

**ESCUELA TÉCNICA SUPERIOR DE INGENIEROS
INDUSTRIALES Y DE TELECOMUNICACIÓN**

UNIVERSIDAD DE CANTABRIA



Proyecto Fin de Carrera

**OPTIMIZACIÓN DEL ALGORITMO DE
MANCHAS ESPECTRALES PARA LA
IDENTIFICACIÓN DE COMPUESTOS**

**(Optimization of the spectral marks algorithm for materials
identification)**

Para acceder al Título de

INGENIERO SUPERIOR DE TELECOMUNICACIÓN

Autor: Rodrigo Nevado Antón

Octubre - 2013

INGENIERÍA TÉCNICA DE TELECOMUNICACIÓN

CALIFICACIÓN DEL PROYECTO FIN DE CARRERA

Director del PFC: Olga María Conde Portilla

Título: “Optimización del algoritmo de manchas espectrales para la identificación de compuestos”

Title: “Optimization of the spectral marks algorithm for materials identification”

Presentado a examen el día:

Para acceder al Título de

INGENIERO SUPERIOR DE TELECOMUNICACIÓN

Composición del Tribunal:

Presidente (Apellidos, Nombre):

Cobo García, Adolfo

Secretario (Apellidos, Nombre):

Conde Portilla, Olga M^a

Vocal (Apellidos, Nombre):

Suárez Rodríguez, Almudena

Este Tribunal ha resuelto otorgar la calificación de:

Fdo.: El Presidente

Fdo.: El Secretario

Fdo.: El Vocal

Fdo.: El Director del PFC
(Sólo si es distinto del Secretario)

Vº Bº del Subdirector

Proyecto Fin de Carrera N°
(a asignar por Secretaría)

Agradecimientos:

Me gustaría expresar en estas líneas mi agradecimiento a todas aquellas personas que han estado a mi lado todos estos años, por su paciencia y comprensión.

En primer lugar quiero expresar mi agradecimiento a la directora de este trabajo de fin de carrera, D^a Olga María Conde Portilla, por el apoyo y dedicación que me ha brindado en estos últimos años que he pasado en la Facultad y muy especialmente en estos últimos meses en la realización de este trabajo.

También quiero agradecer muy especialmente la colaboración de Jean Pierre Román, Lucía Uriarte Calva que han realizado partes precedentes de este estudio sin las cuales yo no podría haber continuado con este proyecto, muchísimas gracias.

Aunque también llegar a realizar un trabajo como este es fruto del reconocimiento y del apoyo vital que nos ofrecen personas que nos estiman, que son las que realmente las que nos dan fuerza y energía día tras día para llegar a ser personas de provecho en un futuro. Por ello quiero transmitir un agradecimiento especial a mi familia; que sin su perseverancia no habría logrado esta meta.

Y como no, a todos mis compañeros de clase de todos estos años, gracias por todos esos buenos momentos que hemos pasado juntos.

Muchas Gracias.

Índice:

1. Introducción

1.1 Introducción y motivación.....	2
1.2 Objetivos.....	6
1.3 Estructura del documento	7

2. Espectroscopia

2.1 Fenómeno de interacción luz-materia.....	9
2.2 Motivo del estudio	10
2.3 Tipos de espectrometría.....	11
2.3.1 Transmisión	11
2.3.2 Reflectancia difusa	12

3. Manchas espectrales

3.1 Introducción a las manchas espectrales	15
3.1.1 Percentiles.....	15
3.1.2 Resolución	17
3.1.3 Dominio espectral y PCA	18
3.1.4 Matriz de confusión	21
3.2 Tipos de validación cruzada	22
3.2.1 Shuffle	22
3.2.2 Lotes	22
3.2.3 Aleatorio	23
3.2.4 Leave-One-Out	23
3.3 Descripción del algoritmo	24

4. Mejoras introducidas

4.1 Procesado del color de una mancha espectral	28
4.2 Validación cruzada, leave-one-out	30
4.3 Espectroscopia hiper-espectral	30
4.4 Aumento del rendimiento	33
4.5 Memory Leaks.....	34
4.6 Otras Mejoras	34

5. Resultados

5.1 Materiales analizados	36
5.1.1 Juego de datos para la espectroscopia de transmisión.....	36
5.1.1.1 Aceite.....	36
5.1.1.2 Café.....	38
5.1.1.3 Vino	39
5.1.2 Juego de datos para espectroscopia híper-espectral	41
5.2 Validación cruzada, Shuffle	44
5.3 Validación cruzada, Leave-one-out.....	54
5.4 Clasificación de datos híper-espectrales.....	56

6. Conclusiones

6.1 Conclusiones.....	69
6.2 Líneas futuras	69

Referencias

Anexos

Anexo 1: Medidas tomadas, Shuffle	73
Anexo 1.1: Aceite	73
Anexo 1.2: Café.....	85
Anexo 1.3: Vino	97
Anexo 2: Medidas tomadas, Leave-one-Out	109
Anexo 2.1: Aceite	109
Anexo 2.1.1: Dominio espectral	109
Anexo 2.1.2: Dominio PCA (Principal Component Analysis).....	112
Anexo 2.2 Café	114
Anexo 2.2.1: Dominio espectral	114
Anexo 2.2.2: Dominio PCA (Principal Component Analysis).....	116
Anexo 2.3: Vino	117
Anexo 2.3.1: Dominio espectral	117
Anexo 2.3.2: Dominio PCA (Principal Component Analysis).....	120

“OPTIMIZACIÓN DEL ALGORITMO DE MANCHAS ESPECTRALES PARA LA IDENTIFICACIÓN DE COMPUESTOS”

Palabras Clave:

Espectroscopia, manchas espectrales, PCA, espectro, absorbancia, transmitancia, validación cruzada, percentiles, resolución, RAW, leave-one.out.

1. Introducción

1.1 Introducción y motivación

Dentro del sector agroalimentario existe un órgano intergubernamental denominado *Comisión del Codex Alimentarius* constituida por más de 170 miembros en el marco del Programa Conjunto FAO/OMS (Organización de las Naciones Unidas para la Agricultura y la Alimentación / Organización Mundial de la Salud) sobre Normas Alimentarias que tiene por objeto proteger la salud de los consumidores y asegurar prácticas equitativas en el comercio de alimentos. La Comisión también promueve la coordinación de todos los trabajos sobre normas alimentarias emprendidos por las organizaciones internacionales gubernamentales y no gubernamentales.

El Codex Alimentarius es el resultado del trabajo de la comisión: un compendio de normas alimentarias, directrices y códigos de prácticas adoptados internacionalmente. Entre ellas se encuentra el etiquetado de los alimentos, que constituye el principal medio de comunicación entre los productores y vendedores de alimentos por una parte, y por otra, sus compradores y consumidores. Las normas y directrices del Codex Alimentarius sobre Etiquetado de los Alimentos son publicadas en formato compacto para permitir su uso y amplio conocimiento por parte de los gobiernos, las autoridades de reglamentación, las industrias de alimentos y minoristas, y los consumidores.

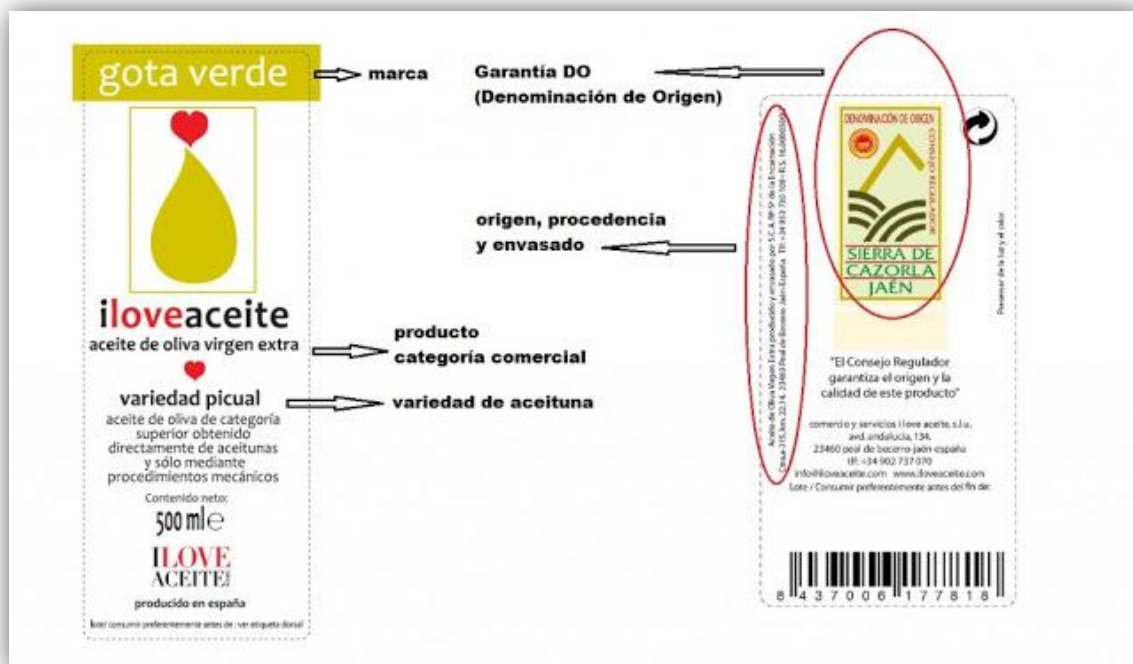


Figura 1: Como leer una Etiqueta de Aceite de Oliva

Existen laboratorios destinados a inspeccionar la correcta composición de los alimentos, indicadas en sus etiquetas, por sus productores. En la Figura 2, se muestra un estudio que hizo la OCU [22] de 40 marcas de aceite, expuestas en el mercado.

ACEITE DE OLIVA						
Marcas	Envase	SE VENDEN COMO... VIRGEN EXTRA			Precio medio por litro	Valoración de la OCU
		Y en realidad son:	Análisis calidad del fruto	prueba de cata		
OLEOESTEPA D.O. Estepa	Plástico	Virgen Extra	■	+	3,35	Buena calidad. ★ Mejor del Análisis
YBARRA. Gran Selección Añonado	Vidrio	Virgen Extra	■	+	4,53	Buena calidad. ★ Mejor del Análisis
HACENDADO (Mercadona).	Plástico	Virgen Extra	■	+	2,76	Buena calidad. 🎯 Compra Maestra
LA ESPAÑOLA Gran Selección	Vidrio	Virgen Extra	■	+	4,02	Buena calidad
DINTEL Selección Especial	Plástico	Virgen Extra	■	+	2,75	Buena calidad. 🎯 Compra Maestra
HACENDADO (Mercadona). Vidrio	Vidrio	Virgen Extra	■	+	3,73	Buena calidad
EL CORTE INGLÉS	Plástico	Virgen Extra	■	+	2,9	Buena calidad. 🎯 Compra Maestra
CARBONELL Gran Selección	Vidrio	Virgen Extra	■	+	4,5	Buena calidad

DÍA Frutado	Vidrio	Virgen Extra	☒	☐	3,46	Calidad media
LA ESPAÑOLA	Plástico	Virgen Extra	☒	☐	3,42	Calidad media
SEÑORÍO DE SEGURA Picual D.O.	Vidrio	Virgen Extra	☒	☐	4,94	Calidad media
AUCHAN (Alcampo)	Vidrio	Virgen Extra	☒	☐	3,42	Calidad media
CARREFOUR	Vidrio	Virgen Extra	☒	☐	3,71	Calidad media
OLIVAR DE SEGURA Picual	Plástico	Virgen Extra	☒	☐	3,59	Calidad media
KOIKE	Plástico	Virgen Extra	+	☐	3,15	Calidad media
BORGES	Plástico	Virgen Extra	☒	☐	3,22	Calidad media
CARBONELL Selección Maestra	Plástico	Virgen Extra	☒	—	3,13	Calidad media
SÚPER (El Árbol)	Plástico	Virgen Extra	+	☐	2,75	Calidad media
CARREFOUR	Plástico	Virgen Extra	☒	—	2,73	Calidad media
HOJIBLANCA	Vidrio	Virgen Extra	☒	—	4,25	Calidad media
SUPERSOL	Plástico	Virgen Extra	☐	☐	2,84	Calidad media
AUCHAN (Alcampo)	Plástico	Virgen Extra	☐	—	2,66	Calidad media
CARREFOUR	Vidrio	Virgen Extra	☒	☐	3,71	Calidad media
OLIVAR DE SEGURA Picual	Plástico	Virgen Extra	☒	☐	3,59	Calidad media
KOIKE	Plástico	Virgen Extra	+	☐	3,15	Calidad media
BORGES	Plástico	Virgen Extra	☒	☐	3,22	Calidad media
CARBONELL Selección Maestra	Plástico	Virgen Extra	☒	—	3,13	Calidad media
SÚPER (El Árbol)	Plástico	Virgen Extra	+	☐	2,75	Calidad media
CARREFOUR	Plástico	Virgen Extra	☒	—	2,73	Calidad media
HOJIBLANCA	Vidrio	Virgen Extra	☒	—	4,25	Calidad media
SUPERSOL	Plástico	Virgen Extra	☐	☐	2,84	Calidad media
AUCHAN (Alcampo)	Plástico	Virgen Extra	☐	—	2,66	Calidad media
CORDOLIVA Frutado	Plástico	Virgen Extra	+	—	3,37	Calidad media
EROSKI	Plástico	Virgen	☒	☹	2,78	No recomendado
HOJIBLANCA	Plástico	Virgen	☒	☹	3,16	No recomendado
COOSUR	Plástico	Virgen	+	☹	3,29	No recomendado
YBARRA	Plástico	Virgen	☒	☹	2,96	No recomendado
CONSUM	Plástico	Virgen	☒	☹	2,75	No recomendado
ARTEOLIVA	Brik	Virgen	☒	☹	3,8	No recomendado
CONDIS	Plástico	Virgen	☒	☹	2,89	No recomendado
OLI SONE (Lidl)	Vidrio	Virgen	☒	☹	3,44	No recomendado
ALIADA (El Corte Inglés)	Plástico	Virgen	☐	☹	2,75	No recomendado
MAEVA	Plástico	Lampante	☹	☹	2,9	No apto para venta
SE VENDEN COMO ACEITE DE OLIVA VIRGEN						
MAR DE OLIVOS (Olivar de Segura)	Plástico	Virgen	☒	☐	2,32	Calidad media
CARREFOUR	Plástico	Virgen	☒	☐	2,38	Calidad media
CORDOLIVA Tradicional	Plástico	Virgen	☒	☐	2,84	Calidad media
AUCHAN Pulgar (Alcampo)	Plástico	Virgen	☐	☐	2,17	Calidad media
SABOR Y ORIGEN (M)	Plástico	Virgen	☹	☹	2,16	No recomendado
OLILÁN	Plástico	Lampante	—	☹	2,35	No apto para venta

Figura 2: Estudio sobre diferentes Marcas de Aceite de Oliva [22]

Los resultados que se obtuvieron fueron: hay nueve productos que se venden como aceite de oliva virgen extra, y no alcanzan los atributos y características que deberían reunir para poder ser considerados tales. Peor todavía es el caso de dos de los productos, que por sus valores analíticos, y por los datos obtenidos en la prueba de cata, demuestran que son aceites denominados “lampantes”, aún no refinados, no deberían estar a la venta

Como se puede observar, estos estudios ponen de manifiesto que muchos productores no siguen fielmente lo que han marcado en las etiquetas, seguramente por temas económicos que pueden repercutir en temas monetarios hacia el consumidor o, peor aún, atentar contra su propia salud. Además, se observa que los métodos de análisis son realizados mediante catas del producto que, aunque necesarias, ofrecen una valoración bastante subjetiva del mismo.

Además de la situación expuesta, también se podrían contemplar estudiar los cambios en la composición de los alimentos. Un caso práctico podría ser la mezcla de cafés de diferentes zonas, y con diferentes tostados, para crear sabores particulares para cada tipo de paladar. Así, en el caso de que una especialidad tenga éxito, su reproducción sea totalmente semejante a su predecesora, ya que tal como ocurre con los vinos, el grano de café ofrece multitud de diferencias de sabor dependiendo del lugar donde se ha realizado su cultivo. Además, en cada región, el café difiere, ampliando aún más la gama de variedades, sabores y aromas.



Figura 3: Tipos de Cafe

En este proyecto se propone emplear técnicas espectroscópicas para asegurar la procedencia de diferentes tipos de alimentos, así como implementar métodos de catalogación automatizados que sean útiles y fiables para el contexto agroalimentario, y extensible a otros ámbitos, que permitan desterrar prácticas no muy aconsejables. La espectroscopia es una ciencia que surgió del estudio de la interacción entre la radiación y la materia como función de la longitud de onda (λ). Para realizar esta técnica, el dispositivo empleado para las medidas espectrales se denomina espectrómetro o espectrógrafo. El análisis espectroscópico se puede plantear para cualquiera de los estados de la materia: sólido, líquido y gaseoso. En este proyecto se analizarán compuestos líquidos mediante espectroscopia de absorción y compuestos sólidos mediante análisis hiper-espectral de la reflectancia difusa. El análisis hiper-espectral difiere de un montaje en transmisión en el tipo y disposición de los dispositivos empleados para obtener el espectro. A partir de aquí, y una vez obtenido el espectro, al análisis del mismo para su catalogación es completamente análogo en las dos situaciones.

Así, este Trabajo Fin de Carrera (TFC) se centra en la ampliación de funcionalidad sobre la interfaz gráfica desarrollada en el TFC previo de la ITTSE [5] dedicado a la identificación de diversos materiales. En el sector agroalimentario se ha estudiado aceite, vino y café. Además se ha realizado un intenso trabajo de clasificación de diversos materiales a partir de imágenes hiper-espectrales (cartón, tabaco, papel) con el fin de poder verificar si la capacidad de clasificación de manchas espectrales es correcta y de paso, comprobar el comportamiento de funcionalidades agregadas. El resultado del estudio será de interés al llevarse al terreno productivo donde se puede emplear para verificar los errores en el etiquetado, la identificación de componentes nocivos, mezclas de diferentes tipos y materiales, etc.

La motivación para realizar este proyecto es sobre todo muy estimulante, puesto que la rama de fotónica es la que siempre más me ha atraído y además de tener un interés increíble en poder realizar un versionado del algoritmo e interfaz creado durante mi proyecto de ITTSE. Ya que se ha tratado de realizar una mejora sobre la tasa de acierto respecto a la discriminación conseguida en trabajos anteriores, mejorando alguna de las técnicas usadas antes. Además se han incorporado funcionalidades y mejoras sobre la interfaz gráfica creada, para que el usuario al recoger los datos obtenidos del espectrómetro no tenga que estar realizando esa ardua tarea de programación y cambio de variables dentro del código fuente, sino que de esta manera se le facilita una opción mucho más rápida y mucho más fácil de visualizar que directamente sobre el código fuente.

La elaboración de este proyecto consiste en realizar las tareas anteriormente mencionadas además de llevar a cabo un estudio con las muestras anteriormente citadas, puesto que se ha podido disponer de unas medidas de absorbancia de tres materiales diferentes y multitud de imágenes hiperespectrales de diversos materiales. Al confeccionar el proyecto se han tenido en cuenta estos materiales, pero la finalidad del programa es de carácter general: consistente en un análisis discriminatorio, eligiendo todo tipo de técnica, de fichero, resolución, percentiles, para cualquier número de categorías, siendo esto muy efectivo e incluso de números de espectros en cada trozo para poder evitar así algún fallo humano a la hora de realizar las medidas, si alguna difiere más de lo deseado. De esta manera no se está condicionado a tener que cambiar el código fuente cada vez que se quiera realizar un estudio sobre diferentes tipos de elementos.

1.2 Objetivos

Este proyecto tiene nueve objetivos claramente definidos, como es en primer lugar, la mejora del procedimiento interno del algoritmo de manchas espectrales.

En segundo lugar, la modificación completa del método de validación cruzada (*cross-validation*) de tipo “mezcla” (*shuffle*). Se pretende que los resultados de clasificación no estén condicionados por el conjunto de datos iniciales que se emplean para el entrenamiento y así, poder aportar generalidad al proceso. Esta técnica trata de realizar una serie de mezclas, en base a un parámetro introducido en la interfaz, en el cuál se indicará el número de espectros que habrá en cada partición a barajar. Este es un parámetro seleccionado por el usuario y así poder ajustar mejor el grado de acierto en su investigación.

En tercer lugar, la implementación de la nueva funcionalidad de validación cruzada denominada Leave-One-Out, [4]. Este método de validación consiste en entrenar tantas veces como número de muestras se tengan, siempre usando como entrenamiento todas las imágenes menos una, que será la que se evalúe en cada iteración. Una vez hecho esto se pueden evaluar las prestaciones del algoritmo construyendo una matriz de confusión.

En cuarto lugar, optimizar la captura de espectros en el interfaz. Se capturará el nombre del directorio, donde se encuentren cada una de las categorías a investigar, con el fin de incorporar a la leyenda de las manchas espectrales mostradas sobre el interfaz.

En quinto lugar, la implementación de la nueva funcionalidad que ayudará a crear e interpretar las imágenes hiperespectrales.

En sexto lugar, la implementación de un fichero de configuración que introduciendo en él, las rutas donde se encuentran las categorías a analizar, abstraerá al usuario de tener que cargarlas manualmente, puesto que es bastante tedioso.

En séptimo lugar, la implementación de un cartel informativo con todos los parámetros y opciones indicados por el usuario y así salvaguardarse de equivocaciones humanas y sobre todo, de valores que no se haya captado bien o se cojan por defecto.

En octavo lugar, cambio de apariencia de la interfaz, haciendo ésta más atractiva al uso.

En noveno lugar se realizará un estudio sobre los espectros de transmisión del aceite, café y vino, además de un análisis de discriminación a partir de la reflectancia difusa obtenida por un sistema hiperespectral y proveniente de los siguientes materiales; cartón, papel de caramelo, film transparente, hilo, hojas secas, madera, papel aluminio, papel caramelo blanco, papel caramelo opaco, papel transparente plástico, piel y tabaco en 6 estados diferentes.

1.3 Estructura del documento

Este documento está compuesto de seis capítulos. El primer capítulo alude brevemente a los objetivos y motivación del trabajo. En él se explica en qué consiste el proyecto y el porqué de su realización, es decir cuáles son los motivos o intereses que empujan a realizarlo.

El segundo capítulo está destinado a definir el fenómeno luz-materia, mostrando aplicaciones de uso y una explicación de los dos tipos de espectroscopia utilizados; transmisión y reflectancia difusa.

El tercer capítulo consiste en la explicación de los parámetros del algoritmo y de una presentación de la interfaz y de cada uno de sus parámetros, con el fin de familiarizar al usuario final.

En el cuarto capítulo se explica todas las mejoras introducidas desde un punto de vista teórico y de implementación. Destinado al porqué y al cómo de la reimplementación de la funcionalidad de validación cruzada denominada validación cruzada, *shuffle*, e introducción del nuevo parámetro. Realizando un estudio sobre el

aceite, café y vino en función de los parámetros de resolución, partes introducidas en la técnica de Shuffle, grosor, además de realizar el estudio para ambos tipos de técnicas de discriminación. También a explicar la nueva funcionalidad de validación cruzada, Leave-One-Out. Realizando un estudio sobre el aceite, café y vino en función de los parámetros de resolución, partes introducidas en la técnica de Shuffle, grosor, además de realizar el estudio para ambos tipos de técnicas de discriminación. Además será destinado a la implementación de imágenes hiper-espectrales y al estudio de diferentes imágenes, previamente capturadas, de diversos materiales.

El quinto capítulo es el referente a los resultados, mostrando infinitud de análisis, con el fin de obtener la respuesta a la pregunta más clara ¿merece la pena?

Para finalizar se encuentra el capítulo de las conclusiones en el cual se expondrá desde una práctica experimental cual sería el resultado óptimo para los parámetros, en función de su tasa de error y rendimiento del programa.

2. Espectroscopia

2.1 Fenómeno de interacción luz-materia

Antes de empezar a definir que se ha creado en este TFC, se dará una idea general sobre este fenómeno. Resolviendo algunas preguntas clave para entenderlo de una forma clara y precisa. Como: ¿Qué es un espectro?, ¿Qué es la espectroscopia? ¿Qué aplicación tiene?, ¿Desde cuándo se usa?, ¿Qué tipos hay?

Los espectros son una serie de colores que se producen al dividir una luz compuesta por una luz blanca en sus colores constituyentes, estos por orden son: violeta, azul, verde, amarillo, anaranjado y rojo; más tarde en el siglo XIX se descubrió el ultravioleta y el infrarrojo.

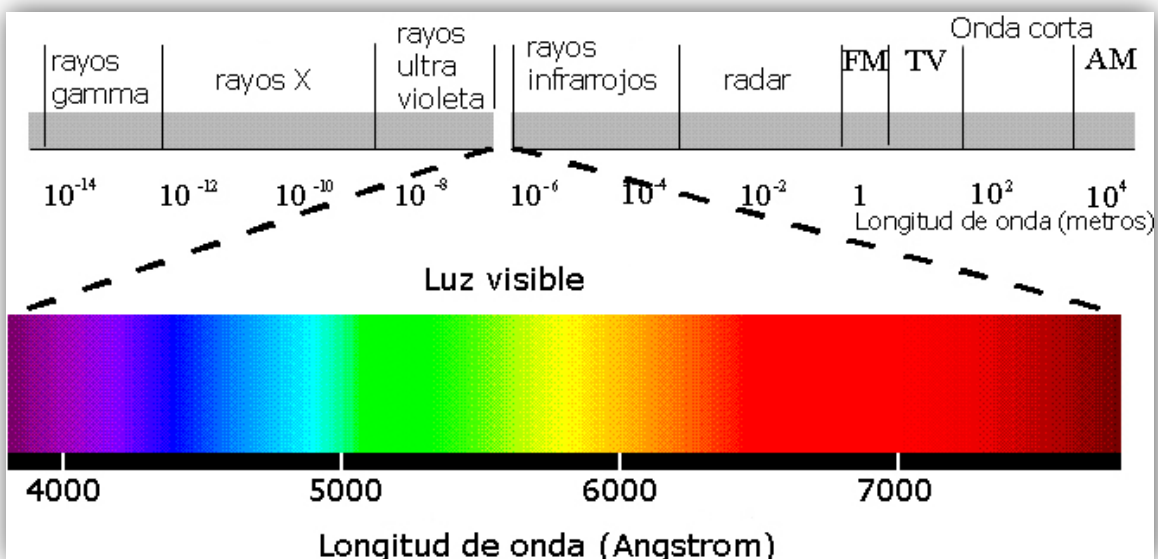


Figura 4: Los espectros

La espectroscopia es una técnica instrumental ampliamente utilizada por los físicos y químicos para poder determinar la composición cualitativa y cuantitativa de una muestra, ya que cada elemento químico tiene su espectro característico, mediante la utilización de patrones o espectros conocidos de otras muestras.

Esta tuvo su origen hacia los siglos XVIII y XIX, al hacer pasar la luz por un prisma usado para descomponer la luz se reforzó con rendijas y lentes telescópicas con lo que se consiguió así una herramienta más potente y precisa para examinar la luz procedente de distintas fuentes, Joseph von Fraunhofer basándose en este espectroscopio creado inicialmente, desarrollo una teoría que decía “el espectro de la luz solar está dividido por una serie de líneas oscuras”, en cambio en el laboratorio utilizando luz generada en base al calentamiento de gases, metales y sales mostraba una serie de líneas estrechas, coloreadas y brillantes sobre un fondo oscuro. Posteriormente se descubrió que cada elemento químico producía un espectro por lo que surgió la idea de huella digital como método de clasificación. El espectroscopio está formado por una rendija por la que entra la luz procedente de una fuente externa, un conjunto de lentes, un prisma y un ocular.

