set against a white background that is partially enclosed by a blue and grey abstract graphic element that curves around it from the top and bottom.

APÍTULO 5

Cuantificación de
adulteraciones y autenticación de
quesos rallados de pasta dura
mediante espectroscopía NIR y
análisis multivariante

CAPÍTULO V: CUANTIFICACIÓN DE ADULTERACIONES Y AUTENTICACIÓN DE QUESOS RALLADOS DE PASTA DURA MEDIANTE ESPECTROSCOPIA NIR Y ANÁLISIS MULTIVARIANTE

5.1. Objetivos

Este trabajo evalúa el potencial del uso de la espectroscopía NIR junto con modelos de regresión y autenticación multivariante para la cuantificación de posibles adulterantes en quesos rallados de pasta dura. La técnica multivariante de regresión empleada es Mínimos Cuadrados Parciales (PLS) y las técnicas empleadas para modelos de autenticación son Mínimos Cuadrados Parciales de una Clase (OC-PLS) y Modelado Independiente y Flexible por Analogía de Clase Impulsado por Datos (DD-SIMCA).

El presente trabajo persigue dos propósitos: (i) la cuantificación de aditivos permitidos (como celulosa microcristalina y dióxido de silicio), en niveles superiores a los reglamentados y (ii) la determinación de agregados no permitidos como harina y sémola de trigo y polvo de madera, a cuatro niveles de adulteración. La idea es proponer una herramienta analítica simple, rápida, de bajo costo y confiable para cuantificar y detectar distintos adulterantes en quesos rallados de pasta dura.

5.2. Antecedentes

En nuestro trabajo anterior (Visconti, Rodríguez, y Di Aníbal, 2020), se utilizó el algoritmo de Análisis Discriminante por Mínimos Cuadrados Parciales (PLS-DA) junto con NIR para la detección de diferentes adulterantes en quesos rallados. PLS-DA es una técnica de análisis discriminante, ya que se ajusta perfectamente a la separación de muestras cuando todas las clases estudiadas están predefinidas. Sin embargo, para resolver problemas de autenticación, especialmente en algunos fraudes alimentarios en particular y autenticación de alimentos, la clasificación de una clase (OCC) es recomendada porque las muestras de la clase objetivo se caracterizan independientemente de las características medidas de otras clases y adulterantes (Rodionova, Titova, y Pomerantsev, 2016). Por otro lado, NIR junto con técnicas de cuantificación como PLS no se han empleado para la determinación cuantitativa de los adulterantes en estudio en quesos de pasta dura rallados.

5.3. Materiales y métodos

5.3.1. Muestras y reactivos

En este trabajo se estudiaron 320 muestras, de las cuales 40 eran quesos de pasta dura rallados originales (no adulterados) y 280 muestras se prepararon añadiendo los siguientes adulterantes a diferentes niveles: celulosa microcristalina, dióxido de silicio, harina y sémola de trigo y polvo de madera, según muestra la Tabla 5.1. Los niveles de adulteración estudiados se seleccionaron en base a las regulaciones vistas en la Tabla 2.1.

Tabla 5.1. Descripción de las muestras en estudio junto al grado de adulteración (% m/m).

Tipo de muestra	Nivel de adulteración (% m/m)
Original (no adulterada)	-
Adulterada con celulosa microcristalina	5, 7, 10, 12
Adulterada con dióxido de silicio	3, 5, 6, 7
Adulterada con harina de trigo	5, 7, 10, 15
Adulterada con sémola de trigo	5, 7, 10, 15
Adulterada con polvo de madera	3, 4, 5, 6

Por otro lado, se adquirieron 10 muestras bajo la denominación "alimento a base de queso rallado y agregados como sémola, sólidos lácteos, harina de arroz, almidones, fibra dietaria, entre otros". Todas las muestras se obtuvieron en puntos comerciales situados en distintas localidades de Argentina (Buenos Aires, Córdoba, Santa Fe y La Pampa). Los aditivos alimentarios permitidos en el queso rallado, utilizados en este trabajo, fueron: celulosa microcristalina (Biopack Argentina) y dióxido de silicio (Merck Argentina). Por otro lado, los agregados no permitidos, como harina y sémola de trigo, se compraron en supermercados locales. Por último, el polvo de madera se obtuvo de aserraderos de la provincia de Misiones (Argentina). Todas las muestras adulteradas se prepararon pesando en balanza analítica y mediante homogeneización manual.

5.3.2. Instrumentación.

Los espectros NIR se adquirieron en un FT-IR Nicolet iS50 equipado con el accesorio de módulo NIR (Thermo Scientific, Madison, WI, EE.UU.) en modo reflectancia difusa. Las muestras se midieron en viales de vidrio en el módulo de esfera integradora por duplicado, rellenando el vial de muestra en cada medición. Se utilizó un material de referencia de teflón como “background”. El rango espectral se situó entre 10.000 y 4.000 cm^{-1} con una resolución de 6 cm^{-1} a temperatura ambiente. Se adquirió un total de 64 barridos o scans por cada medición realizada.

5.3.3. Análisis multivariante

Se evaluaron varios algoritmos de pre-procesamiento espectral para compensar variaciones sistemáticas no deseadas, incluida la variante normal estándar (SNV), la corrección multiplicativa de la dispersión (MSC), la corrección lineal de la línea de base (LBC), la corrección del desplazamiento de la línea de base (BO), el suavizado con la primera derivada empleando un polinomio de segundo orden y un tamaño de ventana de 19 puntos (Savitsky-Golay) (Rinnan, Van Den Berg y Engelsen, 2009).

Todos los procedimientos quimiométricos se realizaron utilizando Matlab 2019a (Mathworks Inc.). Los modelos PCA y PLS se ejecutaron utilizando PLS-Toolbox (Eigenvector Research), mientras que los modelos DD-SIMCA se construyeron utilizando una interfaz disponible en <https://github.com/yzontov/dd-simca> y OCPLS en <https://www.mathworks.com/matlabcentral/fileexchange/47870-ocpls-zip>.

5.3.3.1. Regresión multivariante

Para el análisis cuantitativo se utilizaron las muestras originales y adulteradas descritas en la Tabla 5.1. El conjunto de datos se dividió en calibración (75%) y predicción (25%) mediante el algoritmo de Kennard-Stone. Los modelos de regresión PLS se construyeron y validaron mediante validación cruzada completa para optimizar el número de variables latentes (LV) a incluir en el modelo de calibración. El rendimiento de los modelos se evaluó en términos de los coeficientes de determinación (R^2) y el error cuadrático medio (RMSE) de los conjuntos de calibración y predicción. Además, la capacidad de predicción de los modelos también se evaluó en términos de error relativo de predicción (REP). Por último, se empleó la prueba de la región de confianza conjunta elíptica (EJCR) para investigar la presencia o ausencia de sesgo, es decir, para evaluar la precisión de los modelos en el conjunto de predicción. En esta prueba, en primer lugar, se realiza un ajuste por mínimos

cuadrados ordinarios (MCO); a continuación, los valores estimados de intercepción (a) y pendiente (b) se comparan finalmente con sus valores ideales de $a=0$ y $b=1$, respectivamente. Los detalles de los cálculos de todos los procedimientos quimiométricos y sus respectivas fórmulas están bien descritos en Fernandes et al. (2019).

