

UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA



Profesor Guía:
Sebastián Godoy Medel, Ph.D.

Profesor Co – Guía:
Juan Pablo Staforelli, Ph.D.

Tesis para optar al grado de:
**Magíster en Ciencias de la Ingeniería con mención en Ingeniería
Eléctrica.**

Aprendizaje automático aplicado a Espectroscopía por Plasma Inducido con Láser (LIBS) y
Visión Artificial Infrarrojas en muestras endémicas de miel de la Región del Bío – Bío,
Chile.

Universidad de Concepción
Facultad de Ingeniería
Departamento de Ingeniería Eléctrica

Profesor Guía:
Sebastián Godoy Medel, Ph.D.
Profesor Co – Guía:
Juan Pablo Staforelli, Ph.D.

Aprendizaje automático aplicado a Espectroscopía por Plasma Inducido con Láser (LIBS) y Visión Artificial Infrarrojas en muestras endémicas de miel de la Región del Bío – Bío, Chile.

Por:

Rodrigo Andrés Jofré Cerda

Tesis para optar al grado:

Magíster en Ciencias de la Ingeniería, mención Ingeniería Eléctrica

Mayo 2024, Concepción

© 2024, Rodrigo Andrés Jofré Cerda. ORCID: 0009-0004-4429-9902.

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo siempre la cita bibliográfica del presente documento y su autor.

RESUMEN

La autenticidad y calidad de la miel han sido temas de creciente interés a nivel mundial debido a su importancia económica, nutricional y medicinal. Los consumidores y reguladores demandan métodos más precisos y rápidos para garantizar la pureza de la miel, especialmente ante el aumento de casos de adulteración con azúcares y otras sustancias. Además, determinar el origen botánico de la miel es crucial para su clasificación y comercialización, lo que presenta desafíos significativos en términos de tiempo y recursos necesarios para el análisis palinológico tradicional.

Los métodos convencionales de análisis palinológico, que se utilizan para identificar el origen botánico de la miel mediante la identificación de polen bajo microscopía, son procedimientos que demandan una gran cantidad de tiempo en estudios y requirieron el conteo de 1200 a 1800 granos de polen de una sola especie para clasificar una muestra en una clase específica. Este proceso puede extenderse de días a semanas, dependiendo de la experiencia del profesional, lo cual representa una significativa limitación para la industria apícola nacional. Además, la detección de adulterantes en la miel, como diferentes azúcares (fructosa, glucosa, maltosa), es otro desafío debido a la necesidad de métodos rápidos y mínimamente destructivos.

Para abordar estos desafíos, se han integrado tecnologías de Espectroscopía de Plasma Inducido por Láser (LIBS) y visión artificial, asistidas por algoritmos de aprendizaje automático, aplicadas a muestras endémicas del sector Alto Bío – Bío, Región del Bío – Bío, Chile. La técnica LIBS permite obtener firmas espectrales específicas de muestras de miel y adulterantes, logrando detectar adulteraciones en tiempos significativamente reducidos, en general de minutos a algunas horas, sin requerir una preparación compleja de las muestras. Simultáneamente, se ha empleado redes neuronales convolucionales para analizar imágenes microscópicas de polen, con el fin de determinar el origen botánico de las muestras mediante el análisis de características morfológicas específicas de cada especie de polen, como tamaño, forma, geometría y textura.

Los resultados obtenidos mediante el uso combinado de LIBS y visión artificial demuestran una mejora significativa en la eficiencia del proceso de autenticación y clasificación de la miel. En la detección de adulterantes, los errores de clasificación se mantienen bajos, con resultados como un error mínimo de 3.7% para glucosa en *Quillaja saponaria* y 9.2% para fructosa en *Eucryphia glutinosa*. Por otro lado, los modelos de visión artificial han mostrado eficacias de clasificación superiores al 90% en la mayoría de las muestras de la base de datos local, destacando especialmente en la clasificación de la especie endémica *Eucryphia glutinosa*. Así, el método propuesto proporciona una ventaja significativa para los apicultores locales, permitiéndoles acceder a resultados de calidad y origen, que son factores críticos para la industria apícola, garantizando el cumplimiento de los estándares de calidad en tiempos menores en comparación con los métodos convencionales, ya contando con los modelos de aprendizaje entrenados.

AGRADECIMIENTOS

En primer lugar, agradezco la paciencia y valores inculcados por mi madre Teresa, en perseverancia y esfuerzo, que, sin su apoyo incondicional, no podría ver resultados de esta investigación.

Deseo expresar también mi gratitud a los profesores que creyeron en mis capacidades y trabajo, profesores Juan Pablo Staforelli PhD. y Sebastián Godoy PhD., y también a todas mis amistades y colegas que me han alentado en el pensamiento crítico y científico, apoyándome con sus palabras de aliento.

TABLA DE CONTENIDOS

1. INTRODUCCIÓN.....	1
1.1. CONTEXTO.....	2
1.2. HIPOTESIS.....	4
1.3. OBJETIVOS.....	4
1.3.1. OBJETIVO PRINCIPAL.....	4
1.3.2. OBJETIVOS ESPECÍFICOS.....	4
1.3.3. ALCANCES.....	5
1.4. METODOLOGÍA.....	6
2. ESTADO DEL ARTE.....	10
2.1. DISCUSIÓN ESTADO DEL ARTE.....	13
2.2. DISCUSIÓN BIBLIOGRÁFICA.....	14
3. MARCO TEÓRICO.....	15
3.1. LASER INDUCED BREAKDOWN SPECTROSCOPY (LIBS).....	17
3.1.1. ADULTERANTES: Fructosa, Glucosa, Maltosa.....	20
3.2. VISIÓN ARTIFICIAL: PRINCIPIOS Y APLICACIONES.....	24
3.3. ALGORITMOS DE APRENDIZAJE AUTOMÁTICO.....	28
3.3.1. Análisis de Componentes Principales (<i>Principal Component Analysis, PCA</i>).....	29
3.3.2. Análisis Discriminante Lineal (<i>Linear Discriminant Analysis, LDA</i>).....	30
3.3.3. Descenso de Gradientes Estocástico (<i>Stochastic Gradient Descent, SGD</i>).....	31
3.3.4. Bosque Aleatorio (<i>Random Forest, RF</i>).....	32
3.3.5. Máquina de Soportes Vectoriales (<i>Support Vector Machine, SVM</i>).....	33
3.4. ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PROFUNDO.....	34
3.5. MÉTRICAS DE EVALUACIÓN.....	36
3.5.1. Métricas aplicadas a LIBS.....	37
3.5.2. Métricas aplicadas a Visión Artificial.....	38
3.6. RESUMEN DEL CAPÍTULO.....	40
4. DISEÑO EXPERIMENTAL Y ADQUISICIÓN DE DATOS.....	41
4.1. METODOLOGÍA PARA MUESTRAS Y BASE DE DATOS.....	42
4.1.1. PREPARACIÓN DE MUESTRAS.....	42
4.2. ESPECTROSCOPÍA LIBS.....	45
4.3. VISIÓN ARTIFICIAL.....	51
4.3.1. METODOLOGÍA DE ENTRENAMIENTO CLASIFICADOR YOLOv8.....	52
5. ANÁLISIS DE DATOS.....	56
5.1. ESPECTROSCOPÍA LIBS.....	56
5.2. VISIÓN ARTIFICIAL.....	66
6. DISCUSIÓN DE RESULTADOS.....	76
7. TRABAJO FUTURO.....	78
8. CONCLUSIONES.....	79
REFERENCIAS.....	80

1. INTRODUCCIÓN

El fraude alimentario, un problema persistente desde los inicios de la producción de alimentos [1], ha ganado importancia en la era de la globalización. Esta situación se complica con los sistemas de producción y distribución de alimentos cada vez más específicos, donde la adulteración puede ocurrir en múltiples etapas de la cadena alimentaria, implicando serias consecuencias para la salud humana y la seguridad alimentaria. A nivel internacional, se han adoptado enfoques regulatorios diversos para combatir este problema, destacando la necesidad de un enfoque colectivo y coordinado. Esto involucra la identificación y certificación de los interesados en la cadena de suministro de alimentos [2], el seguimiento en tiempo real de los productos alimenticios y la exclusión de aquellos que no cumplen con los estándares.

En Chile, el fraude alimentario representa una oportunidad significativa para realzar el valor de sus productos de exportación, productos especiales como el vino, el aceite de oliva y, particularmente, la miel de abejas [3].

La miel, un producto natural apreciado por sus propiedades culinarias y medicinales, es elaborada por las abejas a partir del néctar de las flores. Este proceso, que involucra transformaciones enzimáticas y cambios en la composición química, contribuye a la singularidad de la miel en términos de sabor, aroma y valor nutricional. La melisopolinología, centrada en el estudio del polen [4], emerge como un agente esencial para comprender la relación entre la miel y las especies vegetales circundantes.

A lo largo de la literatura, diversas metodologías, como el análisis de polen mediante técnicas avanzadas de aprendizaje automático [5] y la caracterización de firmas espectrales [6] han permitido avances significativos en la identificación y clasificación de la miel. Sin embargo, las limitaciones asociadas con la complejidad de las técnicas de análisis y la necesidad de preparación intensiva de muestras han motivado la búsqueda de enfoques más eficientes y mínimamente invasivos.

En este contexto, esta investigación propone la aplicación de técnicas ópticas mínimamente invasivas, específicamente visión artificial por imágenes de polen y espectroscopía por plasma inducido con láser (LIBS) [7], asistidas por algoritmos de aprendizaje automático desde la adquisición cruda de los datos. Estos métodos, que permiten dicha implementación en diversos experimentos, buscan caracterizar compuestos en la miel y detectar adulteraciones en tramos, centrándose en la identificación de diferentes azúcares y elementos químicos presentes en las muestras.

La melisopolinología desempeña un papel fundamental en la identificación de la procedencia botánica de la miel [8]. Sin embargo, en este proceso, nos enfrentamos a desafíos significativos relacionados con la digitalización de datos y la identificación eficaz de diferentes clases de miel por el volumen de datos que esto conlleva. En este contexto, es importante destacar la evolución de técnicas como LIBS, que se han utilizado en diversas disciplinas científicas. Esta técnica se destaca por su aplicabilidad en la caracterización de muestras de miel y su capacidad para abordar los desafíos mencionados.

En el marco de estos avances, la presente investigación busca superar las limitaciones de los métodos tradicionales ya desarrollados, proponiendo un enfoque que combine técnicas ópticas avanzadas con algoritmos de aprendizaje automático para la caracterización de la miel y sus posibles adulteraciones. Se aborda así la necesidad de utilizar métodos eficaces y eficientes que permitan identificar adulteraciones alimentarias y evaluar la calidad de la miel de manera rápida y confiable. Este enfoque no solo contribuirá al conocimiento científico en el campo, sino que también ofrecerá herramientas prácticas para la industria alimentaria y la salud pública.

1.1. CONTEXTO

En Chile, en octubre 2022, se ha promulgado la “Ley apícola” [9] la cual tiene como objetivo principal “la promoción, protección y fomento del desarrollo sustentable de la apicultura como actividad silvoagropecuaria”. En ella es posible encontrar indicadores y exigencias mínimas las cuales hay que tener en cuenta para poder dar con el análisis de las características fisicoquímicas de la miel, penas por adulteración o falsificación de productos melíferos, entre otras.

La institución autorizada para realizar las mediciones correspondientes es el Instituto de Investigaciones Agropecuarias, siendo también destacable en el desarrollo de tecnología en la agricultura chilena la Fundación de Innovación Agraria, y como ente regulador y fiscalizador Servicio Agrícola y Ganadero. Estos organismos actualmente poseen investigaciones y reportes [10], [11], [12] de diversos productores apícolas, haciendo referencia a métodos como el “análisis sensorial de la miel” como una vía a su caracterización haciendo uso de sentidos como el olfato o el gusto, siguiendo una exigente rúbrica con origen en la Universidad de Salamanca, España, donde este tema es discutido a nivel científico y profesional desde hace algunas décadas.

Por otra parte, otra caracterización realizada es de parámetros físico – químicos con técnicas que miden en distintas muestras, en concentraciones disueltas, conductividad eléctrica, o a través de tablas de colores como referencias directas de sus características visuales. Estos parámetros se consideran como algunos de los más importantes, pero a pesar de que la información obtenida pueda ser relevante, sigue siendo un desafío poder determinar índices de humedad precisos, la cantidad de azúcares que naturalmente provienen del proceso melífero como la diastasa o fructosa, o incluso parámetros que están correlacionados directamente con la calidad o autenticidad de la miel debido a prácticas de almacenamiento adecuadas en largos periodos de tiempo, fermentación de la miel o adulteración de esta con otros compuestos para añadir espesor o dulzor artificiales, llamado hidroximetilfurfural (HMF), los cuales presentan bajas concentraciones al medirlos en muestras frescas [13], [14], [15] pero que se elevan considerablemente al ser adulteradas o estar caducadas.

Estos compuestos, según la literatura expuesta anteriormente en referencias bibliográficas, pueden ser medidos a través de técnicas de espectrometría, y en especial, es posible caracterizar en el rango espectral infrarrojo distintas curvas que permiten contrastar muestras genuinas con muestras adulteradas.

Por ejemplo, cuando una muestra contiene una cantidad significativa de carbono (C) a nivel molecular, esto puede detectarse mediante las líneas de absorción específicas de carbono.

Por otra parte, para poder determinar el grado de adulteración presente en estas muestras y así establecer si estas poseen diferencias relevantes entre ellas o si forman algún patrón base con la respuesta de adulterantes (fructosa, maltosa, glucosa), es necesario hacer uso de técnicas de espectrometría, la cual permita establecer parámetros que logren describir correctamente el análisis e indicar el grado de adulteración. Es aquí donde juega un rol esencial el análisis de una máquina entrenada bajo algoritmos de aprendizaje automático, ya que, dependiendo de su precisión y desempeño para clasificar distintas clases de miel, será posible identificar adulteraciones e indicar sus niveles de adulteración.

1.2. HIPOTESIS

- Es posible desarrollar un método integral para identificar adulteración de azúcar y determinar el origen botánico de mieles endémicas chilenas, utilizando técnicas ópticas de espectroscopía LIBS para análisis elemental y visión artificial por microscopía para obtener características morfológicas de las muestras utilizando algoritmos de aprendizaje automático para obtener y generar modelos clasificadores para características elementales y morfológicas, identificando la autenticidad y el origen botánico de las muestras.

1.3. OBJETIVOS

1.3.1. OBJETIVO PRINCIPAL

- Desarrollar un método que permita automatizar el reconocimiento botánico y la detección de miel adulterada en un banco de datos local, implementando algoritmos de aprendizaje automático a técnicas de procesamiento de imágenes mediante microscopía óptica para el análisis de conteo de polen y análisis de características espectrales mediante espectroscopía LIBS, generando y correlacionando indicadores fundamentales para evaluar la adulteración y calidad de la miel.

1.3.2. OBJETIVOS ESPECÍFICOS

1. Realizar la obtención, registro y etiquetado de muestras de miel con un número mínimo de 25 muestras pertenecientes a la región del Bío – Bío (entre mieles puras y sus respectivas adulteraciones de fructosa, glucosa y maltosa), que serán utilizadas para estudios de espectrometría e imágenes infrarrojas.
2. Preparación de muestras puras y adulteradas de miel, en contenido de azúcar añadido en laboratorio (fructosa, glucosa, maltosa), para estudios comparativos de tramos con distintas cantidades de azúcar (al menos 5 intervalos).
3. Caracterización óptica sobre muestras de miel para medir índices de absorbancia relacionados a sus componentes principales por análisis semiautomático, generando indicadores de referencia y para cada muestra a través de espectrometría infrarroja LIBS.
4. Estandarización y preparación de muestras para visión artificial del banco de datos de muestras florales local, utilizando microscopía de laboratorio para la adquisición de imágenes.
5. Registro de cada muestra de miel a través de visión artificial, obteniendo imágenes que permitan distinguir las concentraciones de polen presentes en ellas, seguido de la elaboración de indicadores que permitan medir dichas concentraciones de polen detectado.
6. Establecer los tramos de contenido, en adulterantes y miel, en las distintas muestras tabulando dichos resultados en tablas comparativas cruzando estos datos con sus firmas espectrales, en preparación para artículo de publicación científica y reporte de tesis.

1.3.3. ALCANCES

El proyecto de investigación está delimitado geográficamente a la Región del Bío – Bío, Chile. Las técnicas aplicadas son de tipo ópticas mínimamente invasivas para análisis de características físico – químicas de muestras de miel en el caso de espectroscopía, y para el caso de técnicas de visión artificial y procesamiento de imágenes son con fines de cuantización y observación de las muestras (polen). Ambas son técnicas realizadas en laboratorios equipados adecuadamente en las instalaciones de:

- I. Laboratorio de Procesamiento de Señales e Imágenes, Facultad de Ingeniería, Universidad de Concepción.
- II. Laboratorio de Trazas Elementales y Especiación, Facultad de Ciencias Químicas, Universidad de Concepción.

La base de datos y muestras a estudiar han sido desarrolladas durante los meses de marzo – abril 2021, en las instalaciones de:

- III. Laboratorio de Palinología y Ecología Vegetal, Facultad de Ciencias Forestales, sede Los Ángeles, Universidad de Concepción.
- IV. Laboratorio de Nano - Biotecnología y Optoelectrónica, Universidad San Sebastián, Campus Las Tres Pascualas, Concepción.

El desarrollo de la base de datos se ha construido a partir de muestras de miel adulteradas, donde la presencia de miel es de [10, 20, 30, 40]% en muestras adulteradas, considerando muestras patrones de sólo adulterantes (*fructosa, glucosa, maltosa*) y mieles puras (*Eucryphia Glutinosa, Quillaja Saponaria*).

1.4. METODOLOGÍA

Para realizar este estudio ha sido necesario recolectar las muestras a utilizar a lo largo de todo su desarrollo, **OE-1**. Estas poseen origen en el sector de Alto del Bío – Bío, región del Bío – Bío, Chile, las cuales sólo serán objeto de análisis, guardando el anonimato de los apicultores locales.

Junto con el apoyo colaborativo de las facultades respectivas de universidades interesadas en este estudio, se ha elaborado un banco de datos normalizado para las muestras, destacando las clases endémicas del sector Alto Bío – Bío, Región del Bío – Bío, Chile “Guindo Santo” y “Quillay”. Estas han sido preparadas bajo protocolos de laboratorio, asegurando la integridad de las muestras.

Dichas muestras han sido adquiridas y registradas, tabulando sus características relevantes de origen y etiquetadas finalmente bajo “*Codigo-miel__Codigo-Adulterante__Porcentaje-de-Miel*”, como parte del **OE-2** (de un total de $N = 29$). De su adquisición, han sido divididas entre muestras puras y muestras alteradas en distintas cantidades porcentuales en azúcar (deciles, 0% a 100%, intervalos de 10%, relación miel/adulterante), midiendo dichas concentraciones en sus últimos 4 deciles de adulteración, y de cada patrón o huella espectral para mieles puras como adulterantes.

Una vez que se ha obtenido el material necesario para las muestras y sus variaciones, se han llevado a cabo 2 tipos de estudios:

1. Laser Induced Breakdown Spectroscopy (LIBS):

Rango de trabajo en el espectro visible e IR cercano (200 ~ 1000) [nm], obtención de información espectroscópica de perfiles de miel o huellas espectrales específicas en **OE-3**.

2. Visión artificial:

Rango de trabajo propuesto ha sido en el intervalo del espectro visible (300 – 750) [nm], con objetivo de la obtención de imágenes digitales en **OE-4**, utilizando para microscopía óptica distintos filtros para longitudes de onda objetivos y asistencia móvil para escaneo de superficie de muestras.

La necesidad de utilizar estas 2 técnicas de análisis de forma paralela, radica en que LIBS obtiene por resultado un perfil de composición elemental para múltiples mediciones, donde cada muestra posee una huella espectral específica, capaz de describir la presencia de distintos elementos presentes en estas, obteniendo así el perfil de una muestra pura, parcial o totalmente adulterada, mientras que, con visión artificial será posible identificar morfológicamente polen con motivos estadísticos de conteo, con interés mayoritario en la distribución espacial y la determinación de origen botánico.