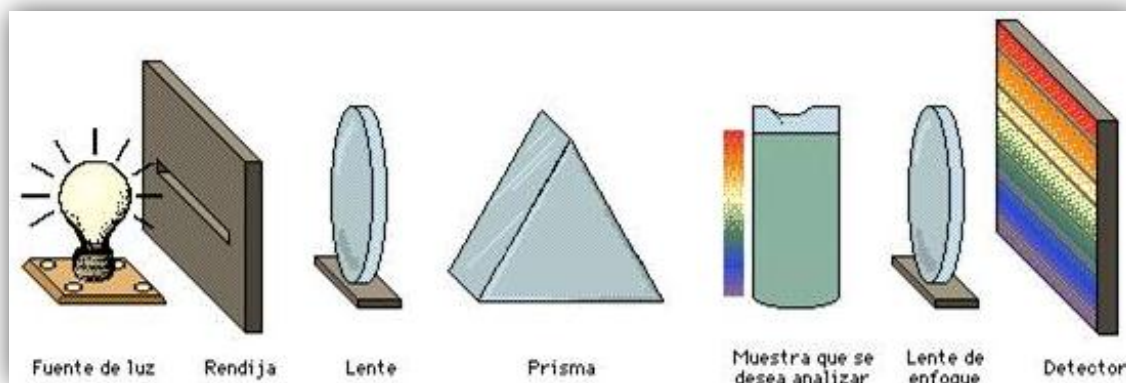


Figura 5: Captura del espectro de una muestra

La técnica permite detectar la absorción o emisión de radiación electromagnética para un rango de longitudes de onda, en este caso de 200nm a 1100nm, de esta forma se pueden relacionar con los niveles de energía implicados en una transición cuántica. Estas radiaciones electromagnéticas realizan una interacción con la materia pudiendo ser esta de tres formas posibles:

- Choque elástico
- Choque inelástico
- Absorción o emisión resonante de fotones.

2.2 Motivo del estudio

Los métodos espectroscópicos tienen mucha importancia en química analítica, y pueden ser bastante específicos a la hora de determinar los componentes de una mezcla y su proporción relativa. En este contexto la espectroscopia es usada ampliamente para ver las transformaciones que se producen en una reacción química de los reactivos a los productos y como se va a mostrar en este proyecto, como a

partir de ellos se podrán clasificar compuestos, observar, cuantificar e identificar de manera relativa materiales incógnitas, mezclas y añadiduras de diferentes compuestos

Cada material deja una huella espectral (un símil es la huella digital en los humanos). A partir de la cual, realizando cruces de diferentes huellas espectrales generadas por compuestos que se estimen, se podrá observar el grado de añadidura o pureza del material analizado.

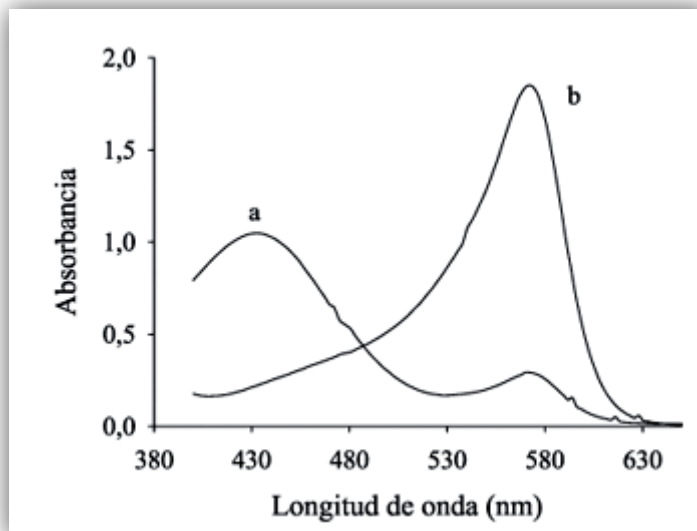


Figura 6: Absorbancias en el espectro. a) agua naranja b) Solucion Zn naranja

Se puede observar en este rango del espectro las diferencias en la huella espectral, basándose en la energía cuantificada en función de la energía absorbida por cada uno de los materiales a cada longitud de onda. Como existe una relativa baja absorbancia para la solución Zn para bajas longitudes de onda y como se encuentra desplazada para el agua esta bajada en las absorbancias, hacia longitudes de onda mayores.

2.3 Tipos de espectrometría

En el presente apartado se explicarán dos tipos de espectrometría muy importantes e interesantes que se han tratado. Como son la espectrometría de transmisión y la espectrometría por reflectancia difusa.

2.3.1 Transmisión

Al igual que en el apartado anterior se comenzará dando una breve explicación de dos conceptos como son el color y transmisión, a la vez de la relación existente entre ellos.

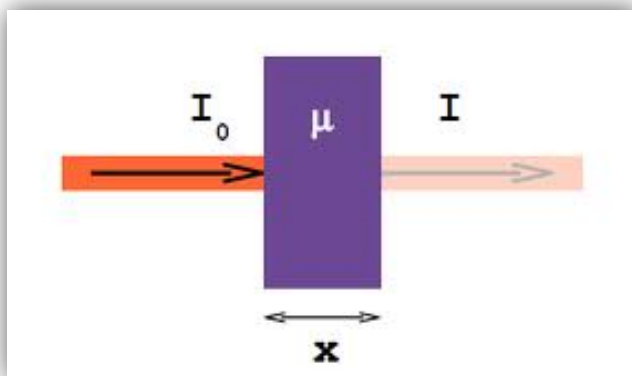
El color de los cuerpos opacos está relacionado con la radiación que dicho cuerpo refleja. Por ejemplo, un objeto que nosotros vemos azul, es un objeto que absorbe todas las radiaciones exceptuando aquellas que nosotros identificamos como color azul, en cambio, el concepto de color por transmisión, es diferente. Cuando

nosotros vemos, por ejemplo, un vidrio de color azul, se debe a que el vidrio es capaz de transmitir la radiación correspondiente a las longitudes de onda que nosotros percibimos como color azul, mientras que son absorbidas las radiaciones restantes. Por tanto, en un medio transparente (por ejemplo, una disolución), el color observado corresponde a la zona del espectro transmitida (no absorbida) por el medio.



Figura 7: Reflexión y absorción del color

La magnitud fundamental con la que se trabaja es la absorbancia, y se basa en la relación entre la intensidad luminosa que se obtiene de medir el montaje con la cubeta de referencia (sin muestra en ella) y la intensidad luminosa que se obtiene al medir el montaje con la muestra problema.



$$T = \frac{I_t}{I_o} \cdot 100 \% \quad (1)$$

A partir de esta relación se puede obtener la absorbancia mediante la siguiente fórmula:

$$A = \log \frac{1}{T} \quad (2)$$

Figura 8: Definición de absorbancia de un compuesto

Por lo tanto, se puede definir el espectro de absorción de una sustancia como la relación entre la absorbancia y la longitud de onda.

2.3.2 Reflectancia difusa

La Reflectancia Difusa es una técnica que ha despertado gran interés, ya que muchas sustancias naturales se comportan como reflectores difusos y tienen superficies que presentan dificultades si se las quiere estudiar con otros métodos.

Además los espectros de la mayoría de las muestras sólidas pueden medirse sin ningún tipo de preparación. Pero, ¿Qué es?, ¿Cómo se puede medir?, ¿Ajustes a considerar?

Se comenzará dando una explicación de reflectancia difusa. Al incidir un rayo de luz en una muestra sólida, una parte de la luz es reflejada (reflectancia especular) en la superficie de la muestra mientras que otra parte de la radiación penetra en la misma (sobre 2 mm aproximadamente) pudiendo ser absorbida dentro de ella. La radiación que no se absorbe puede ser transmitida a través de la muestra o reflejarse en ella de forma parcial (reflectancia difusa) [2]. La suma de la reflectancia especular más la reflectancia difusa se llama reflectancia total

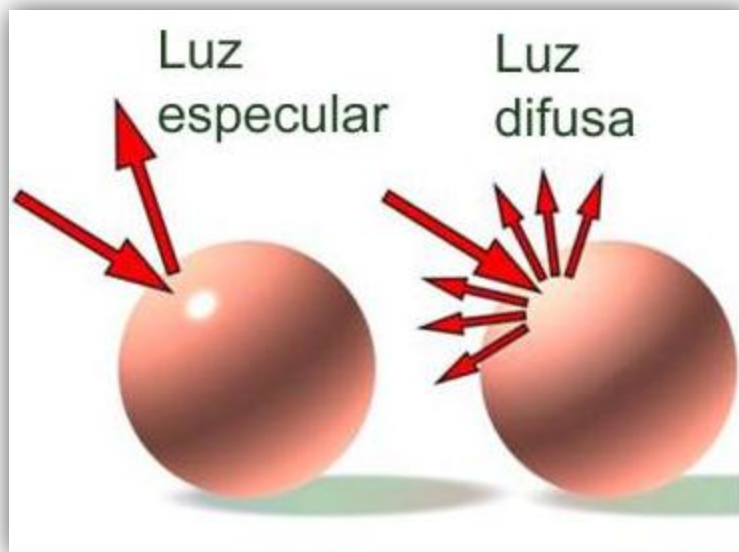


Figura 9: Representación de los dos tipos de reflectancia

La espectroscopia [7] por reflectancia difusa es una técnica de análisis directa y no destructiva, basada en la cantidad de reflectancia especular y de reflectancia difusa de la superficie del objeto. Además la técnica es rápida, económicamente asequible y fácilmente reproducible, ofreciendo la posibilidad de estudiar características tangibles de los efectos auténticos indubitados, frente a aquellas de los que las muestras falsas adolecen.

Para los objetos que tienen superficies muy brillantes, la luz reflejada especularmente es relativamente fuerte y la luz difusa es más débil. En superficies rugosas con bajo brillo, el componente especular es débil y la luz difusa es más fuerte. Cuando se ve un objeto de plástico con una superficie brillante en el ángulo especular, el objeto no parece ser tan azul. Esto se debe a que la reflectancia tipo espejo de la fuente de luz se añade al color de la muestra. Normalmente, una persona observa el color del objeto e ignora la reflexión especular de la fuente de luz. Para medir el color de un espécimen del mismo modo que se ve, la reflectancia especular deberá excluirse y sólo deberá medirse la reflectancia difusa. El color de un objeto puede parecer distinto debido a las diferencias en el nivel de la reflectancia especular.

Para realizar la medición de este tipo de espectroscopia se realiza en base a un espectrógrafo. Los espectrógrafos son instrumentos equipados con una red de difracción o un prisma que se encargan de dispersar la luz y producir el espectro deseado. Este espectro queda recogido como imagen en un CCD y puede ser extraído

posteriormente con los programas apropiados. La región del espectro que podemos obtener al medir está limitada por el tamaño del detector, en nuestro caso, el CCD. Los espectrógrafos son instrumentos equipados con una red de difracción o un prisma que se encargan de dispersar la luz y producir el espectro deseado. Este espectro queda recogido como imagen en un CCD y puede ser extraído posteriormente con los programas apropiados. La región del espectro que podemos obtener al medir está limitada por el tamaño del detector, en nuestro caso, el CCD (200nm – 1100nm).

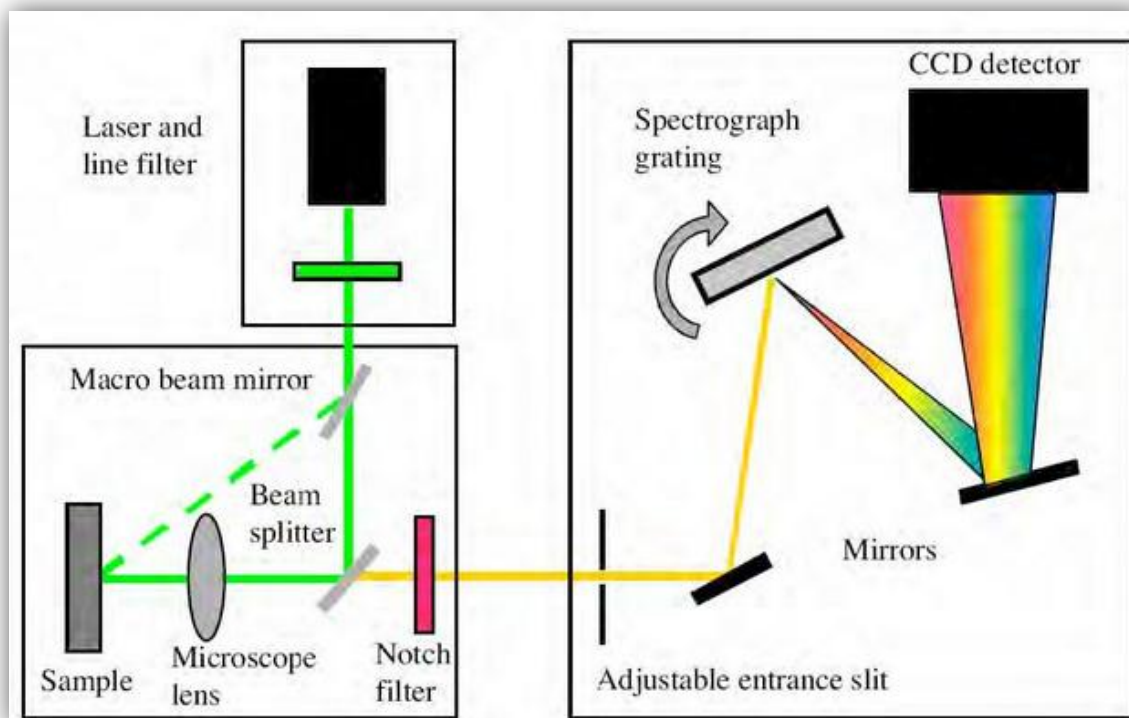


Figura 10: Medición mediante un espectrógrafo

A continuación se pasan los datos obtenidos al ordenador y posteriormente se analiza en Matlab, mediante una interfaz desarrollada, que se encuentra provista de un optimizado algoritmo de manchas espectrales para la identificación de compuestos.

3. Manchas espectrales

3.1 Introducción a las manchas espectrales

Las manchas espectrales son los resultados gráficos de las absorbancias de cada longitud de onda. Estas manchas se pueden representar en el dominio espectral o un dominio PCA (*Principal Component Analysis*). Esta representación se realizara mediante coordenadas polares.

A continuación se pasará a explicar cada uno de los parámetros implicados en la generación de estas manchas espectrales citadas. Incluyendo explicaciones teóricas y graficas sobre la formación de estas manchas en función de cada uno de estos parámetros y su influencia a la hora de realizar el discriminado entre compuestos.

3.1.1 Percentiles

El percentil [8] es una medida de posición no central que dice cómo está posicionado un valor respecto al total de una muestra. Los percentiles son por tanto 99 valores que dividen a la distribución en 100 partes iguales. Por lo tanto cada mancha espectral tendrá un grosor el cual vendrá dado por el percentil inferior y superior deseado por el usuario con el fin de introducir un mayor ruido en la mancha y así poder realizar una discriminación más efectiva de cada una de estas manchas. Un ejemplo con el aceite sería el siguiente:

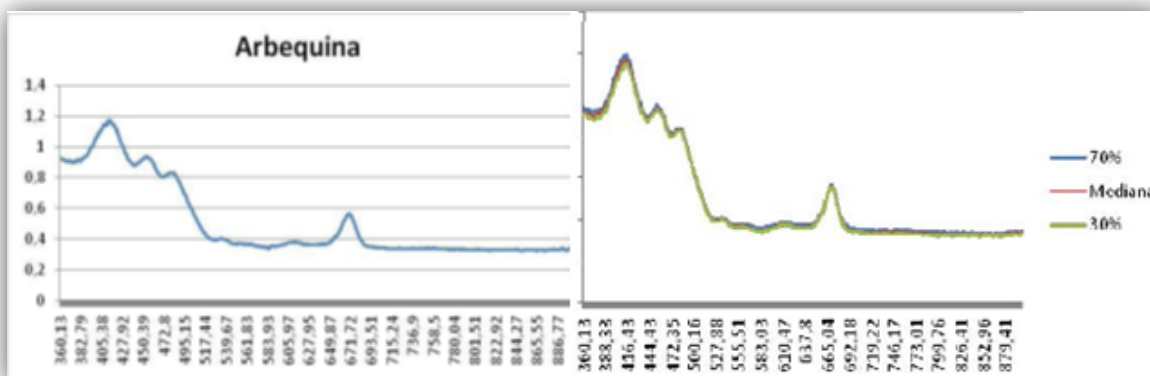


Figura 11: Espectro en eje lineal del aceite Arbequina sin (1) y con (2) percentiles

Las manchas espectrales se forman a partir de la conversión de la absorbancia en cada λ (λ) y del grosor que se esté introduciendo en ésta, también denominado ruido en la mancha. Todo esto se encuentra representado mediante la figura 12.

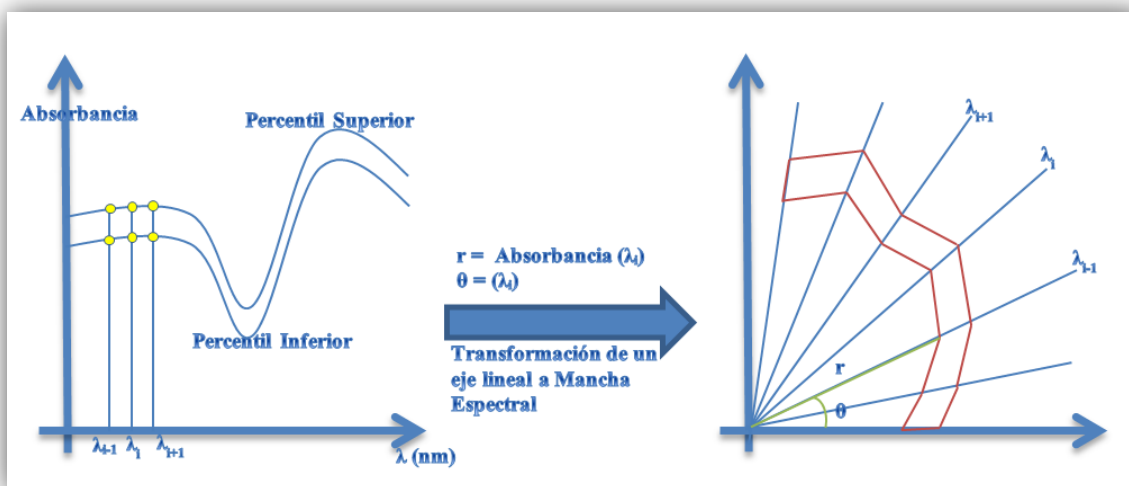


Figura 12: Obtención de manchas espectrales

Para explicar la influencia de este parámetro, se mostrará en la siguiente imagen, una clasificación real de nueve categorías de aceite, variando únicamente este parámetro del grosor. En esta ejecución del algoritmo, se obtiene un error total del 11,89% con un percentil del 40 - 60 %, mientras que para un grosor del 25 - 75 % se captura un error total del 4,22%. Como se observa, este parámetro tiene una influencia significativamente alta en el momento de discriminar. Además se puede intuir que cuanto mayor sea el ruido permitido en la mancha espectral, más fácil es que los espectros con los que se comparan en cada iteración, para sacar este error total, entren dentro de los varemos establecidos por los percentiles.

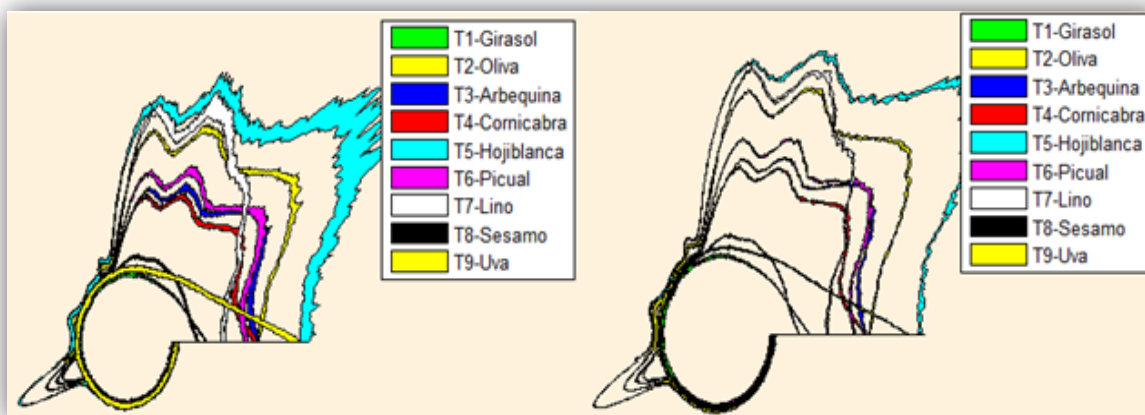


Figura 13: Manchas espectrales en el dominio espectral del aceite con percentiles
a) 25 – 75 % b) 40 - 60 %

3.1.2 Resolución

Este parámetro, suele dar lugar a pensamientos contrarios a la realidad, como que si se aumenta esta resolución el error total obtenido a la hora de clasificar disminuirá. La respuesta es que sucede todo lo contrario como se va a tratar de explicar en los siguientes párrafos.

Si se hace mucho zoom sobre dos puntos que aparentemente están en el mismo espacio, se puede observar que al final estos dos se separan pudiéndose dar como resultado final, que estos no son iguales. Pues lo mismo nos ocurre en este caso, por lo tanto como se ofrecerá en el estudio de puntos posteriores, el error aumenta con la resolución de una forma logarítmica llegando a ser constante al acercarse a una resolución infinita. Esto se explica gráficamente en la figura 14. Claro, eso sí, esto ocurrirá en este proyecto microscópicamente.

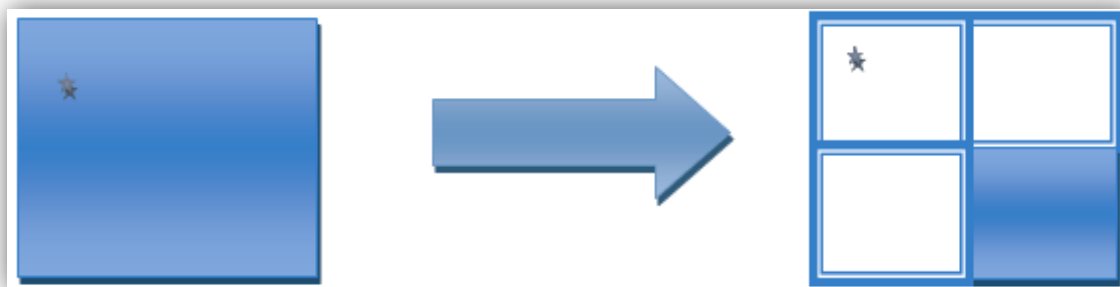


Figura 14: Gráfica de un punto que pertenece y luego no, al subir la resolución

En esta explicación gráfica se quiere dar a entender, que en un principio el punto que queremos observar, sí pertenece a esa mancha espectral pintada, en el entrenamiento formado por los otros dos subgrupos restantes, pero que al aumentar la resolución está deja de pertenecer. Por lo que en este caso, lo que ocurre, es que los puntos que acertaban bajan, mientras que los que fallan aumentan, dando lugar a un aumento del error total. Este problema habrá un momento en el que dejará de existir y éste será en función de los decimales dados en cada punto.

Por lo tanto, la resolución aunque en un principio nos da pie a pensar, que se trata de un método para reducir el error total, es en realidad un método para

caracterizar cualquier categoría desde un punto de vista de exactitud, ya que aumentando la resolución, aumentaríamos directamente la exactitud de nuestro trabajo. Además nos damos cuenta también, que si lo que se quiere, es reducir el error total sin pensar en ningún momento en la exactitud, habría que reducir notablemente la resolución, eso sí, dentro de unos límites para los cuales se habría que estudiar en cada caso, porque llegara un momento que todos los puntos pertenezcan a todas las manchas espectrales, con el consiguientes aumento del error total, debido a los errores de indeterminación entre manchas.

Así que, una gráfica general que podría expresar el error total frente a la resolución sería la de la figura 15, indicando con un círculo rojo el lugar donde se encontraría el menor valor del error, este sería para una resolución alrededor de 0.01-0.04 Mpx y con un círculo anaranjado el lugar donde se hallaría una mayor exactitud situada sobre una resolución aproximada mayor de 25 Mpx.

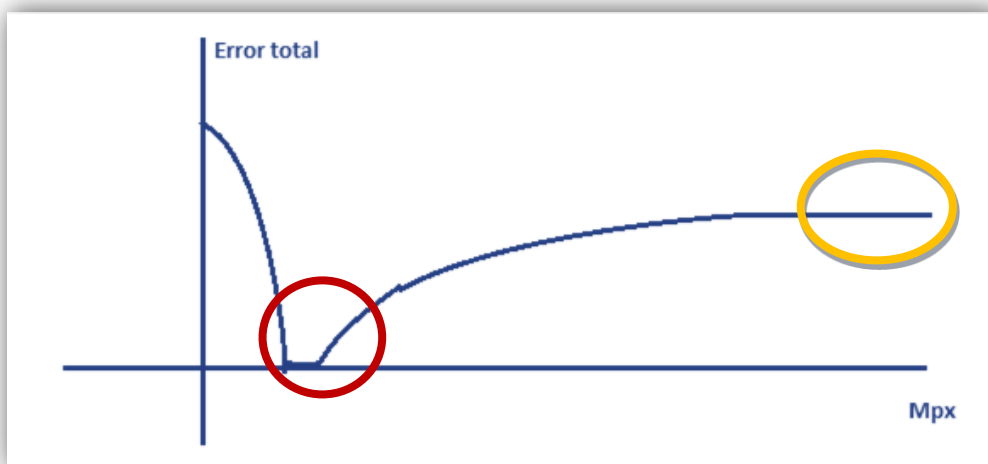


Figura 15: Gráfica error vs resolución

Cabe destacar que en base a este parámetro se realizará el pixelado de la mancha espectral, produciendo con ello matrices con las que trabajar de tamaños variopintos indicador por el usuario. Mencionar que cuanto más resolución, las matrices de las manchas son mayores y por lo tanto el rendimiento computacional disminuye.

3.1.3 Dominio espectral y PCA

Comenzaré haciendo un breve esquema de los pasos a seguir, para realizar esta representación. En él se puede observar un apartado dedicado al dominio en el que se va a realizar el análisis, es decir en el dominio espectral o en el dominio PCA. En el esquema se muestra mediante un relleno amarillo en que parte del análisis opera esta mejora.