5.3.3.2. Clasificación multivariante

Para el análisis cualitativo de los datos se aplicó el análisis de componentes principales (PCA) como análisis exploratorio para evaluar tendencias y variabilidad natural de las muestras y eliminar posibles valores atípicos (Bartholomew, 2010). A continuación, se utilizó el algoritmo de Kennard-Stone (KS) para dividir las muestras de cada clase en entrenamiento (75%) y prueba (25%) (Kennard y Stone, 1969). Para construir una regla de clasificación correcta es necesario disponer tanto de muestras genuinas como de un conjunto de muestras adulteradas. Teniendo en cuenta que la adición de las sustancias estudiadas en los niveles de concentración descritos a los quesos rallados se considera adulteración, la clase objetivo se define exclusivamente por muestras auténticas o genuinas de quesos rallados, mientras que las muestras adulteradas se consideran únicamente para probar la capacidad predictiva de los modelos. Los modelos OCC utilizados con fines de autenticación no involucran la identificación de los adulterantes, porque ese tipo de clasificación es definida con modelos discriminantes. Por consiguiente, los modelos OCC se emplean para determinar si una muestra de queso rallado incógnita es reconocida como auténtica.

El conjunto de entrenamiento está compuesto por 30 muestras de quesos rallados auténticos o genuinos para construir los modelos. Las otras 10 muestras de quesos rallados auténticos se incluyeron en el conjunto de prueba, junto con 80 muestras adulteradas (que corresponden a muestras adulteradas en la máxima concentración de cada adulterante), para comprobar la robustez de los modelos. El enfoque quimiométrico evaluado en este trabajo consistió en emplear técnicas de reconocimiento de patrones en las que sólo se modela la clase objetivo, OC-PLS y DD-SIMCA, asemejándose a situaciones reales en las que no es posible caracterizar el tipo y/o la naturaleza de los adulterantes. El rendimiento de los modelos se evaluó en términos de sensibilidad, especificidad y eficiencia, según se describió en la sección 2.6 de la presente tesis. Para la construcción de los modelos OC-PLS, se adoptó un nivel de significancia de 0,05 para las distancias computadas (computed scores distances) (SD) y valores absolutos de residuales centrados. (ACR) También, para los modelos DD-SIMCA se consideraron los errores de tipo I y II y la significancia de valores

atípicos. Para elegir el número óptimo de variables latentes (LV) o componentes principales (PC) en OC-PLS y DD-SIMCA, respectivamente, se escogió el criterio del mejor compromiso entre el número de casos True Positives (TP) obtenidos en los conjuntos de entrenamiento y prueba o, (Lu Xu et al., 2013; Zontov, Rodionova, Kucheryavskiy, y Pomerantsev, 2017).

5.4. Resultados y discusión

5.4.1. Espectros NIR

La Figura 5.1 (a) muestra los espectros NIR de algunas muestras de quesos auténticos y la 5.1 (b) de una muestra aleatoria auténtica y adulterada con todas las sustancias estudiadas. Los máximos de absorbancia en el espectro medio se localizan en $4258,94/4332,76/5154,64/5675,37/5787,04/6858,71/8264,46 \text{ cm}^{-1}$. Las absorbancias de estos números de onda se han atribuido al agua (O-H: $5154,64$ y $6858,71 \text{ cm}^{-1}$), lípidos (bandas de deformación CH_2 : $4258,94$ y $4332,76$, C-H: $5675,37/5787,04$ y $8264,46 \text{ cm}^{-1}$) y proteínas (N-H: 6000 cm^{-1} ; C-H: $5675,37/5787,04$, CH_2 : $4258,94$ y $4332,76$) (G. Downey, 2005). Considerando todo el espectro, se pueden observar algunas diferencias espectrales entre las muestras de queso no adulteradas, principalmente en el rango de 4000 y 5000 cm^{-1} y entre 6000 y 7000 cm^{-1} . Además, los espectros contienen variabilidad no deseada (desplazamientos de la línea de base y efectos de dispersión), por lo que fue necesario aplicar un pre-procesamiento para eliminar y/o minimizar información no deseada. En la Figura 5.1 (b) no se aprecian diferencias entre las muestras auténticas y adulteradas mediante inspección visual, lo que evidencia la aplicación de herramientas quimiométricas para lograr la discriminación entre las distintas muestras.

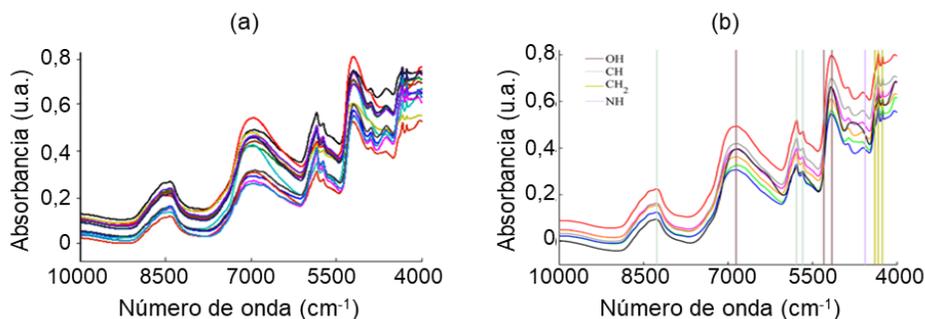


Figura 5.1. Espectros NIR de **(a)** quesos rallados originales y **(b)** una muestra original aleatoria (línea roja) y la misma muestra adulterada con celulosa microcristalina 5% (línea gris), dióxido de silicio 3% (línea azul), harina de trigo 5% (línea magenta), sémola de trigo 5% (línea negra) y polvo de madera 3% (línea verde).

5.4.2. Análisis multivariante

La figura 5.2 resume el procedimiento quimiométrico aplicado a los datos NIR. Como se mencionó anteriormente las muestras adulteradas se preparan a partir de las auténticas añadiendo las cantidades adecuadas de cada adulterante para alcanzar los niveles de concentración estudiados. A continuación, se adquieren todos los espectros NIR y se evalúan varios algoritmos de pre-procesamiento. Finalmente, se construyen los conjuntos de datos para implementar modelos de clasificación con fines de autenticación y modelos de calibración para la cuantificación. Antes de aplicar los modelos multivariantes, el conjunto de datos se divide en conjuntos de calibración y predicción.