Los resultados LIBS presentan una alta dimensionalidad en representación de datos, lo que implica una tarea de clasificación multivariable intensiva. Para abordar dicha tarea, se implementaron 2 enfoques o estrategias principalmente. Primero, se realizan análisis exploratorios para identificar longitudes de onda

relevantes para la clasificación, lo que implica un análisis preliminar de características, las cuales pueden estar orientadas al objetivo tanto como a otras señales del contexto de medición.

Luego de realizada la etapa exploratoria, en conjunto con la identificación de longitudes de onda para elementos químicos específicos relevantes para la diferenciación de azúcares (como carbono, hidrógeno, oxígeno o nitrógeno, entre otros elementos) se implementó un modelo con enfoque en sus características químicas más relevantes (**OE-3**), generando indicadores cuantitativos específicos para dichos grupos elementales.

Para visión artificial, la base de datos utilizada contiene 15 clases utilizables, entre ellas 2 procesadas también por LIBS. El trabajo de analizar, elaborar limpieza de datos, corregir errores y preparación de dicha base de datos es crucial para contar con información fiable y diferenciable entre clases para los algoritmos de aprendizaje posteriores:

ID	Especie endémica	Nombre común
0	<i>Gevuina Avellana</i>	<i>Almendro</i>
1	<i>Medicago Sativa</i>	<i>Alfalfa</i>
2	<i>Anthemis cotula</i>	<i>Manzanilla</i>
3	<i>Aristotelia chilensis</i>	<i>Maqui</i>
4	<i>Luma apiculata</i>	<i>Arrayán</i>
5	<i>Luma chequen</i>	<i>Chequén</i>
6	<i>Conium maculatum</i>	<i>Cicuta</i>
7	<i>Convolvulus arvensis</i>	<i>Correhuela</i>
8	<i>Eucryphia glutinosa</i>	<i>Guindo Santo</i>
9	<i>Escallonia pulverulenta</i>	<i>Madroño</i>
10	<i>Mentha pulegium</i>	<i>Poleo</i>
11	<i>Quillaja saponaria</i>	<i>Quillay</i>
12	<i>Trifolium pratense</i>	<i>Trébol rojo</i>
13	<i>Cissus striata</i>	<i>Voqui</i>
14	<i>Colletia hystrix</i>	<i>Yaqui</i>

Tabla 1.4.1: Clases correspondientes a la base de datos para visión artificial.

Los algoritmos de aprendizaje profundo implementados han sido construidos a través de arquitecturas de aprendizaje automático profundo (red neuronal convolucional, algoritmo “You Only Look Once”) permitiendo establecer indicadores estadísticos respecto a la cantidad de polen presente en cada muestra, una vez obtenido un modelo clasificador efectivo para dicha tarea (**OE-5**).

Una vez obtenidos los resultados finales de cada clasificador, se evaluaron dos aspectos: por un lado, los rendimientos efectivos de LIBS en la identificación de tramos de adulteración; y por otro, el uso de visión artificial para el conteo estadístico del polen presente en las muestras de miel. Esta evaluación permitió elaborar informes y tablas de resultados, visualizando de forma paralela las diferencias de características relevantes.

Utilizando ambos resultados, se determinó el grado de adulteración en cada muestra a través de los índices de composición elemental y, paralelamente, con indicadores de la concentración de polen en cada muestra, identificar su origen botánico.

El modelo de visión artificial entrenado ha sido validado tanto internamente (set de validación interno) como externamente en laboratorios de Palinología y Ecología Vegetal (incluyendo comparación estadística). Las validaciones internas han permitido ajustar y optimizar el modelo, mientras que las externas, realizadas en colaboración con laboratorios especializados, compararon los resultados reales de métodos palinológicos. Estas pruebas definieron la precisión del modelo en el conteo de polen y su capacidad para diferenciar entre diversas fuentes botánicas, validando su utilidad en el análisis de adulteración de miel.

Es natural que cada técnica posea resultados que dependan o varíen porcentualmente de cómo se ha medido en la práctica una muestra (ruido base instrumental), por lo que para disminuir el ruido electrónico se debe corregir medidas por interferencias externas y reproducir intensidades mínimas/máximas, procesos que se deben respetar en el uso de sistemas ópticos, con objetivo de no perder fidelidad ni generalidad respecto a la forma original de los datos obtenidos.

Una vez realizado correctamente todo el proceso, se ha tabulado los indicadores de autenticidad y calidad en las muestras de miel (**OE-6**) obtenidos de las mediciones anteriores, con el objetivo de representar e identificar perfiles o anomalías espectrales relacionadas con posibles adulteraciones. En esta etapa, se correlacionaron los resultados para establecer si una muestra estaba adulterada o no, asegurando así una evaluación detallada de la autenticidad de las muestras de miel analizadas.

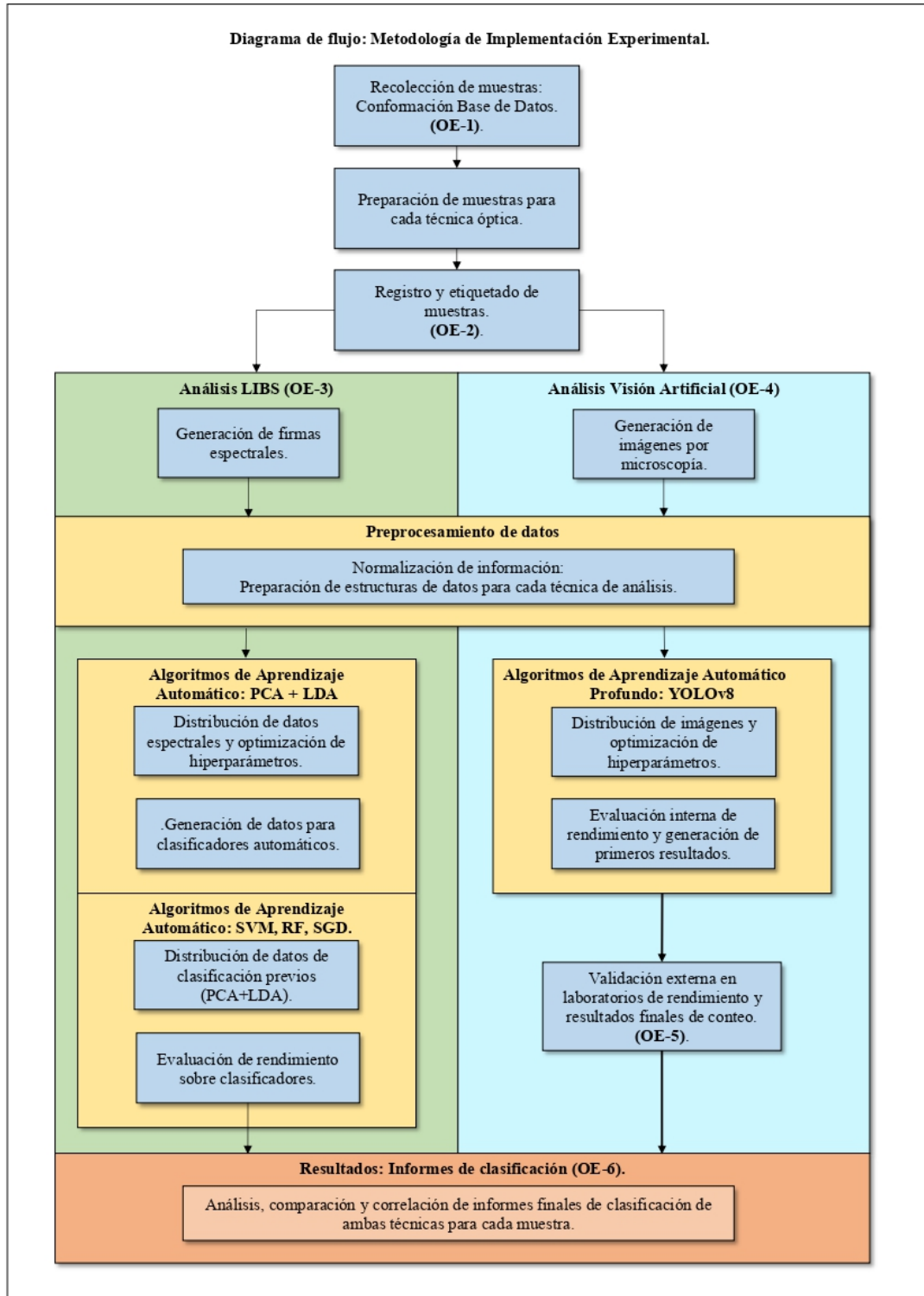


Figura 1.4.1: Diagrama de flujo para la implementación experimental de algoritmos de aprendizaje automático a ambas técnicas: Espectroscopía LIBS y Visión Artificial en imágenes de polen.

2. ESTADO DEL ARTE

La miel es un producto natural elaborado por las abejas a partir del néctar de las flores o de secreciones vegetales. Una vez extraído el néctar de las flores, el proceso de elaboración de la miel continúa cuando las abejas ingieren el néctar y lo almacenan en su abdomen, denominado “buche melario”. Una vez ingerido, las abejas modifican el pH de la miel al combinarlo con enzimas específicas.

Posteriormente, las abejas vuelven a la colmena, donde el proceso de elaboración de la miel continúa. Es ahí donde estas traspasan el néctar a otras abejas obreras más jóvenes, que extraen y descomponen el néctar a través de diversas digestiones enzimáticas en compuestos más sencillos, ricos principalmente en fructosa y glucosa.

Esto le aporta el sabor dulce a la miel y su gran aporte calórico, características aportadas por su composición química de azúcares naturalmente presentes como fructosa, glucosa, maltosa, entre otros [16]. Las abejas mastican este néctar durante aproximadamente media hora, hasta que la humedad del néctar pasa del 70% hasta un 20%. Asimismo, estas enzimas modifican el pH de la miel, situándose en torno al 3.9, es decir, es un producto muy ácido. Transcurrido el tiempo, manteniendo este proceso natural de digestión, el néctar se acaba convirtiendo en miel apta para el consumo humano, valorada por su sabor dulce y que contiene nutrientes como azúcares naturales, vitaminas, minerales y antioxidantes, siendo utilizada históricamente con propósitos culinarios y medicinales en diversas culturas.

La melisopalinología es un campo de investigación que desempeña un papel fundamental en la comprensión de la miel [17]. Se dedica al estudio y la caracterización de polen, el cual está directamente relacionado con las diversas especies vegetales presentes de una región. Esta relación entre el polen y las plantas circundantes es crucial, ya que el polen es una de las principales fuentes de componentes y propiedades de la miel. A través del análisis de las semillas de polen, podemos identificar las fuentes botánicas que contribuyen a la miel en una región particular [18], lo que tiene un impacto significativo en su sabor, aroma y valor nutricional. Por lo tanto, la melisopalinología desempeña un papel esencial en la caracterización y la calidad de la miel, permitiendo una comprensión más profunda de su origen y sus propiedades únicas.

Así, en los esfuerzos para lograr recabar más información respecto a su morfología o composición, se han diseñado métodos por expertos palinólogos dedicados principalmente a determinar y caracterizar polen, y así representarlos gráficamente. Sus tamaños rondan el orden de (10 ~ 60) [μm], a través de técnicas de microscopía donde la optoelectrónica de estos sistemas [19] suelen generar desafíos a abordar respecto al ruido generado durante los procesos de digitalización, principalmente en sensores ópticos con un ruido de corriente oscura relevante, que, si no se posee el conocimiento de rigor, puede ser un problema aún más complejo si no es compensado su efecto en la respuesta espectral deseada.

En diversos estudios, se ha logrado caracterizar la composición polínica mediante el uso de modelos automáticos de aprendizaje de máquina y aprendizaje profundo. Estos modelos se han entrenado con bases de

datos públicas notables, como POLEN73S (una base de datos denominada "Savana brasileña" con múltiples clases) o POLEN13K (que comprende 13,000 fotos distribuidas en 4 clases) [20], [21]. Estas bases de datos se distinguen por contar con un preprocesamiento especializado, lo que permite establecer estándares de trabajo uniformes entre diversas fuentes. Esta metodología proporciona diferentes parámetros que facilitan la correcta inferencia de la clasificación del polen, identificando también si se trata de información relacionada con material particulado indeseado u otros elementos orgánicos residuales. El logro de este objetivo por parte de los equipos de investigación es significativo y esencial para el continuo desarrollo de tecnologías de análisis capaces de proporcionar información relevante sobre el polen local.

Durante los últimos años, se han visto distintos algoritmos y redes de aplicación en algoritmos de aprendizaje automático en redes neuronales, tales como AlexNET, ResNET, YOLO ("You Only Look Once", uno de los algoritmos recientes más potentes y eficientes a la fecha) [22], [23], [24], [25], [26], [27], que desde sus inicios han marcado grandes precedentes. Es necesario notar que, en el desarrollo de diferentes arquitecturas, los rendimientos en la actualidad respecto a la identificación eficaz y eficiente rondan entre el [95 ~ 99]%, y tratándose específicamente de sus rendimientos, el trabajo conjunto con técnicas de preprocesamiento convencionales puede dar con el 99% de validación de datos a través de indicadores de "*Intersection over Union (IoU)*".

En este aspecto, también es destacable mencionar que aquellos estudios que han considerado obtener resultados de preprocesamiento óptimos u obtención de muestras lo más fiel posibles a las morfologías a representar [28], como es el caso de concentrados de miel, sus indicadores IoU presentan desafíos a sus procesos finales de identificación.

Existen otros tipos de acercamientos a la inspección de polen que aplicado redes neuronales convolucionales profundas, como algoritmos de clasificación convencionales (SVM, RF) en muestras aéreas como líquidas, sistemas como "RAPID – E particle analyzer" [29] en Rumania, pueden actuar como instrumentos que permiten identificación de nuevas especies o determinación de especies locales. El problema que suscita en tal caso, es que es un sistema cerrado que no propone validaciones en base a estudios de laboratorios bajo supervisión de expertos palinólogos, por lo que si bien es posible identificar 10.000 muestras por segundo como según lo indica su fabricante, los datos obtenidos no han sido corroborados por estudios complementarios, y los resultados obtenidos respecto a las clases correspondientes, oscilan entre 57% y 90% en redes de análisis para aprendizaje automático; otros estudios [30] que han optado por analizar muestras líquidas a través de micro fluidos e identificación electro-óptica de semillas de polen, señalan que es posible identificar hasta 150 muestras por segundo, incrementando la eficiencia de sus modelos de aprendizaje automático en alrededor 97% a través de algoritmos de . En este caso, se ha subdividido en 3 métodos para lograr contrastes de resultados relevantes, tales como entrenamiento en base a sólo mediciones eléctricas a través de las impedancias medidas por micro canales; entrenamientos en base a sólo mediciones ópticas haciendo uso de imágenes digitales. Por último, el 3er método o "multimodal" es haciendo uso de estos datos

y cruzando la información. En este último caso, es posible hablar de desempeños importantes cercanos al 99% en algunas clases, pero que no logra un resultado uniforme con respecto a otras especies en comparación.

Dentro de las aplicaciones encontradas, estudios relacionados a la determinación y caracterización ambiental sobre calidad de aire, que miden principalmente la sobrecarga de sistemas de ventilación o filtrado de aire [31], [32], han derivado en estudios diferenciadores clave para la detección y conteo de granos de polen, elaborando índices de concentración y pureza del aire, criterios que también son aplicados sobre distintos productos alimentarios, como la miel [33].

Los métodos planteados anteriormente solucionan la problemática de poder caracterizar correctamente distintas clases de miel haciendo uso de complejos métodos para procesar imágenes obtenidas por microscopía, e incluso utilizando e implementando múltiples algoritmos de aprendizaje profundo. Es por esta característica, que se propone utilizar métodos ópticos mínimamente invasivos, utilizando técnicas como LIBS y visión artificial por imágenes de microscopía haciendo uso de algoritmos de aprendizaje automático para procesar dichos datos. La portabilidad de muestras es crucial para el uso de estas técnicas, ya que podrían ser rápidamente implementables en los distintos experimentos, caracterizando así distintos compuestos o polen presentes en las muestras, como también indicar qué tipo de elementos químicos pueden encontrarse en dichas muestras de miel, aquellos pertenecientes a grupos orgánicos relevantes como carbono (C), calcio (Ca), magnesio (Mg), potasio (K), sodio (Na), hidrógeno (H) u oxígeno (O).

Uno de los aspectos importantes dentro de la literatura es la caracterización de las huellas espectrales específicas [34], [35], [36], [37] con respecto a distintas muestras de miel, las cuales permitan determinar o dilucidar compuestos que generan la adulteración de alimentos tal como sucede en otras industrias (como es el caso de análisis en concentrados de cobre [38] con presencia de elementos químicos que merman su calidad) y así lograr la detección e identificación eficaz de distintos tipos de productos que puedan ser desarrollados, analizándolos en base a técnicas de espectrometría, haciendo posible el estudio de concentrados de miel, determinando el nivel de jarabes artificiales u otros compuestos no pertenecientes a la miel producida por abejas, en casos de determinación de adulteración alimentaria.

Así es como resulta imperativo trabajar con bases de datos que posean la mayor cantidad de parámetros identificables indicados en sus pasos de preprocesamiento de datos, que, al ingresar a algoritmos de aprendizaje automático, consideren parámetros indicados por profesionales y expertos de áreas palinológicas, cruzando datos que puedan verificarse a través de otros métodos para la correcta identificación de clases de miel [39]. Si bien este proceso puede resultar relativamente sencillo de leer, en la práctica contiene extensas jornadas de identificación y digitalización; diversos estudios de diferenciación de especies; y no menos importante, tiempos de entrenamiento de uso extensivo a nivel computacional en sistemas de aprendizaje automático, por lo que consecuentemente se habla de optimización en la energía utilizada para la generación de modelos descriptivos.

2.1. DISCUSIÓN ESTADO DEL ARTE

El análisis espectral y espacial permite la caracterización de muestras endémicas de miel en distintos aspectos relevantes para la determinación de su calidad: molecular y origen botánico. Inicialmente, para lograr el objetivo de identificar compuestos adulterantes y determinar el tipo de miel u origen botánico de procedencia [2], es necesario abordar el tema con especial consideración de la Ley 21.489 que tipifica, define, regula y sanciona las actividades de adulteración o fraude alimentario [9] en la industria apícola nacional, haciendo uso de diversos métodos de forma paralela para la tarea de clasificación y regulación de la actividad productora apícola. Esta actividad regulatoria, actualmente es delegada a instituciones y laboratorios certificados que realizan métodos físico – químicos para el análisis apícola, que, en su proceso de estudio y determinación, el tiempo y preparación de muestras son considerables (días, semanas) [3], [10].

Las tecnologías utilizadas, LIBS y visión artificial por microscopía, son utilizadas de forma que sus resultados puedan ser procesados por algoritmos de aprendizaje automático, con objetivo de identificar características distintivas entre diferentes clases de miel o detectar adulterantes utilizando distintos tipos de elementos químicos (magnesio, potasio, calcio, sodio, entre otros) [42] que son de gran relevancia para determinar el tipo de elementos mayoritarios que contenga dicha clase, posibilitando la exploración e identificación de trazas que puedan aportar a discernir y caracterizar estructuras químicas relevantes, como también determinar el origen botánico [43] ya que los terrenos donde se produzca dicha actividad poseen rasgos propios de formaciones terrestres y vegetales. Así, se puede lograr caracterizar la mineralogía presente en dichas muestras, utilizando esta información de referencia con respecto a distintos lugares geográficos, pero, aunque los rendimientos para dichos clasificadores sean destacables, se debe dejar en claro que las especies de polen cambian en todo el mundo, y el desafío de distinguir entre clases se vuelve mucho más variado, elevando la complejidad para tareas de clasificación. Para alcanzar este objetivo, es crucial tener un entendimiento cabal del dominio espectral para poder considerar distintas estructuras elementales que puedan describir el contenido de azúcar en una muestra [40], e incluso considerando contenidos minerales los que pueden ser representados según las necesidades a identificar [41], como la ubicación geográfica u origen botánico, pero que aunque las firmas espectrales puedan dar indicios de presencia elemental o mineralógica relevantes, es una técnica que debe utilizarse en complemento con otras técnicas de análisis.