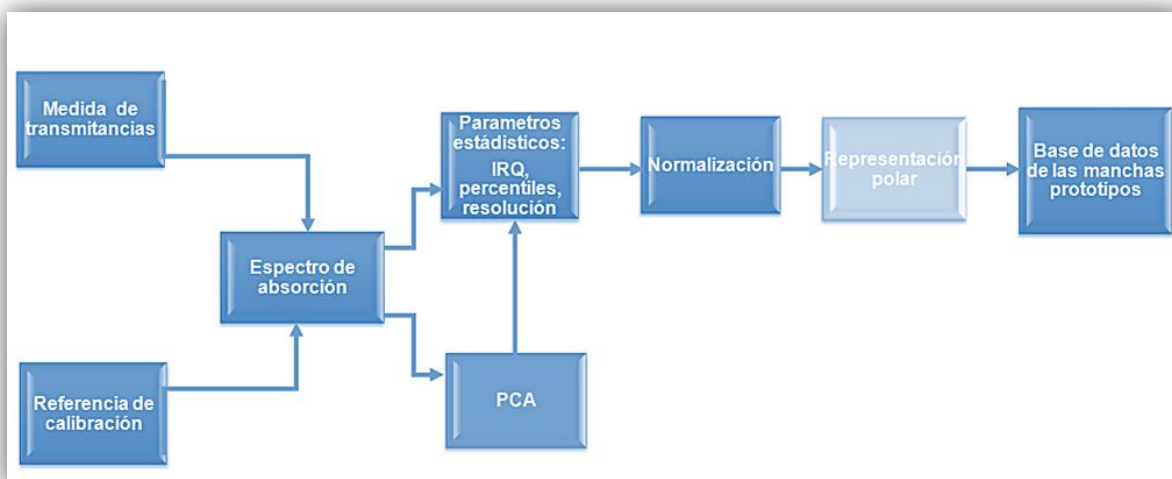


Figura 16: Generación de manchas espectrales

Como se puede observar, la “cajita” de representación polar viene remarcada, ya que esta se aplicará en ambos dominios y es la sección más revolucionaria de este tipo de análisis, explicado en puntos anteriores como se aplicaría. También se puede percibir, como en el caso de elegir realizar el análisis en el dominio PCA, se harán unas operaciones previas con los espectros, previamente a la captura en base a los parámetros y su paso a coordenadas polares.

Por consiguiente se dará una explicación del significado y aplicación de este método de análisis, para finalizar este apartado mostrando la influencia de elegir este dominio a la hora de discriminar.

Lo primero que hay que tener claro es que en este método no tenemos el mismo sistema de coordenadas que anteriormente, sistema de coordenadas espectral; sino que se trata de clasificar de una forma más clara cuales son las principales componentes de cada elemento, por lo que el nuevo sistema de coordenadas estará formado según el tamaño de las varianzas de cada elemento, es decir, el que tenga la varianza más grande estará representado en el primer eje y así sucesivamente. De esta forma se juega con la dimensionalidad del conjunto de datos, pudiendo reducir ésta a nuestro antojo. Para construir esta transformación lineal debemos construir primero la matriz de covarianza, siendo a partir de ésta de donde obtenemos los autovectores y autovalores. A continuación en este proyecto se da a elegir al usuario dos métodos de corte de datos:

- Épsilon: Este método consiste en encontrar un índice que cumple la exigencia del usuario. Cuando el vector de autovalores está ordenado, se cogerá desde el primero hasta el último en orden, dividiéndolos uno a uno por el máximo de este vector, hasta que este número sea menor o igual que el número de corte dado por el usuario, es decir, épsilon. Esto se rige mediante la siguiente fórmula matemática:

$$Epsilon \geq \frac{auval}{\max(auval)} \quad (3)$$

- Varianza conservada: En este caso se trata de que te devuelva lo mismo, ordenando primeramente el vector de autovalores, este método consiste en que ahora la varianza conservada tiene que ser menor igual que el sumatorio desde el primero hasta el autovalor que sea, que al dividirlo por el sumatorio del vector de autovalores se cumpla la inecuación. La fórmula matemática sería la siguiente:

$$\text{Var}_{\text{conservada}} \geq \frac{\sum_{i=1}^M \text{auval}}{\sum_{i=1}^N \text{auval}} \times 100 \quad (4)$$

Siendo M el índice del vector que cumple la inecuación y N la longitud total del vector.

Seguidamente se construye un vector nuevo de autovectores y autovalores utilizando el índice devuelto, como longitud total de los dos vectores, de esta forma se eliminan partes insignificantes a la hora de la discriminación. Luego convertiremos la expresión de los datos iniciales en la nueva base de autovectores. Para finalizar, se recuperan los datos originales mediante la expresión en la nueva base de autovectores. Como la expresión de los espectros de absorbancia en función de las componentes principales pueden tener componentes positivas y negativas. Se introducirá un offset, con la finalidad de pasar a positivos todos los valores y así poder obtener su representación polar. Esquemático en la figura 17.

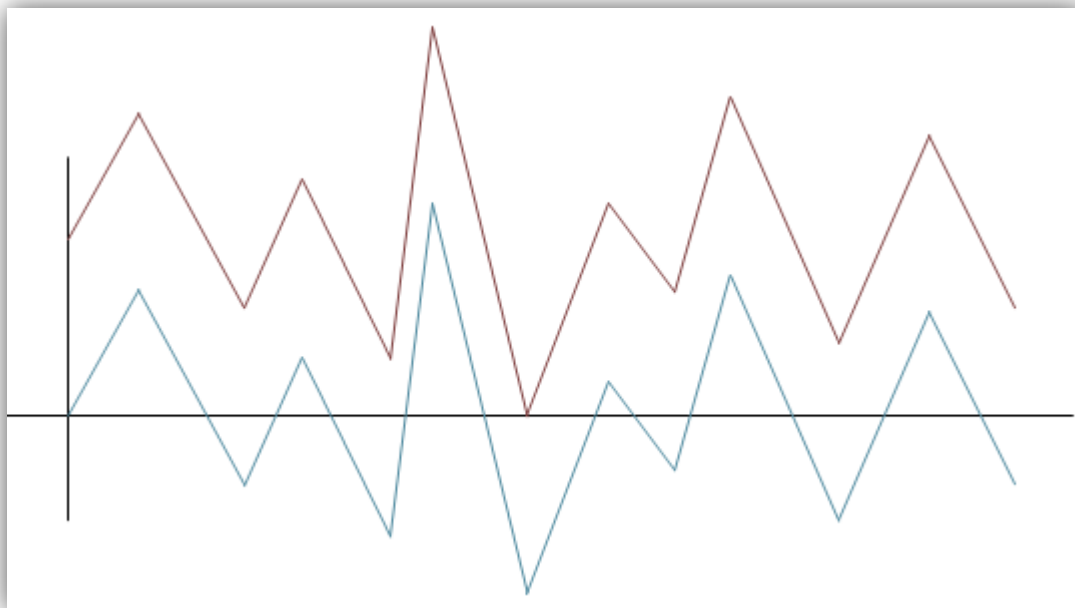


Figura 17: Offset de las componentes principales

Al igual que se hizo para explicar la influencia del parámetro del grosor, se mostrará una imagen, variando el dominio en el que se ha capturado la mancha espectral. En esta ejecución del algoritmo, se obtiene un error total del 4,22% con un dominio espectral mientras que para un dominio PCA (utilizan 4 PCAs para el análisis, [5] ya que es el parámetro óptimo para esta sustancia) se captura un error total del 0,98%. Como se observa, este parámetro tiene una influencia significativamente alta en el momento de discriminar. Además se puede intuir que cuanto menos número de PCAs se tengan, mejor es para el análisis y que siempre

será mejor analizarlo en este dominio, en detrimento del dominio espectral. Esto no siempre es así ya que está muy ligado al tipo de material y se podrá observar en el análisis de puntos posteriores.

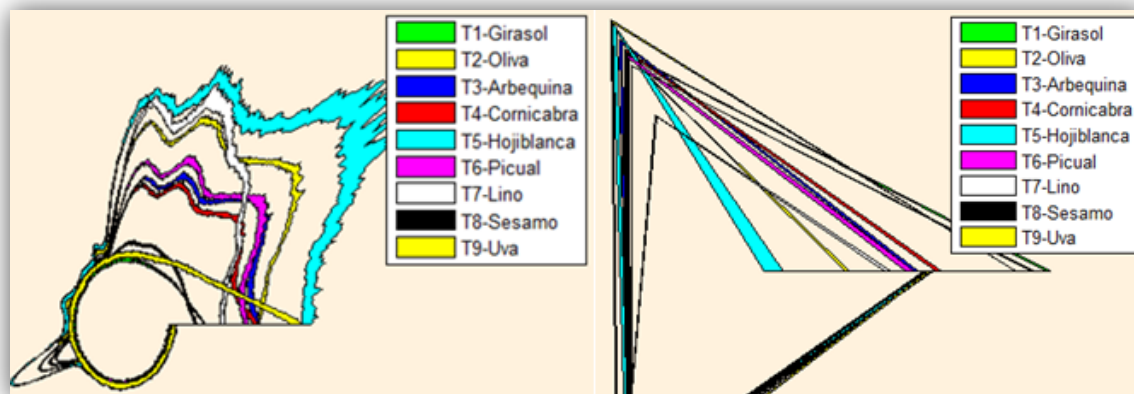


Figura 18: Manchas espectrales del aceite a) dominio espectral y b) dominio PCA

3.1.4 Matriz de confusión

El análisis de la calidad de los trabajos de clasificación se acometerá mediante el uso de matrices de confusión, las cuales presentan una visión general de las asignaciones correctas (en la diagonal) e incorrectas (en los lugares fuera de la diagonal), que nos darán una idea relativa del porcentaje de aditivo hay de un compuesto en otro.

En la siguiente tabla se muestra un ejemplo de tabla de confusión de 5 tipos de café, con el fin de entender ésta. Posteriormente en los siguientes apartados se discutirán resultados con los diferentes juegos de datos analizados para los distintos tipos de espectroscopia utilizados.

Matriz de Confusión	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	69.00	14.00	7.00		10.00	
Suave	4.00	12.00	21.00	12.00	49.00	2.00
Costa Rica	24.00	4.00	34.00		38.00	
Kenya				70.00	30.00	
Descafeinado	14.00		12.00	30.00	44.00	

Tabla 1: Matriz de confusión obtenida para el café con un grosor de 25 – 75% y 4 Mpx

Se observa (mirando la diagonal de la tabla) como existe un 69% de aciertos en la clasificación del café Forza respecto a los otros 4 tipos, un 12% del café suave, un 34% del café Costa Rica, un 70% del café Kenya y un 44% del café descafeinado. También se puede concluir que el café descafeinado tiene un aditivo que lo poseen los otros 4 tipos de café. El café Forza en un 10%, el café suave en un 49%, el café Costa Rica en un 38% y el café Kenya en un 30%, por lo que podemos determinar que el café suave es el que más se asemeja al café descafeinado.

3.2 Tipos de validación cruzada

En este apartado se presentarán desde un marco teórico-explicativo, las cuatro técnicas de validación cruzada que tiene desplegadas la interfaz.

3.2.1 Shuffle

Esta técnica de validación cruzada con posterior división en tres entrenamientos, realiza un mezclado controlado de espectros obtenidos en diferentes momentos del tiempo, con el fin de no influir negativamente en el análisis si algunas medidas difieren de otras en exceso.

Esta funcionalidad se ha modificado respecto a la implementada en el proyecto [5], debido a que el mezclado era sobre un número de espectros fijos (bastante alto), además de dejar un gran bloque en la parte sobrante (debiendo ser esta despreciable), siendo ésta, aplicada en el test sobre los tres entrenamientos consecutivamente. Esto es un gran problema si las mejores medidas (las que más se acercan a la mediana) se hicieron en los últimos días, propiciando que la tasa de acierto sea realmente alta, siendo esto un engaño, como ocurre en el caso del análisis del aceite. O, si por el contrario, son las peores muestras las que tenemos en último lugar, propiciando un grave incremento del error total.

Para corregirlo, se ha introducido un parámetro, denominado “Partes” que aparecerá únicamente si el usuario final selecciona esta técnica, además de realizar una división de lotes divisible por tres y así tener como sobrante algo insignificante y no inducir a engaños.

3.2.2 Lotes

Este método consiste en dividir las absorbancias de un fichero en tres partes iguales. Cuando no se puede dividir, porque al dividir el número de absorbancias entre tres no queda un número entero, se dejarían otra parte de absorbancias denominadas como sobrantes, que se unirán a la matriz test en cada caso.

Esta división de las absorbancias a la hora de realizar el análisis espectral es apropiada siempre y cuando no haya una gran diferencia entre las medidas de unos días y otros; ya que como se puede observar en el estudio del aceite, las primeras absorbancias difieren mucho de las demás.



Figura 19: Esquema gráfico de división por lotes

3.2.3 Aleatorio

Esta forma de dividir las absorbancias se realiza de una manera totalmente aleatoria (sin seguir ningún patrón), es decir, que cuando las medidas difieren mucho de unos días a otros, esta técnica se puede utilizar al igual que en el apartado anterior, lo único que en este caso al ser de forma aleatoria el error no será siempre el mismo, utilizando el mismo ámbito de análisis, por lo que se trataría de obtener un valor promedio de todos los errores que vayamos recogiendo cada vez que realicemos el análisis espectral en las mismas condiciones.

3.2.4 Leave-One-Out

Esta técnica es la más novedosa dentro de las técnicas de validación cruzada, por lo que se explicará detenidamente de una forma clara y precisa, con la finalidad de mostrar las ventajas e inconvenientes que ofrece.

Por lo tanto, leave-one-out (dejando uno fuera), se trata de una técnica de validación cruzada que implica separar los datos de forma, que para cada iteración tengamos una sola muestra para los datos de prueba y todo el resto conformando los datos de entrenamiento. La evaluación viene dada por el error, y en este tipo de validación cruzada el error es muy bajo, pero en cambio, a nivel computacional es muy costoso, puesto que se tienen que realizar un elevado número de iteraciones, tantas como número de muestras se tenga y para cada categoría, analizar los datos tanto de entrenamiento como de prueba. Como se expresa en la figura 20, que trata de clarificar como se clasifican dos compuestos mediante esta técnica, en función de bolas rojas y verdes, adecuando una esfera de color por cada muestra de la categoría que represente.

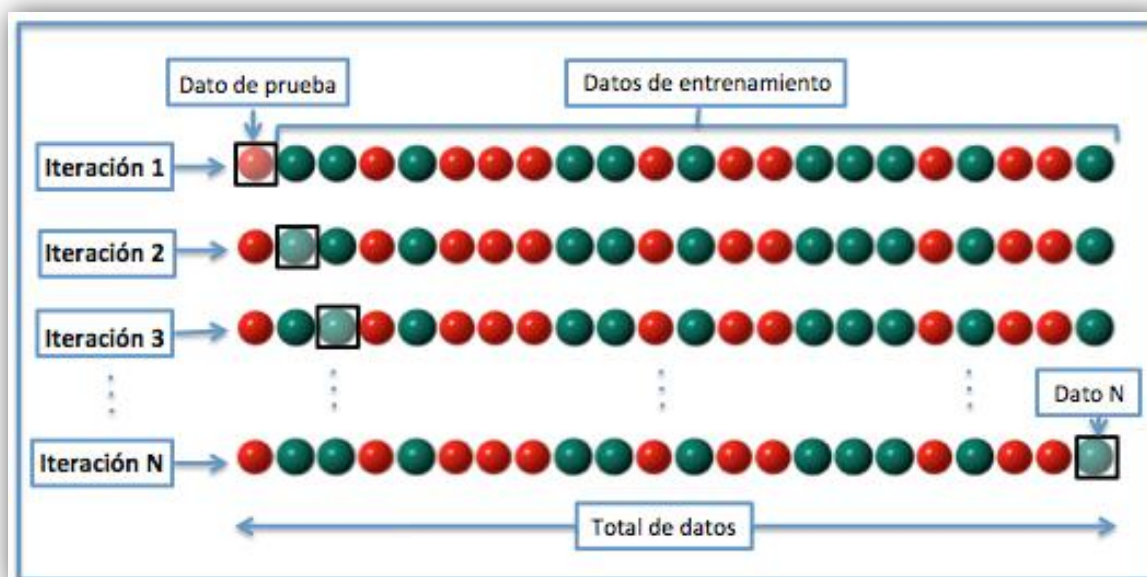


Figura 20: Representación de leave one out

Por lo tanto esta técnica, enriquece la funcionalidad de la aplicación, dando por ello una posibilidad adicional al usuario final de ésta, de obtener un análisis discriminatorio más acertado en función de los espectros obtenidos previamente. Como se ha expresado en la definición, con esta técnica se formará una única matriz de confusión, a partir de la cual, se medirá el error total. Este método es realmente costoso a nivel computacional, por lo que se ha depurado de una forma muy exhaustiva, como se podrá observar en el capítulo 4, ya que generaba muchísimos problemas de tiempos y espacios en el heap de Java del que dispone MATLAB. Con esta técnica se crean multitud de objetos y figuras que residen en esta memoria y que se han tenido que optimizar.

3.3 Descripción del algoritmo

A la hora de implementar estos objetivos, se ha tratado de intentar que el usuario no necesite saber programar para realizar un estudio de sus datos obtenidos en el laboratorio. En cualquier caso, para ello se ha llevado a cabo una interfaz gráfica mediante el *GUI* de *MATLAB* [15] para que de esta manera, introduciendo sus datos y parámetros deseados manualmente y de una manera fácil y visual, pueda representar y recibir su grado de acierto o error de sus mediadas obtenidas. Así el usuario dedicará más tiempo en realizar correctamente dichas medidas dejando de malgastar tiempo y esfuerzo en tareas de programación.

En el interfaz desarrollado se ofrecen diversas alternativas, en base a diferentes parámetros, para conseguir aumentar la tasa de acierto, a la hora de identificar materiales. A continuación se marcará una guía de lectura de la interfaz, en función de marcadores numéricos, introducidos en la figura 21.

The interface is titled "MATERIAL IDENTIFICATION" and is divided into several functional areas:

- Top Left:** Buttons for "Nuevo Análisis" (1) and "Generar Mancha" (2).
- Top Center:** A section for "Método de Identificación" with radio buttons for "Spectral Marks" (3) and "PCA". It also includes input fields for "Epsilon:" and "Varianza conservada:". Below this is a "Cross Validation" dropdown set to "Shuffle" (4) and a "Partes:" input field (5).
- Top Right:** A section for "Tipo de Fichero de Entrada" with radio buttons for "Absorbancias", "Transmitancias", and "Raw". There is also a "BMP" checkbox.
- Left Panel:** Labeled "Entrenamiento S1S2", it contains a large plot area with a diagonal cross (6, 7) and a small data table below it.
- Center Panel:** Labeled "Entrenamiento S1S3", it contains a large plot area with a diagonal cross (8, 9) and a small data table below it. A point is marked on the cross (14, 15).
- Right Panel:** Labeled "Entrenamiento S2S3", it contains a large plot area with a diagonal cross (10, 11, 12, 13) and a small data table below it.
- Bottom Center:** A "Zoom" button (4) and a "Mover imagen" button (5).

Data Tables:

	1	2
1		
2		
3		
4		

Figura 21: Interfaz gráfica

- (1) Este botón lanza el análisis, capturando previamente los espectros, mediante un fichero plano, o las imágenes, para posteriormente realizar todo el proceso discriminatorio. Esta captura se podrá realizar, a través de un buscador, el usuario final seleccionará cada una de las categorías, sobre las que desee obtener el análisis.
- (2) Se trata de un botón similar al apuntado anteriormente (1), con la salvedad que en este caso no se capturan de nuevo, las categorías que se van a analizar, con la finalidad que se pueda obtener un mejor rendimiento, en el momento de realizar un mejor ajuste a través de parámetros como los percentiles, número de PCA's, resolución, etc. Los parámetros se explican en los siguientes puntos.
- (3) Es un texto estático, que al final de la ejecución marcará la tasa de fallo que ha tenido el análisis realizado.
- (4) Este punto está marcado sobre una lista desplegable, que dará al usuario final la posibilidad de realizar una de las técnicas de validación cruzada.
- (5) Este apunte se encuentra sobre un texto editable, que emergerá en el caso de seleccionar el método de validación cruzada tipo *shuffle*. Este parámetro rige el grado de mezclado de los espectros. Siendo éste, realmente significativo en el caso de tener lotes de muestras bastante diferenciadas en cuanto a su precisión. El número de partes que marque el usuario final, se corresponderán con el número de espectros, que se cogerán para realizar el barajeo.
- (6) El siguiente texto editable, es el encargado de capturar el número de categorías o materiales a analizar por el usuario.
- (7) Se encuentra puntualizando el parámetro de la resolución. Esta se introducirá por medio de los píxeles o cuadros que el usuario desee, en función del eje de las abscisas y ordenadas.
- (8) Es el parámetro encargado de aplicar el 'grosor' a la mancha espectral, este se define a través de los percentiles inferior y superior.
- (9) Este punto es quizás el más importante, en cuanto al análisis espectral se refiere. Aquí, el usuario puede elegir, uno de los siguientes métodos de discriminación:
 - Utilizando un sistema de coordenadas espectral.
 - Empleando coordenadas provenientes del análisis PCA.
- (10) Remarca un botón de 'ZOOM' o ampliación que al pulsar sobre él, aparecen dos botones, para realizar una ampliación o disminución en función de los intereses del usuario.
- (11) Señala al botón que tiene por funcionalidad, mover las imágenes, este es bastante útil si se ha realizado previamente zoom sobre alguna parte.
- (12) Esta reseña se realiza sobre un panel con tres opciones, a través de las cuales el usuario indicará al programa que tipo de datos va a recibir, para saber cómo tratarlos, éstos son:
 - Absorbancias.

- Transmitancias.
 - Datos RAW, según se obtienen del espectrómetro.
- (13) Este checkbox es el encargado de definir si los datos a capturar, provienen de un fichero plano o si por el contrario se trata de una imagen.
- (14) Delimita el cuerpo de la interfaz, donde se pintarán las manchas espectrales, de cada uno de los tres entrenamientos, si los tuviera. Además vienen acompañadas de una leyenda para identificar cada una de las categorías analizadas, ésta se obtendrá del nombre del directorio que contenga el material en cuestión. La figura 22, muestra a modo de ejemplo un entrenamiento de 9 categorías de aceites diferentes.

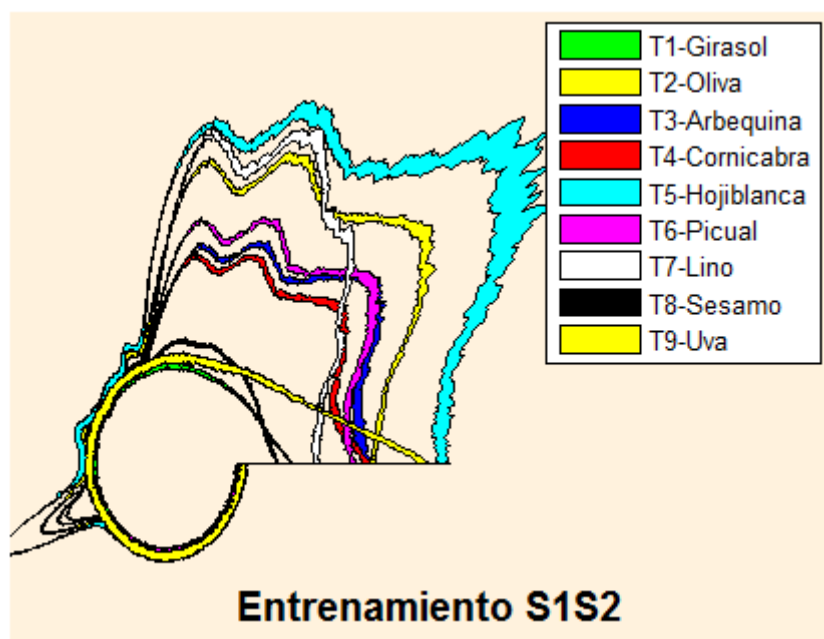


Figura 22: Ejemplo de Mancha espectral de 9 categorías de aceite

- (15) Este apartado al pie de la interfaz, es el encargado de mostrar la matriz de confusión generada. Ésta es, una matriz usada como clasificador de dos o más clases. Contiene información acerca de las clasificaciones actuales y predicciones hechas. La matriz tiene una dimensión n por m , donde n es el número de compuestos y m es el número de compuestos más una columna adicional para las indeterminaciones. A partir de las cuales se guiará la clasificación de las diferentes sustancias introducidas.

4. Mejoras introducidas

4.1 Procesado del color de una mancha espectral

Este apartado se encuentra dedicado a una parte del análisis discriminatorio que se encarga de rellenar las manchas espectrales en función de los percentiles, grosor, introducidos por el usuario. Se ha introducido este punto en el proyecto, ya que en proyecto anteriores [1] se alude a una posible mejora en este aspecto.

El algoritmo pinta dos veces cada categoría con el fin de utilizar, una para mostrársela al usuario y otra para realizar la captura de ésta, en función de parámetros como la resolución. Lo que realmente implica en este apartado explicativo, es la captura de estas categorías donde se va a hacer hincapié a partir de ahora. Esta captura se realiza a través de la función ‘imcapture’, que tiene como salida una celda con tres matrices, correspondientes a los tres planos RGB.

En un principio este algoritmo trataba de capturar estas manchas espectrales, pasándoles como parámetro de color, el color que ésta tuviera a la hora de mostrarlas en pantalla (aleatorio). En la figura 23 se tratará de expresar como se trataban estas matrices y el porqué de su posterior corrección.



Figura 23: Operación que realizaba el algoritmo tras la captura.

El algoritmo trata de realizar una superposición, mediante una operación OR, de los tres planos, RGB. Por lo tanto, al realizar el pintado del relleno en verde, negro, azul ... se marcará un '1' lógico en la posición concreta donde se encuentra rellena de alguna de las matrices y la matriz resultante tendrá marcado correctamente ese relleno. El problema surgía en el momento que el relleno de la mancha espectral era de color blanco, produciendo un '0' lógico en todos los planos y en consecuencia, se capturaba únicamente el borde de la mancha (éste siempre pintado de color negro), produciendo por ello un error exponencial a la hora de detectar ésta en concreto que variaba en función del grosor de la mancha.

Para solventar el problema expuesto anteriormente se ha optado por realizar siempre la captura con un relleno en color negro, esto significa que nos aseguramos que el algoritmo capta el relleno, ya que este color marcará con un '1' lógico los tres planos RGB, quedando como en la figura 24.

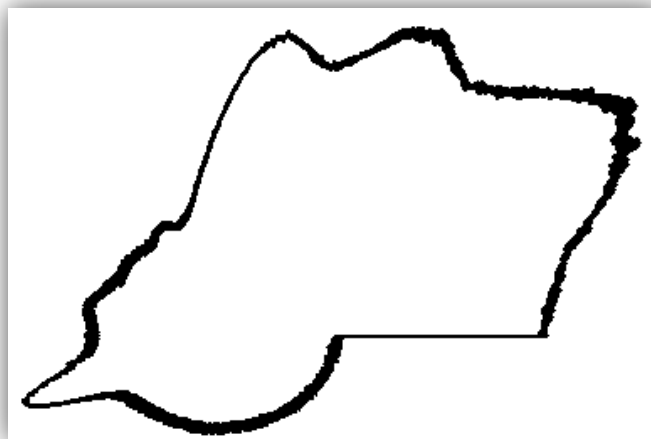


Figura 24: Ejemplo de captura de una categoría de aceite



Figura 25: Operación que realiza el algoritmo tras la captura

Por lo tanto, queda demostrado que aunque en el proyecto [1] mencionado se sugiriese que aplicando esta técnica el error mejoraría sustancialmente, no es correcto ya que no se tuvo en cuenta la operación 'OR' de los tres planos que realiza el algoritmo. En definitiva, con esta medida se obtiene una corrección de los casos en los que el relleno era de color blanco.

4.2 Validación cruzada, leave-one-out

Esta técnica es quizá la más importante dentro de las técnicas de validación cruzada, ya que se realiza un entrenamiento con cada uno de los espectros de todas las categorías a analizar. Esto provoca una independencia total de una posible discrepancia de espectros tomados en diferentes momentos, además de producir un cálculo más preciso que mediante las otras técnicas. Como punto detractor se puede tomar como referencia, su aspecto costoso computacionalmente.

En el siguiente capítulo se dedicará a la exposición de medias y cruces con diversos materiales y así poder observar el grado de mejora que supone esta técnica implantada.