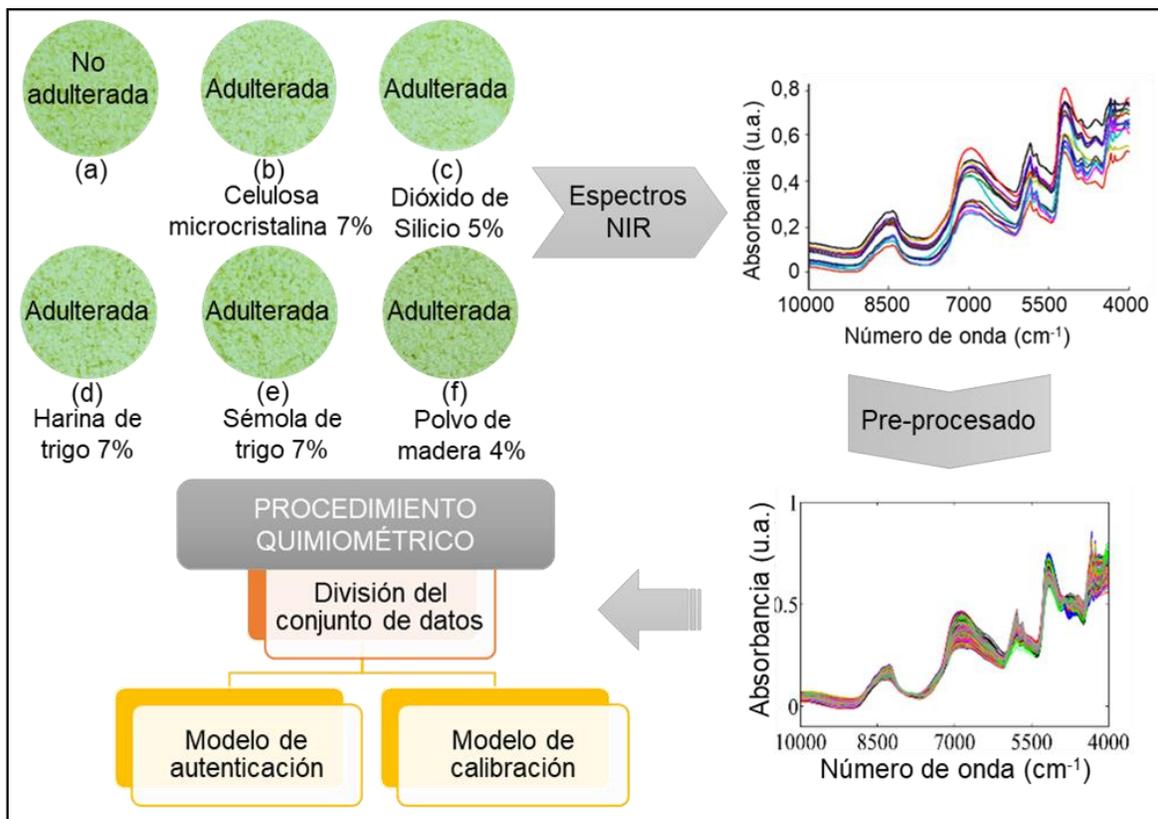


Figura 5.2. Procedimiento quimiométrico

5.4.3. Análisis exploratorio

La Figura 5.3 ilustra los espectros de los quesos rallados pre-procesados (columna derecha) y los gráficos Scores de PCA (columna izquierda). Los pre-procesados empleados en este trabajo fueron los descritos en la sección 2.6.

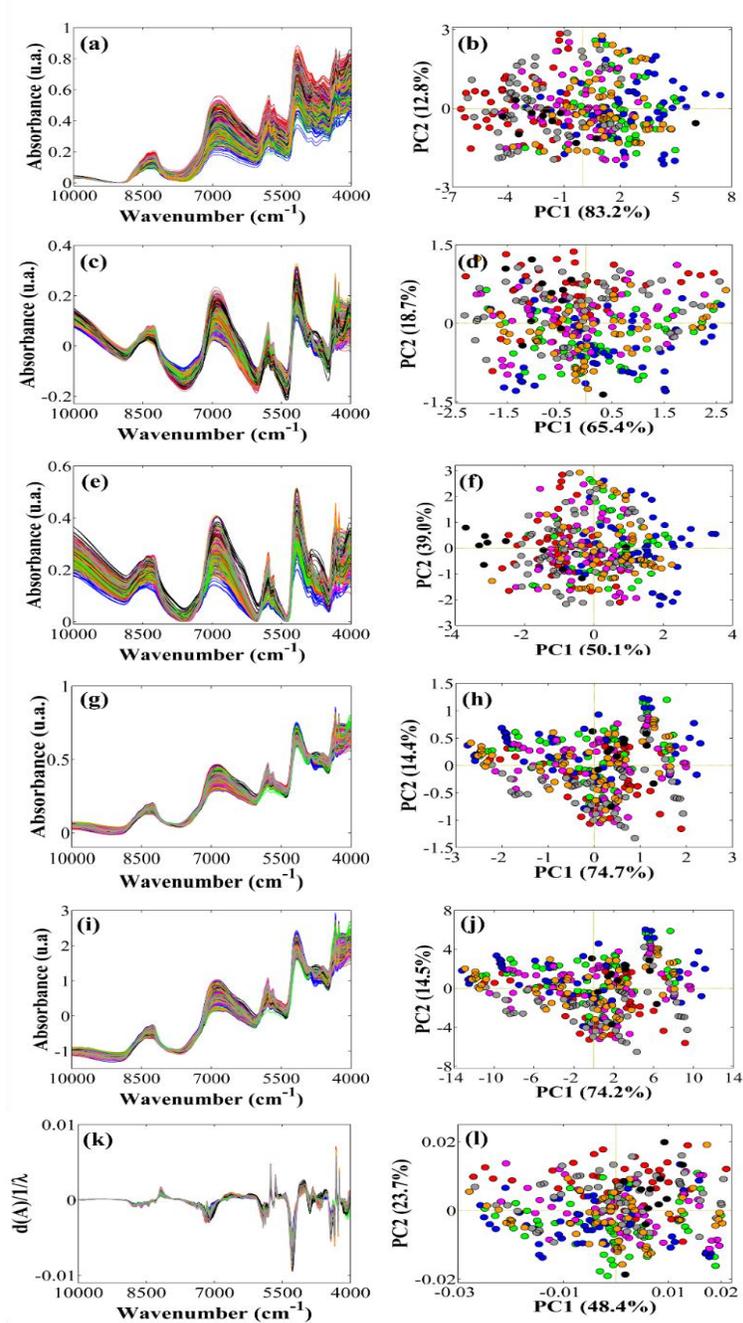


Figura 5.3. Secuencia de las figuras de espectros originales y pre-procesados (izquierda) y gráficos Scores de PCA (derecha). Se muestra PC1 frente a PC2: (a) originales, (c) Corrección lineal de base, (e) Corrección de desplazamiento, (g) MSC, (i) SNV y (k) suavizado de la primera derivada con un polinomio de segundo orden y un tamaño de ventana de 19 puntos; Quesos rallados auténticos (●), quesos con: celulosa microcristalina (●), dióxido de silicio (●), harina de trigo (●), sémola de trigo (●), polvo de madera (●); “alimentos a base de...” (●).

Se puede observar que no es posible una adecuada discriminación debido a un elevado solapamiento entre las muestras de todas las clases en los gráficos de PCA, por lo que a continuación se emplearon técnicas de autenticación para continuar con la problemática estudiada.

5.4.4. Cuantificación de adulterantes

La Tabla 5.2 muestra los resultados para las diferentes cifras de mérito al aplicar la técnica PLS con diferentes pre-procesamientos espectrales. Los modelos de regresión PLS se construyeron y validaron mediante validación cruzada completa para optimizar el número de variables latentes (LVs) que debían incluirse en el modelo de calibración.

Tabla 5.2. Resultados PLS para celulosa microcristalina, dióxido de silicio, harina y sémola de trigo, y polvo de madera, en quesos rallados usando diferentes técnicas de pre-procesamiento espectral.