Por otra parte, a nivel espacial [21], [22] es esencial describir o poseer la capacidad algorítmica de detectar características relevantes para analizar por visión artificial, tales como bordes de objetos o grano de polen donde la morfología de estas muestras aporta de manera contundente a distintos algoritmos de visión y redes convolucionales profundas, principalmente aplicadas en bases de datos muy diversas (19.000 imágenes para 40 clases) [44], donde si se dispone de equipos para generar imágenes con una alta densidad de elementos de píxeles y múltiples bandas de representación óptica (RGB o HSV), los algoritmos de reconocimiento y redes convolucionales profundas son capaces de procesar dichas características y entregar clases diferenciadoras basadas en características físico – ópticas, las cuales permiten realizar inferencias más efectivas.

Es por esto que, poseer un nivel de comprensión profundo de las propiedades de la luz permite enfocar el análisis en características espectrales específicas asociadas con elementos de compuestos químicos y morfologías a distintas clases de polen. Como estrategia principal, se propone entrenar modelos clasificadores utilizando técnicas avanzadas de espectrometría y visión artificial.

Cada acercamiento óptico posee también estrategias que optimizan sus rendimientos, tal como sería el caso de un proceso complejo que cuente con múltiples filtros dedicados a una traza en especial en la preparación de muestras [14], o considerando distintas alternativas a nivel óptico [30]. Por otra parte, configurar correctamente los parámetros de distintas arquitecturas de aprendizaje y clasificación para mejorar etapas de entrenamiento [36], o la validación iterativa para distintas fuentes de datos internas y externas, con objetivo de obtener un modelo aplicable a procesos reales e implementables de clasificación automática, son parte de algunos procesos de selección que definen distintos subprocesos, con el fin de mejorar el rendimiento final de inferencia del modelo clasificador.

2.2. DISCUSIÓN BIBLIOGRÁFICA

La autenticación y control de calidad de productos apícolas han sido temas recurrentes en la literatura científica debido a la creciente preocupación por el fraude alimentario y la necesidad de métodos más eficientes y precisos para su detección [1], [2], donde se cumplan las regulaciones demandadas por ley. Estudios previos han explorado diversas técnicas de laboratorio [49], incluyendo métodos palinológicos tradicionales que utilizan instituciones regulatorias (Servicio Agrícola Ganadero; Fundación Innovación Agraria) [11], que, aunque efectivos, presentan limitaciones en términos de tiempo y recursos debido a la necesidad de analizar manualmente grandes volúmenes de datos.

Por otro lado, la espectroscopía LIBS ha sido ya ampliamente utilizada en diferentes disciplinas para el análisis elemental de muestras biológicas [33], permitiendo una identificación rápida y mínimamente invasiva de componentes químicos [50]. Sin embargo, su aplicación en la autenticación de miel es relativamente nueva e incluso de difícil acceso debido a los instrumentos ópticos y sistemas de control necesarios, pero enfatizando estudios recientes que han demostrado su potencial para diferenciar entre mieles puras y adulteradas [40].

En paralelo, la visión artificial ha mostrado un avance significativo en la clasificación y conteo de polen [39], lo cual es esencial para la identificación del origen botánico de la miel. La combinación de estas técnicas con algoritmos de aprendizaje automático profundo ha permitido mejorar la precisión y velocidad de estos análisis.

No obstante, pocos estudios han abordado la integración de ambas técnicas en un solo enfoque para la autenticación de miel, lo que representa una oportunidad única para optimizar los procesos de análisis y ofrecer una herramienta más robusta y precisa ante este problema de clasificación. Así, la investigación se posiciona en este vacío, proponiendo un método integrado que reduce significativamente los tiempos de análisis y aumentar la precisión en la detección de adulterantes específicos, contribuyendo así a la literatura existente y ofreciendo un avance metodológico relevante para la industria alimentaria.

3. MARCO TEÓRICO

El espectro electromagnético representa la totalidad de las frecuencias posibles de la radiación electromagnética, abarcando desde las más bajas hasta las más altas. Constituye un espectro continuo, comprendiendo diversas categorías de radiaciones, cada una definida por su frecuencia o longitud de onda específica. En el extremo de menor energía del espectro se sitúan las ondas de radio, caracterizadas por su baja frecuencia y largas longitudes de onda (extremo derecho figura 3.1). Progresando a lo largo del espectro, se encuentran las microondas, seguidas por la radiación infrarroja, zona de energía la cual será utilizada en este estudio.

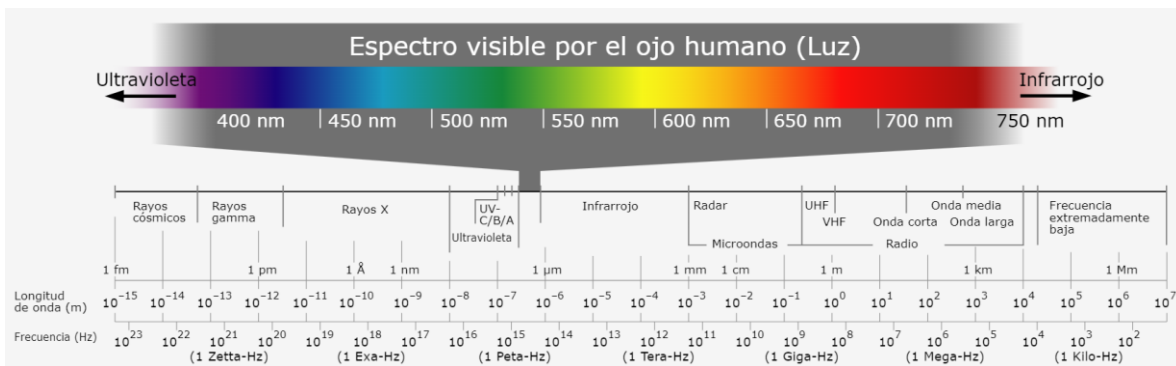


Figura 3.1: Espectro electromagnético detallado en longitud de onda y frecuencia para distintos tramos que lo componen.

En una posición central dentro de este espectro se halla la luz visible, que, aunque representa solo una fracción muy pequeña del espectro total, es de gran importancia, ya que es la única parte que el ojo humano es capaz de detectar. Esta región visible del espectro incluye el rango de longitudes de onda asociadas con el fenómeno óptico del espectro de dispersión, comenzando con la banda correspondiente al violeta, que tiene la longitud de onda más corta dentro de los colores perceptibles por el ojo humano y, consecuentemente, la mayor energía entre los colores visibles finaliza con el rojo, que tiene la longitud de onda más extensa y la menor energía dentro de este rango. Prosiguiendo más allá del espectro visible, se encuentran la radiación ultravioleta, los rayos X y, en el extremo de mayor energía del espectro, los rayos gamma.

El espectro infrarrojo (IR) es una parte importante del espectro electromagnético que se sitúa justo más allá de la luz visible, en longitudes de onda más largas y frecuencias más bajas que la luz roja. Este espectro se subdivide en las siguientes regiones:

División del Infrarrojo	Rango de Longitud de Onda [μm]	Características y Aplicaciones Generales
Infrarrojo Cercano (NIR)	0.75 - 1.4	Usado en telecomunicaciones, fotogrametría, medicina y visión nocturna.
Infrarrojo de Onda Corta (SWIR)	1.4 - 3	Aplicado en aplicaciones industriales, militares y de imágenes térmicas.
Infrarrojo Medio (MIR)	3 - 8	Importante en espectroscopía, detección de gases y comunicaciones a larga distancia.
Infrarrojo de Onda Larga (LWIR)	8 - 15	Utilizado en imágenes térmicas y sistemas de seguridad.
Infrarrojo Lejano (FIR)	15 - 1000	Empleado en astronomía y aplicaciones científicas.

Tabla 3.1: División del espectro electromagnético infrarrojo y sus principales características.

3.1. LASER INDUCED BREAKDOWN SPECTROSCOPY (LIBS)

Como técnica de análisis, la espectroscopía por inducción con láser LIBS [40] es una técnica analítica que ha ganado reconocimiento en diversos campos debido a su capacidad para proporcionar información rápida y precisa sobre la composición elemental de una muestra.

El principio fundamental de LIBS se basa en la generación de una nube de plasma en la superficie de la muestra mediante la focalización de un pulso láser de alta intensidad de energía. A medida que el pulso láser interactúa con la muestra, se produce una ablación y una posterior ionización del material, generando un plasma de alta temperatura que emite radiación característica de los elementos presentes.

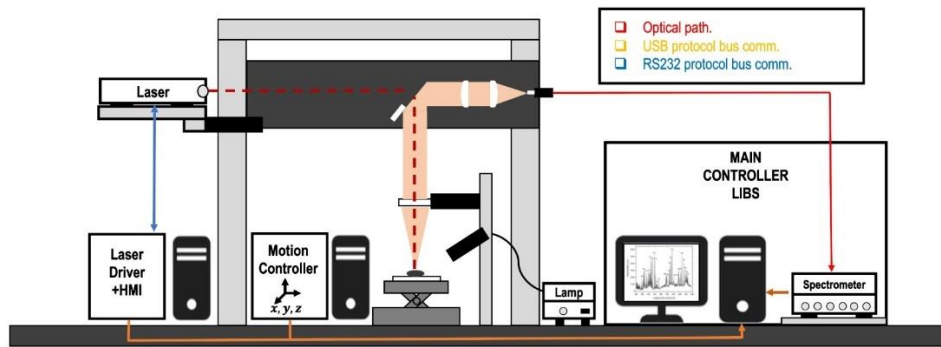


Figura 3.1.1: Configuración de referencia de técnica “Laser Induction Breakdown Spectroscopy (LIBS)”.

LIBS ofrece varias ventajas, como la capacidad de análisis sin necesidad de una preparación especial de la muestra y da la posibilidad de analizar muestras en condiciones ambientales. Además, es una técnica con aplicación no destructiva, lo que permite el análisis repetido de una misma muestra y la preservación de la integridad del material.

En la última década, la técnica LIBS ha experimentado un rápido desarrollo y se ha aplicado en numerosos campos, como la geología, la arqueología, la ciencia de materiales, la medicina y la agricultura. En la geología, por ejemplo, se utiliza para el análisis elemental de minerales y rocas [41], lo que permite la identificación de su composición y la caracterización de yacimientos. En la arqueología se emplea para el análisis de pigmentos y materiales utilizados en artefactos antiguos, lo que proporciona información sobre su origen y autenticidad [45].

En cuanto a la ciencia de materiales, LIBS se ha utilizado para el análisis de recubrimientos, la caracterización de aleaciones metálicas y la detección de contaminantes en materiales semiconductores [46]. En medicina, se ha explorado el uso de la LIBS para el análisis de tejidos biológicos y la detección de enfermedades, como el cáncer. En el campo de la agricultura, la LIBS se ha aplicado para el análisis elemental de suelos y plantas, lo que permite el monitoreo de nutrientes y la evaluación de la calidad del suelo.

LIBS se basa en un fenómeno físico conocido como la emisión atómica inducida por láser. Este fenómeno ocurre cuando un láser de alta energía incide sobre una muestra y provoca la formación de un plasma.

El plasma es un estado de la materia que se forma cuando se suministra suficiente energía a un material, lo que produce la ionización de los átomos y la formación de iones y electrones libres. En el caso de LIBS, el láser de alta energía vaporiza y excita los átomos de la muestra, creando un plasma temporal. Cuando los átomos en el plasma vuelven a su estado fundamental, liberan energía en forma de radiación electromagnética. Esta radiación es característica de cada elemento químico presente en la muestra y se puede analizar para determinar su composición química.

Elemento	Longitud de onda [nm]
H (I)	656.3, 486.1
C (I)	247.9
N (I)	742.4, 744.2, 746.8, 818.6, 821.8, 824.2, 856.8, 859.4, 862.8, 868.2, 871.3, 938.7, 939.3
O(I)	777.2, 777.4, 777.5, 844.6, 926.6
Na (I)	589.0, 589.6
Mg (I)	285.2
Mg (II)	279.6, 280.3
K (I)	766.5, 769.9
Ca (I)	422.7
Ca (II)	393.3, 396.8
CN ($\Delta v = -1$)	359.0 (1-0), 358.59 (2-1)
CN ($\Delta v = 0$)	388.3 (0,0), 387.1 (1,1), 386.2 (2,2)

Tabla 3.1.1: Tabla de longitudes de onda de emisión para cada elemento atómico relevante en la clasificación de miel [40] en LIBS.

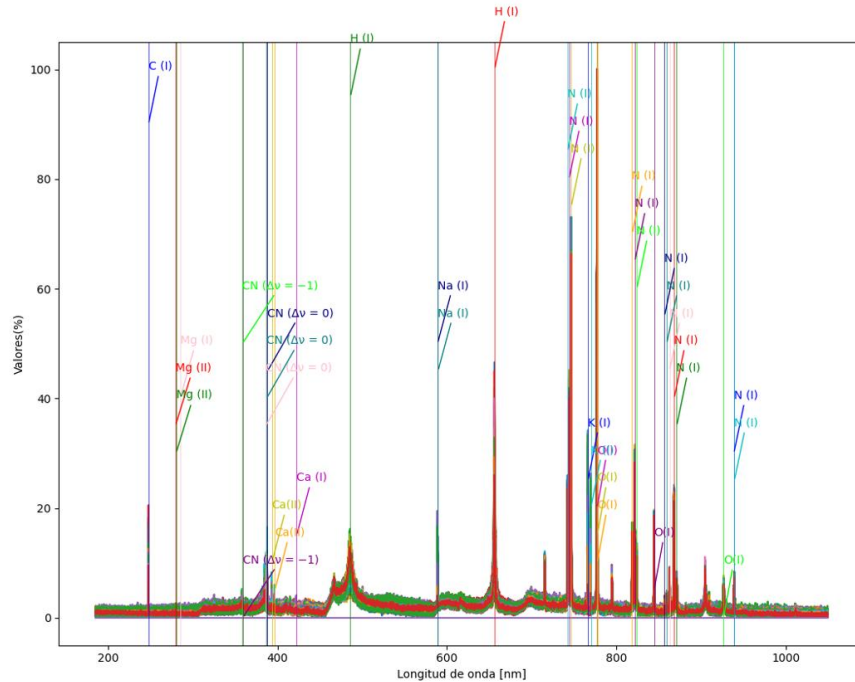


Figura 3.1.2: Representación visual de longitudes de onda de emisión para cada elemento relevante en la clasificación de miel [40] en LIBS, rango UV – NIR [200 - 1000] [nm].

La ablación es un proceso fundamental en la técnica LIBS, que se refiere a la remoción selectiva de material de la superficie de la muestra mediante el impacto del láser de alta energía. La ablación se produce cuando la intensidad del láser es lo suficientemente alta como para vaporizar, ionizar y excitar los átomos de la muestra, formando así el plasma.

Durante la ablación, el láser penetra en la superficie de la muestra y genera un frente de vaporización que avanza hacia el interior del material. A medida que el frente de vaporización se propaga, arrastra consigo los átomos, iones y moléculas biatómicas presentes, los cuales son excitados y eventualmente liberados como especies atómicas y moleculares en el plasma. La ablación por LIBS puede ser influenciada por varios factores, como la energía y la duración del pulso láser, el tipo de láser utilizado, la composición y estructura de la muestra, y las propiedades ópticas y térmicas del material. La optimización de estos parámetros es esencial para lograr una ablación eficiente y reproducible.

La radiación emitida por el plasma se recoge mediante un sistema óptico, generalmente fibra óptica, que la dirige hacia un detector. El detector, que suelen ser arreglos de distintos semiconductores sensibles a distintas longitudes de onda (sensores CMOS/CCD), registran la radiación en forma de un espectro electromagnético, que muestra la intensidad de la radiación en función de la longitud de onda. Cada elemento químico tiene líneas de emisión atómica características en su espectro, lo que permite identificar y cuantificar los elementos presentes en la muestra en sistemas de alta precisión.

La técnica LIBS ofrece varias ventajas en comparación con otras técnicas de análisis químico. Es un método mínimamente invasivo, lo que significa que no se requiere una preparación especial de la muestra y se puede analizar directamente en su estado original. Además, LIBS es rápido y puede proporcionar resultados en tiempo real, lo que lo hace adecuado para aplicaciones en campo y en líneas de producción.

Sin embargo, la técnica LIBS también tiene algunas limitaciones. La sensibilidad y precisión pueden verse afectadas por factores como la matriz de la muestra, la presencia de interferencias y la homogeneidad de la distribución de los elementos en la muestra. Además, la calibración y la interpretación de los espectros requieren experiencia y conocimientos especializados.

3.1.1. ADULTERANTES: Fructosa, Glucosa, Maltosa.

La fructosa, la glucosa y la maltosa son azúcares esenciales en productos alimenticios, influyendo significativamente en su sabor y textura. La fructosa se utiliza por su alta dulzura, la glucosa es común en confitería y productos horneados, y la maltosa se emplea en la producción de malta, cerveza, y ciertos tipos de pan y dulces.

Investigaciones han identificado tendencias de adulteración en alimentos mediante la adición de fructosa y glucosa, lo que puede alterar la calidad y autenticidad del producto. Detectar estas adulteraciones es crucial para garantizar la integridad de los alimentos.

La espectroscopía de ruptura inducida por láser (LIBS) es una técnica analítica avanzada que permite identificar y cuantificar la presencia de elementos como carbono, hidrógeno y oxígeno en estos azúcares. Esta técnica es especialmente útil para analizar la composición elemental de los azúcares y detectar adulteraciones en productos alimenticios. Mediante el uso de LIBS, es posible asegurar la calidad y autenticidad de los productos alimenticios, facilitando la detección de prácticas fraudulentas.

Fructosa:

La fructosa es un monosacárido con la fórmula molecular $C_6H_{12}O_6$, conocido por su alta dulzura. Su estructura química puede presentarse como un anillo de cinco miembros (furanosa) o de seis miembros (piranosa). Es altamente soluble en agua y tiene un punto de fusión bajo, lo que la hace ideal para la industria alimentaria.

Debido a su intensa dulzura y solubilidad, la fructosa se utiliza comúnmente como sustituto de otros azúcares, lo que puede llevar a la adulteración de productos alimenticios. La detección de la fructosa es crucial para asegurar la calidad y autenticidad de los alimentos.

Glucosa:

La glucosa es un monosacárido con la fórmula molecular $C_6H_{12}O_6$, fundamental en la bioquímica humana como fuente primaria de energía. Su estructura química se presenta

principalmente en forma de un anillo de seis miembros (piranosa). Es un componente clave en numerosos alimentos y tiene un papel esencial en varios procesos metabólicos.

La glucosa es altamente soluble en agua y tiene un punto de fusión moderado. Estas propiedades físicas la hacen comúnmente utilizada en productos alimenticios. Sin embargo, al igual que la fructosa, su uso indebido puede llevar a la adulteración de alimentos, donde se sustituye por otros azúcares para reducir costos de producción.

Maltosa:

La maltosa es un disacárido compuesto por dos unidades de glucosa, con la fórmula molecular $C_{12}H_{22}O_{11}$. Es menos dulce que la fructosa y la glucosa, y se encuentra comúnmente en productos de malta y ciertos alimentos procesados. Su estructura química se caracteriza por un enlace glicosídico entre dos moléculas de glucosa.

En términos de propiedades físicas, la maltosa es soluble en agua y tiene un punto de fusión relativamente bajo. Estas propiedades la hacen útil como endulzante y aditivo en diversas aplicaciones alimentarias. Sin embargo, la maltosa también puede ser utilizada para adulterar productos alimenticios, reemplazando otros azúcares más costosos.