4.3 Espectroscopia hiper-espectral

Este punto es quizá, lo más novedoso del proyecto. Ya que hasta ahora se han tratado ficheros planos con espectros, tomados en diferentes espacios del tiempo.

Este análisis se basa en un sistema óptico de visualización hiper-espectral [2] que permite identificar y clasificar materiales en función de su reflectancia difusa. Con

ayuda de dispositivos ópticos específicos se pueden analizar muestras de materiales en diferentes rangos espectrales: ultravioleta, visible o infrarrojo cercano. Si se opta por el rango visible-infrarrojo cercano, se podría analizar el área de visión y la zona del infrarrojo cercano, es decir, aproximadamente desde una longitud de onda de 400nm a los 1160nm.

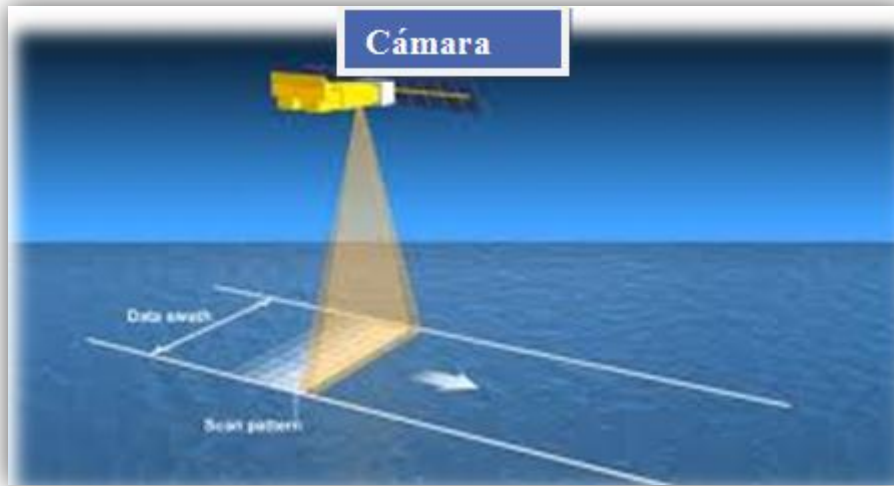


Figura 26: Captura de imagen hiperespectral

A partir de la captura obtenida por este sistema óptico citado, se formará un cubo hiper-espectral en función de las dos variables espaciales y la longitud de onda, que capturará el rango del espectro sensible por éste, en cada posición del espacio como se observa en la Figura 27.

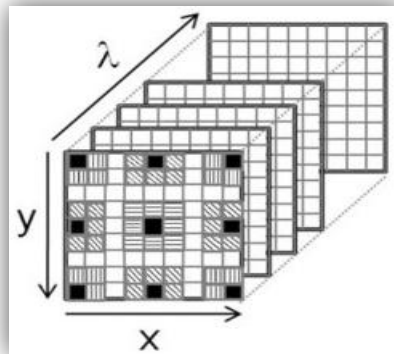


Figura 27: Cubo hiper-espectral

Este cubo se cargará en el algoritmo desarrollado, ya que se le ha adaptado para dotarlo de capacidad de proceso de imágenes en 3D, y así procesar capturas obtenidas a partir de este sistema óptico citado.

Este espectrógrafo captura los espectros en datos RAW, que se deben tratar. Un archivo RAW contiene todos los datos en bruto, que captura el sensor y ocupa mucho más espacio que otro en formato .jpg. Absolutamente todos los datos se conservan en el archivo, y esto nos permite trabajar con la imagen, ya sea en color o en blanco y negro.

Este espectrógrafo deberá capturar adicionalmente dos imágenes adicionales; una imagen negra y otra blanca, para poder adecuar los datos y calibrar espectralmente las medidas tomadas, mediante la siguiente operación:

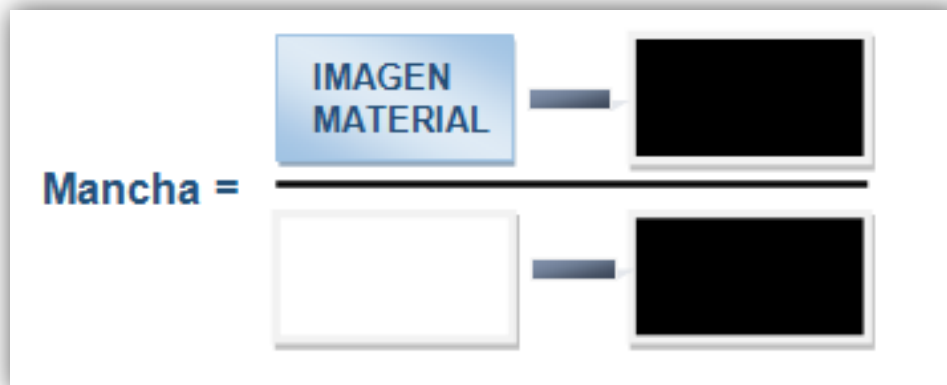


Figura 28: Operación con datos RAW

Las imágenes que se leerán son de 1208 x 1024, pero como se aprecia en la siguiente Figura, los límites inferior y superior son dependientes de la sensibilidad del ImSpector y la resolución de la cámara, por lo que se recortará para introducir un menor error en la medición, además de aligerar el peso computacional que este método exige a la máquina.

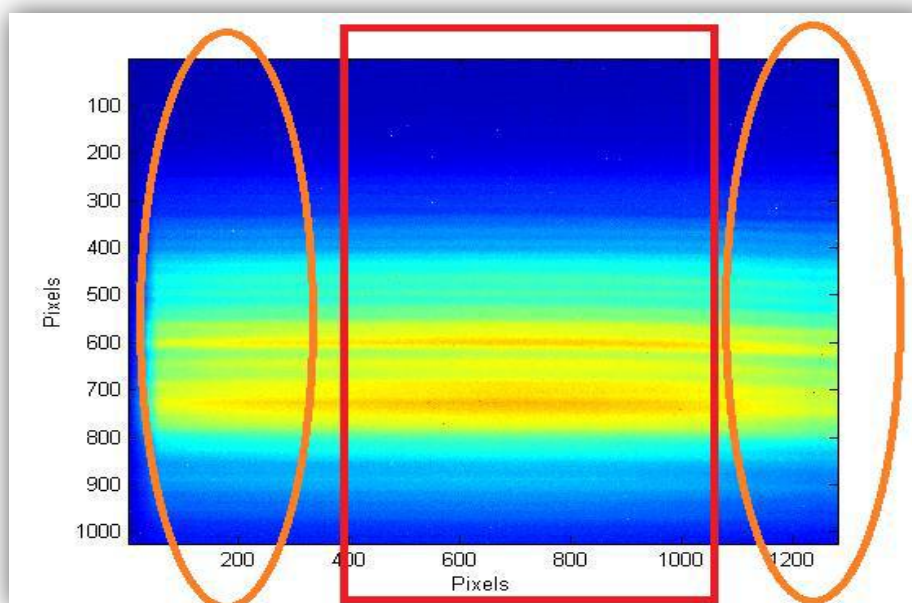


Figura 29: Imagen blanca marcada con óvalos las zonas eliminadas

Este sistema se podría instalar en situaciones varias. En un vehículo para ir tomando muestras de la carretera e identificar posibles irregularidades del terreno en directo o realizar un despliegue cuidadoso en un laboratorio con el fin de identificar compuestos y añadiduras en estos, etc.

4.4 Aumento del rendimiento

Para llevar a cabo una medición de esta mejora técnica se va a utilizar unas mediciones de tiempos utilizando los comandos de Matlab “tic & toc”. Esta funcionalidad permitirá dar tiempos de ejecución de la aplicación, utilizando para ello en ambas ejecuciones los mismos parámetros y ficheros de entrada.

La primera medida que se tomó, fue considerar una preasignación de variables, arrays y matrices antes de entrar en bucles mediante una asignación de zeros. La asignación previa de memoria o “*Preallocating*” evita la necesidad de MATLAB de copiar los datos de una matriz a otra dentro del bucle, además de evitar cambiar el tamaño de una matriz cada iteración, agrandando o disminuyendo sus dimensiones, siendo esto muy costoso computacionalmente. Un ejemplo de esta medida utilizada en el código sería lo siguiente:

```
matriz_ar=[];
for i=3:length(ficheros)
    path=strcat(camino,'\ ',ficheros(i).name);
    mat=dlmread(path,'\t',2);
    matriz_ar=[matriz_ar mat(:,2)];
end
```

Figura 30: Captura de espectros sin preallocating

```
[f_mat ~]=size(mat);
matriz_ar=zeros(f_mat,length(ficheros)-3);
for i=3:length(ficheros)
    path=strcat(camino,'\ ',ficheros(i).name);
    mat=dlmread(path,'\t',2);
    matriz_ar(:,i-2)=mat(:,2);
end
```

Figura 31: Captura de espectros tras preallocating

Como se puede observar, funcionalmente ambas implementaciones son exactamente iguales, no siendo así en términos de tiempos y aprovechamiento de la CPU. Esta limpieza del código se ha plasmado en todo el algoritmo, a partir de la cual se mostrarán posteriormente los resultados de los tiempos de ejecución.

Tiempos de Ejecución				
	Shuffle sin mejora	Shuffle con mejora	Leave-one-out sin mejora	Leave-one-out con mejora
Nuevo Análisis	670,45 seg	63.95 seg	2462.57 seg	393.83 seg
Generar Mancha	597.55 seg	36.67 seg	2388.89 seg	363.91 seg

Tabla 2: Tiempos de ejecución c/s preallocating

Comparando los tiempos de ejecución, de los dos métodos de validación cruzada plasmados en este estudio con esta mejora de rendimiento y sin ella, se ha ganado un aumento de velocidad de proceso en la técnica Shuffle, reduciendo los tiempos entorno al 90% , y en la técnica de leave-one-out entorno al 85%. Además se puede concluir que sin volver a realizar la captura de datos, es decir pulsando sobre el botón de ‘Generar Mancha’ se reducen los tiempos de ejecución aproximadamente a 30 segundos.

4.5 Memory Leaks

Este epígrafe está destinado a la explicación de los grandes problemas de memoria que surgían por causa de la implementación en años anteriores de las diversas funcionalidades de esta aplicación. Matlab se basa sobre JAVA y por lo tanto dispone de una JVM (Máquina Virtual Java) que interpreta las instrucciones de código y de una *memory heap*, que es la encargada de guardar en memoria los objetos creados durante la ejecución. Un usuario programador de Matlab no suele tener este hecho en consideración gracias a las grandes dimensiones de esta memoria.

El problema surgió al desplegar el nuevo método de validación cruzada leave-one-out. El nuevo método de validación realiza muchísimos más cruces, en definitiva esto genera multitud de objetos que sin tener en cuenta lo anteriormente explicado, carga la memoria heap de Matlab, provocando un error de out-of-memory y por consiguiente, aunque el código realice la funcionalidad prevista, el PC no soportaba esa carga, constituyendo que nunca llegase al fin de la ejecución. Así que se procedió a optimizar el algoritmo eliminando en cada parte del código de la aplicación, no sólo en la funcionalidad que lo requería, cada uno de los objetos guardados en memoria que ya no se usaban.

Este problema si se hubiese generado la aplicación sobre JAVA, no existiría ya que existe un proceso denominado Garbage Collector, que se encarga de esto mismo.

4.6 Otras Mejoras

En esta sección se comentarán todas aquellas mejoras desplegadas, que no provocan grandes evolutivos pero si ayudas en diferentes puntos.

Una de ellas es la gran modulación que ha sufrido el código, con el propósito de dejar al siguiente investigador de esta ciencia, un algoritmo claro y legible y por lo tanto fácil de mejorar y amplificar.

Otra, es la introducción de un mensaje, que salta al pulsar sobre uno de los dos botones de ejecución; ‘Nuevo Análisis’ o ‘Generar Mancha’ que mostrará al usuario los parámetros, que utilizará la aplicación para capturar los análisis propuestos por éste. Esta medida surge por problema de persistencia de datos durante la sesión al ir de una funcionalidad a otra, pudiendo evitar tener que esperar hasta el final de la ejecución para observar que no se ha analizado tal como se esperaba.

Otra medida introducida y quizá la más importante de este bloque, es la ejecución de las figuras utilizadas para las capturas (por la función *“imcapture”*, mencionada en puntos anteriores) en background, produciendo un aumento en la velocidad de ejecución y un mejor efecto de cara al usuario de la aplicación, como

consecuencia de no tener que visualizar infinitas imágenes abriendo y cerrando en cuestión de décimas de segundo.

Como última medida introducida, se trata de un fichero de configuración a partir del cual, se pueden introducir las rutas de los espectros a analizar y así evitar la tediosa tarea de seleccionar categoría a categoría mediante un buscador.

5. Resultados

5.1 Materiales analizados

En este capítulo se van a mostrar resultados obtenidos en diferentes cruces realizados para cada tipo de espectroscopia. Para ello como se ha visto hasta ahora, se van a dividir en dos juegos de datos generales en función de si es utilizado para la espectroscopia de transmisión o de reflectancia difusa.

5.1.1 Juego de datos para la espectroscopia de transmisión

En concreto para este tipo de espectroscopia se han realizado análisis para tres tipos de materiales.

5.1.1.1 Aceite

La palabra aceite es un término genérico para designar numerosos líquidos grasos de orígenes diversos que no se disuelven en el agua y que tienen menor densidad que ésta. Dentro de los comestibles, los aceites vegetales aportan ácidos grasos insaturados y son ricos en vitamina E.

En este proyecto se estudiarán medidas de 9 categorías distintas de aceite de las que se dispone, todas ellas procedentes de dos fuentes; la semilla y el aceite de oliva.

En este TFC se ha realizado un análisis para la discriminación de estos 9 tipos de aceite entre ellos. Una aplicación muy útil, sería para comprobar el estado de un tipo de aceite e incluso comprobar el grado de pureza de éste. Como se ha podido entender en la introducción actualmente se venden, incluso aceites “lampantes” siendo estos perjudiciales para la salud. Ya que aplicando espectroscopia de transmisión, seremos capaces de obtener variaciones entre los diferentes componentes de los que dispone, cada variedad de aceite.

Aceites provenientes de semillas



Figura 32: Aceite de girasol



Figura 33: Aceite de lino



Figura 34: Aceite de sesamo



Figura 35: Aceite de uva

Variedades dentro del aceite de oliva



Figura 36: Aceite de oliva virgen



Figura 37: Aceite de oliva virgen Arbequina



Figura 38: Aceite de oliva virgen Cornicabra



Figura 39: Aceite de oliva virgen Hojiblanca



Figura 40: Aceite de oliva virgen Picual

5.1.1.2 Café

El café es la bebida que se obtiene a partir de las semillas tostadas y molidas de los frutos de la planta de café o cafeto.

En este proyecto se estudiarán medidas de las 5 clases distintas de variedades de café de las que se dispone, todas ellas procedentes de la semilla tostada.

En este TFC se ha realizado un análisis para la discriminación de estos 5 tipos de café entre ellos. Una aplicación muy útil, sería para comprobar el estado de un tipo de café e incluso comprobar el grado de pureza de éste, comprobando que el color oscuro que posee no es debido a azúcares añadidos. Ya que aplicando espectroscopia de transmisión, seremos capaces de obtener variaciones entre los diferentes componentes de los que dispone, cada variedad de café.



Figura 41: Café Forza



Figura 42: Café Suave



Figura 43: Café Costa Rica



Figura 44: Café Kenya



Figura 45: Café Descafeinado

5.1.1.3 Vino

El vino es una bebida obtenida de la uva mediante la fermentación alcohólica de su mosto o zumo. La fermentación se produce por la acción metabólica de levaduras que transforman los azúcares del fruto en alcohol etílico y gas en forma de dióxido de carbono

En este proyecto se estudiarán medidas de las ocho clases distintas de variedades de vinos de las que se dispone, todas ellas procedentes de la uva y agrupadas así en tres grupos: vinos *tintos*, *blancos* y *rosados*.

En este TFC se ha realizado un análisis para la discriminación de estos 7 tipos de vinos entre ellos. Una aplicación muy útil, sería para comprobar el estado de un

tipo de vino en comparación con un espectro de otro que se sepa que es de buena calidad. Ya que aplicando espectroscopia de transmisión, seremos capaces de obtener variaciones entre los diferentes componentes de los que dispone, cada variedad de vino

Blancos



Figura 46: Vino Riesling

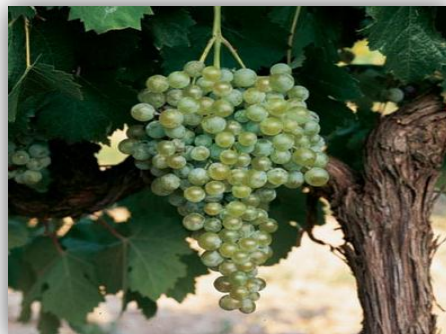


Figura 47: Vino Verdejo

Rosados



Figura 48: Vino Garnacha



Figura 49: Vino Lambrusco

Tintos



Figura 50: Vino Gabernet



Figura 51: Vino Mencia



Figura 52: Vino Tempranillo

5.1.2 Juego de datos para espectroscopia híper-espectral

Dentro del análisis de la espectroscopia híper-espectral se van a utilizar combinaciones de diversos tipos de materiales, desde el cartón hasta un hilo pasando por el tabaco. Lo interesante de este tipo de espectroscopia tan novedosa en la aplicación, es que se puede analizar de una manera rápida y sencilla, cualquier material que tenga reflectancia difusa.

En este TFC se ha querido realizar diversos análisis, con diferentes combinaciones. Incluyendo en muchas de ellas el tabaco, con la finalidad de ver si el algoritmo sería capaz de identificarlo entre diferentes cubos híper-espectrales de otros compuestos.

A continuación se mostraran las manchas hiperespectrales, sin tratar, de cada uno de los materiales utilizados en el análisis.



Figura 53: Mancha híper-espectral del cartón

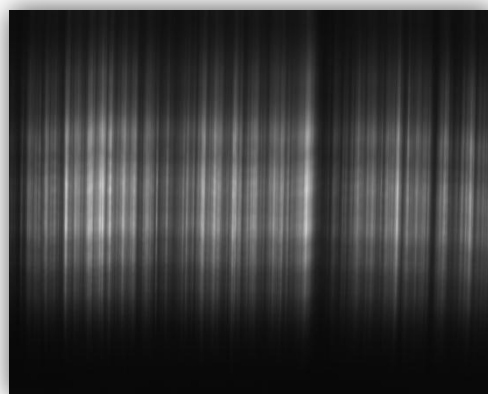


Figura 54: Mancha híper-espectral del papel aluminio



Figura 55: Mancha hiper-espectral del film transparente



Figura 56: Mancha hiper-espectral de piel



Figura 57: Mancha hiper-espectral de madera

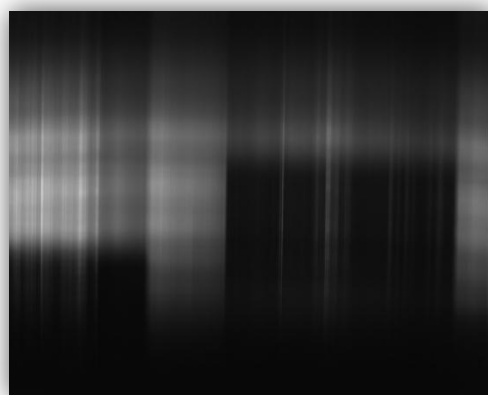


Figura 58: Mancha hiper-espectral de caramelo opaco

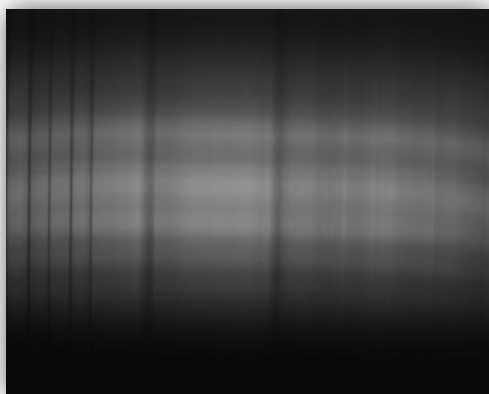


Figura 59: Mancha hiper-espectral de papel caramelo blanco



Figura 60: Mancha hiper-espectral de papel transparente de plástico

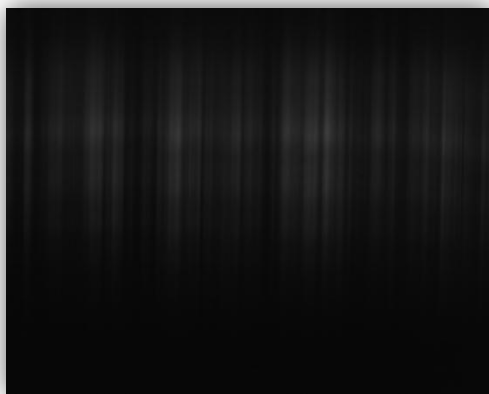


Figura 61: Mancha hiper-espectral de tabaco estado 1

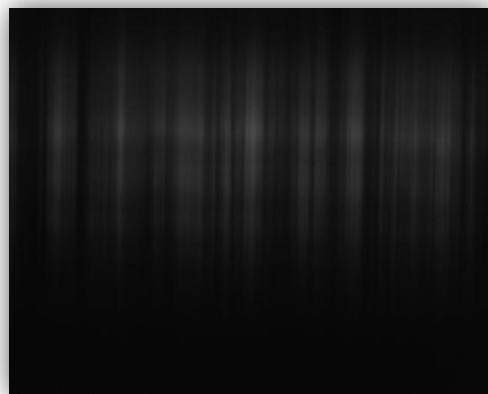


Figura 62: Mancha hiper-espectral de tabaco estado 2

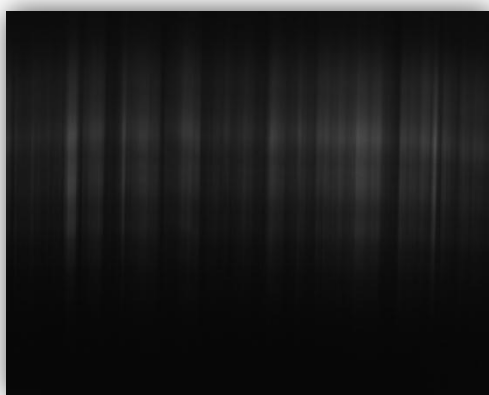


Figura 63: Mancha hiper-espectral de tabaco estado 3

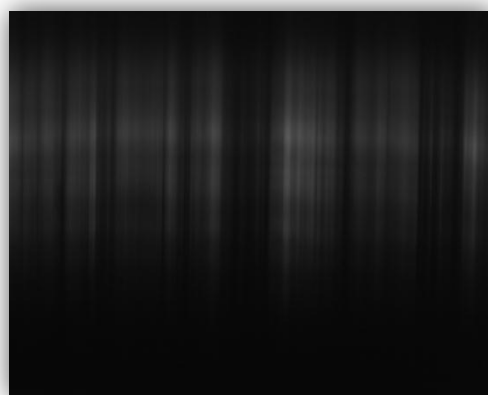


Figura 64: Mancha hiper-espectral de tabaco estado 4



Figura 65: Mancha hiper-espectral de tabaco estado 5

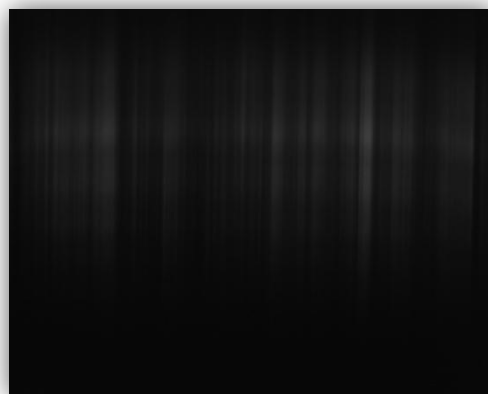


Figura 66: Mancha hiper-espectral de tabaco estado 6

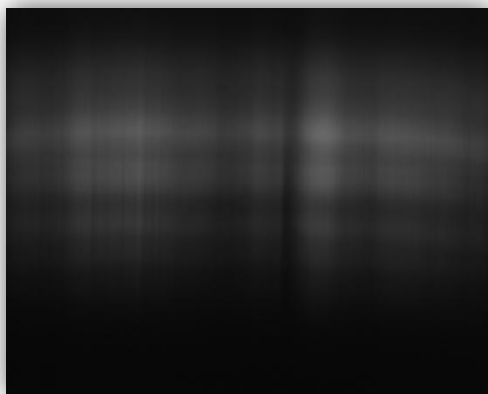


Figura 67: Mancha hiper-espectral de hojas secas

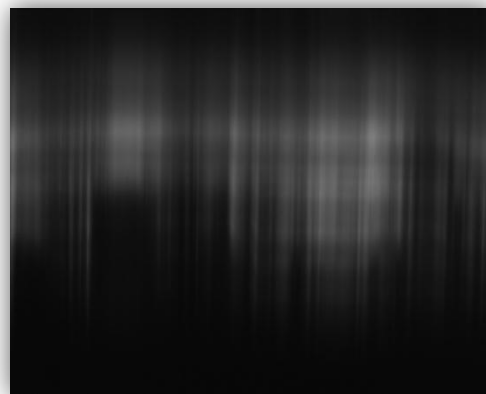


Figura 68: Mancha hiper-espectral de hilo

5.2 Validación cruzada, Shuffle

En el anexo 1, se plasma el análisis exhaustivo que se ha realizado entre las diferentes categorías de aceite, café y vino. Con unos parámetros de entrada:

- Resolución: 0,25 Mpx
- Percentiles: 40 – 60 y 25 – 75
- Dominio espectral
- Técnica de validación cruzada: Shuffle
- Número de partes: 3,5,8,10,12,15

Con el fin de buscar el valor más óptimo, para el parámetro partes, que se debe indicar al elegir como método de validación cruzada, shuffle.

En el presente apartado, se va a intentar dar una cierta coherencia a estos resultados. Para ello, se muestran en la tabla 3, los errores totales por cada uno de los anteriores análisis marcados.