	RMSECV (mg kg⁻¹)	R²_{CV}	RMSEP (mg kg⁻¹)	R²_{Pred}	RPD	LV
Celulosa microcristalina						
BO	1.5548	0.8830	1.2121	0.9245	3.6383	8
LBC	1.5107	0.8783	1.2202	0.9361	3.9556	8
BO + LBC	1.4114	0.9006	1.4176	0.9027	3.2061	6
MSC	1.5285	0.8867	1.5210	0.9030	3.2106	7
SNV	1.4501	0.8787	1.2166	0.9455	4.2852	10
SG-1d-2o-19p	1.2586	0.9274	1.3629	0.8947	3.0814	11
Dióxido de silicio						
BO	1.2782	0.7497	1.9542	0.6232	1.6290	7
LBC	1.4766	0.6782	1.5785	0.7240	1.9036	10
BO + LBC	1.6236	0.6484	1.6670	0.6276	1.6387	10
MSC	1.8705	0.5009	1.6390	0.7022	1.8325	3
SNV	1.6250	0.6104	2.0567	0.4967	1.4096	3
SG-1d-2o-19p	1.4714	0.6835	1.7382	0.6686	1.7371	7
Harina de trigo						
BO	1.3696	0.9307	1.1335	0.9432	4.1950	11
LBC	1.2205	0.9481	1.3428	0.9398	4.0763	13
BO + LBC	1.2625	0.9432	1.3902	0.9433	4.1979	14
MSC	0.8524	0.9730	1.3172	0.9452	4.2709	13
SNV	1.2957	0.9390	1.1214	0.9677	5.5684	12
SG-1d-2o-19p	1.3299	0.9401	1.2344	0.9393	4.0598	10
Sémola de trigo						
BO	2.0656	0.8574	3.4034	0.5835	1.5495	11
LBC	2.5347	0.7747	3.0952	0.7518	2.0074	8
BO + LBC	2.7600	0.7429	2.3437	0.7923	2.1943	11
MSC	2.8510	0.7550	2.2969	0.7395	1.9591	8
SNV	2.8784	0.7088	2.4156	0.8150	2.3252	8
SG-1d-2o-19p	2.3952	0.7867	2.7550	0.7980	2.2250	7
Polvo de madera						
BO	0.8325	0.8483	0.6994	0.9245	3.6398	10
LBC	0.7290	0.8846	0.7106	0.9243	3.6345	11
BO + LBC	0.8237	0.8541	0.5481	0.9585	4.9085	11
MSC	0.7258	0.8711	0.6560	0.9447	4.2530	11
SNV	0.7144	0.8751	0.5618	0.9572	4.8356	11
SG-1d-2o-19p	0.6215	0.9209	0.7582	0.8997	3.1579	7

RMSECV: error cuadrático medio de validación cruzada; R²CV: coeficiente de determinación para la validación cruzada; RMSEP: error cuadrático medio de predicción; R²Pred: coeficiente de determinación para la predicción; RPD: relación entre el rendimiento y la desviación para la

predicción; REP: error relativo de predicción; LV: variables latentes utilizadas en el modelo. BO: corrección de desplazamiento, LBC: corrección lineal de la línea de base, BO+LBC: corrección de desplazamiento + corrección lineal de la línea de base, MSC: corrección multiplicativa de la dispersión; SNV: variación normal estándar; y SG-1d-2o-19p: suavizado de Savitzky-Golay de primera derivada con un polinomio de segundo orden y un tamaño de ventana de 19 puntos.

En la tabla anterior se observa que los valores de **RMSECV** (Error cuadrático medio de validación cruzada) para los modelos PLS aplicados a celulosa microcristalina (SNV; 1,45), dióxido de silicio (LBC; 1,48), harina de trigo (MSC; 0,85) y polvo de madera (SNV; 0,71) fueron aceptables. En cambio, para el caso de la sémola de trigo (SG-1d-2o-19p; 2,40) el valor obtenido no fue satisfactorio. Asimismo, se obtuvieron valores satisfactorios de **RMSEP** (Error cuadrático medio de predicción) para celulosa microcristalina (SNV; 1,22), dióxido de silicio (LBC; 1,58), harina de trigo (MSC; 1,32) y polvo de madera (SNV; 0,56), no siendo así para sémola de trigo (SG-1d-2o-19p; 2,76). Este comportamiento para el caso de la sémola de trigo puede deberse a la granulometría y a la heterogeneidad que presenta este alimento, lo que podría influenciar en gran medida los espectros NIR obtenidos.

En cuanto al **R²CV** (R-cuadrado de validación cruzada) se obtuvieron resultados favorables (valores entre 0,79 y 0,97) para todos los adulterantes, con excepción del dióxido de silicio (0,68). Esto último, puede deberse a que la técnica espectroscópica NIR es poco sensible a moléculas inorgánicas como el SiO₂, por lo cual es más difícil la cuantificación de las muestras adulteradas en el rango de concentración seleccionado (3%, 5%, 6% y 7% m/m), para este adulterante. De manera análoga los **R²pred** resultan muy buenos para todos los modelos, siendo el dióxido de silicio el valor más bajo por lo anteriormente explicado.

En cuanto al **RPD** (Relación entre rendimiento y desviación), se observan valores no satisfactorios con los modelos de dióxido de silicio y sémola de trigo, lo que indica una capacidad predictiva limitada en relación con la variabilidad de los datos para estos dos adulterantes.

Por otro lado, la Figura 5.4 ilustra las curvas de valores predichos vs valores de referencia (a,c,e,g,i) y los gráficos de elipses de confianza (b,d,f,h,j). Se puede notar que las elipses de confianza de los modelos que contienen dióxido de silicio y sémola de trigo son más amplias lo que indican mayor incertidumbre en la estimación de los parámetros. Estas curvas indican menor precisión y mayor variabilidad. Por otro lado, se observan elipses más estrechas para el resto de adulterantes.