Para identificar y clasificar adulteraciones en alimentos a través de LIBS, es crucial entender las respuestas espectrales de los elementos constituyentes. Los elementos C, H, y O son fundamentales en la composición de azúcares como fructosa, glucosa y maltosa, tal como se puede ver en sus ecuaciones químicas respectivas. A continuación, se presentan los espectros obtenidos para estos elementos:

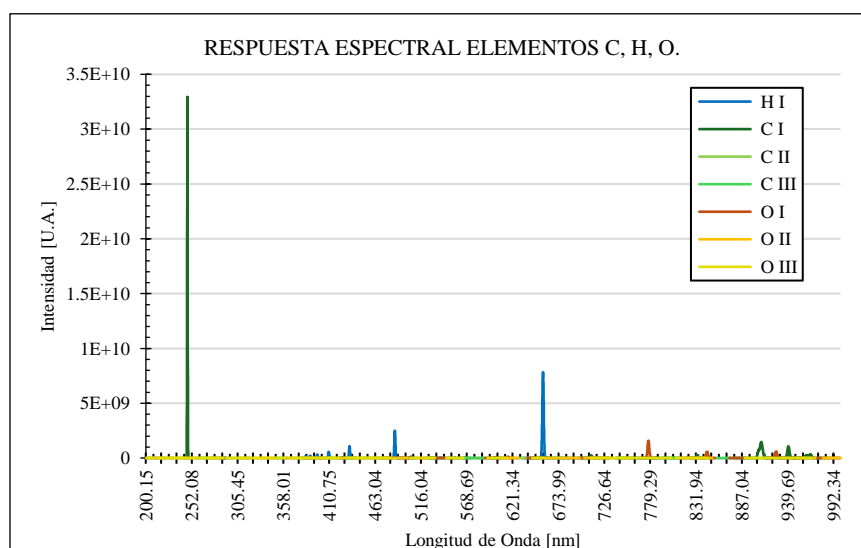


Figura 3.1.3: Representación espectral elementos químicos (C – H – O), fuente Instituto de Nacional de Estándares y Tecnología (NIST, Estados Unidos).

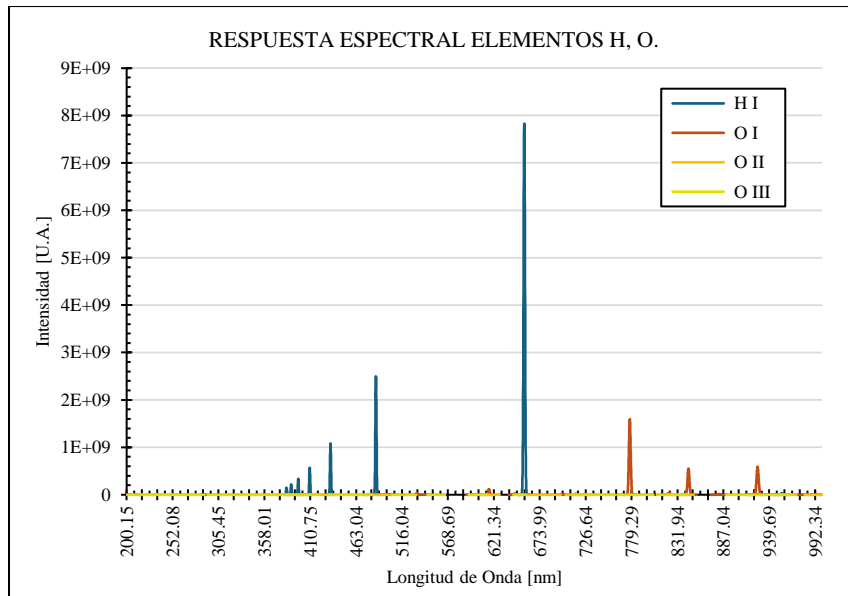


Figura 3.1.4: Representación espectral elementos químicos (H – O), fuente Instituto de Nacional de Estándares y Tecnología (NIST, Estados Unidos).

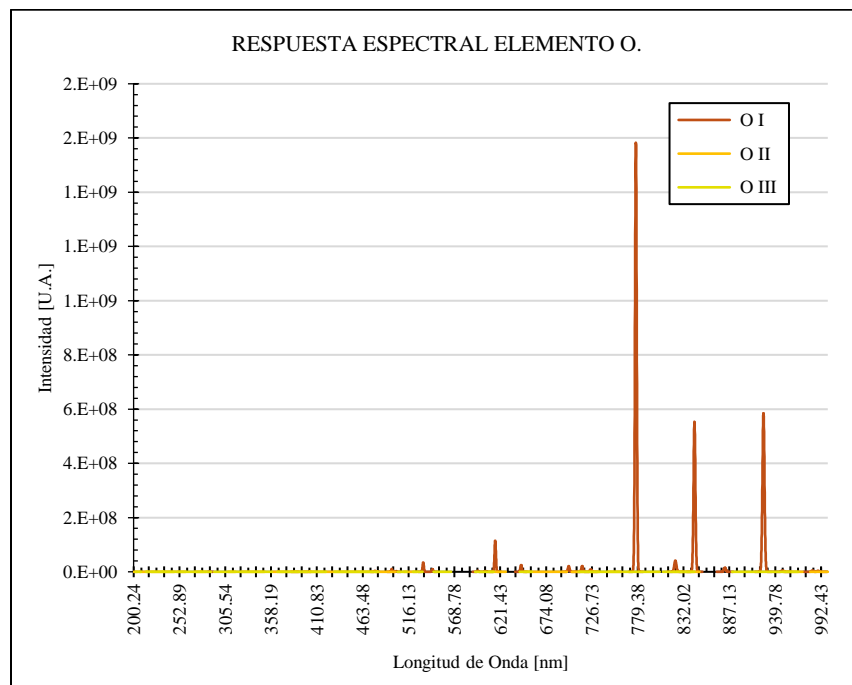


Figura 3.1.5: Representación espectral elemento químico (O), fuente Instituto de Nacional de Estándares y Tecnología (NIST, Estados Unidos).

De los gráficos anteriores, se puede desprender que uno de los elementos con mayor presencia para el análisis de adulterantes, correspondería a Carbono, luego Hidrógeno y ya en menor proporción Oxígeno. Estos grupos pueden definir y clasificar el universo de muestras teóricamente, aunque en la práctica es posible encontrar más elementos que puedan aportar distintas o más características a las clases a través de rastros minerales, depositados en la miel por la trayectoria seguida por las abejas en su recolección según características geográficas.

Así, la fructosa, glucosa y maltosa que son azúcares compuestos principalmente por los elementos carbono, hidrógeno y oxígeno, pueden desempeñar un rol clasificador. Estos azúcares pueden ser identificados y analizados mediante LIBS, técnica que se basa en la detección de altas intensidades espectrales características, correspondientes a sus elementos constituyentes.

LIBS permite la identificación de elementos presentes en una muestra mediante la ablación inducida por una alta concentración de energía láser sobre un material específico, lo que provoca la emisión de luz en longitudes de onda específicas. Para la fructosa, glucosa y maltosa, las altas intensidades de carbono se observan típicamente en 247.86 [nm] y 283.71 [nm], donde la mayor intensidad de hidrógeno se encuentra en 656.28 [nm], y las intensidades máximas de oxígeno aparecen en 777.19 [nm] y 844.68 [nm].

Estas señales son fundamentales para la identificación de los azúcares mencionados en muestras de miel. De esta forma, LIBS proporciona espectros dependientes de longitudes de onda a escala nano o micrométrica, que pueden ser comparados con referencias obtenidas de bases de datos de investigación, como las proporcionadas por el NIST, sección LIBS. Esto permite la validación de la presencia de estos azúcares y la detección de posibles adulteraciones en productos apícolas.

3.2. VISIÓN ARTIFICIAL: PRINCIPIOS Y APLICACIONES

En ingeniería, la visión artificial se establece como un subcampo dinámico y en constante evolución, situado en la intersección de la inteligencia artificial, la óptica y el procesamiento digital de señales. Su propósito fundamental es facultar a las máquinas con la capacidad de interpretar y analizar datos provenientes de la estimulación de sensores ópticos, tales como CCD o CMOS.

Este campo se sustenta en el desarrollo y la implementación de algoritmos avanzados ciencias de estadística cuantitativa para análisis y/o aprendizaje automático, registrando, procesando y analizando imágenes y secuencias de video, con el objetivo de extraer información significativa y utilizable.

Un aspecto crítico de la visión artificial en el contexto de la melisopalinología es su capacidad para contar y clasificar el polen presente en las muestras de miel derivado de su capacidad de detección de contornos de granos, lo que facilita la identificación de las especies de plantas y sus orígenes botánicos. Estas transiciones se capturan mediante la detección de bordes, una técnica que resalta los cambios de intensidad en los píxeles cercanos a los bordes de los objetos. La detección de bordes puede verse como el cálculo de la primera o segunda derivada de la intensidad de la imagen, donde la primera derivada identifica cambios abruptos, y la segunda derivada señala los picos en esos cambios.

Este análisis se apoya en la potencia de la visión artificial para discernir patrones y características específicas del polen, que son indicativos de su procedencia y tipo, permitiendo así trazar un perfil detallado del origen botánico de la miel.

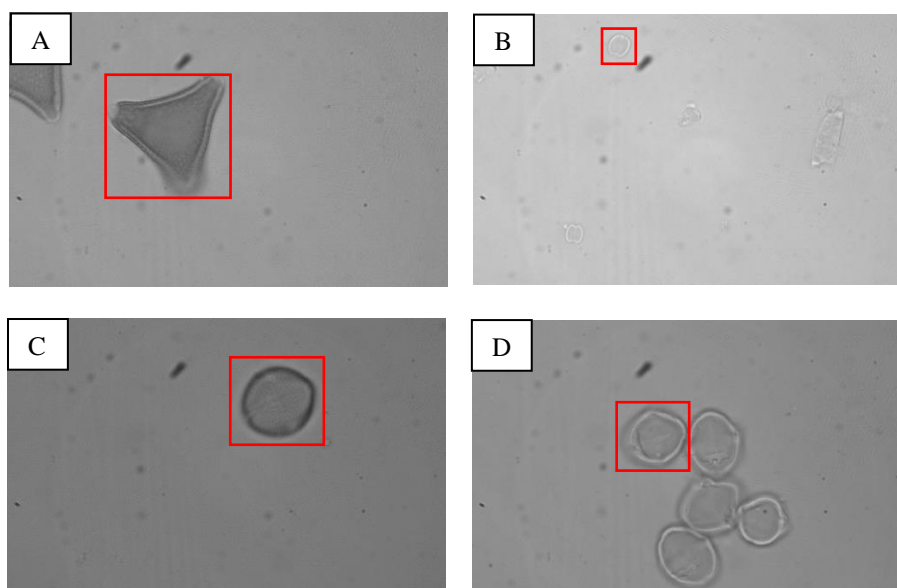


Figura 3.2.1: Descripción imágenes superiores (granos de polen en regiones de interés o “Bounding Boxes ROI”): (A) *Gevuina Avellana*, (B) *Eucryphia Glutinosa*, (C) *Cissus striata*, (D) Grupo de granos *Quillaja Saponaria*.

Tanto desde la adquisición de los datos, como hasta el final del proceso, se debe tener en consideración variables como la longitud de onda a utilizar como objetivo en un estudio, preparando así las herramientas ópticas y filtros pasa – banda (BPF) para intervalos espectrales acotados, así enmarcar el desempeño a que corresponda con una mejor descripción de bordes para los objetos de estudio en cuestión.

Una vez que se cuenta con la información digitalizada, dichas imágenes pueden ser interpretadas como matrices de datos, que generalmente se normalizan en valores de 8/16 bits de profundidad en escala de grises, con el fin de poder trabajar con una base de datos compacta sin perder, en lo posible, demasiada información que pueda ser utilizada como entrada al momento de entrenar un modelo. Esto es particularmente relevante para los clasificadores de polen, donde la precisión en la identificación y cuantificación de los granos de polen es esencial para realizar inferencias correctas sobre la calidad y el origen de la miel.

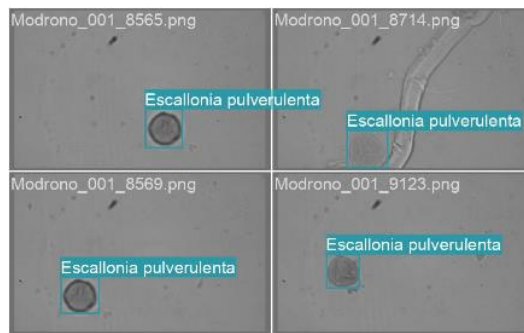


Figura 3.2.2: Identificación de objetos, ejemplo de clase “Escallonia pulverulenta”.

Entre las técnicas más utilizadas para la detección de bordes están el filtro de Sobel, que calcula las derivadas en las direcciones horizontal y vertical, y el algoritmo de Canny, que proporciona una detección de bordes más robusta mediante un proceso de varios pasos que incluye suavizado, cálculo de gradientes, y supresión de no-máximos. Estos métodos se enfocan en destacar las áreas de cambio rápido en la intensidad de los píxeles, que corresponden a los contornos de los objetos dentro de la imagen.

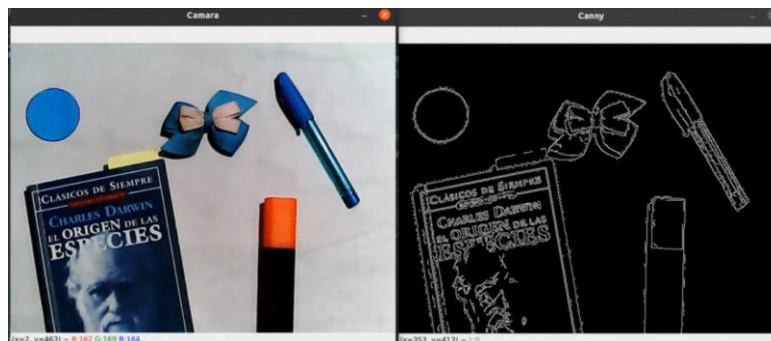


Figura 3.2.3: Ejemplificación de algoritmos detectores de bordes en objetos.

de eficiencia. Estos módulos operan en varias capas o "stages", procesando la imagen en resoluciones decrecientes, lo que permite a la red identificar patrones y contornos en diferentes escalas.

2. Neck:

Está basado en la estructura YOLOv8PAFPN (Path Aggregation Feature Pyramid Network), es responsable de fusionar características extraídas en diferentes escalas. Esto es esencial para que la red pueda detectar objetos de diversos tamaños dentro de una imagen. En esta sección, las operaciones como **Concat**, **Upsample**, y **Downsample** desempeñan un rol esencial al combinar la información proveniente de diferentes niveles del "Backbone", permitiendo una mejor comprensión del contexto general de la imagen y la preservación de detalles finos. Esta combinación de información a múltiples escalas es clave para una detección precisa y robusta.

3. Head:

Es la sección final de la red, donde se realizan las inferencias finales de los "bounding boxes" (rectángulos que delimitantes de objetos) y las clases de los objetos detectados. A diferencia de modelos anteriores que utilizaban "anclas" (anchor boxes) predefinidas para ajustar la predicción de los bounding boxes, YOLOv8 adopta un enfoque "**anchor-free**". Esto significa que el modelo predice directamente la ubicación de los objetos en lugar de ajustar estas cajas predefinidas, simplificando el proceso y reduciendo la carga computacional. Los módulos dentro del Head, como **Conv2d**, están diseñados para realizar estas predicciones de manera eficiente, mientras que los módulos de pérdida (**Loss**), como **Bbox Loss** y **Cls Loss**, optimizan el modelo durante el entrenamiento para mejorar su precisión.

3.3. ALGORITMOS DE APRENDIZAJE AUTOMÁTICO

En el ámbito académico, específicamente en las áreas de la ingeniería y la ciencia de la computación, los algoritmos de aprendizaje automático se conceptualizan como métodos computacionales avanzados diseñados para aprender de los datos y mejorar el rendimiento en tareas específicas. Estos algoritmos son un pilar fundamental en el campo del análisis de datos, proporcionando la base para sistemas que pueden realizar desde simples tareas de clasificación hasta complejas predicciones y toma de decisiones. La investigación y desarrollo en aprendizaje automático integran principios de estadística, matemáticas, inteligencia artificial y teoría de la información, buscando crear modelos que puedan adaptarse y mejorar a medida que se exponen a más datos. Esta capacidad para aprender y adaptarse es lo que diferencia a estos algoritmos de los métodos tradicionales de programación, marcando una evolución significativa en la forma en que las máquinas procesan la información y toman decisiones.

En general, los algoritmos se dividen según la entrada de datos y etiquetado como:

- I. **Supervisados:** Estos algoritmos se utilizan cuando los datos de entrenamiento están etiquetados. Significa que cada entrada de datos en el conjunto de entrenamiento está asociada con la respuesta correcta. El algoritmo aprende de estos datos y luego puede hacer predicciones sobre datos nuevos basándose en este aprendizaje. Ejemplos comunes incluyen la regresión lineal y la clasificación.
- II. **No Supervisados:** A diferencia de los supervisados, estos algoritmos se aplican a datos que no están etiquetados. El objetivo aquí es explorar la estructura de los datos para encontrar algún tipo de organización o patrón. Los algoritmos no supervisados son útiles para la segmentación de grupos ("*clustering*") o la reducción de dimensionalidad. Ejemplos populares son "*K-means*" y el análisis de componentes principales ("*PCA*").
- III. **Semi-Supervisados:** Estos algoritmos representan un punto medio entre los métodos supervisados y no supervisados. Se utilizan cuando se tiene una gran cantidad de datos no etiquetados y una pequeña cantidad de datos etiquetados. El algoritmo aprende tanto de la estructura inherente de los datos no etiquetados como de la información proporcionada por los datos etiquetados. Esto puede mejorar significativamente el rendimiento del aprendizaje en comparación con los métodos puramente supervisados o no supervisados.

3.3.1. Análisis de Componentes Principales (*Principal Component Analysis, PCA*)

Es una técnica estadística utilizada para simplificar la complejidad en conjuntos de datos de alta dimensión. Su objetivo principal es reducir el número de variables en un conjunto de datos, manteniendo al mismo tiempo la mayor cantidad posible de información original. Esto se logra identificando las direcciones a lo largo de las cuales varían los datos.

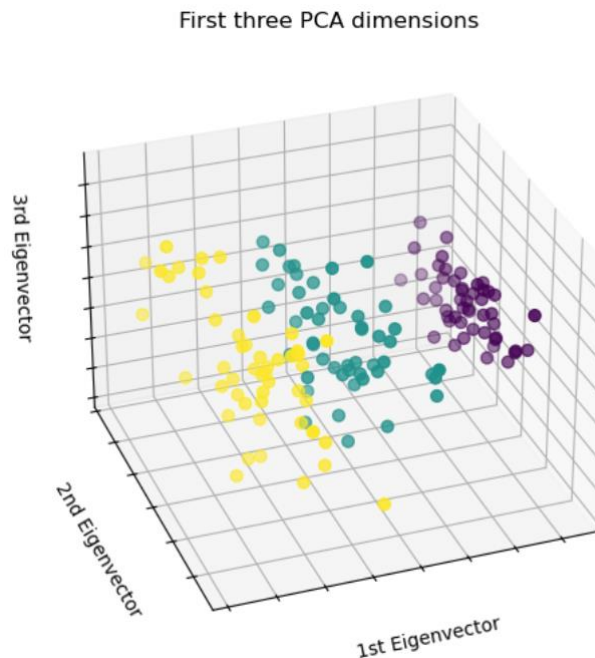


Figura 3.3.1: Representación visual de 3 planos de componentes principales de n -dimensiones, para 3 grupos de datos distintos, referencia "Iris Dataset, SciKit Learn ToolKit".

En PCA, los datos originales se transforman en un nuevo conjunto de variables, las componentes principales, que son combinaciones lineales de las variables originales. Estas componentes se ordenan de forma que la primera captura la mayor variabilidad, y cada componente subsiguiente, siendo ortogonal a las anteriores, explica la mayor parte de la varianza restante.

Uno de los puntos clave de PCA reside en el uso de autovectores y autovalores derivados de la matriz de covarianza de los datos. Los autovalores representan la varianza que cada componente principal retiene, y los autovectores definen las direcciones de estas componentes.