Material	Percentil (%)	Número de partes					
		3	5	8	10	12	15
Aceite	40 - 60	11.55	10.67	11.65	9.80	9.69	15.36
	25 - 75	4.68	5.66	4.68	4.47	4.79	6.86
Café	40 - 60	56.47	57.65	58.23	56.86	58.82	59.02
	25 - 75	56.67	53.72	55.29	55.88	58.04	56.86
Vino	40 - 60	66.67	63.09	63.09	63.09	63.09	63.09
	25 - 75	61.31	60.12	60.71	60.71	60.71	60.71

Tabla 3: Tabla con los errores relativos referentes a los análisis discriminatorios del anexo 1

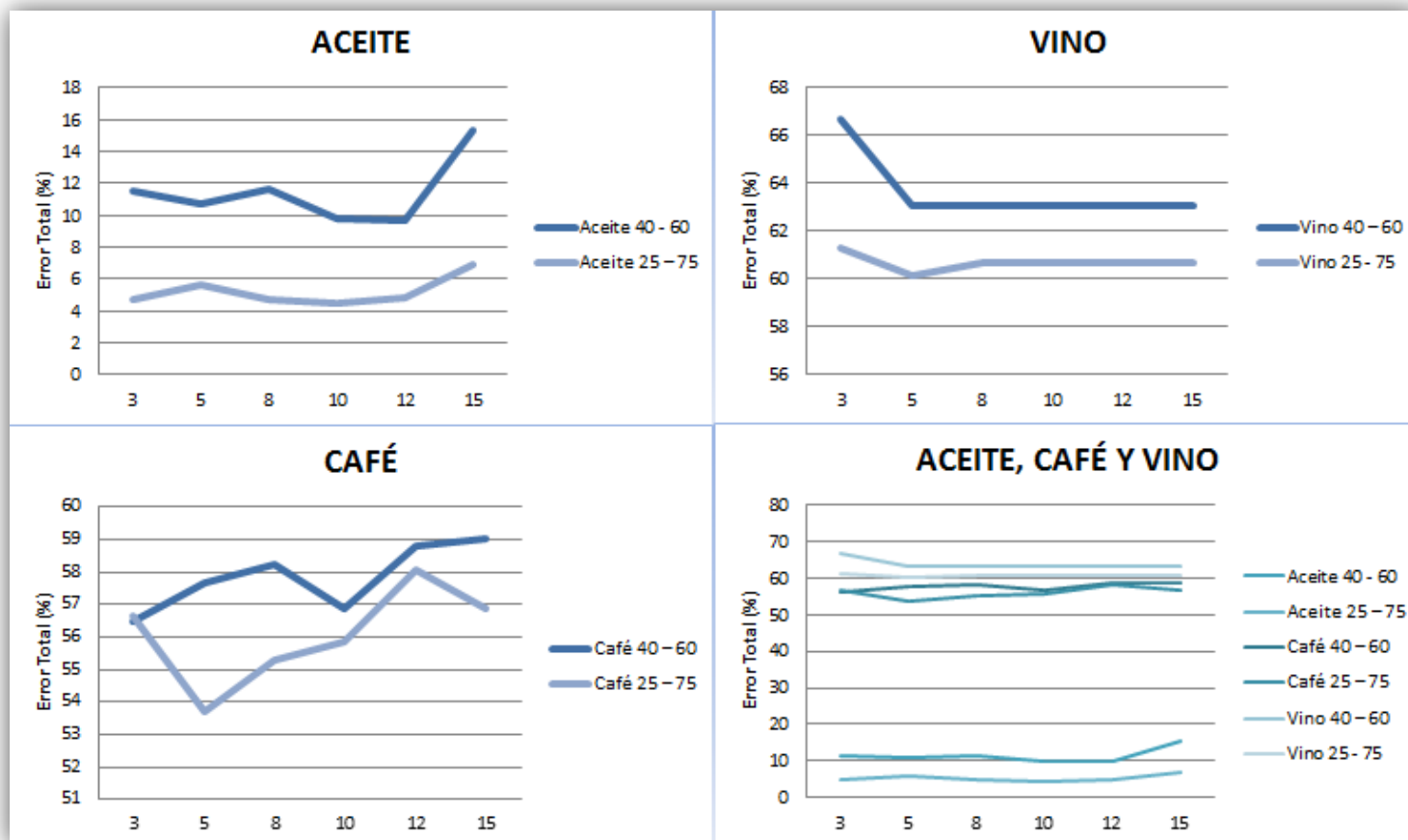


Figura 69: Gráficas con los errores relativos referentes a los análisis discriminatorios del punto 4.3

Como se puede observar en las gráficas, este parámetro influye de manera significativa, variando en mayor grado tanto y cuanto, los valores capturados en diferentes días, difieran mucho de la mediana. Como se expone al comienzo de este apartado, dedicado a esta técnica, se han mantenido el resto de parámetros, variando únicamente el indicador de ‘partes’ de la técnica, Shuffle, con la finalidad de poder dar a entender el valor de este parámetro estudiado.

En las figuras 70 – 81, se muestran las manchas espectrales que corresponderán con las de cada uno de los casos, en las que el error obtenido ha sido mayor y menor, en función de dos valores de percentiles diferentes. Como se puede apreciar, el error aumenta en función de la diferencia de datos apartados en cada uno de los entrenamientos y viceversa. Por lo que el usuario final, tendrá que variar este parámetro hasta el punto en el que se distribuyan los espectros, que desea analizar, equitativamente en función de sus pesos, absorbancias o transmitancias a analizar en los tres entrenamientos.

En conclusión, se observa que este parámetro difiere, en función de los espectros a analizar y el grosor o percentil que se introduzca en la aplicación. Además se puede intuir que si el valor introducido es alto, el error aumenta, por lo que el parámetro óptimo, se encontrará entre tres y doce partes. Asimismo, este parámetro influye en gran medida en función de la discrepancia que exista en las medidas entre unos días y otros. Esta última conclusión se contempla muy bien en la figura 69, en ella se expone como realmente los espectros capturados para las siete categorías de vino no difieren en exceso de unos días a otros y por tanto no influye demasiado este parámetro, no siendo así en las medidas tomadas para las categorías de café. En este último caso,

se puede observar como el usuario deberá buscar con exactitud, la efectividad de este parámetro o volver a realizar las medidas ya que nos servirá de indicativo, de que estas medidas no son muy correctas.

A continuación se va a realizar una comparación entre resultados obtenidos en el TFC de Lucia Uriarte Calva y los obtenidos en éste, para intentar determinar dónde y cuánto es el grado de mejora en esta nueva implantación.

Por lo tanto se comenzaran mostrando dos tablas con los resultados que obtuvo sobre las 9 categorías de aceite, analizadas también en este proyecto. Los parámetros de entrada son los mismos para ambos análisis a excepción de la resolución, por ello se tomaran las muestras que se obtuvieron con una resolución inferior y superior, a la analizada en el presente proyecto.

Lambda	c1	c2	c3	c4	c5	c6	c7	c8	c9	No clasificada	Totales
c1	100	0	0	0	0	0	0	0	0	0	0
c2	0	89,17	0	0	10	0	0	0	0	0,83	10,83
c3	0	7,5	92,5	0	0	0	0	0	0	0	7,5
c4	0	0	0	100	0	0	0	0	0	0	0
c5	0	0,83	0	0	99,17	0	0	0	0	0	0,83
c6	0	0	0	0	0	69,17	0	0	30,83	0	30,83
c7	0	0	0	0	0	0	100	0	0	0	0
c8	0	0	0	0	0	11,67	0	88,33	0	0	11,67
c9	0	0	0	0	0	11,67	0	0	88,33	0	11,67
										Media	8,15

Tabla 4: Matriz de confusión de Lucia para el aceite con una resolución de 0,1 Mpx y un grosor del 25 - 75 %

Lambda	c1	c2	c3	c4	c5	c6	c7	c8	c9	No clasificada	Totales
c1	100	0	0	0	0	0	0	0	0	0	0
c2	0	100	0	0	0	0	0	0	0	0	0
c3	0	0	100	0	0	0	0	0	0	0	0
c4	0	0	0	100	0	0	0	0	0	0	0
c5	0	0,83	0	0	99,17	0	0	0	0	0	0,83
c6	0	0	0	0	0	82,5	0	10	7,5	0	17,5
c7	0	0	0	0	0	18,33	5,83	75,83	0	0	94,17
c8	0	0	0	0	0	16,67	0	83,33	0	0	16,67
c9	0	0	0	0	0	0	0	0	100	0	0
										Media	14,35

Tabla 5: Matriz de confusión de Lucia para el aceite con una resolución de 1 Mpx y un grosor del 25 - 75 %

Como se puede observar, como se ha explicado en puntos anteriores, el error total es directamente proporcional a la resolución. Por lo que el resultado esperado para una resolución de 0,25 Mpx que es la utilizada para el análisis en este proyecto, debe estar en un varemos de 8,15% – 14,35% para considerarse una medida aceptable, más de 14,35% para considerarla como una peora y menos de 8,15% para considerarse una mejora. A continuación se expondrán mediante tres matrices de confusión el resultado del análisis utilizando el valor óptimo en partes, 10 en este caso.

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	91.18								8.82	
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						94.12				5.88
Lino							100.00			
Sésamo								100.00		
Uva									100.00	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	88.24							8.82		2.94
Oliva		100.00								
Arbequina			100.0							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	2.94							97.06		
Uva	20.59								79.41	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	100.00									
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	35.29							64.71		
Uva	35.29								64.71	

Tabla 6: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75%, 0.25Mpx y 10 partes

El resultado del análisis ha sido un error total de 4,47%, por lo que supone una mejora realmente notable, ya que no solo se ha bajado el error incluso por debajo de situaciones con menos resolución, sino que además se ha bajado el lote de sobrantes por lo que la medida es más exacta, haciendo ésta independiente de las discrepancias en los datos obtenidos. Además como se puede observar para el aceite de oliva Picual, hay una gran dependencia de datos, produciendo por ello una gran disminución en la tasa de fallos en su confusión con el aceite de Lino.

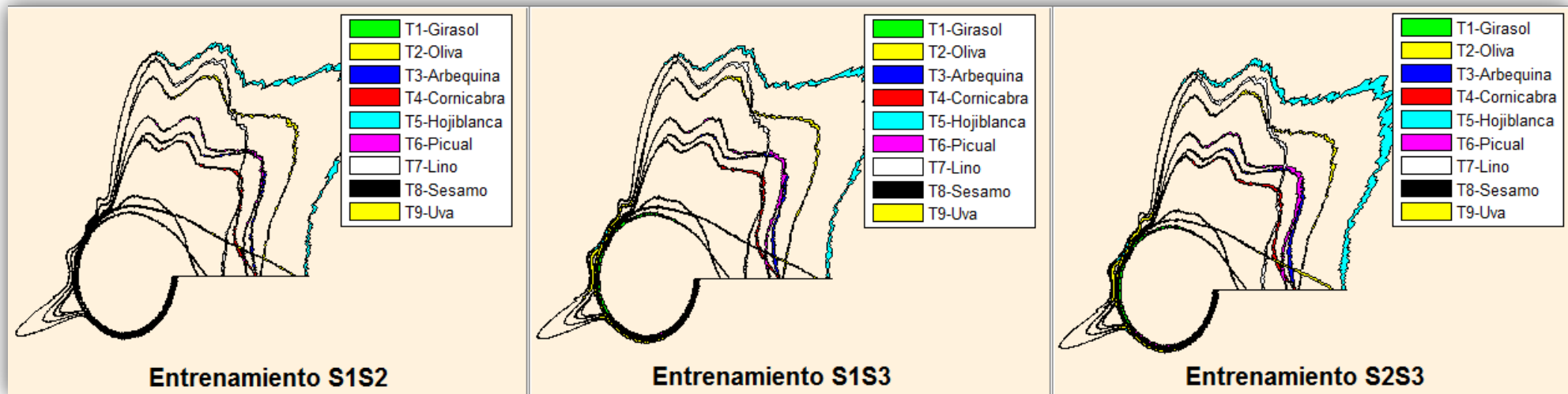


Figura 70: Manchas espectrales de aceite 40 – 60% 12 partes (error menor)

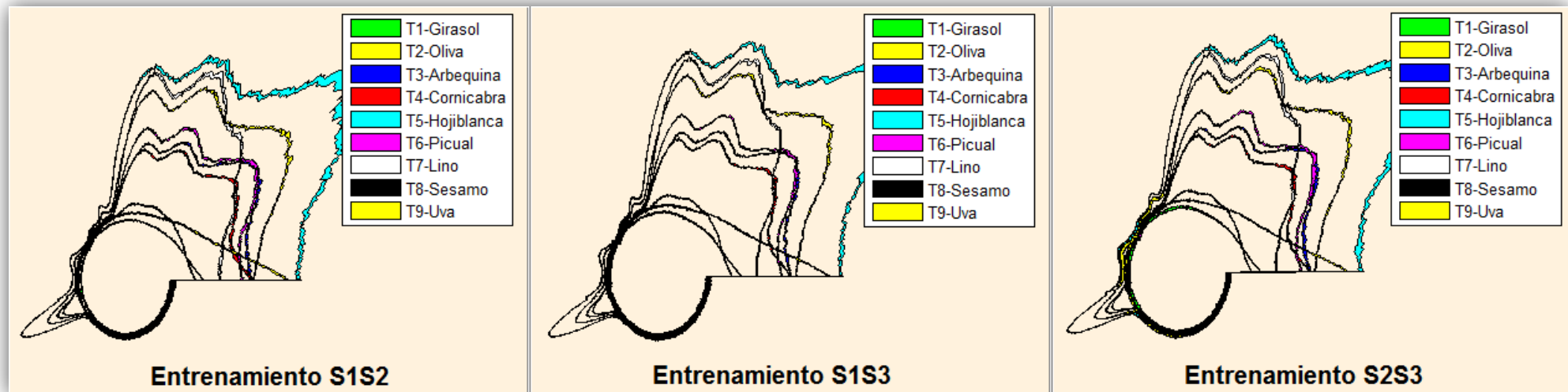


Figura 71: Manchas espectrales de aceite 40 – 60% 15 partes (error mayor)

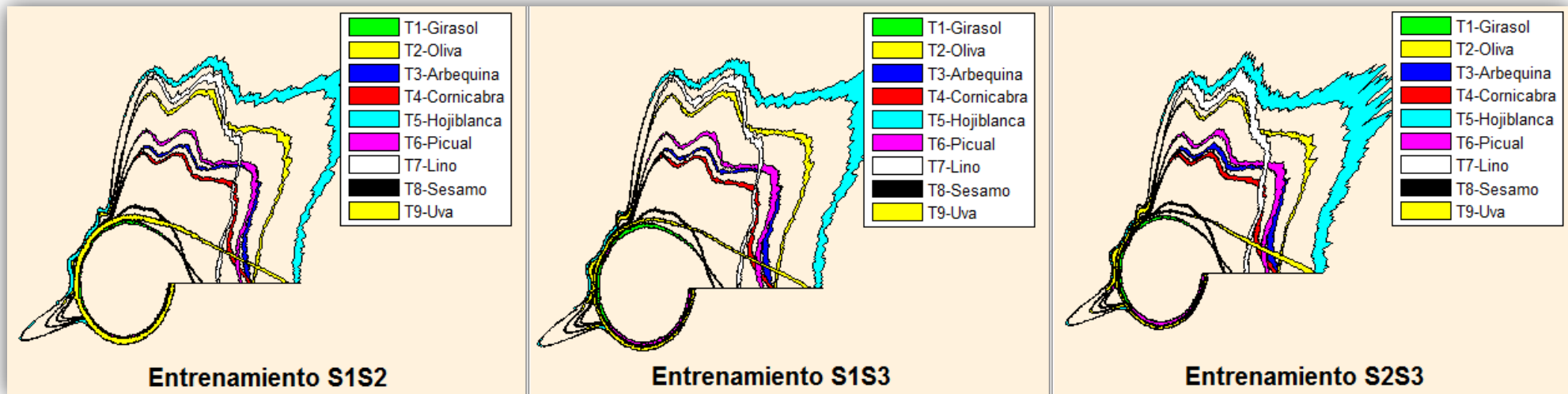


Figura 72: Manchas espectrales de aceite 25 – 75% 10 partes (error menor)

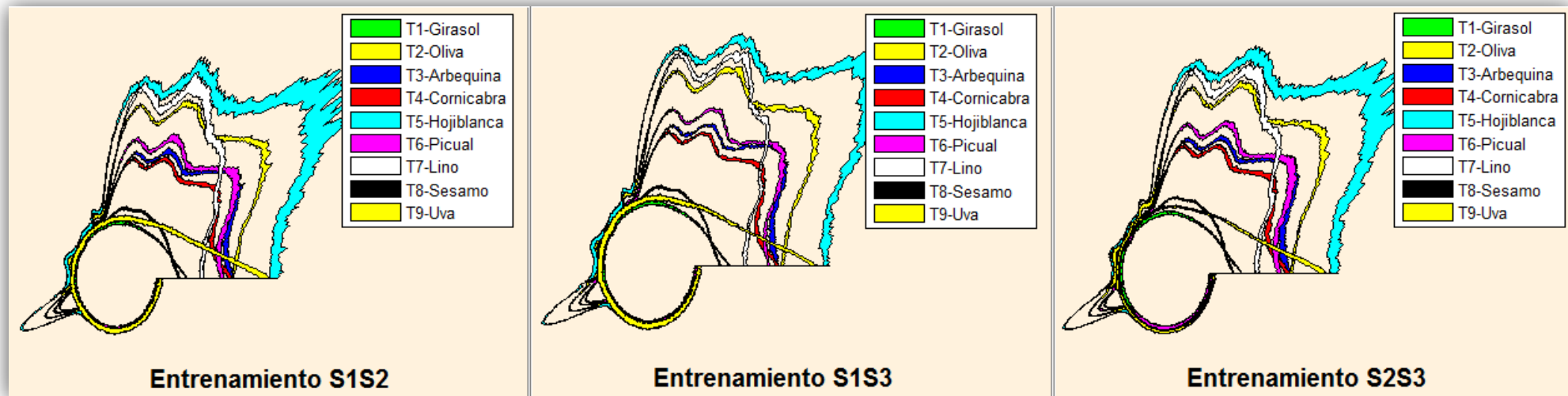


Figura 73: Manchas espectrales de aceite 25 – 75% 15 partes (error mayor)

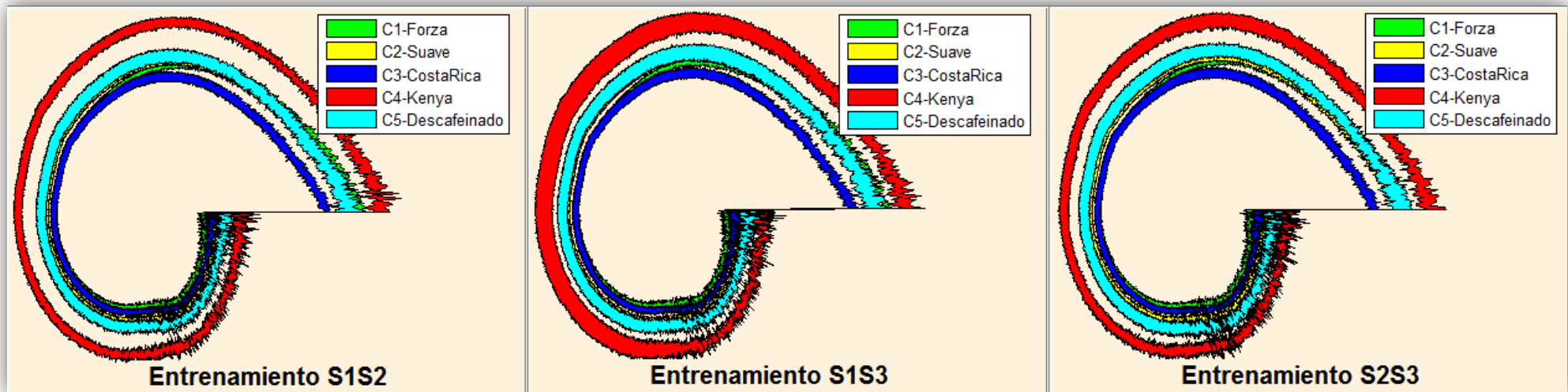


Figura 74: Manchas espectrales de café 40 – 60% 3 partes (error menor)

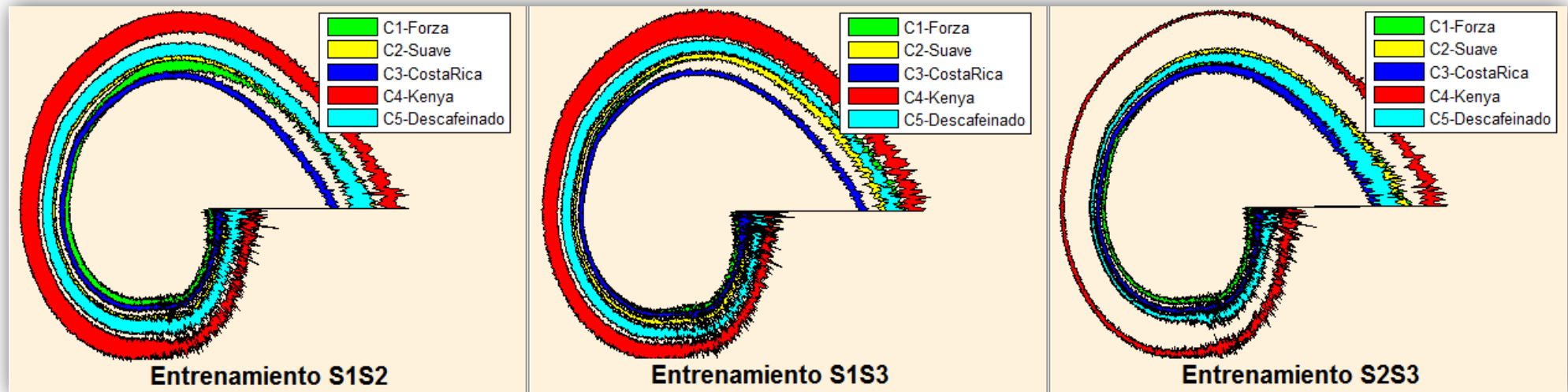


Figura 75: Manchas espectrales de café 40 – 60% 15 partes (error mayor)

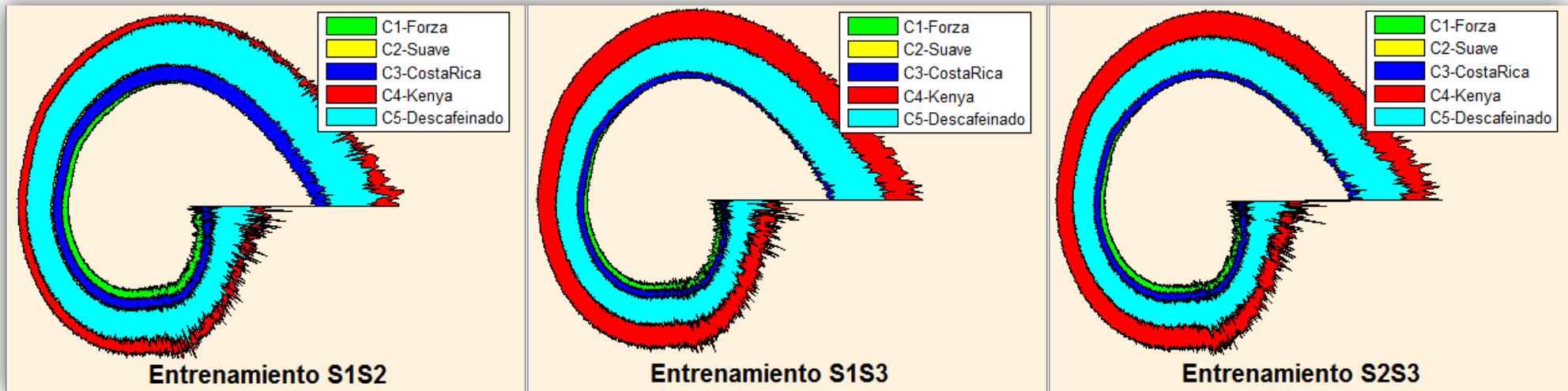


Figura 76: Manchas espectrales de café 25 – 75% 5 partes (error menor)

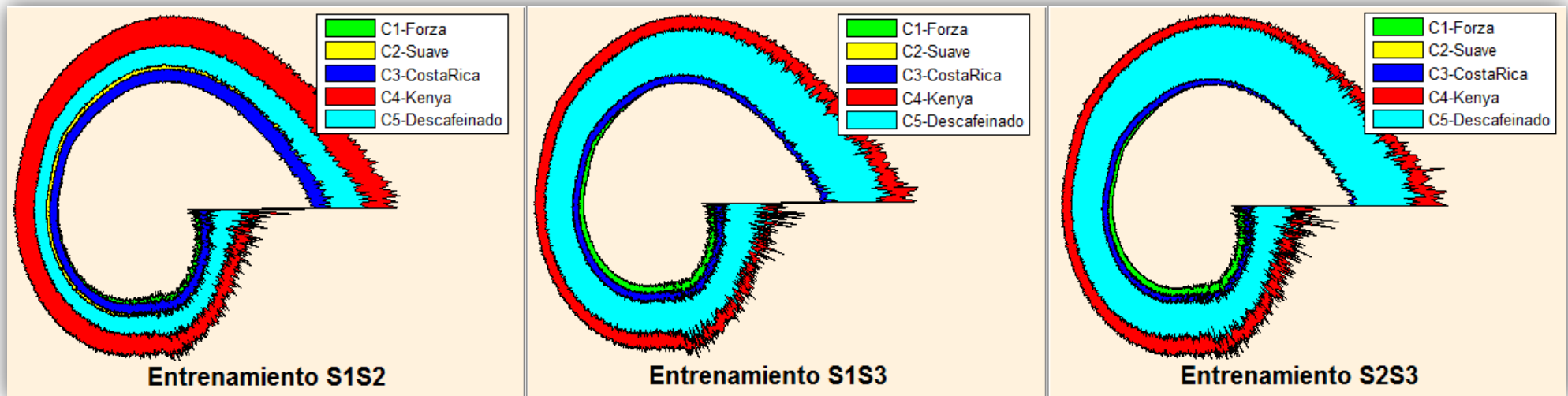


Figura 77: Manchas espectrales de café 25 – 75% 12 partes (error mayor)

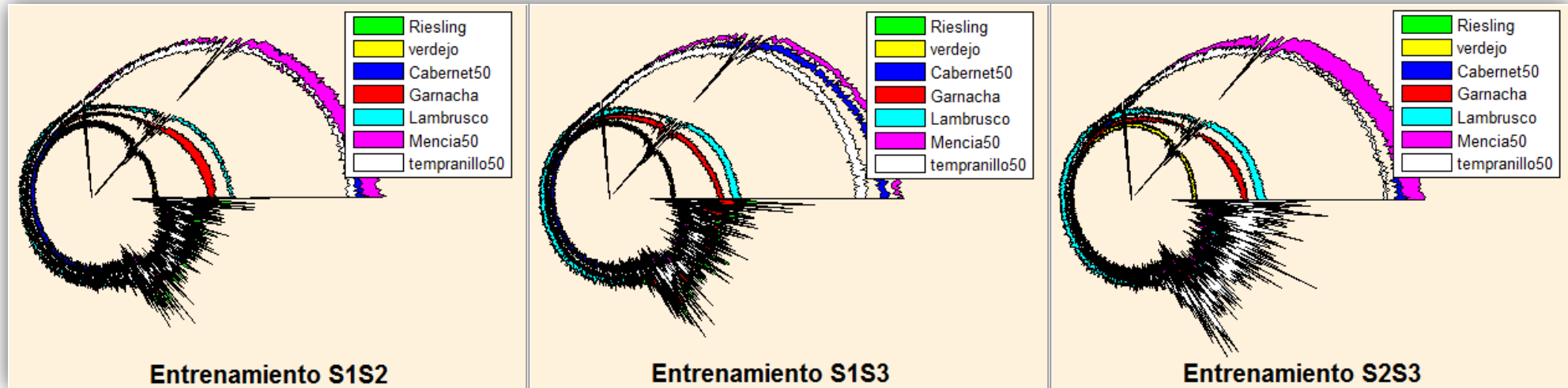


Figura 78: Manchas espectrales de vino 40 – 60% 5 partes (error menor)