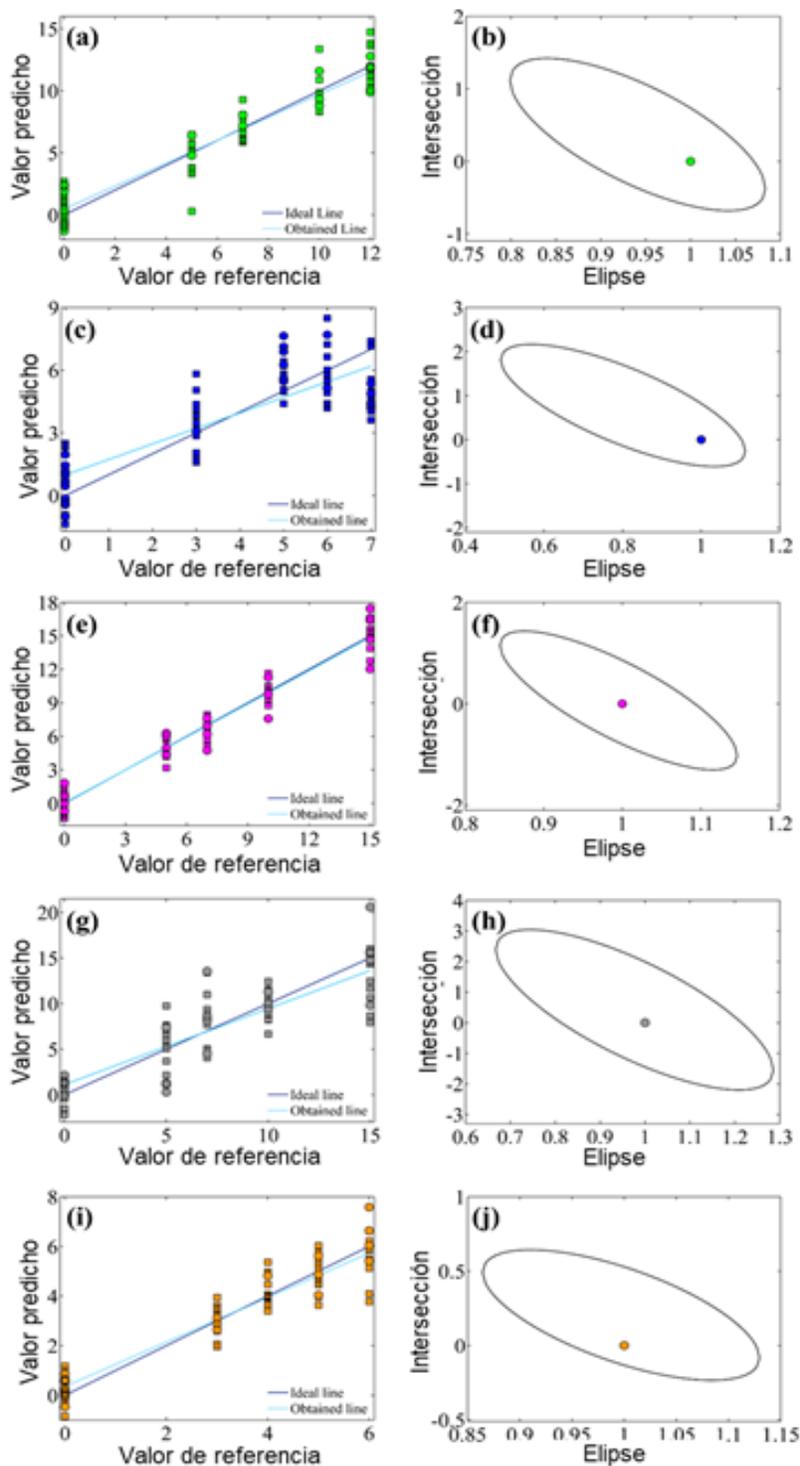


Figura 5.4. Valores predichos frente a valores de referencia en el conjunto de predicciones (a, c, e, g, i) y la respectiva elipse de confianza que incluye el punto ideal teórico (b, d, f, h, j). Muestras con celulosa microcristalina (●), dióxido de silicio (●), harina de trigo (●), sémola de trigo (●) y polvo de madera (●).

5.4.5. Modelos de autenticación

Debido a los resultados obtenidos en regresión, y en aquellos casos donde no podemos cuantificar con alta exactitud ciertos adulterantes como dióxido de silicio y sémola de trigo, se evaluó la implementación de modelos de autenticación como OC-PLS y DD-SIMCA para asegurar de manera alternativa que los quesos rallados no contienen estos adulterantes. La Tabla 5.3 muestra los resultados obtenidos con las técnicas *one-class* en términos de sensibilidad, especificidad y eficacia de los modelos aplicados.

Tabla 5.3. Matriz de confusión, con sensibilidad, especificidad, y eficiencia, obtenidas con DD-SIMCA y OC-PLS para autenticación de muestras no adulteradas de quesos rallados de pasta dura usando diferentes técnicas de pre-procesamiento.

Técnicas de pre-procesamiento	Muestras	DD-SIMCA ($\alpha = 0,05$)				OC-PLS ($\alpha = 0,05$)			
		PCs	Entrenamiento	Testeo		LVs	Entrenamiento	Testeo	
			Target	Target	Adulteradas		Target	Target	Adulteradas
BO	Target	6	28	10	-	11	30	9	1
	Adulteradas		2	24	56		-	31	49
	Sensibilidad		0,93	1,00			1,00	0,90	
	Especificidad		-	0,71			-	0,82	
	Eficiencia		0,87				0,82		
LBC	Target	7	29	9	1	7	30	10	-
	Adulteradas		1	13	67		-	58	22
	Sensibilidad		0,97	0,90			1,00	1,00	
	Especificidad		-	0,84			-	0,28	
	Eficiencia		0,90				0,65		
BO + LBC	Target	7	29	9	1	15	30	10	-
	Adulteradas		1	19	61		-	23	57
	Sensibilidad		0,97	1,00			1,00	1,00	
	Especificidad		-	0,76			-	0,71	
	Eficiencia		0,90				0,89		

MSC	Target	9	30	10	-	8	30	10	-
	Adulteradas		-	3	77		-	14	66
	Sensibilidad		1,00	1,00			1,00	1,00	
	Especificidad		-	0,96			-	0,82	
	Eficiencia		0,99				0,94		
SNV	Target	3	30	10	-	10	30	10	-
	Adulteradas		-	32	48		-	37	43
	Sensibilidad		1,00	1,00			1,00	1,00	
	Especificidad		-	0,60			-	0,54	
	Eficiencia		0,84				0,81		
dX_19	Target	6	29	10	-	8	30	14	-
	Adulteradas		1	13	67		-	48	32
	Sensibilidad		0,97	1,00			1,00	1,00	
	Especificidad		-	0,84			-	0,40	
	Eficiencia		0,93				0,74		

SNV: Variación Normal Standard; MSC: Corrección Multiplicativa de la Dispersión; LBC: Corrección Lineal de la Línea de Base; BO: corrección de desplazamiento; dX_19: suavizado de Savitzky-Golay de primera derivada con un polinomio de segundo orden y un tamaño de ventana de 19 puntos; PCs: Número óptimo de Componentes Principales; LVs: Número óptimo de Variables Latentes

Los mejores resultados obtenidos fueron empleando el pre-procesamiento MSC tanto para el modelo DD-SIMCA como para el OC-PLS. Se obtuvieron excelentes resultados en el caso de DD-SIMCA, con una sensibilidad de 100% y una especificidad y eficacia superiores a un 95%, en ambos conjuntos de datos (entrenamiento y prueba). Los resultados obtenidos por OC-PLS pueden considerarse bastante satisfactorios, con una sensibilidad del 100% (conjunto de entrenamiento y de prueba), una especificidad del 82% y una eficacia del 94%. La máxima sensibilidad obtenida denota la excelente capacidad de ambas técnicas *one-class* para reconocer las muestras de la clase objetivo. Por otro lado, la alta especificidad de DD-SIMCA denota la capacidad del modelo de rechazar correctamente las muestras extrañas, ya que las muestras adulteradas contienen información no incluida en la construcción del modelo de clase objetivo: la contribución de los adulterantes en los espectros NIR.

La Figura 5.5 muestra el área de aceptabilidad obtenida por los modelos DD-SIMCA y OC-PLS construidos utilizando MSC como pre-procesamiento espectral para un nivel de significancia (alfa) de 0,05 en los conjuntos de entrenamiento (a y c) y de prueba (b y d).