Matemáticamente, PCA se puede realizar mediante la descomposición en valores singulares (SVD), una técnica que factoriza la matriz de datos en componentes que facilitan la extracción de las direcciones de máxima varianza (autovectores) y la cantidad de varianza capturada (autovalores).

3.3.2. Análisis Discriminante Lineal (*Linear Discriminant Analysis, LDA*)

Es una técnica estadística y un método de clasificación utilizado en aprendizaje automático y estadística. La meta principal de LDA es encontrar una combinación lineal de características que caractericen o separen dos o más clases de objetos o eventos. La idea es proyectar los datos en un espacio de menor dimensión con buena capacidad de clasificación, preservando al máximo la separabilidad entre las distintas clases.

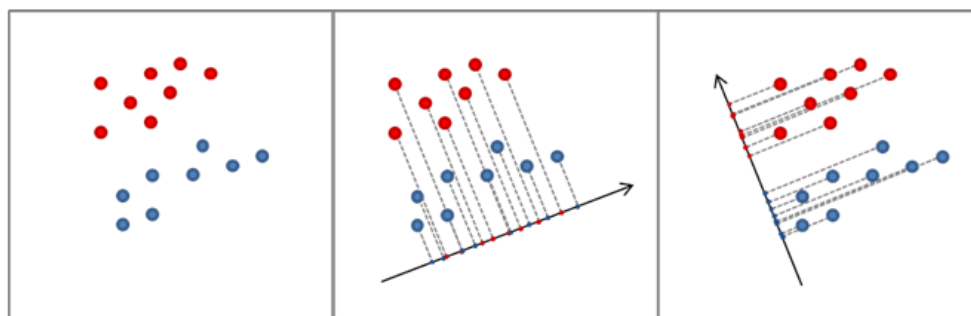


Figura 3.3.2: Ejemplificación de aplicar a un conjunto de datos ya procesado con PCA la maximización entre grupos por LDA.

En LDA, se busca maximizar la distancia entre las medias de las diferentes clases, mientras se minimiza la variación (o dispersión) dentro de cada clase. Esto se logra mediante la creación de un criterio, conocido como el criterio de Fisher, que busca maximizar la razón entre la varianza entre clases y la varianza dentro de las clases.

Matemáticamente, LDA implica calcular los vectores medios para cada clase, la matriz de dispersión entre clases y la matriz de dispersión dentro de las clases. Luego se calculan los vectores que maximizan el criterio de Fisher. Estos vectores forman un nuevo espacio en el que los datos pueden ser proyectados.

3.3.3. Descenso de Gradientes Estocástico (*Stochastic Gradient Descent, SGD*)

Es un algoritmo de optimización fundamental en aprendizaje automático para minimizar una función objetivo, típicamente una función de pérdida, mediante actualizaciones iterativas de los parámetros del modelo. A diferencia del descenso de gradiente convencional, que utiliza el conjunto de datos completo para el cálculo del gradiente, el SGD emplea un subconjunto aleatorio en cada iteración, lo que resulta en una aproximación estocástica del gradiente.

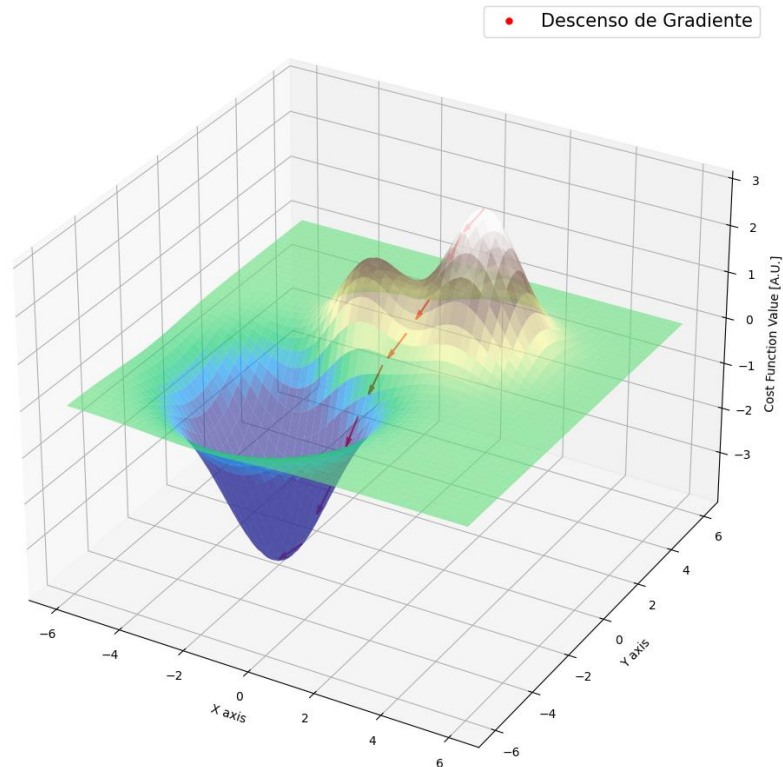


Figura 3.3.3: Representación visual de máximos y mínimos, globales y locales, en un plano de características según el método SGD.

Esta aproximación introduce un elemento de ruido en las actualizaciones de los parámetros, lo que puede facilitar la evasión de mínimos locales en espacios de parámetros complejos. La eficiencia del SGD es notablemente superior en grandes volúmenes de datos, ofreciendo ventajas en términos de tiempo de cómputo y uso de memoria.

La convergencia del SGD depende críticamente de la elección de la tasa de aprendizaje y otros hiperparámetros. Variantes como SGD con impulso y algoritmos adaptativos (p. ej., Adam, RMSprop) han sido desarrolladas para optimizar la trayectoria de convergencia, equilibrando la magnitud y la dirección de las actualizaciones de parámetros.

3.3.4. Bosque Aleatorio (*Random Forest, RF*)

Es una técnica avanzada de aprendizaje automático, clasificada como un algoritmo de ensamblaje. Su propósito principal es mejorar la precisión predictiva y el control del sobreajuste a través de la combinación de múltiples árboles de decisión durante el proceso de entrenamiento. Cada árbol en el bosque es construido usando una muestra aleatoria de los datos de entrenamiento (conocido como "*bootstrap sample*") y una selección aleatoria de las características en cada división, lo que garantiza la diversidad dentro del modelo.

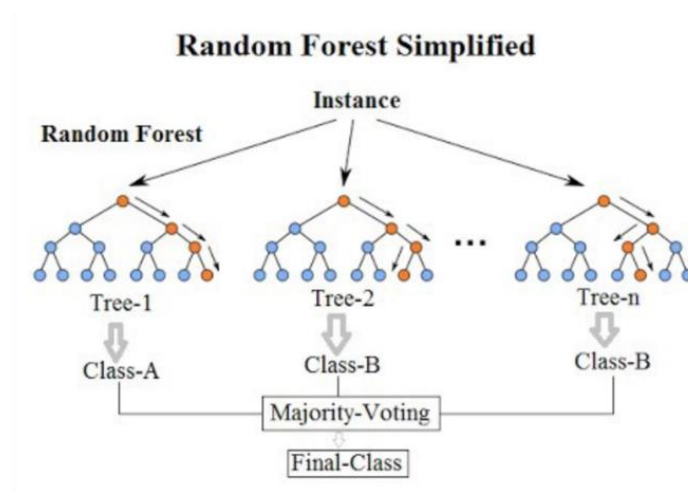


Figura 3.3.4: Representación visual de decisiones tomadas por cada rama de un árbol según el método Random Forest.

El algoritmo "*Random Forest*" se destaca por su robustez y capacidad de manejar grandes conjuntos de datos con una dimensionalidad elevada, ofreciendo una interpretación intuitiva de la importancia de las variables. La eficiencia de este método proviene de su naturaleza no paramétrica y la habilidad de realizar tanto tareas de clasificación como de regresión. Además, es notablemente resistente al sobreajuste, incluso cuando se utiliza con datos de alta dimensionalidad.

En la práctica, "*Random Forest*" opera mediante la agregación de los resultados de múltiples árboles (votación para clasificación o promedio para regresión) para determinar la salida final.

3.3.5. Máquina de Soportes Vectoriales (*Support Vector Machine, SVM*)

Es un método de aprendizaje supervisado ampliamente utilizado en el campo del aprendizaje automático y la estadística. La principal finalidad de SVM es encontrar un hiperplano óptimo que clasifique distintas clases de datos con un margen máximo de separación.

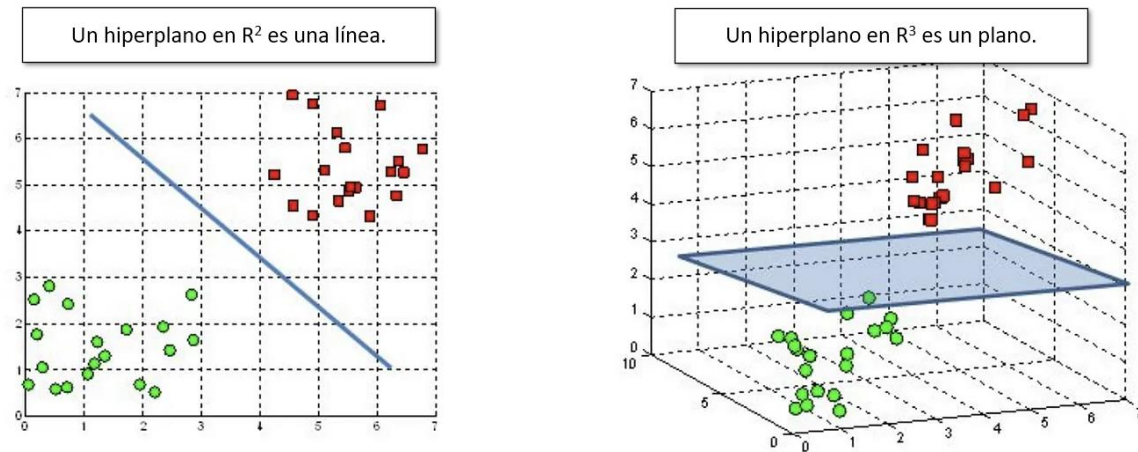


Figura 3.3.5: SVM establece planos entre 2 grupos o clústeres de información, utilizando vectores como base para formar una mantisa de evaluación.

En el núcleo de SVM está el concepto de maximizar el margen entre los conjuntos de datos de diferentes clases, lo cual es crucial para mejorar la capacidad de generalización del modelo. Este margen se define como la distancia entre el hiperplano de decisión y los puntos más cercanos de cada clase, conocidos como vectores de soporte. Matemáticamente, SVM resuelve un problema de optimización para encontrar el hiperplano que maximiza este margen.

SVM es especialmente eficaz en espacios de alta dimensión y en situaciones donde la separación entre las clases no es claramente definible. Además, mediante el uso de funciones kernel, SVM es capaz de resolver problemas no lineales, mapeando los datos a un espacio de mayor dimensión donde pueden ser separados linealmente.

3.4. ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PROFUNDO

El aprendizaje profundo, un subcampo especializado dentro del aprendizaje automático, se basa en el empleo de redes neuronales artificiales para el análisis y modelado de datos complejos. Inspiradas en la estructura de las redes neuronales biológicas, estas redes compuestas por múltiples capas de nodos, conocidas como neuronas artificiales, ejecutan transformaciones secuenciales de datos. Un aspecto crítico en estas transformaciones es la aplicación de funciones de activación, tales como la ReLU (Rectified Linear Unit) o la función sigmoide.

En el entrenamiento de estas redes, conceptos como la retropropagación y el algoritmo de descenso de gradiente son fundamentales. Estos mecanismos ajustan los pesos en las conexiones neuronales para minimizar el error en la salida de la red, lo cual es esencial para el aprendizaje efectivo del modelo.

Dentro de los procesos llevados a cabo por estos algoritmos, se incluye:

1. Funciones de Activación en Redes Neuronales:

Las redes neuronales artificiales utilizan funciones de activación para introducir no linealidades, permitiendo a la red aprender y modelar relaciones complejas en los datos.

Ejemplos comunes incluyen la ReLU, que ofrece una respuesta lineal positiva, y la función sigmoide, que mapea las entradas a un rango entre 0 y 1.

2. Técnicas de Regularización:

El “dropout” es una técnica de regularización en la que, durante el entrenamiento, algunos nodos se "desactivan" aleatoriamente. Esto evita que la red se vuelva demasiado dependiente de cualquier conjunto de nodos, reduciendo el sobreajuste, un concepto muy fuerte para aplicar en soluciones reales.

3. Normalización de Lotes:

Ajusta la distribución de las entradas a cada capa durante el entrenamiento. Esto ayuda a combatir el problema del desplazamiento de la media interna, mejorando la estabilidad y eficiencia del entrenamiento.

4. Optimizadores:

Los optimizadores como el Descenso de Gradiente Estocástico (SGD) y Adam son cruciales para ajustar los pesos de la red. Estos algoritmos determinan cómo se actualizan los pesos en respuesta al gradiente del error de la red.

Algunos ejemplos de arquitecturas, sin entrar en detalles sobre ellas, de algoritmos de aprendizaje profundo:

1. YOLO (You Only Look Once):

Especializada en la detección de objetos, YOLO divide la imagen en una cuadrícula y predice clases y ubicaciones de objetos en una sola pasada, lo que mejora la velocidad de detección.

2. ResNet (Redes Residuales):

Incorpora conexiones residuales que permiten a las señales de datos saltar capas, facilitando el entrenamiento de redes muy profundas y mitigando el problema del gradiente desvaneciente.

3. AlexNet:

Una red convolucional profunda que utiliza dropout y normalización de respuesta local. Fue pionera en el uso de GPU para el entrenamiento de redes profundas.

4. VGGNet:

Se enfoca en incrementar la profundidad de la red utilizando capas convolucionales pequeñas, mejorando la precisión en tareas de clasificación de imágenes.

3.5. MÉTRICAS DE EVALUACIÓN

En esta sección, se aborda como se evaluarán las distancias entre las distintas clases con respecto a los patrones de referencia. Esto vale decir que se tomará como referencia la clase perteneciente al patrón de miel pura y de adulterante como extremos, asumiendo que las muestras aparecerán descendentemente con respecto a su contenido de miel.

Para cumplir con dicho objetivo, se utilizarán centroides para posicionar y simplificar la aplicación de las métricas correspondientes a distancias entre clústeres.

Centroides:

Representan el punto central en el espacio de un clúster, que puede definirse como el valor promedio de todos los puntos pertenecientes a una dimensión o componente. Es una medida central que se utiliza para describir la ubicación promedio de todos los puntos en un clúster y así tener una referencia numérica de su posición con respecto a otros clústeres.

Matemáticamente se representa como:

$$\text{Centroide} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m)$$

Donde:

\bar{x}_m : Representa promedios de las coordenadas de todos los puntos en cada una de las “m” – dimensiones.

Se utilizarán los 3 primeros componentes o dimensiones para representar los gráficos, luego de haber procesado los datos por el análisis discriminante lineal (LDA), estableciendo así un punto central para cada clúster o clase.

$$\text{Centroide} = (\bar{x}_{ij}, \bar{x}_{ij}, \bar{x}_{ij})$$

Expresándose desde la siguiente ecuación:

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$$

Ecuación 3.5.1: Determinación centroide para cada clúster o clase.

Donde:

- \bar{x}_{ij} : Representa la información de todos los puntos “i”, de cada clase o dimensión “j”.
- \bar{x}_j : Es el promedio de todos los puntos para “j”.

3.5.1. Métricas aplicadas a LIBS.

Las distancias son métricas fundamentales en el análisis de grupos (clustering), utilizadas para medir la similitud o disimilitud entre objetos en un espacio de características. Las más comunes son la Euclidiana, Manhattan y Chebyshev, cada una con propiedades únicas que las hacen adecuadas para diferentes tipos de datos y aplicaciones.

I. Distancia Euclidiana:

Es la más comúnmente utilizada, definida como la raíz cuadrada de la suma de las diferencias al cuadrado entre los puntos en cada dimensión. Representa la longitud del segmento más corto o la línea recta entre dos puntos en un espacio euclidiano.

$$D_{Euclidiana}(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Ecuación 3.5.2: Distancia Euclidiana.

II. Distancia Manhattan:

También conocida como distancia del taxista o L1, es la suma de las diferencias absolutas de las coordenadas. Es útil en cuadrículas urbanas donde el movimiento solo es posible a lo largo de ejes rectos, como las calles y avenidas que forman un patrón de cuadrícula.

$$D_{Euclidiana}(p, q) = \sum_{i=1}^n |p_i - q_i|$$

Ecuación 3.5.3: Distancia Manhattan.

III. Distancia Chebyshev:

Esta métrica, conocida también como distancia del ajedrez o L_∞ , es el máximo de las diferencias absolutas entre coordenadas de los puntos. Es relevante en aplicaciones donde el movimiento puede ser en cualquier dirección, pero solo se permite un paso a la vez, como el movimiento del rey en el ajedrez.

$$D_{Euclidiana}(p, q) = \max_i |p_i - q_i|$$

Ecuación 3.5.4: Distancia Chebyshev.

Donde:

- $p = (p_1, p_2, p_3, \dots, p_n)$: Es un punto en el espacio de n-dimensiones, donde cada p_i representa la coordenada del punto p en la i-ésima dimensión.
- $q = (q_1, q_2, q_3, \dots, q_n)$: Es un punto en el espacio de n-dimensiones, donde cada q_i representa la coordenada del punto q en la i-ésima dimensión.

3.5.2. Métricas aplicadas a Visión Artificial

Las métricas de evaluación son formas cuantitativas de medir el rendimiento y la calidad de un modelo o algoritmo. En el campo de la visión artificial, que trata sobre cómo las computadoras pueden ser entrenadas para interpretar imágenes digitales y/o video de la misma manera que un humano. Estas métricas son particularmente importantes, asistiendo a las ciencias de la ingeniería a entender cómo está funcionando un modelo y donde puede necesitar mejoras.

I. Recall (Exhaustividad o Sensibilidad)

Evalúa la habilidad de un modelo para identificar todas las instancias relevantes de una categoría dada, calculada como la razón entre los Verdaderos Positivos (TP) y la suma de TP y Falsos Negativos (FN). Es fundamental en contextos donde la omisión de casos positivos es inaceptable.

$$Recall = \frac{TP}{TP + FN}$$

Ecuación 3.5.5: Sensibilidad o "Recall".

II. Precisión

La precisión mide la proporción de identificaciones positivas que fueron en realidad correctas. En otras palabras, si el modelo está identificando objetos en imágenes, la precisión será la proporción de objetos que el modelo identificó correctamente de todos los que identificó. Un modelo con una precisión alta no produce muchos falsos positivos.

$$Accuracy = \frac{TP}{TP + FP}$$

Ecuación 3.5.6: Precisión o "Accuracy".

III. F1 – Score

Es una métrica de rendimiento que armoniza la precisión y el recall mediante su media armónica, proporcionando una evaluación equilibrada de un modelo clasificador en presencia de distribuciones de clases desiguales. Valora tanto la capacidad del modelo para identificar correctamente las instancias positivas (recall) como para clasificar solo las instancias relevantes como positivas (precisión).

$$F1 = 2 * \frac{Accuracy * Recall}{Accuracy + Recall}$$

Ecuación 3.5.7: Indicador “F1 – Score”.

IV. Mean Average Precision (mAP50)

Es utilizada en la detección de objetos. Calcula el promedio de la precisión promedio de detecciones en diferentes clases, aplicando un indicador umbral de Intersección (de 50% para mAP50) sobre Unión (IoU) entre las áreas de los cuadros delimitadores predichos y los verdaderos. Esta métrica sintetiza tanto la precisión de la localización de las cajas delimitadoras como la exactitud en la clasificación de los objetos detectados.

$$mAP50 = \frac{1}{N} \sum_{i=1}^N AP_i$$

Ecuación 3.5.8: Indicador “Mean Average Precision mAP50”.

Donde:

- N: número de clases.
- AP_i : Precisión media de la clase “ i ”.
- FP : Falsos positivos
- FN : Falsos negativos.
- TP : Verdaderos positivos.