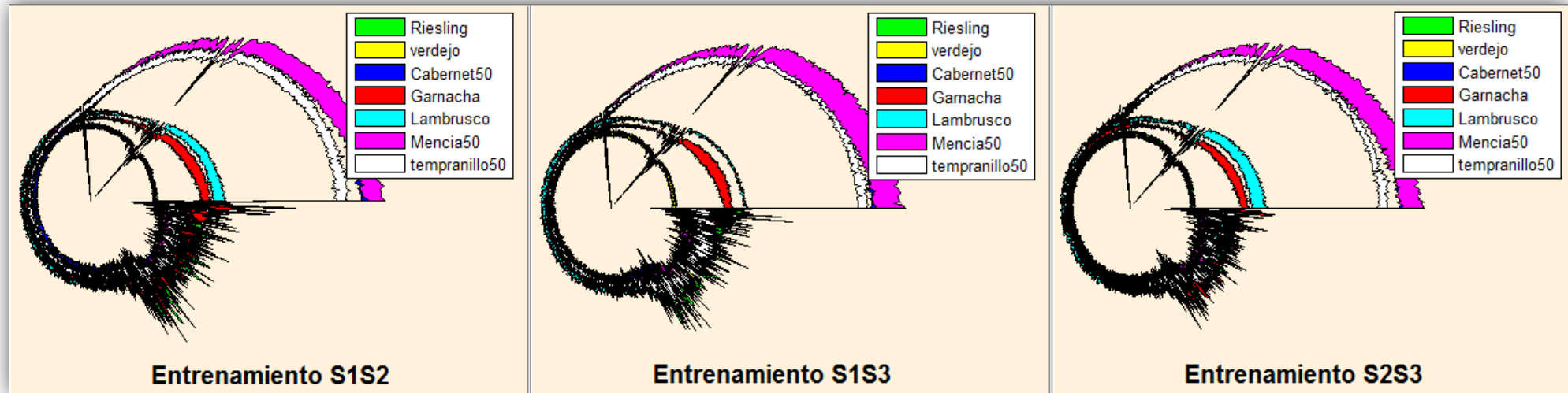


Figura 79: Manchas espectrales de vino 40 – 60% 3 partes (error mayor)

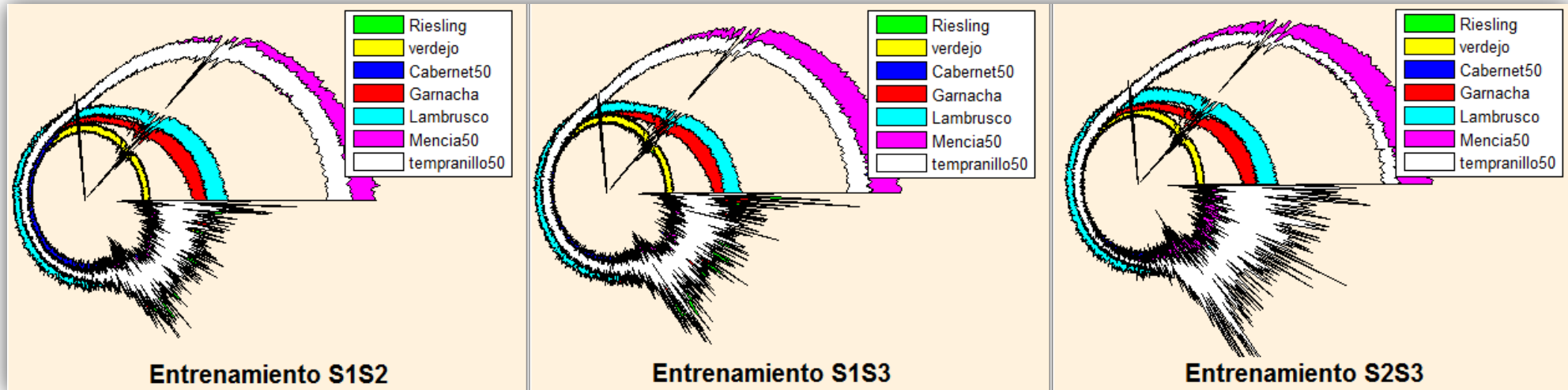


Figura 80: Manchas espectrales de vino 25 – 75% 5 partes (error menor)

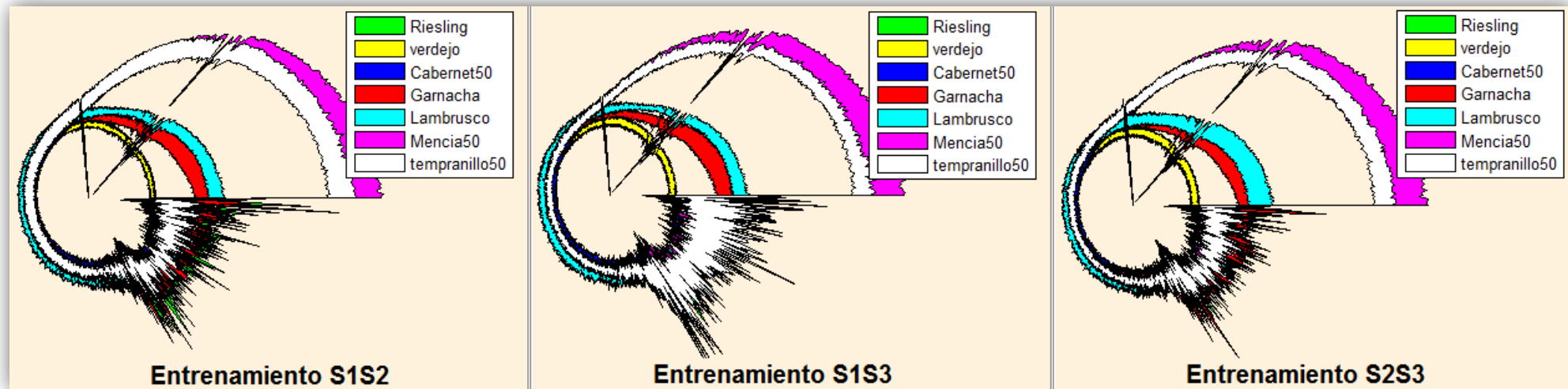


Figura 81: Manchas espectrales de vino 25 – 75% 3 partes (error mayor)

5.3 Validación cruzada, Leave-one-out

Ésta es la técnica de validación cruzada más novedosa desplegada en el algoritmo. Por lo que se va a realizar un análisis en ambos dominios; espectral y PCA. Por lo tanto en este apartado, igual que su homónimo en el punto anterior se va a dedicar a realizar un análisis de los resultados obtenidos, para esta técnica. Realizando para ello una impresión de los errores totales producidos para cada uno de los casos del anexo 2.

Material	Percentil (%)	Resolución			
		0.25 MPx	0.56 MPx	1 MPx	2 MPx
Aceite	40 - 60	11.89	18.44	19.89	23.22
	25 - 75	4.22	6.78	6.89	10.44
Café	40 - 60	53.8	54.4	55.8	57.4
	25 - 75	56.2	53.8	54.2	54.2
Vino	40 - 60	56.43	52.86	53.57	55.71
	25 - 75	49.28	50.71	51.43	57.86

Tabla 6: Tablas con los errores relativos referentes a los análisis discriminatorios del anexo 2 en el dominio espectral

Material	Percentil (%)	Resolución			
		0.25 MPx	0.56 MPx	1 MPx	2 MPx
Aceite	40 - 60	1.89	2.44	4.11	5.56
	(4 PCAs) 25 - 75	0.22	0.78	2.44	4.67
Café	40 - 60	100.00			
	(10 PCAs) 25 - 75				
Vino	40 - 60	56.43	59.28	60.00	60.71
	(10 PCAs) 25 - 75	40.00	42.14	43.57	41.43

Tabla 7: Tablas con los errores relativos referentes a los análisis discriminatorios del anexo 2 en el dominio PCA

A continuación se mostraran las matrices de confusión del punto anterior, correspondiente al análisis de las 9 categorías del aceite con un percentil de 25-75% y una resolución de 0,25 Mpx con la finalidad de poder comparar los resultados, variando únicamente el parámetro de técnica de validación cruzada o “cross-validation”. Para proceder seguidamente a comentar debidamente las ventajas e inconvenientes de cada uno de estos métodos. En la tabla 8 se podrá observar la matriz de confusión para la técnica Shuffle con los parámetros mencionados.

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	91.18								8.82	
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						94.12				5.88
Lino							100.00			
Sésamo								100.00		
Uva									100.00	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	88.24							8.82		2.94
Oliva		100.00								
Arbequina			100.0							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	2.94							97.06		
Uva	20.59								79.41	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	100.00									
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	35.29							64.71		
Uva	35.29								64.71	

Tabla 8: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75%, 0.25Mpx y técnica Shuffle con 10 partes

El error obtenido para esta técnica, Shuffle, es del 4,47%. Se puede observar en las matrices impresas anteriormente, como la tasa de error crece en su mayoría al clasificar el aceite de uva, confundiendo este con el aceite de girasol en exceso. Esto implica que hay componentes que ambos comparten. Seguidamente se muestra la matriz de confusión para la novedosa técnica, leave-one-out, con el fin de observar si ésta, nos puede ayudar a clasificar mejor.

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	95.00							5.00		
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	12.00							87.00		1.00
Uva	20.00								80.00	

Tabla 9: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75% y 0.25Mpx y leave-one-out

Como se puede observar al comparar, estos resultados con los obtenidos para el caso de la técnica de Shuffle controlado, el error en este caso disminuye al 4,22%. Ya que se ha subido el grado de acierto del aceite de uva hasta el 80%, siendo esto una mejora aunque no muy significativa si se compara con su gran coste computacional.

. Por lo que se deduce que con esta técnica se puede simplificar el análisis del usuario y una bajada de rendimiento, debido al grado de multitud de operaciones que genera esta funcionalidad.

Además comparando todos los resultados obtenidos con esta técnica y la técnica Shuffle, se puede concluir, que se estima que el mejor análisis se produce con un parámetro de resolución de 0,25 MPx, 25 – 75 % grosor, y sobre todo mediante la inclusión en el análisis de PCA, (exceptuando en la clasificación del café, como consecuencia de una primera componente principal muy similar entre las diferentes categorías, teniendo que realizar la discriminación en este caso sin PCA) donde se puede observar que con un valor de 4 PCAs como valor óptimo para el aceite y de 10 PCAs para el vino, el error total disminuye drásticamente, además de utilizar como método de validación cruzada, leave-one-out, que aunque computacionalmente sea realmente costoso obtiene el mejor análisis posible y abstrae al usuario de buscar puntos de inflexión, aplicando cualquier otro mecanismo.

5.4 Clasificación de datos híper-espectrales

Este análisis es el más novedoso e importante dentro del algoritmo, ya que hasta ahora se habían realizado análisis del sector alimentario. Con este tipo de espectroscopia se puede capturar cualquier material que emita reflectancia difusa, la gran mayoría. Por lo que se van a analizar 6 agrupaciones de materiales, dividiendo estos en dos grandes grupos:

- Grupo 1: Materiales de la misma familia
 - Estudio 4: Papel aluminio, papel caramelo blanco, papel caramelo opaco, papel transparente de plástico
 - Estudio 5: Tabaco estado 1 - 6

- Grupo 2: Materiales aparentemente inconexos
 - Estudio 1: Cartón y papel aluminio
 - Estudio 2: Cartón, papel aluminio y film transparente
 - Estudio 3: Piel y madera
 - Estudio 6: Hojas secas, hilo, piel y tabaco estado 1

Mediante estos dos grupos, podemos analizar el grado de eficiencia de este algoritmo. Si es capaz de analizar con una tasa de error aceptable el grupo 2 y no el grupo 1, nos encontramos ante un algoritmo capaz de identificar cualquier compuesto sino es de su familia. Si por el contrario se consiguen discriminar ambos grupos, el algoritmo se podría clasificar de imperiosamente bueno. Pudiendo ser un filón comercial.

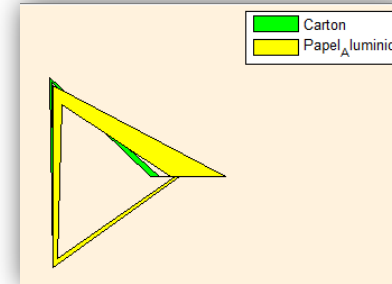
Por lo tanto a continuación se expondrán los estudios, realizando variaciones en los parámetros de análisis:

- Dominio PCA
 - Número de PCAs: 4
 - Número de PCAs: 10
- Dominio espectral

Estudio 1:

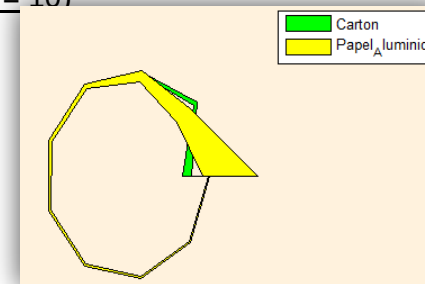
Método de Discriminación: PCA (PCA = 4)

Matriz de Confusión	Cartón	Papel aluminio	No clasificado
Cartón	80.68	19.32	
Papel aluminio	3.28	96.32	0.40
Error total	11.50		



Método de Discriminación: PCA (PCA = 10)

Matriz de Confusión	Cartón	Papel aluminio	No clasificado
Cartón	50.85	48.62	0.53
Papel aluminio		100.00	
Error total	24.57		



Método de Discriminación: Dominio espectral

Matriz de Confusión	Cartón	Papel aluminio	No clasificado
Cartón	100.00		
Papel aluminio	2.10	97.90	
Error total	1.05		

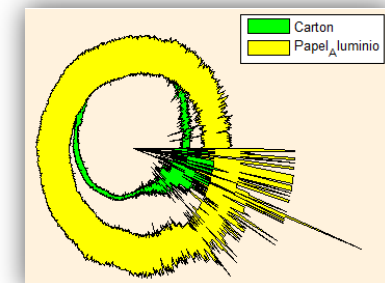
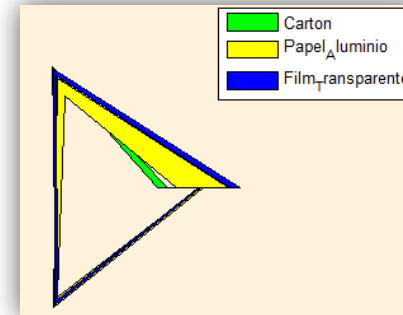


Figura 82: Matriz de confusión obtenida para el cartón y papel aluminio

Estudio 2:

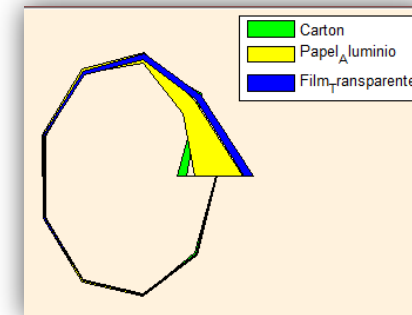
Método de Discriminación: PCA (PCA = 4)

Matriz de Confusión	Cartón	Papel aluminio	Film transparente	No clasificado
Cartón	54.66	30.22	14.45	0.66
Papel aluminio	0.53	93.03	5.52	0.92
Film transparente	0.39	18.40	81.08	0.13
Error total	23.74			



Método de Discriminación: PCA (PCA = 10)

Matriz de Cofusión	Cartón	Papel aluminio	Film transparente	No clasificado
Cartón	48.09	51.64		0.26
Papel aluminio		99.47	0.53	
Film transparente		41.26	58.48	0.26
Error total	31.32			



Método de Discriminación: Manchas Espectrales

Matriz de Confusión	Cartón	Papel aluminio	Film transparente	No clasificado
Cartón	100.00			
Papel aluminio	2.10	84.36	13.53	
Film transparente		18.92	81.08	
Error total	11.519			

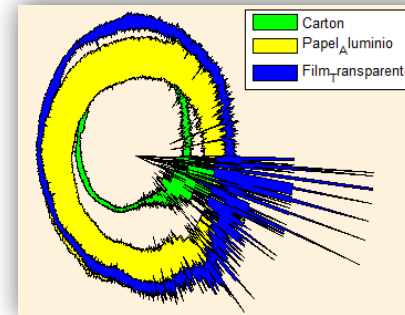
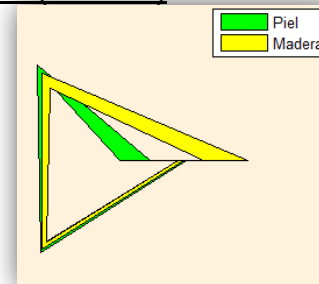


Figura 83: Matriz de confusión obtenida para el cartón, papel aluminio y film transparente

Estudio 3:

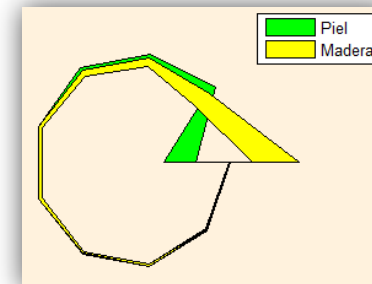
Método de Discriminación: PCA (PCA = 4)

Matriz de Confusión	Piel	Madera	No clasificado
Piel	75.033	24.84	0.13
Madera	28.52	71.35	0.13
Error total	26.81		



Método de Discriminación: PCA (PCA = 10)

Matriz de Confusión	Piel	Madera	No clasificado
Piel	88.83	10.78	0.39
Madera	33.77	65.44	0.79
Error total	22.86		



Método de Discriminación: Manchas Espectrales

Matriz de Confusión	Piel	Madera	No clasificado
Piel	88.30	11.70	
Madera	3.81	96.19	
Error total	7.75		

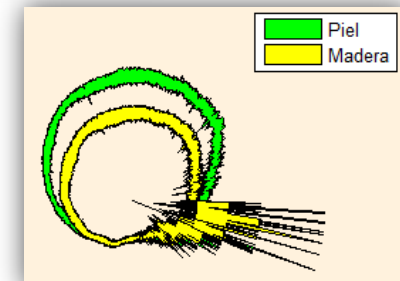
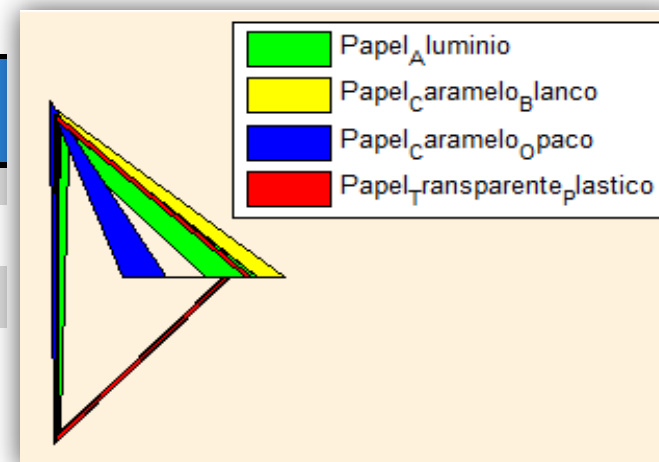


Figura 84: Matriz de confusión obtenida para el piel y madera

Estudio 4:

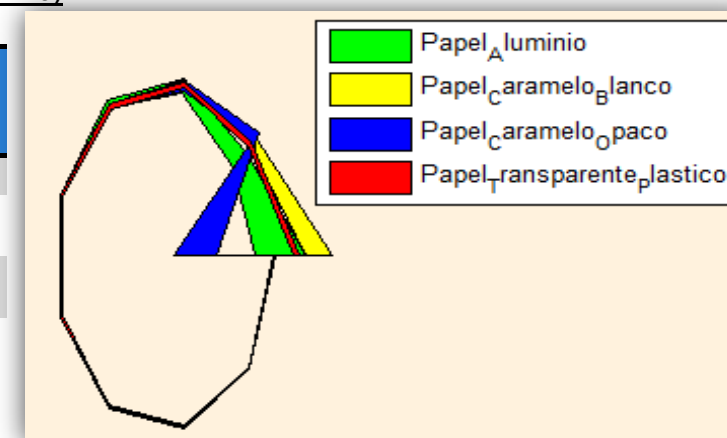
Método de Discriminación: PCA (PCA = 4)

Matriz de Confusión	Papel aluminio	Papel caramelo blanco	Papel caramelo opaco	Papel transparente de plástico	No clasificado
Papel aluminio	81.08	8.80	7.88	2.23	
Papel caramelo blanco	5.52	59.79	21.68	13.01	
Papel caramelo opaco	17.48	5.12	68.86	7.23	1.31
Papel transparente de plástico	24.05	14.85	0.13	60.97	
Error total	32.32				



Método de Discriminación: PCA (PCA = 10)

Matriz de Confusión	Papel aluminio	Papel caramelo blanco	Papel caramelo opaco	Papel transparente de plástico	No clasificado
Papel aluminio	98.82	1.05		0.13	
Papel caramelo blanco	44.55	50.07	0.66	4.20	0.53
Papel caramelo opaco	49.41	1.84	48.23	0.13	0.39
Papel transparente de plástico	53.09	4.73		41.66	0.53
Error total	40.31				



Método de Discriminación: Manchas Espectrales

Matriz de Confusión	Papel aluminio	Papel caramelo blanco	Papel caramelo opaco	Papel transparente de plástico	No clasificado
Papel aluminio	99.01	0.99			
Papel caramelo blanco		81.19		18.81	
Papel caramelo opaco	13.86		86.14		
Papel transparente de plástico		21.78		78.22	
Error total	13.86				

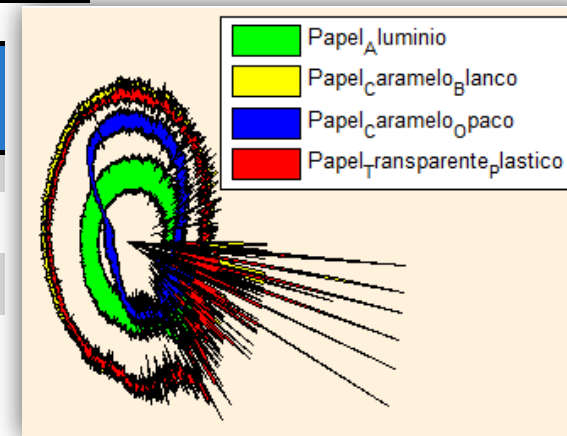
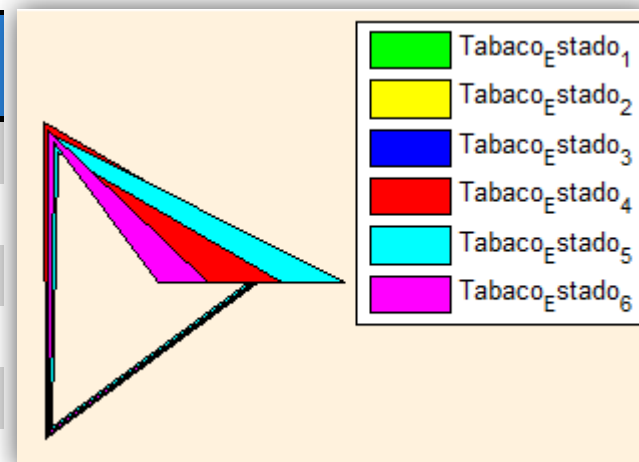


Figura 85: Matriz de confusión obtenida para el papel aluminio, papel de caramelo blanco, papel de caramelo opaco y papel transparente de plástico

Estudio 5:

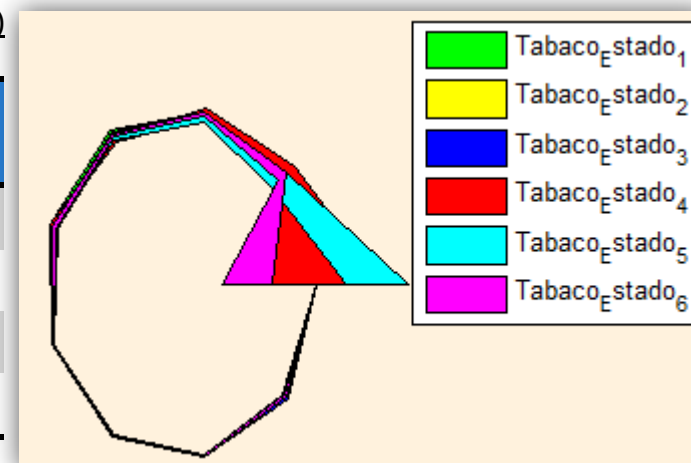
Método de Discriminación: PCA (PCA = 4)

Matriz de Confusión	Tabaco estado 1	Tabaco estado 2	Tabaco estado 3	Tabaco estado 4	Tabaco estado 5	Tabaco estado 6	No clasificado
Tabaco estado 1	29.70	25.74	3.96	14.85	16.83	4.95	3.96
Tabaco estado 2	22.77	29.70	0.99	20.79	11.88	4.95	8.91
Tabaco estado 3	3.96	8.91	7.92	15.84	55.45	1.98	5.94
Tabaco estado 4	4.95	10.89		45.54	37.62	0.99	
Tabaco estado 5	4.95	7.92		17.82	68.32		0.99
Tabaco estado 6	8.91	30.69		0.99	23.76	24.75	10.89
Error total	65.68						



Método de Discriminación: PCA (PCA = 10)

Matriz de Confusión	Tabaco estado 1	Tabaco estado 2	Tabaco estado 3	Tabaco estado 4	Tabaco estado 5	Tabaco estado 6	No clasificado
Tabaco estado 1	50.50	21.78	3.96	4.95	10.89	5.94	1.98
Tabaco estado 2	40.59	24.75	5.94	13.86	9.90	3.96	0.99
Tabaco estado 3	10.89	4.95	51.49	11.88	19.80	0.99	
Tabaco estado 4	8.91	4.95	7.92	55.45	18.81	3.96	



Tabaco estado 5	14.85	9.90	3.96	8.91	62.38		
Tabaco estado 6	26.73	29.70	6.93	0.99	4.95	28.71	1.98
Error total	54.45						

Método de Discriminación: Manchas Espectrales

Matriz de Confusión	Tabaco estado 1	Tabaco estado 2	Tabaco estado 3	Tabaco estado 4	Tabaco estado 5	Tabaco estado 6	No clasificado
Tabaco estado 1	1.98	9.90		77.23	10.89		
Tabaco estado 2	0.99	16.83		80.20	1.98		
Tabaco estado 3			9.90	40.59	48.51		0.99
Tabaco estado 4		12.87	0.99	59.41	25.74	0.99	
Tabaco estado 5				31.68	68.32		
Tabaco estado 6		41.58		55.45		2.97	
Error total	73.43						

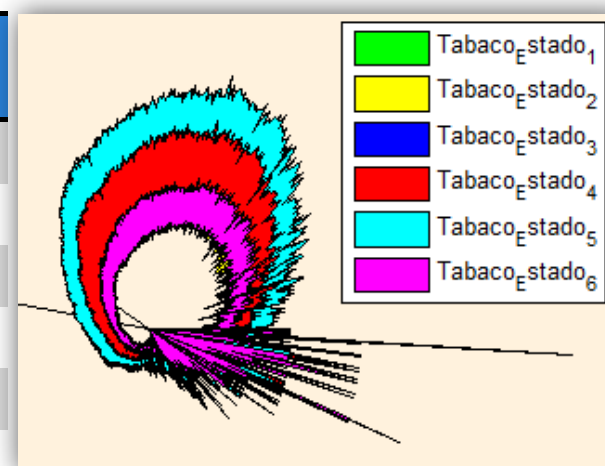
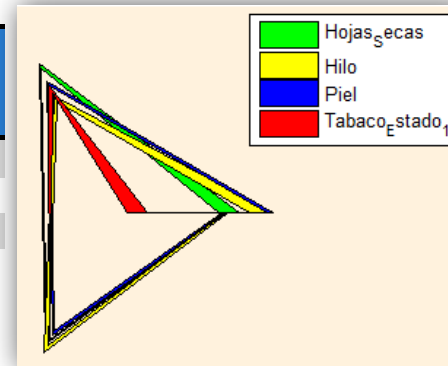


Figura 86: Matriz de confusión obtenida para el tabaco estado 1, tabaco estado 2, tabaco estado 3, tabaco estado 4, tabaco estado 5 y tabaco estado 6

Estudio 6:

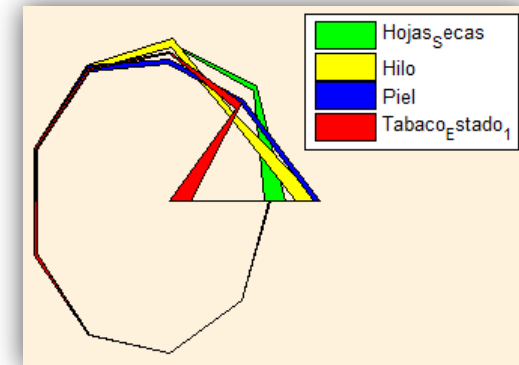
Método de Discriminación: PCA (PCA = 4)

Matriz de Confusión	Hojas secas	Hilo	Piel	Tabaco estado 1	No clasificado
Hojas secas	93.07	6.93			
Hilo		94.06	0.99	4.95	
Piel			100.00		
Tabaco estado 1		4.95	5.94	89.11	
Error total	5.94				



Método de Discriminación: PCA (PCA = 10)

Matriz de Confusión	Hojas secas	Hilo	Piel	Tabaco estado 1	No clasificado
Hojas secas	87.13	12.87			
Hilo		98.02	0.99	0.99	
Piel		7.92	91.09	0.99	
Tabaco estado 1		4.95		94.06	0.99
Error total	7.42				



Método de Discriminación: Manchas Espectrales

Matriz de Confusión	Hojas secas	Hilo	Piel	Tabaco estado 1	No clasificado
Hojas secas	98.02	1.98			
Hilo	4.95	95.05			
Piel	3.96	0.99	95.05		
Tabaco estado 1	1.98			98.02	
Error total	3.46				

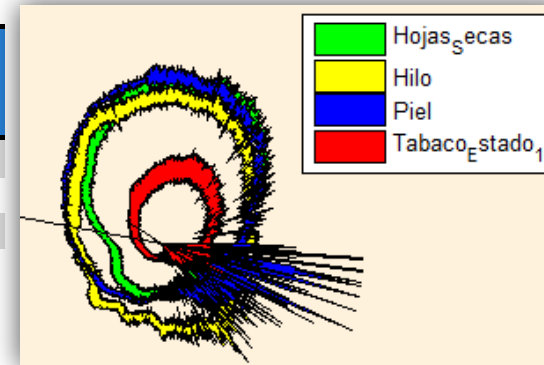


Figura 87: Matriz de confusión obtenida para hojas secas, hilo, piel y tabaco estado 1

A la luz de los resultados obtenidos en el apartado anterior, se puede observar que utilizando como método de discriminación el dominio espectral (excepto en el análisis de los diferentes estados de tabaco, que son manchas muy parecidas, y se podrá obtener un mejor acierto utilizando, la técnica PCA), se están obteniendo grados de aciertos, más que considerables. Por lo que la implementación de esta funcionalidad en la aplicación ha sido todo un éxito. Provocando que cruzando mediante leave-one-out estas capturas, obtenidas a partir del sistema óptico explicado anteriormente, el ingeniero dedicado a esta ciencia pueda analizar de una forma espectacular y rápida cada una de sus investigaciones, siendo éstas de lo más variopintas.