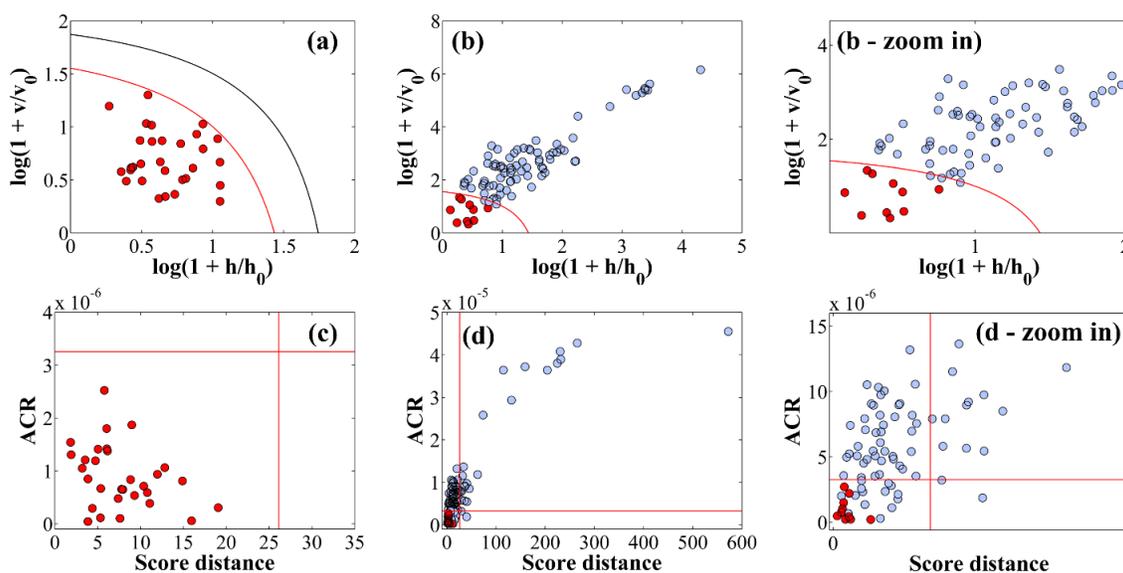


Figura 5.5. Muestras originales (●) Adulteradas (●), DD-SIMCA de (a) entrenamiento y (b) predicción, OC-PLS (c) de entrenamiento y (d) predicción.

Como puede observarse en el paso de entrenamiento para DD-SIMCA y OC-PLS, todas las muestras caen dentro de las áreas de aceptación de ambas técnicas (figuras a y c respectivamente), lo que indica que la clase objeto de estudio (quesos auténticos) es un grupo I que puede modelarse correctamente mediante las estrategias de una clase,

corroborado por la máxima sensibilidad obtenida. Por otro lado, en las figuras con zoom del conjunto de prueba puede verse la superioridad de la técnica DD-SIMCA, ya que muy pocas muestras adulteradas se localizan en la frontera de la clase modelada, lo que se traduce en el alto valor de especificidad obtenido. A partir de los resultados obtenidos, cabe destacar que es posible asegurar la presencia de dióxido de silicio y de sémola de trigo con técnicas *one-class* en quesos rallados, aunque no es posible cuantificarlos de manera satisfactoria en muestras adulteradas.

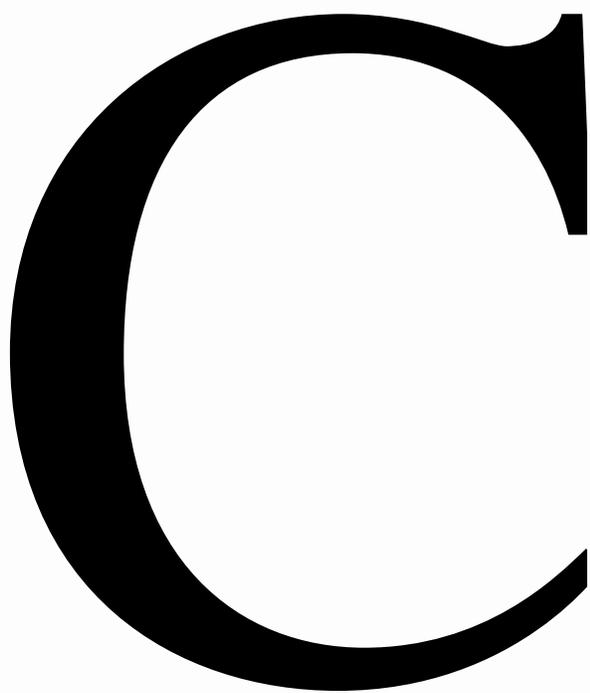
5.5. Conclusiones

Este estudio demostró la viabilidad del uso de la espectroscopía NIR, junto con el análisis multivariante, para cuantificar la adulteración de quesos rallados tanto con aditivos permitidos en cantidades superiores a las reguladas (celulosa microcristalina y dióxido de silicio) como por agregados no permitidos (harina y sémola de trigo y polvo de madera) y, eventualmente, determinar la autenticidad de los mismos.

El análisis cuantitativo PLS determinó de manera satisfactoria diferentes niveles de concentración de celulosa microcristalina, harina de trigo y polvo de madera. Para los casos donde la cuantificación no fue satisfactoria (dióxido de silicio y sémola de trigo), se empleó una estrategia de autenticación del alimento en base a modelos *one-class* como DD-SIMCA, obteniendo excelentes resultados: la máxima sensibilidad (100%) y una excelente especificidad (96%). La principal ventaja del método propuesto es la rápida adquisición de información analítica sin la necesidad de pretratamiento de la muestra. Cada uno de los enfoques propuestos (cuantificación y autenticación) fueron satisfactorios para los diferentes conjuntos de datos estudiados (clases), por lo que representan alternativas complementarias con la estrategia de autenticación, solo hace falta analizar muestras de quesos rallados genuinos, con el consecuente ahorro en tiempo y costos. Por otro lado, la estrategia de cuantificación proporciona una mayor cantidad de información del problema analítico cuantificando el adulterante en cuestión. Por lo tanto, la metodología propuesta resulta una herramienta analítica prometedora para determinar adulteraciones en quesos rallados de pasta dura con el potencial de poder utilizarse para análisis *in-situ* a través de instrumentos NIR portátiles, para ser implementados en laboratorios de control de calidad de alimentos o en la propia cadena de producción.

5.6. Referencias

- Bartholomew, D. J. (2010). Principal components analysis. In P. Peterson, E. Baker, y B. McGaw (Eds.), *International encyclopedia of education*, 374-377). Oxford, UK: Elsevier Science.
- Downey G., Sheehan E., Delahunty C., O'Callaghan D., Guinee T., y Howard V. (2005). Prediction of maturity and sensory attributes of Cheddar cheese using near-infrared spectroscopy. *International Dairy Journal*, 15, 701-709.
- Fernandes D. D. S., Romeo F., Krepper G., Di Nezio M. S., Pistonesi M. F., Centurión M. Araújo E., M. C. U., y Diniz, P. H. G. D. (2019). Quantification and identification of adulteration in the fat content of chicken hamburgers using digital images and chemometric tools. *LWT - Food Science and Technology*, 100, 20–27.
- Kennard R. W., y Stone L. A. (1969). Computer Aided Design of Experiments. *Technometrics*, 11, 137-148.
- Lu Xu, Si-Min Yan, Chen-Bo C., Xiao-Ping Y. (2013). One-class partial least squares (OCPLS) classifier. Chemometrics and Intelligent Laboratory Systems. *Chemometrics and Intelligent Laboratory Systems*, 126, 1-5.
- Rinnan A., Van Den Berg F., y Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28, 1201-1222.
- Rodionova O. Y., Titova A. V., y Pomerantsev A. L. (2016). Discriminant analysis is an inappropriate method of authentication. *TrAC Trends in Analytical Chemistry*, 78, 17–22.
- Visconti L.G., Rodríguez M. S, y Di Anibal C.V. (2020). Determination of grated hard cheeses adulteration by near infrared spectroscopy (NIR) and multivariate analysis, *International Dairy Journal*, 104, 104647,
- Zontov Y.V., Rodionova O.Ye., Kucheryavskiy S.V., y Pomerantsev A.L. (2017). DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach. *Chemometrics and Intelligent Laboratory Systems*, 167, 23-28.