3.6. RESUMEN DEL CAPÍTULO

Desde la perspectiva de la melisopalinología, se destaca la importancia de entender la compleja interacción entre abejas y polen en la producción de la miel. Considerando técnicas actuales al aplicar métodos avanzados de ingeniería para garantizar la autenticidad y calidad de la miel, se utilizan técnicas como la espectroscopía por inducción con láser (LIBS) y la visión artificial aplicada a microscopía. Estas tecnologías permiten un análisis detallado y mínimamente destructivo de las muestras de miel, permitiendo la identificación precisa de su procedencia y calidad, lo cual es clave para combatir la adulteración de la miel y asegurar su pureza.

Se expone cómo los algoritmos de aprendizaje automático y las técnicas de visión artificial están aportando en gran medida a los métodos de análisis de miel, describiendo el uso de bases de datos de imágenes de polen y modelos de aprendizaje profundo, que actualizan las prácticas de clasificación y análisis. Estos avances no solo mejoran la precisión en la identificación de tipos de polen, sino que también establecen nuevos estándares de velocidad y eficiencia en los tiempos de análisis, lo que permite a productores y reguladores tomar decisiones más informadas y efectivas.

Esta discusión se extiende a la relevancia de la legislación, en octubre 2022 a la publicación de la Ley Apícola 21.489, enfatizándose en el fomento a la adopción de tecnologías avanzadas y métodos analíticos rigurosos en la industria apícola nacional. Este marco legal eleva los estándares de producción y análisis de la miel, estimulando tanto la innovación tecnológica como la investigación aplicada, recalcando así la importancia del trabajo colaborativo entre científicos, legisladores y apicultores para asegurar que las prácticas de producción y análisis cumplan con los estándares de calidad y autenticidad, protegiendo así al consumidor y sosteniendo la viabilidad del mercado.

Así, se busca extender una visión comprensiva y multidisciplinaria que integre la ingeniería, tecnología y regulación con el fin de fortalecer y proteger la industria de la miel nacional. Los métodos y tecnologías discutidos son fundamentales para asegurar la calidad y autenticidad de la miel y sirven como modelo para otras áreas de investigación y desarrollo en seguridad alimentaria y tecnología agroindustrial. Este enfoque es crucial para enfrentar los desafíos futuros en la industria alimentaria, promoviendo no solo innovación sino también una colaboración científico – académica que permita mejorar la calidad de vida y la seguridad de los consumidores.

4. DISEÑO EXPERIMENTAL Y ADQUISICIÓN DE DATOS

La espectroscopía LIBS ofrece un análisis detallado de la composición elemental de la miel a través de sus huellas espectrales específicas, las cuales pueden evidenciar la presencia de componentes atípicos de la miel pura (tales como en azúcares de fructosa, glucosa y maltosa respecto a contenido de carbono, hidrógeno y oxígeno), sugiriendo así una posible adulteración. Por otro lado, el uso de la visión artificial, apoyada en cámaras digitales de visión y algoritmos avanzados, permite el conteo y la clasificación precisa del polen contenido en la miel, lo cual es esencial para determinar el origen botánico y asegurar la autenticidad del producto. Desviaciones en las proporciones habituales o la presencia de tipos de polen no característicos pueden ser indicativos de adulteración o de mezclas con mieles de distintas procedencias.

En el ámbito de la melisopalinología, se utilizan estándares específicos que actúan como referencia para los resultados obtenidos mediante la visión artificial. Por ejemplo, discrepancias entre la distribución y tipos de polen detectados en una muestra y los perfiles esperados según los estándares regionales o específicos de un tipo de miel pueden señalar adulteraciones.

La comparación de los datos obtenidos mediante LIBS y visión artificial contra los estándares de melisopalinología permite una evaluación integrada de la autenticidad de la miel. La presencia de huellas espectrales con características similares a azúcares puros, junto con una correspondencia menor en cantidad polínica detectada en el análisis de muestras de miel, apuntan a la posibilidad de adulteración. Posteriormente, al cuantificar las desviaciones de las huellas espectrales específicas y los resultados del conteo de polen respecto a muestras de control conocidas, es posible determinar el grado de adulteración de manera más precisa.

La validación y calibración de estos métodos son cruciales y deben realizarse comparando los resultados con muestras de miel pura y, de manera controlada, con muestras adulteradas. Este proceso es vital para el ajuste fino de los métodos y la definición de umbrales exactos que faciliten la detección confiable de adulteraciones en la miel.

4.1. METODOLOGÍA PARA MUESTRAS Y BASE DE DATOS.

De los objetivos específicos [OE-1], se describen las metodologías asociadas a la preparación de muestras de miel (Guindo Santo y Quillay) mezcladas con diferentes jarabes de uso alimenticio obtenidos mediante hidrólisis del almidón de maíz (Glucosa, Fructosa Alto Maltosa) en distintos deciles, elaborando adulteraciones desde el 0% al 100%, siendo objetivo para este estudio tramos de los contenidos con menor presencia de miel (40% y menor).

4.1.1. PREPARACIÓN DE MUESTRAS.

La metodología de preparación de muestras se basa en procedimientos indicados en publicaciones asociadas al estudio de adulteración controlada de mieles y diferentes sistemas de detección [35 – 39].

Procedimiento:

- 1) En recipientes de plástico transparente y con tapa rosca, se mezcla miel con 3 diferentes jarabes (glucosa, alto maltosa y fructosa) por separado.
- 2) Las proporciones de mezcla empleados son desde 10 a 90% en masa, de tal manera que la suma entre miel y jarabe sea 23 g. Cada ensayo se realiza en duplicado (n = 2) para cada miel, correspondientes a Guindo Santo y Quillay.
- 3) Cada muestra, correspondiente a una proporción de mezcla, se agita manualmente con una cuchara plástica desechable a 20-25°C durante 2 minutos como mínimo o hasta lograr la mezcla homogénea de sus componentes detectable visualmente.
- 4) Se preparan muestras de referencia correspondientes, en el caso de los jarabes, a 100 % de Fructosa, Glucosa y Alto Maltosa, y en el caso de las mieles, a dos muestras (duplicado) de 100% Miel de Guindo Santo y 100% de miel de Quillay.
- 5) Cada muestra preparada se mantiene a 22 °C, en ausencia de luz solar, cerrada con la tapa del recipiente y una película transparente a su alrededor para evitar la evaporación y contaminación hasta su análisis mediante imágenes.

La tabla a continuación describe la composición de cada adulteración controlada y sus respectivos controles, que mantienen el patrón de referencia.

Proporción Miel/Masa Total (% m/m)	Cantidad de Miel (g)	Cantidad de Jarabe (g)
0 (control jarabe utilizado)	0	23
10	2,3	20,7
20	4,6	18,4
30	6,9	16,1
40	9,2	13,8
50	11,5	11,5
60	13,8	9,2
70	16,1	6,9
80	18,4	4,6
90	20,7	2,3
100 (control de miel utilizada)	23	0

Tabla 4.1.1: Valores de porcentajes y masas de las mezclas analizadas para dos diferentes tipos de miel, correspondientes a Guindo Santo y Quillay. Masa total de cada muestra: 23 g.



Figura 4.1.1: Identificación Procedimiento de generación de las muestras de mezclas de miel con jarabes en diferentes proporciones en masa.

Los pasos siguientes a la planificación [OE-2], corresponden a la realización de las muestras totales del banco de datos, procedimiento que es evidenciado en la figura 3.1.1. El resultado final de estos pasos genera un total de 29 muestras, las cuales corresponden a:

1. Mieles puras: Quillay y Guindo Santo (2 clases, patrón de medición).
2. Adulterantes puros: Fructosa, Maltosa y Glucosa (3 clases, patrón de medición).
3. Adulteraciones controladas: Desde [0 – 100]%, en deciles (54 clases).

De este universo total, se utilizarán los tramos con menor presencia (menores o iguales a 40%, 24 clases).

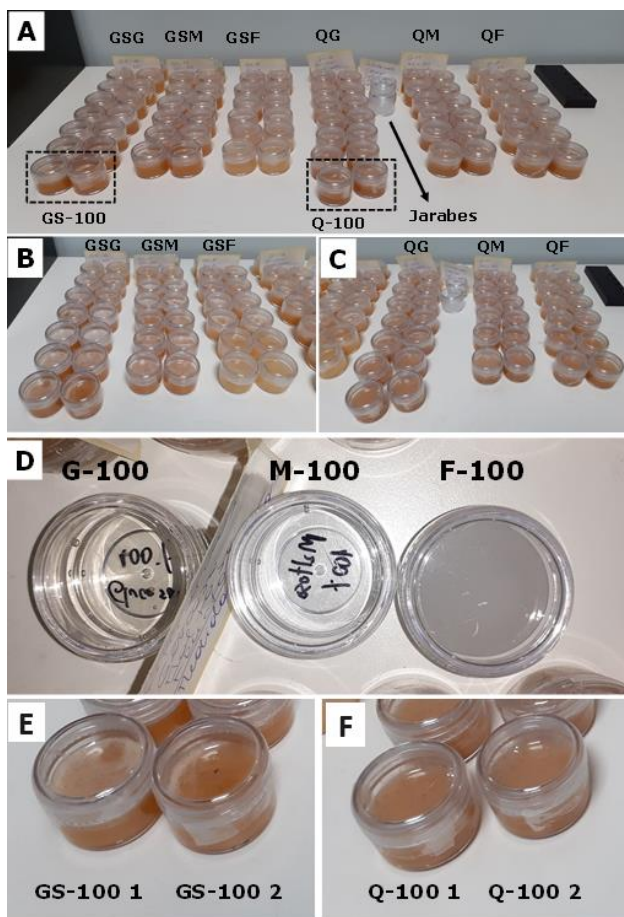


Figura 4.1.2: (A) Vista de todas las muestras. Se indica GSG, GSM y GSF para miel de Guindo Santo mezclada con jarabes de Glucosa, Maltosa y Fructosa, respectivamente y QG, QM y QF para miel de Quillay mezclada con los mismos jarabes. Se indican además las muestras 100 % miel de Guindo Santo (GS-100) y Quillay (Q-100). Acercamiento a las mezclas de miel de **Guindo Santo** (B) y **Quillay** (C) con diferentes jarabes a diferentes proporciones. (D) Acercamiento a patrones de referencia de los jarabes de Glucosa, Maltosa y Fructosa. (E) y (F) Patrones de referencia medidas para Guindo Santo y Quillay.

4.2. ESPECTROSCOPIA LIBS

La realización de caracterización óptica sobre las muestras a estudiar, se debe introducir la configuración del experimento bajo la cual se ha presentado el sistema y sus parámetros adecuados de uso.

Para realizar una serie de mediciones sobre distintas muestras, el sistema puede ser caracterizado, haciendo referencia a la figura 3.2.1, en distintas etapas o fases, donde:

I. Etapa de potencia:

Corresponde a la energización, puesta en marcha y correcta implementación de los dispositivos asociados al sistema láser, extracción de vapores por ablación, interruptores de seguridad (*“security interlocks”*) y aseguramiento de potencia eléctrica mediante respaldo de energía.

II. Etapa Óptica:

En ella ocurren diversos fenómenos ópticos relativos a la forma y trayectoria del haz presente en el camino óptico. Dichos pasos consisten en focalizar el impacto respecto a un punto focal determinados para cada muestra, realizando el mismo proceso, pero inverso, para obtener los datos a través de fibra acoplada al espectrómetro.

III. Portamuestras:

Sección que trabaja sobre el sistema de movimiento tridimensional. Acopla correctamente las muestras sin modificarlas en su composición para ser medidas. Permite la corrección de errores en profundidad o altura, característica importante al momento de realizar una medida.

IV. Control de movimiento tridimensional (*XYZ motion controller*):

Ejecuta los movimientos necesarios para abordar una muestra en una zona o superficie que considere siempre distintos puntos entre medidas. La resolución óptica debe calibrarse según los sistemas de retroalimentación ópticos para establecer una grilla que se ajuste correctamente al punto focal y superficies de medición en la muestra.

Una vez que todo ya ha sido correctamente inicializado, es posible considerar los pasos correspondientes a la adquisición de la información, la cual considera:

V. Recepción de la información:

La información llega al espectrómetro a través de un conversor análogo digital de 14 bits, por lo que se debe ajustar el límite máximo de valores arbitrarios a dicha escala, iterando hasta el

punto de obtener parámetros de tiempo de desfase para producir la ablación, potencia máxima a utilizar por el láser, ventana o tiempo de integración para el sensor, tiempo de emisión para el láser, entre otros. Los nombrados son aquellos más importantes.

VI. Preprocesamiento:

Corresponde a la selección óptima relativa a la cantidad máxima de mediciones a utilizar. Se pueden utilizar distintos enfoques, donde se puede (1) utilizar toda la información haciendo uso extensivo de la memoria computacional, (2) selección de promedios y descartes de disparos para corrección de medidas sobreestimadas y (3) utilizar directamente filtros sobre la información, reduciendo las dimensiones y/o mediciones de salida con longitudes objetivo.

Los algoritmos utilizados en este paso comprenden:

- Corrección de banda base: Corrige desplazamientos específicos del espectrómetro.
- Corrección de ruido oscuro: Elimina el ruido de fondo producido por el detector cuando no hay luz incidente.
- Normalización de intensidad: Ajusta las intensidades de las señales a un valor común para facilitar la comparación entre diferentes espectros
- Centrado y escalado por rangos (mínimo – máximo): Transforma los datos para que los valores de cada característica estén dentro de un rango específico, generalmente entre 0 y 1.
- Suavizado de señal: Reduce el ruido en los datos mediante el ajuste de una función suave sobre los datos originales.

VII. Extracción de datos:

Una vez determinados los parámetros anteriores, es posible utilizar la información desplegada por el espectrómetro, utilizando los algoritmos correspondientes para su transformación, a datos matriciales, los cuales son manipulables a través de MATLAB o Python.

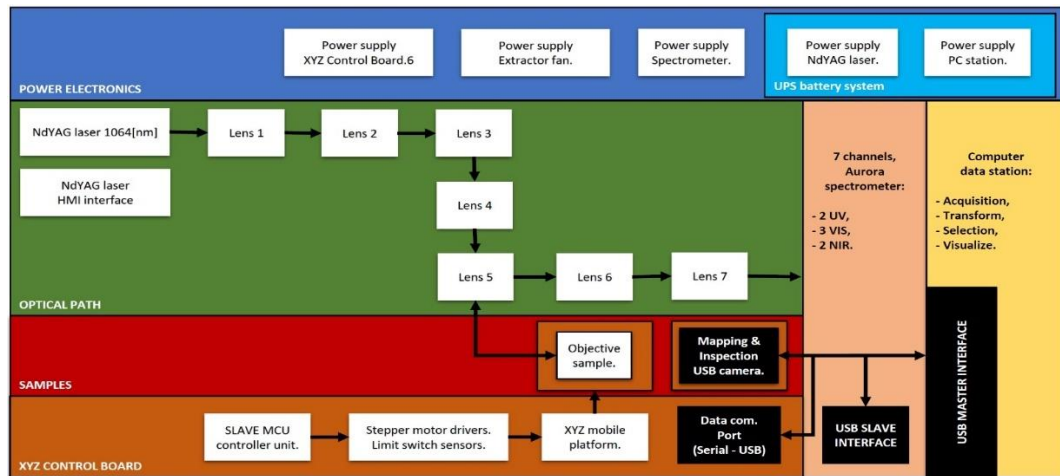


Figura 4.2.1: Diagrama de bloques "Laser Induction Breakdown Spectroscopy (LIBS)".

El proceso estructurado para la aplicación de técnicas de aprendizaje automático en la detección de adulteraciones en muestras de miel mediante la espectroscopia LIBS (Laser-Induced Breakdown Spectroscopy) consta de varias etapas que se desarrollan de manera secuencial y sistemática para maximizar la eficacia en la identificación de adulterantes en la miel.

En detalle:

I. Recopilación de Datos:

Se recopilan las muestras de miel, tanto puras como adulteradas en laboratorio. Se ajusta la configuración del sistema LIBS, determinando parámetros como la potencia del láser y los tiempos de retardo, para optimizar la captura de datos espectrales. Posteriormente, se realizan ablaciones láser sobre las muestras para vaporizar áreas pequeñas y generar un plasma cuya emisión espectral es medida por un espectrómetro.

II. Preprocesamiento de Datos:

Las medidas capturadas en el espectrómetro se agrupan y se procesan inicialmente; específicamente, se toman grupos de medidas, descartando y promediando los disparos correspondientes para minimizar errores y variabilidad. Estos datos crudos son luego normalizados y etiquetados según las características específicas de las muestras, preparándolos para un análisis más detallado.

III. Extracción de Características:

Se emplean técnicas estadísticas para extraer y destacar las características relevantes de los datos preprocesados y filtrados. El análisis de componentes principales (PCA) busca reducir la dimensionalidad de los datos, destacando las variaciones más significativas entre las muestras. Secuencialmente, el análisis discriminante lineal (LDA) se utiliza para identificar las características que mejor discriminan entre diferentes tipos de muestras de miel y sus respectivas adulteraciones. En este paso, se encuentra el desarrollo más costoso respecto a la cantidad de variables necesarias para considerar a estudio, ya que dependerá de la intensidad de variables mínimas para detectar diferencias entre clases, y mientras más sean, más costosos serán los entrenamientos, y mayor riesgo de generar un sobreajuste de datos.

IV. Clasificación Automática:

Habiendo obtenido las características relevantes, se procede a la clasificación de las muestras utilizando distintos algoritmos de aprendizaje automático. Máquinas de Vectores de Soporte (SVC), Descenso de Gradiente Estocástico (SGDC), y Bosques Aleatorios (RF) son aplicados para clasificar las muestras en categorías de adulteración. Cada algoritmo aporta una metodología distinta, desde la construcción de hiperplanos de separación hasta métodos de ensamble, para aumentar la precisión de la clasificación.

V. Evaluación de Resultados:

Finalmente, se obtienen perfiles detallados de los distintos tramos de adulteración y se generan reportes sobre los resultados de la clasificación. Estos reportes detallan cómo cada muestra ha sido clasificada por los algoritmos, proporcionando una base sólida para la toma de decisiones respecto a la calidad e integridad de las muestras de miel analizadas.

Este flujo de trabajo integral no solo facilita la identificación precisa de adulteraciones en la miel, sino que también refina el uso de la espectroscopia LIBS con técnicas de aprendizaje automático, optimizando tanto la recolección de datos como su análisis posterior.