En conclusión cabe destacar, que el algoritmo ha obtenido grados de aciertos desde el 1,05% al 13,86% en prácticamente la totalidad de los estudios. Provocando únicamente, para diferenciar entre los diversos estados del tabaco, un error del 54,45% ya que se introdujo este análisis como algo complejísimo y prácticamente imposible, así que se puede decir el algoritmo es impresionantemente bueno.

Cabe nombrar a la luz de los resultados, que el parámetro óptimo para discriminar este tipo de espectroscopia es en el dominio espectral, a no ser que los materiales a identificar sean muy similares. En este caso, se debe utilizar como dominio, el dominio PCA y buscar en este su valor óptimo en el número de PCAs, ya que como se ha podido ver difiere mucho el acierto con éste.

6. CONCLUSIONES

6.1 Conclusiones

En este apartado dedicado a las conclusiones, me dedicaré a dar un resultado idóneo final, utilizando los errores totales anteriores y algo muy importante hoy en día, como es el rendimiento.

A simple vista al ver este proyecto se nos pueden pasar por la cabeza diferentes preguntas, como pueden ser ¿Qué método de validación cruzada es el idóneo?, ¿Cómo es que se han hecho análisis de materiales tan diversos?

La respuesta a la primera pregunta es muy simple, depende de lo que busque el usuario final, es decir si busca rapidez y una aproximación al grado de error propuesto para el análisis cualitativo que está realizando en un momento dado, el método que debe aplicar es shuffle. Si por el contrario busca una mayor precisión y dispone de tiempo o una CPU más que aceptable su técnica es la validación cruzada leave-one-out, que como se ha demostrado, los resultados que ofrece son más que aceptables en cuanto a discriminación de diversos materiales, llevando esto a un campo muy interesante como es, ¿Hasta dónde puedo identificar? La respuesta es simple, todo aquello que emita o absorba luz en alguna medida, se podrá identificar en esta aplicación.

Como conclusión final, cabe destacar el grado de acierto en la clasificación de manchas híper-espectrales, abriendo una gran vía de investigación y quien sabe, de gran ayuda en cualquier proceso de análisis y discriminación de compuestos.

6.2 Líneas futuras

Las líneas futuras que se pueden proponer son desde un punto de vista práctico y experimental pudiendo ser éstas numerosísimas desde el punto de vista del campo de la espectroscopía.

Ante todo lo primero que habría que realizar en un futuro sería un mayor número de estudios sobre los materiales captados mediante cubos hiperespectrales y así poder definir con un menor grado de incertidumbre, lo exacto de estas medidas tomadas y llevarlo más allá, a la ciencia médica donde creo que puede haber un gran futuro.

Además se podrían desarrollar nuevos métodos de análisis sobre estos dos ya creados, como son el método KNN (estudia el número de muestras que se encuentran lo más cerca posible de la muestra problema) o SAM (mide la similitud entre dos espectros en función del ángulo que forman si son tratados como vectores), para así darle un mayor poder de análisis a este programa, además de ir mejorando partes gráficas con el fin de parecerse cada día más a un programa bonito y sencillo desde el aspecto visual.

Otra vía es el llevar esta aplicación a la nube, mediante su desarrollo web, implementando EJBs (*Enterprise Java Beans*: Beans u objetos del tipo de la clase de la que se este haciendo, gestionados por un contenedor del servidor de aplicaciones para gestionar el acceso a recursos (bases de datos, colas de mensajes, ficheros, etc) y proporcionar servicios (seguridad, transacciones, mensajería, nombres) de forma sistemática y optimizada) que simplifica el desarrollo de aplicaciones web y permite construir aplicaciones escalables, que corran en servidores de aplicaciones, además de utilizar las distintas caches que tendrá el programador a su disposición, desde el servidor de aplicaciones hasta el navegador del usuario final. Esto permitirá versionar la aplicación, llegando a todos los usuarios de ella en el momento, sin tener que bajarse un cliente con la nueva distribución, y aumentar el grado de satisfacción de éste. Además de mejorar el grado de rendimiento de la aplicación en gran medida, claro está en función de la máquina en la que se despliegue, del número de instancias levantadas, etc. Siendo este un punto muy importante. Incluso mejorarla sobre todo a nivel gráfico, gracias a la creación de hojas de estilos mediante CSS3, javascript como lógica ejecutada en el navegador, frameworks o patrones de desarrollo como spring, bootstrap y la gran tecnología web, HTML5.

Para terminar me gustaría expresar que esta técnica de la espectrometría tiene mucho futuro. Y que además de mis puntos de vista, que como bien he dicho con anterioridad, están basados en la experiencia, se pueden dar desarrollos basados en conocimientos más profundos y técnicos.

Referencias

- [1] Proyecto fin de carrera: *Análisis del rendimiento del algoritmo de manchas espectrales para la clasificación de materiales*, Lucia Uriarte Calva, 2012.
- [2] Proyecto fin de carrera: *Interfaz web y análisis hiper-espectral para cuantificar la presencia de sal en carretera*, Jean Pierre Román Paucar, 2012.
- [3] Artículo: *Spectral and Optimized Marks for Qualitative Material Discrimination*, Olga M. Conde, Lucía Uriarte, Pilar B. García-Allende, Ana M. Cubillas, F. Anabitarte, Jose M. Lopez-Higuera *Sensors Journal*, IEEE (Volume: 12 , Issue: 1), páginas:230 – 236, 2012
- [4] Referencia en la web:
http://e-archivo.uc3m.es/bitstream/10016/10958/1/PFC_Guillermo_Marinas_Miranda.pdf
- [5] Proyecto fin de carrera: *Interfaz gráfica y optimización del método de manchas espectrales para la identificación de materiales*, Rodrigo Nevado Antón, 2011.
- [6] Referencia en la web:
http://www.cesareox.com/opinion/articulos/75844/el_concepto_de_percentil.html
- [7] Artículo: *Aplicaciones del uso de la técnica de espectroscopía de reflectancia difusa en Química forense para la caracterización de autenticidad de sellos de correos de España*, Pilar Campins Falcó y Francisco Manuel Querol Piñón, 2012
- [8] Referencia en la web: <http://www.galeon.com/home3/ssolar/sol.html>
- [9] Libro: *Introducción a la espectroscopia Raman*, Dionisio Bermejo, Serie de química, 1988
- [10] Libro: *Problemas de determinación estructural orgánica por espectroscopia IR*, Enrique Pando Ramos, Secretariado de Publicaciones de la Universidad de Sevilla, 1979
- [11] Libro: *Introducción a la espectrometría de masa de sustancias orgánicas*, Otto Richard Gottlieb y Raimundo Braz Filho, Programa Regional de Desarrollo Científico y Tecnológico, Departamento de Asuntos Científicos, Secretaría General de la Organización de los Estados Americanos, 1976
- [12] Referencia en la web:
<http://es.scribd.com/doc/15532859/MANUAL-DE-GUI-EN-MATLAB>
- [13] Referencia en la web: <http://w3.ualg.pt/~jluis/>

- [14] Referencia en la web:
http://es.encarta.msn.com/encyclopedia_761552490/Espectroscopia.html
- [15] Libro: *Graphics and GUIs with MATLAB*, Patrick Marchand, CRC, 2003
- [16] Libro: *Principal components analysis*, George H. Dunteman, SAGE, 1989
- [17] Libro: *Principal component analysis*, I. T. Jolliffe, Springer, 2002
- [18] Guía: *GUI Graphical User Interface*, Matlab R2010
- [19] Libro: *Multivariate Analyses: Principal component analysis and Independent Component Analysis*, Chapter 9, TLFBOOK, Marcel Dekker, 2004
- [20] Referencia en la web: <http://www.fotonostra.com/grafico/coloresobjetos.htm>
- [21] Referencia en la web: <http://www.esi2.us.es/~jaar/Datos/FIA/T9.pdf>
- [22] Referencia en la web: <http://www.ocu.org/alimentacion/alimentos/informe/aceite-oliva>
- [23] Referencia en la web: <http://www.gestiopolis.com/recursos/>

Anexos

Anexo 1: Medidas tomadas, Shuffle

Anexo 1.1: Aceite

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	58.82							41.18		
Oliva		100.00								
Arbequina			100.00							
Cornicabra			2.94	97.06						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	17.65							61.76	17.65	2.94
Uva	20.59								79.41	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	58.82							41.18		
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	14.71							73.53	11.76	
Uva	17.65								82.35	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	58.82							41.18		
Oliva		100.00								
Arbequina			85.29			14.71				
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	20.59							70.59	8.82	
Uva	20.59								79.41	

Tabla 1: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60% , 0.25Mpx y 3 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	58.2							41.18		
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	17.65							67.65	8.82	5.88
Uva	20.59								79.41	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	85.29							14.71		
Oliva		100.00								
Arbequina			82.35			17.65				
Cornicabra			8.82	91.18						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	23.53							73.53	2.94	
Uva	23.53								76.47	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	73.53							26.47		
Oliva		100.00								
Arbequina			76.47			23.53				
Cornicabra			11.76	85.29						2.94
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	23.53							70.59	2.94	2.94
Uva	23.53								76.47	

Tabla 2: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60%, 0.25Mpx y 5 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	41.18							38.24	20.59	
Oliva		100.00								
Arbequina			73.53	20.59						5.88
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo								94.12	5.88	
Uva	8.82								91.18	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	85.29							14.71		
Oliva		100.00								
Arbequina			82.35			17.65				
Cornicabra			20.59	79.41						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	26.47							73.53		
Uva	26.47								73.53	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	79.41							20.59		
Oliva		100.00								
Arbequina			76.47			23.53				
Cornicabra			29.41	70.59						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	26.47							73.53		
Uva	26.47								73.53	

Tabla 3: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60%, 0.25Mpx y 8 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	61.76							38.24		
Oliva		100.00								
Arbequina			97.06			2.94				
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	8.82							70.59	20.59	
Uva	8.82								91.18	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	58.82							41.18		
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	17.65							64.71	14.71	2.94
Uva	20.59								79.41	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	73.53							26.47		
Oliva		100.00								
Arbequina			88.24			8.82				2.94
Cornicabra			26.47	73.53						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	35.29							64.71		
Uva	35.29								64.71	

Tabla 4: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60%, 0.25Mpx y 10 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	38.24							29.41	32.35	
Oliva		100.00								
Arbequina			97.06	2.94						
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo								82.35	17.65	
Uva	5.88								94.12	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	100.00									
Oliva		100.00								
Arbequina			82.35			17.65				
Cornicabra			8.82	91.18						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	17.65							82.35		
Uva	17.65								82.35	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	79.41							20.59		
Oliva		100.00								
Arbequina			100.00							
Cornicabra			2.94	97.06						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	38.24							61.76		
Uva	38.24								61.76	

Tabla 5: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60%, 0.25Mpx y 12 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	26.47							55.88	17.65	
Oliva		100.00								
Arbequina			58.82	41.18						
Cornicabra				100.00						
Hojiblanca					100.00					
Picual			2.94			97.06				
Lino							100.00			
Sésamo								100.00		
Uva	5.88								94.12	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	58.82							41.18		
Oliva		100.00								
Arbequina			73.5			26.47				
Cornicabra			5.88	94.12						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	8.82							50.00	41.18	
Uva	8.82								91.18	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	100.00									
Oliva		100.00								
Arbequina			88.24			11.76				
Cornicabra			55.88	44.12						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	44.12							55.88		
Uva	44.12							5.88	50.00	

Tabla 6: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60%, 0.25Mpx y 15 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	88.24							8.82		2.94
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo								100.00		
Uva	20.59								79.41	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	94.12							5.88		
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo								100.00		
Uva	17.65								82.35	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	91.18							8.82		
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	20.59							79.41		
Uva	20.59								79.41	

Tabla 7: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75% , 0.25Mpx y 3 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	100.00									
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	2.94							94.12		2.94
Uva	20.59								79.41	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	100.00									
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	11.76							85.29		2.94
Uva	23.53								76.47	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificad
Girasol	85.29							14.71		
Oliva		100.00								
Arbequina			94.12			2.94				2.94
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	23.53							76.47		
Uva	23.53								76.47	

Tabla 8: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75%, 0.25Mpx y 5 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	79.41								20.59	
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo								100.00		
Uva									100.00	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	100.00									
Oliva		100.00								
Arbequina			97.06			2.94				
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	23.53							76.47		
Uva	26.47								73.53	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	100.00									
Oliva		100.00								
Arbequina			91.18			8.82				
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	26.47							73.53		
Uva	26.47								73.53	

Tabla 9: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75%, 0.25 Mpx y 8 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	91.18								8.82	
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						94.12				5.88
Lino							100.00			
Sésamo								100.00		
Uva									100.00	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	88.24							8.82		2.94
Oliva		100.00								
Arbequina			100.0							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	2.94							97.06		
Uva	20.59								79.41	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	100.00									
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	35.29							64.71		
Uva	35.29								64.71	

Tabla 10: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75%, 0.25Mpx y 10 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	64.71								35.29	
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo								100.00		
Uva									100.00	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	100.00									
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	8.82							91.18		
Uva	17.65								82.35	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	100.00									
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	38.24							61.76		
Uva	38.24								61.76	

Tabla 11: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75%, 0.25Mpx y 12 partes

Entrenamiento 1	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	26.47								73.53	
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo								100.00		
Uva									100.00	
Entrenamiento 2	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	58.82								41.18	
Oliva		100.00								
Arbequina			52.94			47.06				
Cornicabra				100.00						
Hojiblanca					100.00					
Picual			2.94			97.06				
Lino							100.00			
Sésamo								100.00		
Uva									100.00	
Entrenamiento 3	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasifica
Girasol	100.00									
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	44.12							55.88		
Uva	44.12								55.88	

Tabla 12: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75%, 0.25Mpx y 15 partes

Anexo 1.2: Café

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	67.65	5.88	14.71		11.76	
Suave	2.94	8.82	23.53	11.76	52.94	
Costa Rica	23.53	2.94	38.24		35.29	
Kenya				73.53	26.47	
Descafeinado	5.88		14.71	38.24	41.18	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	50.00	20.59	29.41			
Suave	2.94	29.41	17.65	8.82	41.18	
Costa Rica	20.59	41.18	38.24			
Kenya		2.94		73.53	23.53	
Descafeinado	17.65	26.47	17.65	23.53	14.71	
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	47.06	11.76	41.18			
Suave		29.41	23.53	11.76	35.29	
Costa Rica	26.47	29.41	44.12			
Kenya		2.94		64.71	32.35	
Descafeinado	17.65	20.59	20.59	20.59	20.59	

Tabla 13: Matriz de confusión obtenida para el café con un grosor de 40 – 60% , 0.25Mpx y 3 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	38.24	8.82	38.24		14.71	
Suave	2.94	11.76	32.35	8.82	44.12	
Costa Rica	17.65	44.12	23.53		14.71	
Kenya				61.76	38.24	
Descafeinado	11.76	2.94	17.65	23.53	44.12	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	79.41		11.76		8.82	
Suave		41.18	14.71	5.88	38.24	
Costa Rica	11.76	17.65	50.00		20.59	
Kenya		2.94		85.29	11.76	
Descafeinado	8.82	8.82	14.71	41.18	26.47	
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	64.71	11.76	23.53			
Suave		26.47	29.41	14.71	29.41	
Costa Rica	47.06	17.65	35.29			
Kenya				64.71	35.29	
Descafeinado	17.65	14.71	11.76	35.29	20.59	

Tabla 14: Matriz de confusión obtenida para el café con un grosor de 40 – 60% , 0.25Mpx y 5 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	47.06	2.94	41.18		8.82	
Suave	2.94	14.71	17.65	17.65	47.06	
Costa Rica	32.35	17.65	50.00			
Kenya				67.65	32.35	
Descafeinado	20.59	8.82	8.82	14.71	47.06	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	47.06	17.65	32.35		2.94	
Suave	2.94	5.88	35.29	17.65	38.24	
Costa Rica	32.35		50.00		17.65	
Kenya		2.94		61.76	35.29	
Descafeinado	17.65	5.88	8.82	38.24	29.41	
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	64.71	17.65	17.65			
Suave		29.41	17.65		50.00	2.94
Costa Rica	11.76	64.71	23.53			
Kenya		2.94		76.47	20.59	
Descafeinado	8.82	20.59	23.53	35.29	11.76	

Tabla 15: Matriz de confusión obtenida para el café con un grosor de 40 – 60% , 0.25Mpx y 8 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	50.00	5.88	44.12			
Suave	5.88	17.65	35.12	20.59	20.59	
Costa Rica	32.35	11.76	55.88			
Kenya		2.94		76.47	20.59	
Descafeinado	20.59	11.76	11.76	5.88	50.00	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	41.18	35.29	23.53			
Suave		20.59	26.47		52.94	
Costa Rica	14.71	44.12	41.18			
Kenya		2.94		50.00	47.06	
Descafeinado	2.94	35.29	20.59	35.29	5.88	
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	61.76	5.88	29.41		2.94	
Suave		35.29	11.76	26.47	26.47	
Costa Rica	29.41	14.71	35.29		20.59	
Kenya				94.12	5.88	
Descafeinado	17.65		8.82	61.76	11.76	

Tabla 16: Matriz de confusión obtenida para el café con un grosor de 40 – 60% , 0.25Mpx y 10 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	41.18	17.65	41.18			
Suave		32.35	14.71	29.41	23.53	
Costa Rica	38.24	50.00	11.76			
Kenya				91.18	8.82	
Descafeinado	35.29		14.71	47.06	2.94	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	47.06	5.88	38.24		8.82	
Suave	5.88	2.94	47.06	5.88	38.24	
Costa Rica	20.59		61.76		17.65	
Kenya				50.00	50.00	
Descafeinado	8.82		17.65	20.59	52.94	
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	35.29	20.59	32.35		11.76	
Suave		23.53	8.82		67.65	
Costa Rica	14.71	35.29	50.00			
Kenya				61.76	38.24	
Descafeinado	2.94	17.65	8.82	17.65	52.94	

Tabla 17: Matriz de confusión obtenida para el café con un grosor de 40 – 60% , 0.25Mpx y 12 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	79.41	8.82	11.76			
Suave	2.94	29.41	23.53		44.12	
Costa Rica	11.76	55.88	32.35			
Kenya		5.88		67.65	26.47	
Descafeinado	2.94	41.18	14.71	17.65	23.53	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	61.76	8.82	20.59	5.88	2.94	
Suave		29.41	5.88	29.41	35.29	
Costa Rica	35.29	58.82	5.88			
Kenya				100.00		
Descafeinado	38.24	5.88	17.65	38.24		
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	41.18	17.65	29.41		11.76	
Suave	14.71	5.88	29.41	5.88	44.12	
Costa Rica	55.88		38.24		5.88	
Kenya		8.82		50.00	41.18	
Descafeinado	5.88	2.94	8.82	32.35	50.00	

Tabla 18: Matriz de confusión obtenida para el café con un grosor de 40 – 60% , 0.25Mpx y 15 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	67.65	17.65			14.71	
Suave	2.94		20.59	11.76	64.71	
Costa Rica	17.65		32.35		50.00	
Kenya				73.53	26.47	
Descafeinado	5.88		8.82	38.24	47.06	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	58.82	17.65			23.53	
Suave		17.65	17.65	8.82	55.88	
Costa Rica	8.82	5.88	41.18		44.12	
Kenya				64.71	35.29	
Descafeinado			35.29	23.53	41.18	
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	64.71	23.53			11.76	
Suave		29.41	23.53	11.76	35.29	
Costa Rica	26.47	29.41	38.24		5.88	
Kenya		2.94		64.71	32.35	
Descafeinado	11.76	5.88	26.47	20.59	35.29	

Tabla 19: Matriz de confusión obtenida para el café con un grosor de 25 – 75%, 0.25Mpx y 3 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	50.00				50.00	
Suave			35.29	8.82	55.88	
Costa Rica	17.76		26.47		61.76	
Kenya				61.76	38.24	
Descafeinado	8.82		14.71	23.53	52.94	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	91.18	2.94	5.88			
Suave	5.88	14.71	8.82	8.82	61.76	
Costa Rica			55.88		44.76	
Kenya				85.29	14.71	
Descafeinado			20.59	41.18	38.24	
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	73.53	14.71			11.76	
Suave		17.65	20.59	14.71	47.06	
Costa Rica	35.29	5.88	35.29		23.53	
Kenya				67.65	32.35	
Descafeinado	17.65		11.76	35.29	35.29	

Tabla 20: Matriz de confusión obtenida para el café con un grosor de 25 – 75%, 0.25Mpx y 5 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	64.71	35.29				
Suave	5.88	17.65	14.71	5.88	55.88	
Costa Rica	2.94	17.65	29.41		50.00	
Kenya				76.47	23.53	
Descafeinado			38.24	26.47	35.29	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	61.76	5.88			32.35	
Suave	11.76		26.47	23.53	38.24	
Costa Rica	29.41		38.24		32.35	
Kenya				73.53	26.47	
Descafeinado	11.76		5.88	44.12	38.24	
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	70.59	5.88			23.53	
Suave	2.94	11.76	14.71	17.65	52.94	
Costa Rica	26.47	8.82	44.12		20.59	
Kenya				67.65	32.35	
Descafeinado	17.65		8.82	8.82	64.71	

Tabla 21: Matriz de confusión obtenida para el café con un grosor de 25 – 75%, 0.25Mpx y 8 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	70.59	17.65	11.76			
Suave			11.76	26.47	61.76	
Costa Rica	29.41	2.94	29.42		38.23	
Kenya				94.12	5.88	
Descafeinado	17.65		2.94	58.82	20.59	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	44.12	26.47			29.41	
Suave		5.88	17.65	5.88	70.59	
Costa Rica	5.88		44.12		50.00	
Kenya				70.59	29.41	
Descafeinado			5.88	41.18	52.94	
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	82.35	2.94			14.71	
Suave		14.71	35.29	2.94	47.06	
Costa Rica	8.82	2.94	35.29		52.94	
Kenya				50.00	50.00	
Descafeinado	8.82	2.94	20.59	2.94	64.71	

Tabla 22: Matriz de confusión obtenida para el café con un grosor de 25 – 75%, 0.25Mpx y 10 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	73.53	23.53	2.94			
Suave		35.29	5.88	35.29	23.53	
Costa Rica	47.06	44.12	5.88		2.94	
Kenya				91.18	8.82	
Descafeinado	38.24		20.59	41.18		
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	64.71				35.29	
Suave			47.06		50.00	2.94
Costa Rica	5.88		50.00		44.12	
Kenya				44.12	55.88	
Descafeinado			8.82	20.59	70.59	
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	44.12	14.71			41.18	
Suave			8.82		91.18	
Costa Rica	2.94		55.88		41.18	
Kenya				61.76	38.24	
Descafeinado			2.94	17.65	79.41	

Tabla 23: Matriz de confusión obtenida para el café con un grosor de 25 – 75%, 0.25Mpx y 12 partes

Entrenamiento 1	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	52.94	5.88			41.18	
Suave	2.94		20.59		76.47	
Costa Rica	2.94		35.29		61.76	
Kenya				70.59	29.41	
Descafeinado			8.82	14.71	76.47	
Entrenamiento 2	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	64.71	17.65	14.71	2.94		
Suave		32.35	5.88	35.29	26.47	
Costa Rica	41.18	44.12	8.82		5.88	
Kenya				97.06	2.94	
Descafeinado	38.24		20.59	41.18		
Entrenamiento 3	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	50.00				50.00	
Suave	2.94		41.18	5.88	50.00	
Costa Rica	26.47		50.00		23.53	
Kenya				44.12	55.88	
Descafeinado	5.88		2.94	26.47	64.71	

Tabla 24: Matriz de confusión obtenida para el café con un grosor de 25 – 75%, 0.25Mpx y 15 partes

Anexo 1.3: Vino

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	12.50					37.50	
Verdejo	12.50	12.50	12.50	12.50			50.00	
Cabernet			25.00			37.50	37.50	
Garnacha	12.50		12.50	12.50	12.50		50.00	
Lambrusco	12.50				62.50		25.00	
Mencia			12.50			25.00	62.50	
Tempranillo	37.50		12.50			12.50	37.50	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	12.50	12.50				25.00	
Verdejo	25.00	37.50	12.50				25.00	
Cabernet			12.50			25.00	62.50	
Garnacha	50.00		12.50		25.00		12.50	
Lambrusco					50.00		50.00	
Mencia			25.00			37.50	37.50	
Tempranillo	12.50		25.00		12.50		50.00	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	12.50			87.50				
Verdejo	62.50			37.50				
Cabernet			50.00			37.50	12.50	
Garnacha	25.00			62.50	12.50			
Lambrusco				37.50	62.5			
Mencia			12.50		12.50	25.00	50.00	
Tempranillo	25.00		25.00			25.00	25.00	