A large, bold, black letter 'C' is positioned on the left side of the page. It is partially enclosed by a thick blue curved line that starts from the top left and curves around the letter. The background is white with a blue and grey abstract graphic element on the right side.

APÍTULO 6

Conclusiones finales

CAPÍTULO VI: CONCLUSIONES GENERALES

Respondiendo al objetivo planteado en esta Tesis se desarrollaron herramientas analíticas para determinar la adulteración de quesos de pasta dura rallados, por el uso de aditivos en cantidades superiores a las establecidas como celulosa microcristalina o dióxido de silicio o por la adición de agregados no permitidos como harina y sémola de trigo o polvo de madera. Se evaluaron diferentes técnicas para la adquisición de la información analítica, como la espectroscopía de infrarrojo cercano (NIR) e imágenes digitales. En relación al análisis multivariante, se evaluaron técnicas de clasificación de diferente naturaleza: discriminantes y de modelado de clase, y dentro de las discriminantes se emplearon técnicas clásicas y de aprendizaje supervisado (*machine learning*). Por otro lado, se plantearon diferentes estrategias para abordar el problema de adulteración estudiado en esta tesis: clasificación discriminante, autenticación o cuantificación. Para este último enfoque se emplearon técnicas de regresión multivariante.

De los trabajos presentados se concluye:

1) El empleo de la espectroscopía NIR junto con el modelo de clasificación de Análisis Discriminante por Mínimos Cuadrados Parciales (PLS-DA) permitió obtener excelentes resultados de clasificación, siendo posible la identificación de los adulterantes estudiados. Cabe destacar que se obtuvo una muy alta especificidad de la clase de muestras originales (sin adulterar) al testear el modelo con muestras adulteradas reales encontradas en locales comerciales. Esto es sumamente positivo, ya que el riesgo de asignar muestras adulteradas como originales es prácticamente nulo.

2) El dispositivo de toma de imágenes desarrollado permitió adquirir las imágenes digitales, de manera reproducible. Este trabajo demostró la necesidad de utilizar técnicas no lineales como las Máquinas de Soporte Vectorial (SVM), que se encuentran dentro del grupo de algoritmos de *machine learning*. Se obtuvieron resultados de clasificación en términos de sensibilidad, exactitud y precisión, por encima del 80% (resultados globales), con una especificidad de la clase sin adulterar de un 89%. De esta manera, el dispositivo construido es novedoso y original y su uso representa una segunda alternativa más económica para identificar los adulterantes estudiados.

3) El uso de la espectroscopía NIR junto con modelos de regresión multivariante permitió no sólo identificar posibles adulterantes, sino también cuantificarlos. Su importancia radica especialmente para el estudio de los aditivos permitidos, cuando son utilizados en concentraciones superiores a las establecidas en la reglamentación vigente.

Por otro lado, el enfoque de autenticación a través del empleo de técnicas multivariantes *one-class* como DD-SIMCA, permitió la correcta identificación de todas las muestras modeladas de la clase objetivo (clase de muestras sin adulterar). Cabe destacar también la elevada especificidad obtenida (96%), por lo que el riesgo de asignar muestras adulteradas como originales es muy bajo. Este enfoque posee la ventaja del ahorro de tiempo y costos de análisis ya que sólo deben modelarse las muestras provenientes de la clase de interés.

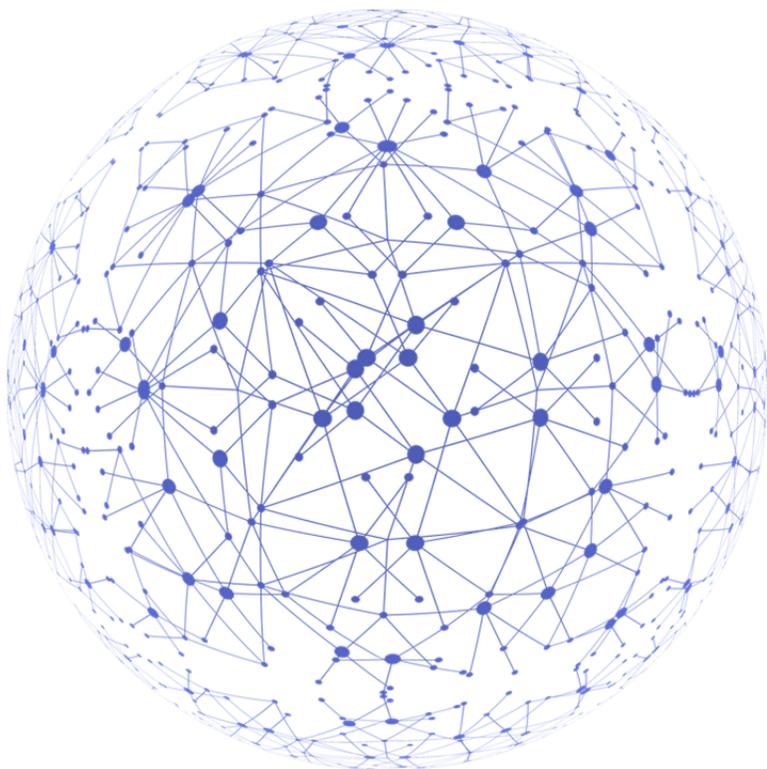
De las técnicas analíticas empleadas se concluye que la espectroscopía NIR y las imágenes digitales son aplicables a la problemática estudiada donde sobresalen sus ventajas como la no destrucción de la muestra, no requerir procesos de extracción y no generar residuos, siguiendo los principios de la química verde. Por otro lado, presentan el valor adicional de poder emplearse para análisis *in-situ*, debido a sus condiciones de portabilidad.

Con respecto al análisis multivariante:

- El análisis exploratorio de datos es una herramienta útil para tener una visualización previa de comportamiento inherente de los datos, lo que posibilita un mejor entendimiento de los mismos.
- Los algoritmos de pre-procesamiento espectral son necesarios para extraer la información relevante de espectros NIR y obtener resultados de clasificación/cuantificación satisfactorios.
- La técnica de clasificación que proporcione los mejores resultados de clasificación depende de la naturaleza de los datos. En este sentido, existen numerosas técnicas en bibliografía que se diferencian matemáticamente entre ellas y que se pueden adaptar a la problemática en estudio. Por otro lado, también poseen la versatilidad de poder emplearse para un enfoque de discriminación o de autenticación. En el primer caso, es necesario tener un conocimiento previo de los posibles adulterantes y se debe contar con un conjunto completo de muestras, es decir, muestras sin adulterar y adulteradas. En el segundo caso, se pueden emplear sin tener un conocimiento previo de los posibles adulterantes, ya que sólo se focaliza en el modelado de la clase de interés (en este caso muestras auténticas o no adulteradas), con el consecuente ahorro en tiempo y costos al analizar sólo una clase de muestras, pero no se cuenta con la cantidad de información del problema analítico que proporciona el enfoque discriminante.

- Las técnicas de regresión son útiles para cuantificar muestras que han sido clasificadas como adulteradas por un análisis de screening previo. De esta manera, resulta muy útil implementar un screening previo de todas las muestras para identificar los posibles adulterantes y posteriormente cuantificarlos.