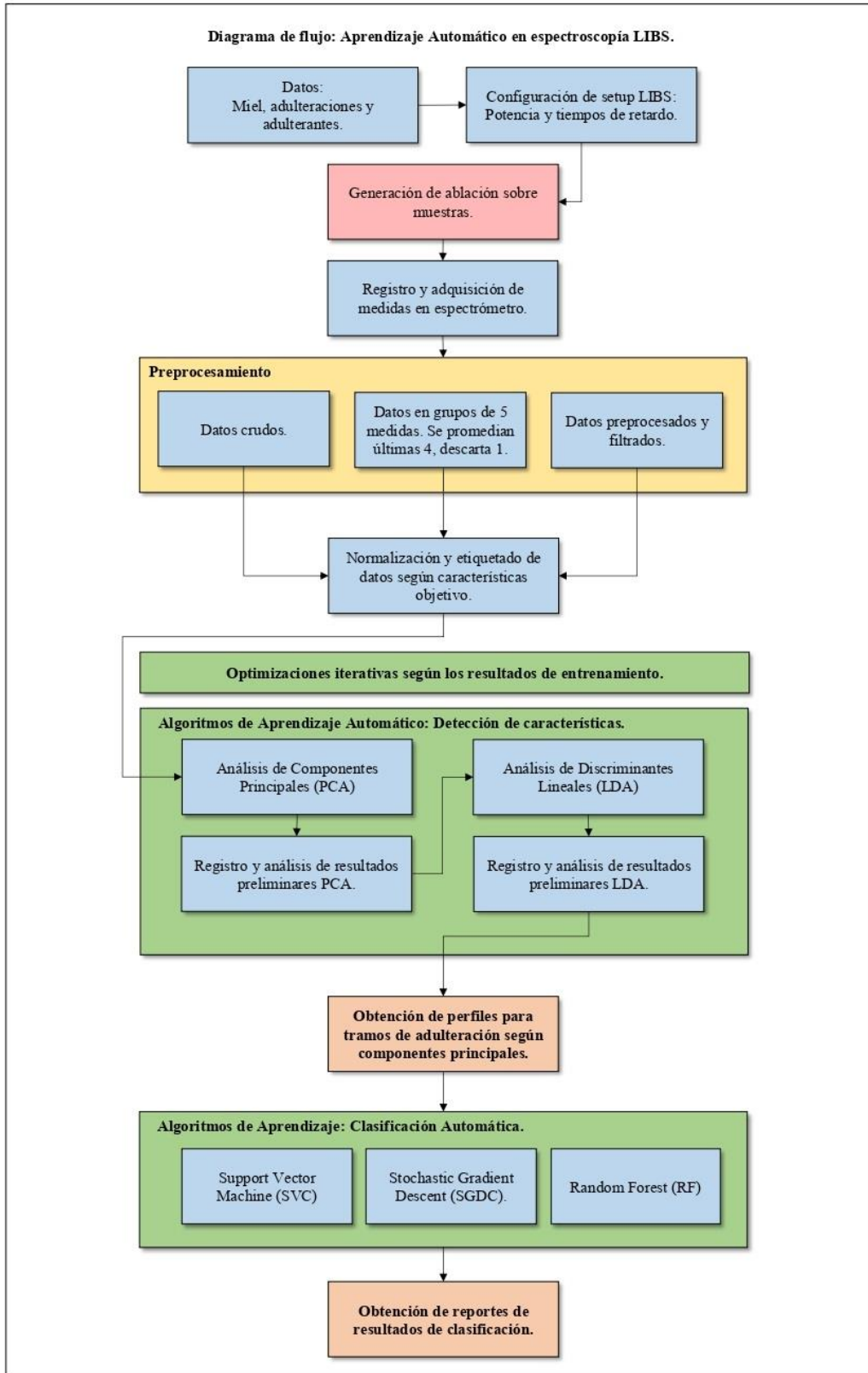


Figura 4.2.2: Diagrama de flujo "Laser Induction Breakdown Spectroscopy (LIBS)".

4.3. VISIÓN ARTIFICIAL

Abordando el objetivo específico [OE-4], la configuración experimental para adquirir datos de imágenes mediante microscopio ha sido diseñado para seguir el flujo de trabajo propuesto en la figura 3.3.1. Se compone de:

- Base de datos: Está compuesta por el universo total de muestras de miel pura, respectivas adulteraciones y adulterantes sintetizados.
- Microscopio: Sistema compuesto de tipo opto-mecánico, necesario para estudiar las muestras, utilizando un sistema de control de movimiento para el recorrido de distintos puntos de medición en las muestras.
- Filtros ópticos: Para el setup se ha considerado una cámara digital montada sobre un filtro de cristal líquido sobre microscopio para adquirir los datos de las muestras.
- Adquisición de datos: Mediante el programa desarrollado por el fabricante, se envían a la estación de recepción o unidad de registros.
- Estación de control: Genera el banco de datos según longitudes de onda objetivos.
- Preprocesamiento: Optimizaciones de tamaño, formato, claves y tipo para imágenes que conformarán la base de datos.

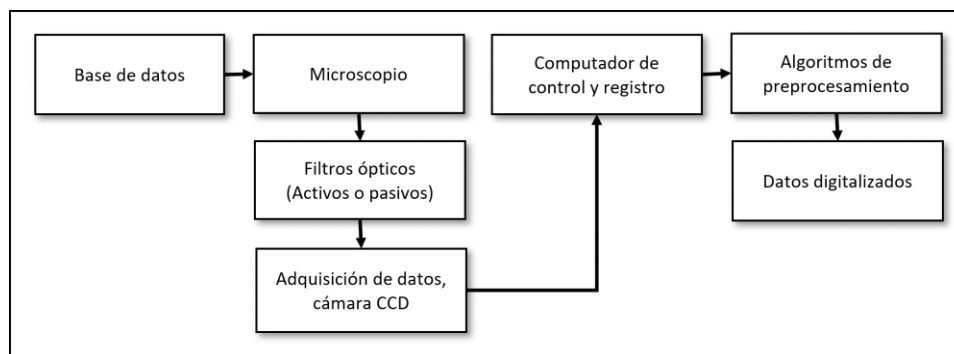


Figura 4.3.1: Esquema general para adquisición de datos para visión artificial.

4.3.1. METODOLOGÍA DE ENTRENAMIENTO CLASIFICADOR YOLOv8.

I. Preparación:

- Adquisición de datos.

Se recopilan datos desde la base de datos local, que posee 15 clases de muestras de polen. Esta base de datos debe ser completa y representativa para asegurar que el modelo pueda poseer una amplia cantidad de muestras para las diversas clases a considerar, así obtener un conjunto de datos diversificado y representativo que permita al modelo reconocer distintas clases de polen, mejorando así su capacidad de generalización.

- Limpieza de datos.

Se eliminan datos duplicados, erróneos o irrelevantes para garantizar la calidad del conjunto de datos. Esto incluye la corrección de errores, el manejo de imágenes faltantes o la eliminación de imágenes borrosas, asegurando que las imágenes sean precisas, consistentes y útiles para el entrenamiento del modelo, lo cual es fundamental para el rendimiento y la fiabilidad del modelo.

- Etiquetado y definición de Regiones de Interés (ROI's).

Se asignan etiquetas a las diferentes clases de polen presentes en la base de datos, y se definen áreas específicas en las imágenes que contienen polen en un margen (x, y) cuadrado indicado en píxeles. Se provee al modelo con datos etiquetados y estructurados que faciliten el aprendizaje supervisado, permitiendo una correcta identificación y clasificación de las diferentes clases de polen.

II. División de datos.

Los datos se dividen en tres subconjuntos: entrenamiento (70-80%), validación (10-15%) y prueba (10-15%). Se busca garantizar que el modelo sea evaluado de manera imparcial, permitiendo ajustes durante el entrenamiento mediante distintos hiperparámetros, validando así su capacidad de generalización a datos no vistos previamente.

III. Entrenamiento y validación del modelo.

- Entrenamiento.

Se utiliza el conjunto de datos de entrenamiento ya declarado para ajustar los parámetros del modelo YOLOv8. En esta etapa, se han generado distintas instancias con el fin de lograr rendimientos que no posean altos costes de penalización. Para ello, se apunta a ajustar los pesos del modelo para minimizar el error y mejorar su capacidad de inferencia.

- Validación.

Se evalúa el modelo generado durante el entrenamiento utilizando el conjunto de validación. En esta etapa, se generan distintas métricas que permitirán visualizar el rendimiento respecto al conjunto de datos interno, como precisión o la sensibilidad, para tomar en cuenta al momento de calibrar y ajustar hiperparámetros, previniendo un caso de sobreajuste hacia las clases.

- Prueba.

Evaluación final del modelo usando el conjunto de prueba, distinto al de la etapa de validación, para medir su rendimiento y su capacidad de generalización. Se requiere validar la efectividad del modelo en datos no vistos y obtener una medida precisa de su desempeño.

IV. Optimización del modelo.

- Configuración inicial.

Se definen parámetros iniciales, como el tamaño de la red, el número de épocas de entrenamiento o la tasa de aprendizaje. Estos hiperparámetros se configuran basándose en pruebas preliminares. Se busca establecer la base para el entrenamiento del modelo, asegurando que los parámetros iniciales sean apropiados para la clasificación.

- Selección de hiperparámetros.

Algunos hiperparámetros clave que afectan en el rendimiento del modelo pueden variar de los ya mencionados, para adecuar las imágenes a entornos donde se asemejen a variaciones de la realidad (giros, traslaciones, superposición, entre otras). Se prueban diferentes combinaciones para encontrar la configuración óptima. El objetivo es mejorar el rendimiento del modelo encontrando la mejor combinación de hiperparámetros, optimizando así su capacidad de inferencia para casos reales.

V. Entrenamiento iterativo y optimización.

Consiste en un proceso continuo de entrenamiento, ajuste de hiperparámetros y evaluación de resultados entre los distintos modelos generados a modo de comparación interna. Se realizan múltiples iteraciones para refinar el modelo, mejorando progresivamente el rendimiento de los modelos generados, mediante ajustes iterativos basados en los resultados previamente obtenidos.

VI. Implementación y validación final.

- Obtención del modelo

Después del anterior proceso de optimización, se obtiene el modelo final que será utilizado en aplicaciones prácticas. Este modelo ha sido ajustado y validado para asegurar su precisión y efectividad. Generando así un modelo aplicable a situaciones reales para realizar tareas de clasificación.

- Implementación en Casos Reales

Utilizando un set externo de imágenes de polen generado por el mismo sistema (buscando una situación real de análisis), se aplica el modelo para clasificar nuevas muestras de polen. Se prueban muestras no etiquetadas para evaluar la capacidad de generalización del modelo. Con ello se busca medir la capacidad de inferencia del modelo en aplicaciones externas, utilizando la cantidad de detecciones en el banco de datos no etiquetado, que posea las mismas clases, buscando validar el desempeño en condiciones reales y no controladas.

- Validación con Expertos

Comparación de los resultados estadísticos de detecciones del modelo con las técnicas de expertos en palinología y ecología vegetal para validar su fiabilidad. Se debe asegurar que el modelo tenga un desempeño comparable al de los expertos en el campo y que sus resultados sean confiables, proporcionando una validación final que certifique la calidad y utilidad del modelo.

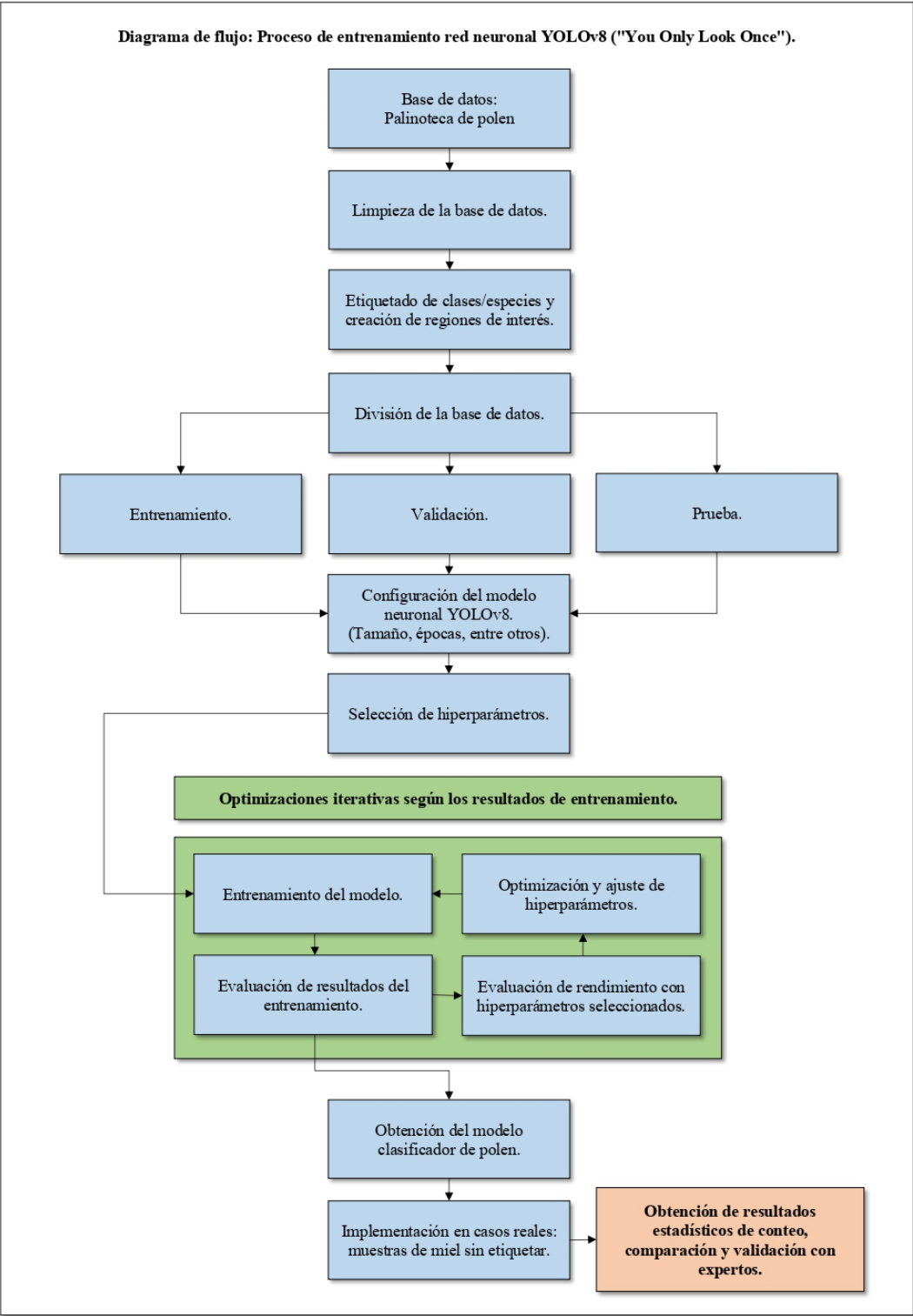


Figura 4.3.2: Proceso principal para procesamiento de datos a través de visión artificial.

5. ANÁLISIS DE DATOS

5.1. ESPECTROSCOPIA LIBS

En virtud del objetivo específico [OE-3], se utilizaron las distintas muestras de miel y adulteraciones, considerando entre los 4 últimos deciles. A continuación, se expone una tabla resumen de dicha información. Cabe mencionar que el volumen de datos generados por el equipo de espectrometría posee una alta dimensionalidad y complejidad en su estructura de datos:

Donde:

- “n”: Número de mediciones.
- “p”: Número de características (escala en longitud de onda nanométricas).

Muestras y adulterantes puros					
Clase	Guindo Santo	Quillay	Fructosa	Glucosa	Maltosa
= (n, p)	(177,12288)	(261,12288)	(162,12288)	(780,12288)	(162,12288)
Especie (Clases)	Adulterante	Mediciones			
		Miel al 10%	Miel al 20%	Miel al 30%	Miel al 40%
Guindo Santo (“ <i>Eucryphia</i> <i>Glutinosa</i> ”)	Fructosa	(221,12288)	(187,12288)	(200,12288)	(171,12288)
	Glucosa	(152,12288)	(190,12288)	(252,12288)	(162,12288)
	Maltosa	(192,12288)	(209,12288)	(190,12288)	(180,12288)
Quillay (“ <i>Quillaja</i> <i>Saponaria</i> ”)	Fructosa	(255,12288)	(255,12288)	(255,12288)	(255,12288)
	Glucosa	(255,12288)	(255,12288)	(255,12288)	(255,12288)
	Maltosa	(255,12288)	(255,12288)	(255,12288)	(255,12288)

Tabla 5.1.1: Clases de miel y adulterantes preprocesadas con objetivo de clasificación, detallando su estructura.

El instrumento adquiere 12288 muestras, distribuido en un rango de longitud de onda correspondiente de (185.363 – 1049.423) [nm], con un paso entre cada una de ~0.07[nm]. Esto da por resultado un cubo hiperespectral que sólo puede ser procesado bajo algoritmos computacionales.

Las estrategias utilizadas para enfrentar dicho escenario de alta dimensionalidad, corresponde a considerar los elementos más relevantes o diferenciables entre las distintas firmas espectrales a través del análisis de componentes principales “PCA”, donde su objetivo es lograr reducir la dimensionalidad de los datos en función de su varianza, considerando así los más relevantes. Luego de ello, se aplican técnicas de discriminación lineal “LDA”, para que, a partir del análisis anterior relativo a la varianza entre clases, se obtenga la mayor distancia entre puntos de clases que puedan distinguir cada firma espectral una de otra.

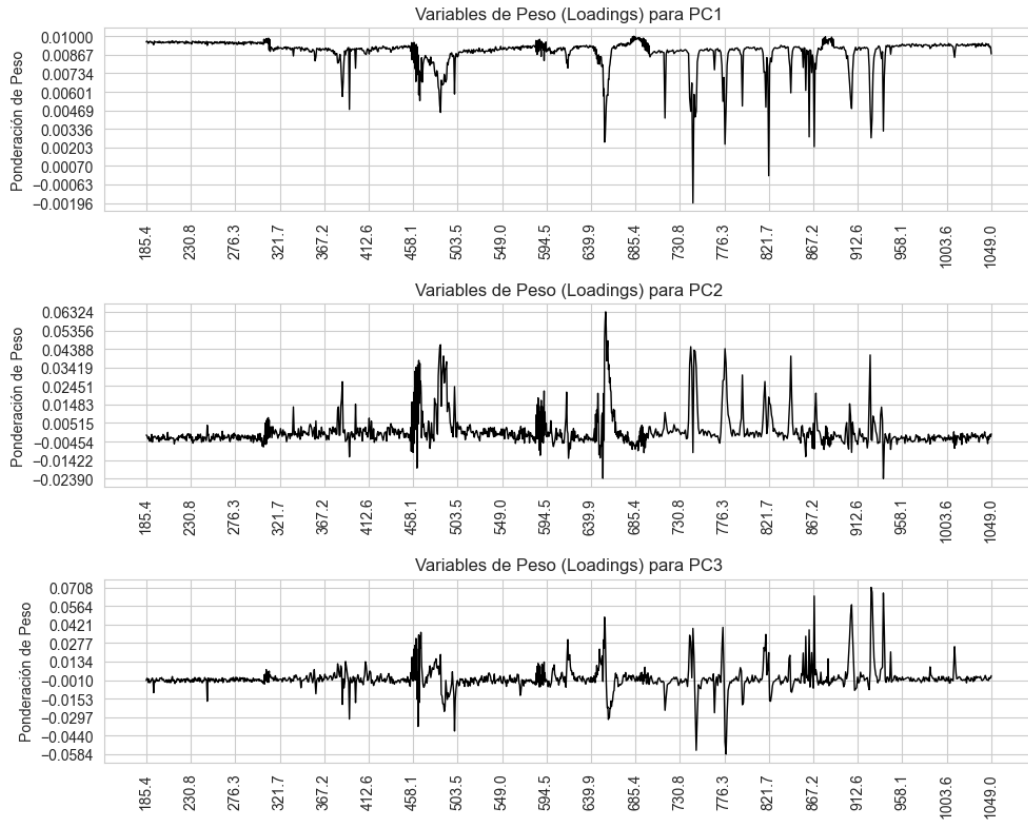


Figura 5.1.1: Variables de peso (3 primeras componentes principales) y su ponderación normalizada de pesos.

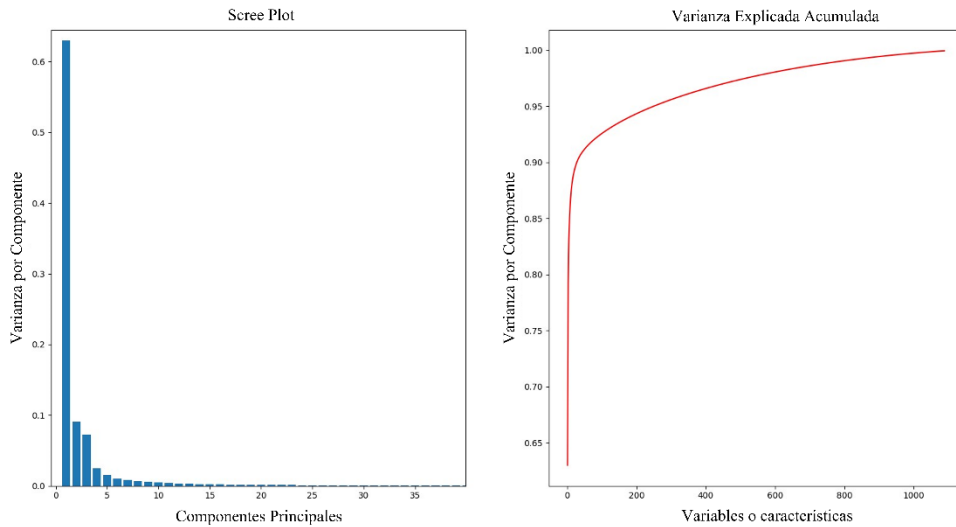


Figura 5.1.2: Varianza explicada para cada longitud de onda proveniente del análisis PCA. Se logra distinguir la importancia de los primeros componentes principales (**izquierda**) y la cantidad de componentes necesarias para diferenciar correctamente las clases (**derecha**).