Tabla 24: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% , 0.25Mpx y 3 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	75.00						25.00	
Verdejo	62.50	25.00		12.50				
Cabernet			62.50				37.50	
Garnacha	50.00		12.50	25.00	12.50			
Lambrusco			12.50		62.50		25.00	
Mencia			12.50		12.50	37.50	37.50	
Tempranillo	25.00		25.00			25.00	25.00	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	37.50						62.50	
Verdejo		12.50		25.00			62.50	
Cabernet			37.50				62.50	
Garnacha	25.00				37.50		37.50	
Lambrusco	12.50				50.00		37.50	
Mencia			25.00			25.00	50.00	
Tempranillo	25.00		12.50		12.50		50.00	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling		25.00	12.50	12.50			50.00	
Verdejo	25.00	37.50	25.00				12.50	
Cabernet			62.50			37.50		
Garnacha				25.00			75.00	
Lambrusco				25.00	75.00			
Mencia			37.50		12.50	50.00		
Tempranillo			75.50		12.50	12.50		

Tabla 25: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% , 0.25Mpx y 5 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	12.50					37.50	
Verdejo	12.50	25.00	37.50	12.50			12.50	
Cabernet			12.50				87.50	
Garnacha	50.00			25.00	12.50		12.50	
Lambrusco					75.00		25.00	
Mencia			25.00				75.00	
Tempranillo	37.50		25.00				37.50	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	62.50		12.50	25.00				
Verdejo	37.50	12.50	25.00	25.00				
Cabernet			62.50			12.50	25.00	
Garnacha	12.50		12.50	37.50	12.50		25.00	
Lambrusco	37.50		12.50	12.50	25.00			12.50
Mencia					25.00	25.00	50.00	
Tempranillo	12.50		25.00			25.00	37.50	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	62.50		12.00		12.50		12.50	
Verdejo	37.50	12.50	25.00				25.00	
Cabernet			12.50			37.50	50.00	
Garnacha	12.50		12.50	50.00	12.50		12.50	
Lambrusco		12.50			37.50		50.00	
Mencia			25.00			50.00	25.00	
Tempranillo			37.50				62.50	

Tabla 26: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% , 0.25Mpx y 8 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	12.50					37.50	
Verdejo	12.50	25.00	37.50	12.50			12.50	
Cabernet			12.50				87.50	
Garnacha	50.00			25.00	12.50		12.50	
Lambrusco					75.00		25.00	
Mencia			25.00				75.00	
Tempranillo	37.50		25.00				37.50	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	62.50		12.50	25.00				
Verdejo	37.50	12.50	25.00	25.00				
Cabernet			62.50			12.50	25.00	
Garnacha	12.50		12.50	37.50	12.50		25.00	
Lambrusco	37.50		12.50	12.50	25.00			12.50
Mencia					25.00	25.00	50.00	
Tempranillo	12.50		25.00			25.00	37.50	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	62.50		12.00		12.50		12.50	
Verdejo	37.50	12.50	25.00				25.00	
Cabernet			12.50			37.50	50.00	
Garnacha	12.50		12.50	50.00	12.50		12.50	
Lambrusco		12.50			37.50		50.00	
Mencia			25.00			50.00	25.00	
Tempranillo			37.50				62.50	

Tabla 27: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% , 0.25Mpx y 10 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	12.50					37.50	
Verdejo	12.50	25.00	37.50	12.50			12.50	
Cabernet			12.50				87.50	
Garnacha	50.00			25.00	12.50		12.50	
Lambrusco					75.00		25.00	
Mencia			25.00				75.00	
Tempranillo	37.50		25.00				37.50	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	62.50		12.50	25.00				
Verdejo	37.50	12.50	25.00	25.00				
Cabernet			62.50			12.50	25.00	
Garnacha	12.50		12.50	37.50	12.50		25.00	
Lambrusco	37.50		12.50	12.50	25.00			12.50
Mencia					25.00	25.00	50.00	
Tempranillo	12.50		25.00			25.00	37.50	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	62.50		12.00		12.50		12.50	
Verdejo	37.50	12.50	25.00				25.00	
Cabernet			12.50			37.50	50.00	
Garnacha	12.50		12.50	50.00	12.50		12.50	
Lambrusco		12.50			37.50		50.00	
Mencia			25.00			50.00	25.00	
Tempranillo			37.50				62.50	

Tabla 28: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% , 0.25Mpx y 12 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	12.50					37.50	
Verdejo	12.50	25.00	37.50	12.50			12.50	
Cabernet			12.50				87.50	
Garnacha	50.00			25.00	12.50		12.50	
Lambrusco					75.00		25.00	
Mencia			25.00				75.00	
Tempranillo	37.50		25.00				37.50	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	62.50		12.50	25.00				
Verdejo	37.50	12.50	25.00	25.00				
Cabernet			62.50			12.50	25.00	
Garnacha	12.50		12.50	37.50	12.50		25.00	
Lambrusco	37.50		12.50	12.50	25.00			12.50
Mencia					25.00	25.00	50.00	
Tempranillo	12.50		25.00			25.00	37.50	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	62.50		12.00		12.50		12.50	
Verdejo	37.50	12.50	25.00				25.00	
Cabernet			12.50			37.50	50.00	
Garnacha	12.50		12.50	50.00	12.50		12.50	
Lambrusco		12.50			37.50		50.00	
Mencia			25.00			50.00	25.00	
Tempranillo			37.50				62.50	

Tabla 29: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% , 0.25Mpx y 15 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	12.50					37.50	
Verdejo	25.00	12.50	12.50				50.00	
Cabernet			37.50				62.500	
Garnacha	12.50		12.50	12.50	12.50		50.00	
Lambrusco	12.52				50.00		37.50	
Mencia			12.50			12.50	75.00	
Tempranillo	37.50						62.50	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	25.00	12.50	12.50				50.00	
Verdejo	12.50	50.00	12.50				25.00	
Cabernet			37.50				62.50	
Garnacha				25.00	12.50		62.50	
Lambrusco				25.00	50.00		25.00	
Mencia			25.00			25.00	50.00	
Tempranillo			25.00				75.00	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	25.00			75.00				
Verdejo	62.50			37.500				
Cabernet			25.00			12.50	62.50	
Garnacha		12.50		75.00	12.50			
Lambrusco				25.00	75.00			
Mencia			12.50			25.00	62.50	
Tempranillo	12.50			12.50		12.50	62.50	

Tabla 30: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% , 0.25Mpx y 3 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	62.50	12.50					25.00	
Verdejo	50.00	25.00					25.00	
Cabernet			37.50			12.50	50.00	
Garnacha	12.50		12.50	37.50			37.50	
Lambrusco				12.50	62.50		25.00	
Mencia			12.50			25.00	62.50	
Tempranillo			25.00				75.00	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	37.50						62.50	
Verdejo		25.00					75.00	
Cabernet			37.50				62.50	
Garnacha	25.00			37.50	12.50		25.00	
Lambrusco					62.50		37.50	
Mencia			12.50				87.50	
Tempranillo	12.50		12.50				75.00	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling		37.50	12.50				50.00	
Verdejo	50.00	25.00	25.00					
Cabernet			37.50			50.00	12.50	
Garnacha				25.00			75.00	
Lambrusco				12.50	87.50			
Mencia			37.50			37.00	25.00	
Tempranillo			50.00		12.50	12.50	25.00	

Tabla 31: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% , 0.25Mpx y 5 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	25.00					25.00	
Verdejo	12.50	25.00	12.50				50.00	
Cabernet			12.50				87.50	
Garnacha	50.00			25.00	12.50		12.50	
Lambrusco				12.50	62.50		25.00	
Mencia			25.00				75.00	
Tempranillo	12.50		12.50				75.00	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	75.00	12.50		12.50				
Verdejo	62.50	25.00		12.50				
Cabernet			37.50			25.00	37.50	
Garnacha	12.50		12.50	37.50	12.50		25.00	
Lambrusco	12.50			50.00	25.00		12.500	
Mencia						25.00	75.00	
Tempranillo	25.00		12.50				62.50	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling			12.50				87.50	
Verdejo	37.50	12.50					50.00	
Cabernet						37.50	62.50	
Garnacha				50.00	12.50		37.50	
Lambrusco				12.50	75.00		12.50	
Mencia			12.50			50.00	37.50	
Tempranillo							100.00	

Tabla 32: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% , 0.25Mpx y 8 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	25.00					25.00	
Verdejo	12.50	25.00	12.50				50.00	
Cabernet			12.50				87.50	
Garnacha	50.00			25.00	12.50		12.50	
Lambrusco				12.50	62.50		25.00	
Mencia			25.00				75.00	
Tempranillo	12.50		12.50				75.00	

Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	75.00	12.50		12.50				
Verdejo	62.50	25.00		12.50				
Cabernet			37.50			25.00	37.50	
Garnacha	12.50		12.50	37.50	12.50		25.00	
Lambrusco	12.50			50.00	25.00		12.500	
Mencia						25.00	75.00	
Tempranillo	25.00		12.50				62.50	

Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling			12.50				87.50	
Verdejo	37.50	12.50					50.00	
Cabernet						37.50	62.50	
Garnacha				50.00	12.50		37.50	
Lambrusco				12.50	75.00		12.50	
Mencia			12.50			50.00	37.50	
Tempranillo							100.00	

Tabla 33: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% , 0.25Mpx y 10 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	25.00					25.00	
Verdejo	12.50	25.00	12.50				50.00	
Cabernet			12.50				87.50	
Garnacha	50.00			25.00	12.50		12.50	
Lambrusco				12.50	62.50		25.00	
Mencia			25.00				75.00	
Tempranillo	12.50		12.50				75.00	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	75.00	12.50		12.50				
Verdejo	62.50	25.00		12.50				
Cabernet			37.50			25.00	37.50	
Garnacha	12.50		12.50	37.50	12.50		25.00	
Lambrusco	12.50			50.00	25.00		12.50	
Mencia						25.00	75.00	
Tempranillo	25.00		12.50				62.50	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling			12.50				87.50	
Verdejo	37.50	12.50					50.00	
Cabernet						37.50	62.50	
Garnacha				50.00	12.50		37.50	
Lambrusco				12.50	75.00		12.50	
Mencia			12.50			50.00	37.50	
Tempranillo							100.00	

Tabla 34: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% , 0.25Mpx y 12 partes

Entrenamiento 1	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	25.00					25.00	
Verdejo	12.50	25.00	12.50				50.00	
Cabernet			12.50				87.50	
Garnacha	50.00			25.00	12.50		12.50	
Lambrusco				12.50	62.50		25.00	
Mencia			25.00				75.00	
Tempranillo	12.50		12.50				75.00	
Entrenamiento 2	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	75.00	12.50		12.50				
Verdejo	62.50	25.00		12.50				
Cabernet			37.50			25.00	37.50	
Garnacha	12.50		12.50	37.50	12.50		25.00	
Lambrusco	12.50			50.00	25.00		12.500	
Mencia						25.00	75.00	
Tempranillo	25.00		12.50				62.50	
Entrenamiento 3	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling			12.50				87.50	
Verdejo	37.50	12.50					50.00	
Cabernet						37.50	62.50	
Garnacha				50.00	12.50		37.50	
Lambrusco				12.50	75.00		12.50	
Mencia			12.50			50.00	37.50	
Tempranillo							100.00	

Tabla 35: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% , 0.25Mpx y 15 partes

Anexo 2: Medidas tomadas, Leave-one-Out

Anexo 2.1: Aceite

Anexo 2.1.1: Dominio espectral

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	60.00							40.00		
Oliva		100.00								
Arbequina			91.00			9.00				
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	20.00							62.00	18.00	
Uva	20.00								80.00	

Tabla 36: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60% y 0.25Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	60.00							40.00		
Oliva		100.00								
Arbequina			84.00	1.00		14.00			1.00	
Cornicabra			15.00	85.00						
Hojiblanca					100.00					
Picual		2.00	13.00			85.00				
Lino							100.00			
Sésamo	20.00							40.00		
Uva	20.00									

Tabla 37: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60% y 0.56Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	60.00							40.00		
Oliva		100.00								
Arbequina			85.00	6.00		9.00				
Cornicabra			18.00	81.00						1.00
Hojiblanca					100.00					
Picual		6.00	19.00			75.00				
Lino							100.00			
Sésamo	20.00							40.00	40.00	
Uva	20.00								80.00	

Tabla 38: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60% y 1 Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	59.00					1.00		40.00		
Oliva		100.00								
Arbequina		2.00	69.00	19.00		10.00				
Cornicabra			20.00	79.00						1.00
Hojiblanca					100.00					
Picual		16.00	20.00			64.00				
Lino							100.00			
Sésamo	20.00							40.00	40.00	
Uva	20.00								80.00	

Tabla 39: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60% y 4 Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	95.00							5.00		
Oliva		100.00								
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	12.00							87.00		1.00
Uva	20.00								80.00	

Tabla 40: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75% y 0.25Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	80.00							20.00		
Oliva		100.00								
Arbequina			99.00			1.00				
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	20.00							80.00		
Uva	20.00								80.00	

Tabla 41: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75% y 0.56Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	80.00							20.00		
Oliva		100.00								
Arbequina			98.00			2.00				
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	20.00							80.00		
Uva	20.00								80.00	

Tabla 42: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75% y 1 Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	80.00							20.00		
Oliva		100.00								
Arbequina			85.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo	20.00							61.00	19.00	
Uva	20.00								80.00	

Tabla 43: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75% y 4 Mpx

Anexo 2.1.2: Dominio PCA (Principal Component Analysis)

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	100.00									
Oliva		92.00				8.00				
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					98.00				2.00	
Picual		7.00				93.00				
Lino							100.00			
Sésamo								100.00		
Uva									100.00	

Tabla 44: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60% y 0.25Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	100.00									
Oliva		91.00				9.00				
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					99.00					
Picual		12.00				88.00				
Lino							100.00			
Sésamo								100.00		
Uva									100.00	

Tabla 45: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60% y 0.56Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	100.00									
Oliva		86.00				14.00				
Arbequina		9.00	91.00							
Cornicabra				100.00						
Hojiblanca	1.00				98.00				1.00	
Picual		7.00				93.00				
Lino							95.00		5.00	
Sésamo								100.00		
Uva									100.00	

Tabla 46: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60% y 1 Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	100.00									
Oliva		87.00				13.00				
Arbequina		8.00	92.00							
Cornicabra			3.00	97.00						
Hojiblanca	1.00				98.00					
Picual		11.00				89.00				
Lino							87.00		13.00	
Sésamo								100.00		
Uva									100.00	

Tabla 47: Matriz de confusión obtenida para el aceite con un grosor de 40 – 60% y 4 Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	100.00									
Oliva		98.00				2.00				
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.0				
Lino							100.00			
Sésamo								100.00		
Uva									100.00	

Tabla 48: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75% y 0.25Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	100.00									
Oliva		93.00				7.00				
Arbequina			100.00							
Cornicabra				100.00						
Hojiblanca					100.00					
Picual						100.00				
Lino							100.00			
Sésamo								100.00		
Uva									100.00	

Tabla 49: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75% y 0.56Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	100.00									
Oliva		91.00				9.00				
Arbequina			94.00		6.00					
Cornicabra				100.00						
Hojiblanca					100.00					
Picual		4.00				96.00				
Lino							97.00		3.00	
Sésamo								100.00		
Uva									100.00	

Tabla 50: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75% y 1 Mpx

Matriz de Confusión	Girasol	Oliva	Arbequina	Cornicabra	Hojiblanca	Picual	Lino	Sésamo	Uva	No clasificado
Girasol	100.00									
Oliva		87.00				13.00				
Arbequina			95.00							
Cornicabra				100.00						
Hojiblanca					99.00				1.00	
Picual		15.00				85.00				
Lino							92.00		8.00	
Sésamo								100.00		
Uva									100.00	

Tabla 51: Matriz de confusión obtenida para el aceite con un grosor de 25 – 75% y 4 Mpx

Anexo 2.2 Café

Anexo 2.2.1: Dominio espectral

Matriz de Confusión	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	57.00	12.00	31.00			
Suave	2.00	26.00	23.00	12.00	37.00	
Costa Rica	24.00	28.00	46.00		2.00	
Kenya		2.00		73.00	25.00	
Descafeinado	14.00	12.00	14.00	31.00	29.00	

Tabla 52: Matriz de confusión obtenida para el café con un grosor de 40 – 60% y 0.25 Mpx

Matriz de Confusión	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	56.00	12.00	32.00			
Suave	2.00	26.00	23.00	12.00	37.00	
Costa Rica	24.00	32.00	44.00			
Kenya		2.00		71.00	27.00	
Descafeinado	14.00	10.00	14.00	31.00	31.00	

Tabla 53: Matriz de confusión obtenida para el café con un grosor de 40 – 60% y 0.56 Mpx

Matriz de Confusión	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	53.00	12.00	35.00			
Suave	2.00	26.00	23.00	12.00	37.00	
Costa Rica	26.00	32.00	42.00			
Kenya		2.00		69.00	29.00	
Descafeinado	14.00	10.00	14.00	31.00	31.00	

Tabla 54: Matriz de confusión obtenida para el café con un grosor de 40 – 60% y 1 Mpx

Matriz de Confusión	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	51.00	12.00	37.00			
Suave	2.00	24.00	23.00	12.00	39.00	
Costa Rica	30.00	32.00	38.00			
Kenya		2.00		71.00	27.00	
Descafeinado	14.00	12.00	14.00	31.00	29.00	

Tabla 55: Matriz de confusión obtenida para el café con un grosor de 40 – 60% y 4 Mpx

Matriz de Confusión	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	67.00	12.00			21.00	
Suave	4.00	8.00	21.00	12.00	55.00	
Costa Rica	24.00	4.00	30.00		42.00	
Kenya				70.00	30.00	
Descafeinado	12.00		12.00	30.00	44.00	2.00

Tabla 56: Matriz de confusión obtenida para el café con un grosor de 25 – 75% y 0.25 Mpx

Matriz de Confusión	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	69.00	14.00	1.00		16.00	
Suave	4.00	14.00	21.00	12.00	49.00	
Costa Rica	24.00	4.00	32.00		40.00	
Kenya				70.00	30.00	
Descafeinado	14.00		10.00	30.00	46.00	

Tabla 57: Matriz de confusión obtenida para el café con un grosor de 25 – 75% y 0.56 Mpx

Matriz de Confusión	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	69.00	16.00	3.00		12.00	
Suave	4.00	14.00	21.00	12.00	49.00	
Costa Rica	24.00	4.00	32.00		40.00	
Kenya				70.00	30.00	
Descafeinado	14.00		12.00	30.00	44.00	

Tabla 58: Matriz de confusión obtenida para el café con un grosor de 25 – 75% y 1 Mpx

Matriz de Confusión	Forza	Suave	Costa Rica	Kenya	Descafeinado	No clasificado
Forza	69.00	14.00	7.00		10.00	
Suave	4.00	12.00	21.00	12.00	49.00	2.00
Costa Rica	24.00	4.00	34.00		38.00	
Kenya				70.00	30.00	
Descafeinado	14.00		12.00	30.00	44.00	

Tabla 59: Matriz de confusión obtenida para el café con un grosor de 25 – 75% y 4 Mpx

Anexo 2.2.2: Dominio PCA (Principal Component Analysis)

Las manchas espectrales del café, tienen la primera componente principal muy parecida, siendo en esta en la que realmente se diferencian normalmente las categorías y por las que se les puede discriminar. En este caso como es imposible (como se puede observar en la figura 1), para el café se mostrará únicamente su discriminación por manchas espectrales.

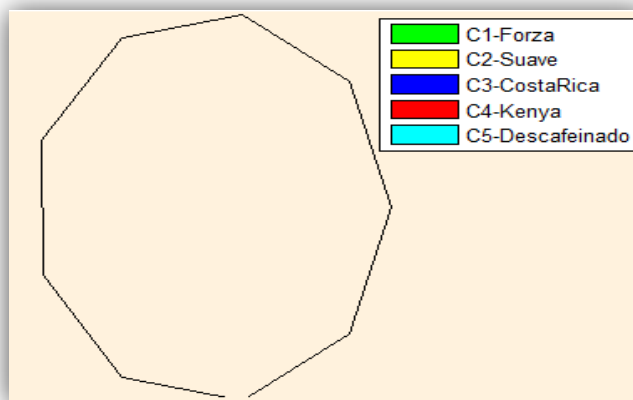


Figura 1: Manchas Espectral del Café con PCA

Anexo 2.3: Vino

Anexo 2.3.1: Dominio espectral

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	65.00	5.00	5.00				25.00	
Verdejo	35.00	25.00	20.00	10.00			10.00	
Cabernet			35.00			25.00	40.00	
Garnacha	25.00	5.00	10.00	45.00	10.00		5.00	
Lambrusco	10.00		5.00		45.00		40.00	
Mencia			20.00		5.00	35.00	40.00	
Tempranillo	10.00		20.00			10.00	55.00	5.00

Tabla 60: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% y 0.25Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	60.00	5.00	5.00				30.00	
Verdejo	25.00	35.00	20.00	5.00			15.00	
Cabernet			50.00			20.00	30.00	
Garnacha	25.00		10.00	45.00	10.00		10.00	
Lambrusco	10.00		5.00		40.00		45.00	
Mencia			20.00		5.00	35.00	40.00	
Tempranillo	5.00		20.00			10.00	65.00	

Tabla 61: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% y 0.56Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	60.00	5.00	5.00				30.00	
Verdejo	20.00	35.00	10.00	10.00			25.00	
Cabernet			45.00			15.00	40.00	
Garnacha	20.00		15.00	45.00	10.00		10.00	
Lambrusco	5.00		5.00		35.00		55.00	
Mencia			20.00			35.00	45.00	
Tempranillo	10.00		15.00			5.00	70.00	

Tabla 62: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% y 1 Mpx

Matiz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	60.00		5.00				35.00	
Verdejo	20.00	25.00	10.00	5.00			40.00	
Cabernet			40.00			15.00	45.00	
Garnacha	20.00		5.00	45.00	5.00		25.00	
Lambrusco	5.00		5.00		35.00		55.00	
Mencia			20.00			30.00	50.00	
Tempranillo	5.00		15.00			5.00	75.00	

Tabla 63: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% y 4 Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	55.00	10.00	5.00				30.00	
Verdejo	30.00	25.00	10.00				35.00	
Cabernet			40.00			5.00	55.00	
Garnacha			5.00	50.00	5.00		40.00	
Lambrusco				5.00	70.00		25.00	
Mencia			20.00			25.00	55.00	
Tempranillo			10.00				90.00	

Tabla 64: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% y 0.25Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	5.00	5.00				40.00	
Verdejo	25.00	20.00	10.00				45.00	
Cabernet			50.00			5.00	45.00	
Garnacha			5.00	50.00	5.00		40.00	
Lambrusco					65.00		35.00	
Mencia			20.00			25.00	55.00	
Tempranillo			15.00				85.00	

Tabla 65: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% y 0.56Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00	5.00	5.00				40.00	
Verdejo	20.00	20.00	10.00				50.00	
Cabernet			50.00				45.00	
Garnacha			5.00	50.00	5.00		40.00	
Lambrusco					60.00		40.00	
Mencia			20.00			25.00	55.00	
Tempranillo			15.00				85.00	

Tabla 66: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% y 1 Mpx

Matiz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	50.00		5.00				45.00	
Verdejo		10.00	10.00				80.00	
Cabernet			50.00				50.00	
Garnacha			5.00	45.00	5.00		45.00	
Lambrusco					35.00		65.00	
Mencia			20.00			20.00	60.00	
Tempranillo			15.00				85.00	

Tabla 67: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% y 4 Mpx

Anexo 2.3.2: Dominio PCA (Principal Component Analysis)

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	60.00	5.00	15.00			20.00		
Verdejo	10.00	60.00	10.00	10.00		5.00	5.00	
Cabernet	15.00	15.00	30.00	5.00		30.00		5.00
Garnacha	5.00	15.00	15.00	55.00	5.00		5.00	
Lambrusco	20.00	15.00	10.00	5.00	30.00	5.00	15.00	
Mencia	5.00		15.00	10.00	10.00	40.00	20.00	
Tempranillo	10.00		25.00	5.00	5.00	25.00	30.00	

Tabla 68: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% y 0.25Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	55.00	5.00	15.00		5.00	20.00		
Verdejo	20.00	50.00	10.00	15.00			5.00	
Cabernet	15.00	5.00	40.00	10.00		25.00	5.00	
Garnacha	10.00	15.00	10.00	40.00	10.00	5.00	10.00	
Lambrusco	20.00	5.00	5.00	5.00	30.00	5.00	30.00	
Mencia	15.00		10.00	10.00	5.00	40.00	20.00	
Tempranillo	15.00		15.00	5.00	10.00	20.00	30.00	5.00

Tabla 69: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% y 0.56Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	55.00	5.00	15.00			25.00		
Verdejo	20.00	45.00	15.00	20.00				
Cabernet	25.00	5.00	40.00	5.00		25.00		
Garnacha	15.00	10.00	15.00	40.00	10.00	5.00	5.00	
Lambrusco	20.00	10.00	15.00	5.00	40.00	5.00	5.00	
Mencia	5.00		15.00	10.00	5.00	40.00	25.00	
Tempranillo	10.00		20.00	5.00	15.00	25.00	20.00	

Tabla 70: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% y 1 Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	55.00	5.00	15.00			25.00		
Verdejo	20.00	50.00	20.00	10.00				
Cabernet	25.00	5.00	40.00	5.00		25.00		
Garnacha	15.00	10.00	20.00	35.00	10.00	5.00	5.00	
Lambrusco	30.00	10.00	10.00		35.00	5.00	10.00	
Mencia	10.00		10.00	10.00	5.00	45.00	20.00	
Tempranillo	20.00		15.00	5.00	10.00	35.00	15.00	

Tabla 71: Matriz de confusión obtenida para el vino con un grosor de 40 – 60% y 4 Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	70.00	10.00				5.00	15.00	
Verdejo	10.00	70.00		10.00			10.00	
Cabernet		10.00	25.00			5.00	60.00	
Garnacha		5.00		70.00			25.00	
Lambrusco	5.00				65.00	5.00	25.00	
Mencia		5.00	5.00			30.00	60.00	
Tempranillo			5.00			5.00	90.00	

Tabla 72: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% y 0.25Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	80.00	10.00					10.00	
Verdejo	15.00	65.00	5.00	10.00			5.00	
Cabernet	5.00	5.00	20.00				70.00	
Garnacha		5.00		65.00			30.00	
Lambrusco	5.00				60.00	5.00	30.00	
Mencia	10.00	5.00	5.00			25.00	55.00	
Tempranillo			10.00				90.00	

Tabla 73: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% y 0.56Mpx

Matriz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	70.00	10.00					20.00	
Verdejo	15.00	60.00		10.00			15.00	
Cabernet	5.00	5.00	30.00				60.00	
Garnacha	5.00	5.00		65.00			25.00	
Lambrusco	5.00				60.00	5.00	30.00	
Mencia	5.00	5.00	5.00			25.00	60.00	
Tempranillo			15.00				85.00	

Tabla 74: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% y 1 Mpx

Matiz de Confusión	Riesling	Verdejo	Cabernet	Garnacha	Lambrusco	Mencia	Tempranillo	No clasificado
Riesling	75.00	10.00					15.00	
Verdejo	20.00	55.00	5.00	10.00			10.00	
Cabernet	10.00	5.00	30.00	5.00			50.00	
Garnacha	5.00	5.00	5.00	70.00			15.00	
Lambrusco	5.00				60.00	5.00	30.00	
Mencia	5.00	5.00				30.00	60.00	
Tempranillo			10.00				90.00	

Tabla 75: Matriz de confusión obtenida para el vino con un grosor de 25 – 75% y 4 Mpx