Los métodos analíticos desarrollados representan una herramienta muy útil para determinar la adulteración de alimentos de elevado valor comercial como son los quesos rallados, los cuales son susceptibles de sufrir adulteraciones para obtener beneficios comerciales. De esta manera se presentan alternativas para proteger al consumidor de estas prácticas fraudulentas y evitar el detrimento nutricional de los quesos. Asimismo, se destaca su rapidez, simplicidad, fiabilidad y su relativo bajo costo económico comparado con métodos empleados habitualmente para determinar los adulterantes en estudio. Finalmente, presentan el potencial de poder implementarse en laboratorios de control de calidad de alimentos y en cualquier estadio de control de calidad de la industria láctea, por todas las ventajas anteriormente mencionadas.



APÉNDICE

PRODUCTIVIDAD

Licenciado en Química Visconti Lucas Gabriel – Universidad Nacional del Sur

APENDICE

A.1. Artículos científicos presentados en revistas internacionales.

Determination of grated hard cheeses adulteration by near infrared spectroscopy (NIR) and multivariate analysis • 2020 •

Article type: Research Article

Scientific Journal: International Dairy Journal, Vol. 104 (2020), page 104647

Impact Factor: 2,735

Index Q: Q1

First Author: Lucas Visconti, PhD student

Order of Authors: Lucas Visconti, PhD student; M. Susana Rodríguez, PhD; Carolina Di Anibal, PhD.

Abstract: Cheese is a dairy product of massive consumption and is commonly commercialized as grated cheese. Adulteration of cheese reduces its quality and consequently, decreases nutritional value. This work proposes an analytical tool to determine the adulteration of cheeses with higher levels of additives than those permitted (cellulose and silicon dioxide) and nonauthentic substances used to increase weight and volume such as wheat-flour, wheat-semolina and sawdust. The proposed methodology is based on Near Infrared Spectroscopy (NIR) and multivariate classification analysis. Excellent classification results were obtained with Partial Least Squares Discriminant Analysis (PLS-DA) as adulterated samples were discriminated from original (unadulterated) samples. The prediction ability was assessed with adulterated commercial grated cheeses. The screening tool represents a fast, reliable and affordable way to perform quality control of grated cheeses in order to avoid consumer deception based on nutritional detriment and misleading labelling.

Keywords: Cheese adulteration, food fraud, quality control, NIR, multivariate analysis.

DOI: <https://doi.org/10.1016/j.idairyj.2020.104647>

Determination of grated hard cheeses adulteration by digital image analysis and multivariate analysis • 2022•

Article type: Research Article

Scientific Journal: International Dairy Journal, Vol. 138, Pages 105539

Impact Factor: 2,735

Index Q: Q1

First Author: Lucas Visconti, PhD student

Order of Authors: Lucas Visconti, PhD student; Steven Martínez Vargas, PhD student; M. Susana Rodríguez, PhD in Chemistry; Carolina Di Anibal, PhD in Chemistry; Claudio Delrieux, PhD in Computer Science.

Abstract: Hard-cheeses are foods with high commercial value, thus fraudulent practices are frequent and difficult to detect, which reduces hard-cheese quality and consequently, decreases nutritional value. This work presents a novel and feasible tool to identify the adulteration of grated-cheese with higher levels of additives than those permitted (cellulose and silicon dioxide), and foreign substances used to increase weight and volume (wheat-flour, wheat-semolina and sawdust). The methodology is based on the use of colour

histograms (obtained from digital images) and multivariate classification analysis. Satisfactory results were obtained with support vector machines with average values of accuracy and sensitivity above 82%, as adulterated samples were discriminated from original (unadulterated) samples. The proposed screening methodology represents a fast, simple, reliable and affordable way to perform quality control of grated-cheeses to avoid food fraud and nutritional detriments of worldwide consumption foods such as hard-cheeses.

Keywords: Digital Image Processing, Food Adulteration, Multivariate Analysis

DOI: <https://doi.org/10.1016/j.idairyj.2022.105539>

A.2. Artículos científicos actualmente en redacción.

Grated hard cheese authentication and determination of its adulteration with noncompliance and compliance substances using FT-NIR spectroscopy and multivariate analysis • 2023•

Lucas Gabriel Visconti, Cynthia Giselle Reile, David Douglas de Sousa Fernandes, Paulo Henrique Gonçalves Dias Diniz, María Susana Rodríguez y Carolina Vanesa Di Anibal.

A.3. Contribuciones en reuniones científico-tecnológicas

• 2018 • Determinación de la adulteración de quesos rallados mediante técnicas espectroscópicas y análisis multivariante.

Evento: VIII Congreso de Alimentos Siglo XXI: Alimentos, Nutrición y Salud.

Lugar: Mendoza, Mendoza, Argentina

Tipo de trabajo: Disertación.

ISSN: 2525-1783

Autores: Lic. Lucas Visconti, Dra. Liliana Albertengo, Dra. M. Susana Rodríguez, Dra. Carolina Di Aníbal.

• 2019 • Espectroscopía NIR junto con clasificación multivariante para la determinación de la adulteración de quesos rallados.

Evento: X Congreso Argentino de Química Analítica

Lugar: Santa Rosa, La Pampa, Argentina

Tipo de trabajo: Resumen – Póster

ISBN 978-950-863-375-0

Autores: Lic. Lucas Visconti, Dra. M. Susana Rodríguez, Dra. Carolina Di Anibal

• 2021 • Determinación de la adulteración de quesos rallados mediante imágenes digitales y clasificación basada en aprendizaje automático.

Evento: IX Congreso de Alimentos Siglo XXI: Alimentos, Nutrición y Salud

Lugar: Virtual

Tipo de trabajo: Resumen – Póster

ISBN: 978-987-83-5227-5

Autores: Lic. Lucas Visconti, Lic. Steven Martínez Vargas, Dr. Claudio Delrieux, Dra. M. Susana Rodríguez, Dra. Carolina Di Aníbal

• 2021 • Análisis cuantitativo de adulterantes en quesos rallados por espectroscopía de infrarrojo cercano (NIR) y regresión PLS

Evento: XI Congreso Argentino de Química Analítica

Lugar: Virtual

Tipo de trabajo: Resumen – Póster

ISBN: 978-987-88-5110-5

Autores: Lic. Lucas Visconti, Lic. Cynthia Reile, Dra. M. Susana Rodríguez, Dra. Carolina Di Anibal, Dr. David de Sousa Fernandes, Dr. Paulo Dias Diniz

• 2021 • Novedoso dispositivo de captura de imágenes digitales junto con técnicas de aprendizaje automático para determinar la adulteración de quesos rallados.

Evento: XI Congreso Argentino de Química Analítica

Lugar: Virtual

Tipo de trabajo: Resumen – Póster

ISBN: 978-987-88-5110-5

Autores: Lic. Lucas Visconti, Lic. Steven Martínez Vargas, Dr. Claudio Delrieux, Dra. M. Susana Rodríguez, Dra. Carolina Di Anibal

• 2022 • Metodología basada en imágenes digitales para el análisis de nanopartículas bimetalicas de Au/Ag obtenidas mediante síntesis verde.

Evento: XXI Encuentro de Superficies y Materiales Nanoestructurados

Lugar: Rio cuarto, Córdoba (Virtual)

Tipo de trabajo: Resumen – Póster

Autores: Lic. Lucas Visconti, Lic. Juan Pablo Arrué, Lic. Cynthia Reile, Dr. Marcelo Pistonesi, Dra. Valeria Springer