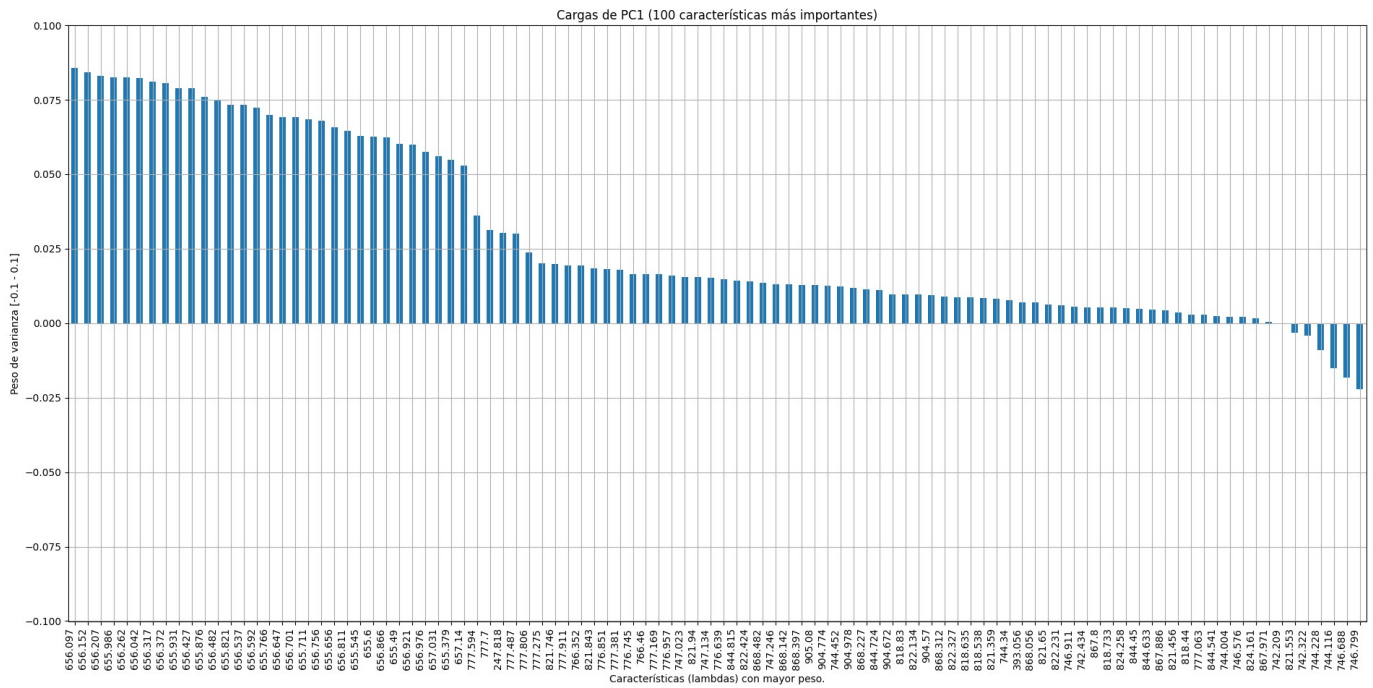


Figura 5.1.3: 100 primeras componentes principales de una firma espectral de la primera dimensión de PCA, indicando las longitudes de onda como distintas componentes, determinando su peso de varianza (elemento de Hidrógeno presente a conjunto 656 [nm] en primeras variables).

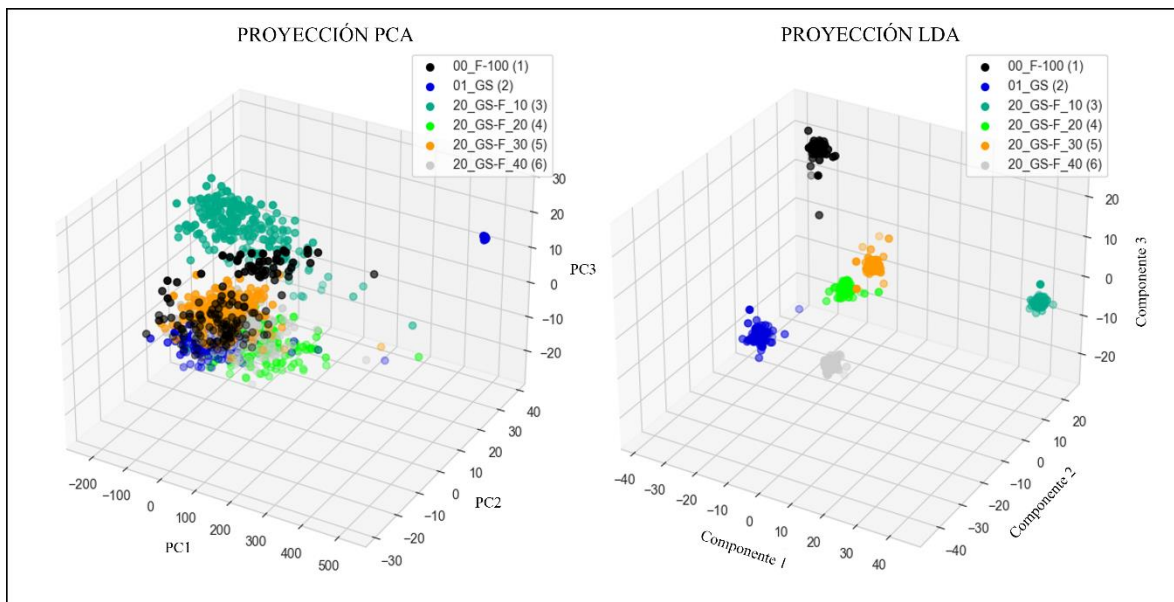


Figura 5.1.4: Representación visual de (Izquierda): Proyección realizada de los datos con PCA; (Derecha) Proyección realizada con LDA a partir de resultados de PCA.

En la figura 4.4, es posible visualizar diferencias notables entre distintas clases, las cuales son:

- I. Fructosa pura.
- II. Guindo Santo – Sin adulteración.
- III. Guindo Santo – Miel al 10% con fructosa.
- IV. Guindo Santo – Miel al 20% con fructosa.
- V. Guindo Santo – Miel al 30% con fructosa.
- VI. Guindo Santo – Miel al 40% con fructosa.

En (b) es especialmente distinguible el adulterante de fructosa, como también la miel pura respecto a las adulteraciones realizadas.

Para poder establecer una relación de comparación entre las clases adulteradas, la clase y el adulterante puros, se ha planteado tomar el centroide de cada clase analizada de las componentes principales según PCA con la maximización pertinente de LDA, reduciendo a un clúster más acotado las variables entre clases, realizando así la medición de distancias utilizando los centroides respectivos de cada clúster.

Este indicador se correlaciona con cada muestra y su porcentaje de adulteración con la muestra de control pura. Esto puede ser usado con distintos grupos que se desprendan del algoritmo inicial PCA/LDA, que a modo de representación visual será de máximo 3 dimensiones.

Representación de Centroides de los Clusters en 3 Componentes Principales

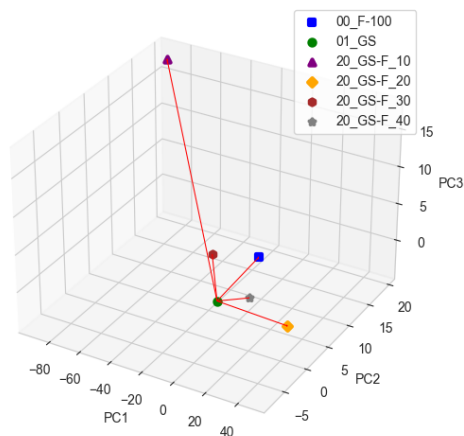


Figura 5.1.5: Centroides de cada clúster utilizados para calcular distancias entre ellos para clases de guindo santo adulterado en fructosa.

Distancias de las series al centroide de referencia				
	Clúster	Euclidiana	Manhattan	Chebyshev
1	01_GS	0.00	0.00	0.00
4	20_GS-F_30	15.26	25.24	10.93
5	20_GS-F_40	15.57	17.61	15.47
3	20_GS-F_20	32.45	40.34	31.89
0	00_F-100	48.74	66.74	46.42
2	20_GS-F_10	93.92	127.09	90.22

Tabla 5.1.2: Distancias medidas entre centroides de cada clúster perteneciente a una clase con respecto al clúster objetivo (Guindo Santo sin adulterar). Idealmente la representación debe ser gradual, desde la muestra patrón auténtica hasta el adulterante puro, lo cual varía según distintas estrategias a utilizar.

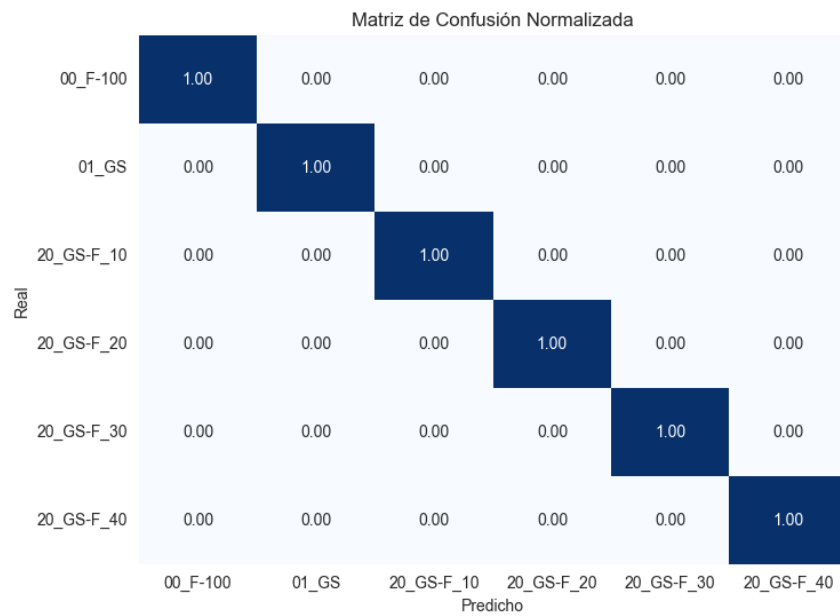


Figura 5.1.3: Matriz de confusión normalizada para diferencia entre clases preprocesadas (Fructosa en Guindo Santo) para algoritmo "Support Vector Machine". Se logra visualizar que, a pesar de diferir en la cantidad de datos para cada clase, no existen correlaciones que afecten negativamente la clasificación con respecto a otras clases existentes. Resultados para "Random Forest" y "Stochastic Gradient Descent" poseen rendimientos similares.

Accuracy:	1.00000			
F1-score:	1.00000			
	Precision	Recall	F1-Score	Support
00_F-100	1.00	1.00	1.00	50
01_GS	1.00	1.00	1.00	56
20_GS-F_10	1.00	1.00	1.00	68
20_GS-F_20	1.00	1.00	1.00	57
20_GS-F_30	1.00	1.00	1.00	54
20_GS-F_40	1.00	1.00	1.00	51
accuracy			1.00	336
macro avg	1.00	1.00	1.00	336
weighted avg	1.00	1.00	1.00	336

Tabla 5.1.3: Resultados para clasificación por algoritmo "Support Vector Machine Classifier". Es posible observar un desempeño eficaz respecto a la identificación de las clases.

La tarea de clasificación puede volverse trivial mientras mayor sea la cantidad de componentes a utilizar desde PCA para LDA, pero esto no sugiere una tarea de clasificación efectiva, ya que, al utilizar todos los componentes, casi no se hace reducción de dimensiones.

Desde LIBS es posible registrar sobre 12000 dimensiones con distintos valores para distintas medidas, pero no todas las dimensiones son requeridas. Por ello, el propósito de PCA, de lograr reducir a escala de cientos o incluso decenas de componentes principales es crucial que sea llevado a cabo de la forma más óptima y adecuada posible.

Para este punto, se pueden considerar 3 escenarios:

I. Datos crudos sin preprocesar:

Este escenario propone utilizar toda la información proveída por el sistema de espectroscopía, utilizando información en longitudes de onda muy relevantes como también información no relevante, provocando que el algoritmo de análisis tome costosamente más tiempo y recursos, al utilizar las 12.288 dimensiones disponibles para realizar el análisis no supervisado por PCA, y a partir de la cantidad de clases para un análisis semi supervisado por LDA. Este resultado considera todas las dimensiones disponibles, siendo costoso e impreciso, ya que existen variaciones que implican una variación en la señal respecto a ruido (SNR) que no aportan en profundidad al modelo. De todas maneras, entrega un clasificador efectivo en los 3 tipos de algoritmos disponibles (SVM, RF, SGD).

II. Datos preprocesados:

Los datos utilizados son los mismos en origen, o datos crudos, pero que, seguido a llevar a grupos de medidas (clustering), por un concepto de cantidad de disparos vs permanencia de señal respecto al tiempo de integración del sensor, es que en primera instancia se descartan y agrupan por sección de muestras, reduciendo la cantidad de muestras representativas de los datos. Sobre este grupo, es aplicado el análisis PCA, el cual toma computacionalmente menos de tiempo, representando una interpolación de medidas. Secuencialmente, se aplican los modelos discriminantes LDA que permitirán considerar las dimensiones más relevantes para una posterior clasificación.

A partir de esta estrategia, se aplican 2 conceptos relevantes en la información generada al momento de trabajar con LIBS: Normalización de amplitud (*“Peak normalization”*) y corrección de ruido oscuro. Debido al preprocesamiento, se pierde información relevante, pero se logra generalizar modelos de clasificación.

La normalización se enfoca en acentuar los valles o picos de amplitud de las firmas espectrales, con el fin de poder localizar magnitudes significativas de un elemento en particular, y en especial hablando de la señal SNR, y el segundo es utilizado para reducir el ruido de fondo u optoelectrónico que es base del instrumento.

Al igual que para los datos anteriores, se lleva a cabo el proceso de entrenamiento de modelos para SVM, SGDC y RF, generando los reportes de agrupamiento y clasificación correspondientes.

III. Datos filtrados:

Considerando que se poseen distintas longitudes de onda en el cubo hiperespectral, y que es posible utilizar longitudes de onda objetivos previamente estudiadas, tal como se indica en la tabla 2.1.1:

Elemento	Longitud de onda [nm]
H (I)	656.3, 486.1
C (I)	247.9
N (I)	742.4, 744.2, 746.8, 818.6, 821.8, 824.2, 856.8, 859.4, 862.8, 868.2, 871.3, 938.7, 939.3
O(I)	777.2, 777.4, 777.5, 844.6, 926.6
Na (I)	589.0, 589.6
Mg (I)	285.2
Mg (II)	279.6, 280.3
K (I)	766.5, 769.9
Ca (I)	422.7
Ca (II)	393.3, 396.8
CN ($\Delta v = -1$)	359.0 (1-0), 358.59 (2-1)
CN ($\Delta v = 0$)	388.3 (0,0), 387.1 (1,1), 386.2 (2,2)

Tabla III.1: Tabla de longitudes de onda de emisión para cada elemento atómico relevante en la clasificación de miel [40] en LIBS.

Se puede aplicar tanto a datos crudos como preprocesados dichos filtros, dependiente directamente de la resolución y precisión del espectrómetro (todo instrumento posee un desplazamiento en la práctica) para considerar la respuesta espectral más precisa posible, obteniendo una disminución clara en sus componentes de estudio, logrando una reducción de tiempos de procesamiento y de rendimientos similares.

Para realizar este filtrado de información, se consideró toda la información proveniente del espectrómetro y se le ha aplicado, en forma de filtro de datos, cada longitud de onda objetivo para los elementos en cuestión, con una tolerancia no mayor a $\pm (0.2 - 0.8)$ [nm], para definir el tramo de análisis de cada elemento. Cabe mencionar que las características a estructuras moleculares de carbono, hidrógeno y oxígeno aportan en gran medida a la diferenciación entre clases, puesto que la estructura molecular de este tipo de azúcares está compuesta por estos elementos.

La cantidad de componentes necesaria para realizar la identificación y clasificación es al menos 2 órdenes menor a las dimensiones disponibles desde LIBS (de 12.000

dimensiones, es posible utilizar 200 o menos). En este punto, se comienza con el análisis PCA y LDA, que plantearán los modelos de agrupación de datos antes de la tarea de clasificación. Es relevante mencionar que, ajustados a estos datos, si bien es posible ver una clasificación efectiva, también es notorio el borde para la tarea de reconocimiento de entre modelos al aplicarlos a validaciones con datos externos al proceso.

Para contrastar los resultados obtenidos a partir de estos 3 enfoques, se utilizará como regla, que, desde la descomposición hasta la clasificación, los clústeres o clases, sean correctamente identificados por su contenido de miel legítima desde el 100% de pureza a un 0%, o sea, totalmente adulterada y de forma gradual, esperando que correlacione dicha estructura patrón con lo obtenido por los clasificadores.

La tabla resumen siguiente define el error de los modelos generados por PCA/LDA respecto a las tareas de clasificación, y su rendimiento al evaluar un set de datos de validación para las muestras de miel:

ESPECIE	ADULTERANTE	ERROR DEL CLASIFICADOR %		
		SIN PROCESAR	PREPROCESADO	FILTRADOS
QUILLAJA SAPONARIA (QUILLAY)	FRUCTOSA	3.70%	5.56%	11.11%
	GLUCOSA	3.70%	16.67%	12.96%
	MALTOSA	24.70%	22.22%	11.11%
	TOTAL	10.70%	14.82%	11.73%
EUCRYPHIA GLUTINOSA (GUINDO SANTO)	FRUCTOSA	24.07%	9.26%	11.11%
	GLUCOSA	20.37%	16.67%	22.22%
	MALTOSA	22.22%	33.33%	25.93%
	TOTAL	22.22%	19.75%	19.75%

Tabla 5.1.4: Resultados del procesamiento de datos, considerando los rendimientos de cada clasificador para los distintos adulterantes y el promedio total para las distintas estrategias.

Analizando brevemente la tabla 4.1.4, se pueden asumir algunos resultados preliminares, como, por ejemplo, que utilizar señales crudas sin procesar, en general da buenos resultados, pero estos toman un tiempo considerable de análisis para PCA/LDA, recursos de memoria que pueden ser elevados para la unidad procesadora. Por ello, se utiliza el acercamiento de preprocesamiento, utilizando grupos de 5 mediciones, descartando la primera y promediando las siguientes 4. Por último, utilizar filtros de información para bandas específicas de interés.

Como resultado de la última estrategia, es posible disminuir de la totalidad de longitudes de onda o dimensiones a unas decenas o centenas de componentes principales, disminuyendo en al menos 2 órdenes de magnitud el uso de información diferenciadora.

Por otra parte, es destacable mencionar que dichos resultados son muy susceptibles a como se ha tomado la información desde el espectrómetro, y si la medida fue realizada correctamente, evitando problemas de ruido de fondo o desfases en la intensidad captada por configuraciones aleatorias en la potencia administrada por el sistema láser.

5.2. VISIÓN ARTIFICIAL

Para iniciar este proceso, es necesario cumplir con el objetivo específico [OE-4]. Dicha tarea está dedicada a retirar, limpiar y estructurar correctamente la base de datos con el foco en poder utilizar correctamente el algoritmo, y no perder resolución espacial de las imágenes tal que pueda afectar a los algoritmos de detección de la red neuronal convolucional. Una vez resuelta esta necesidad, es posible avanzar al siguiente objetivo secundario [OE-5].

Se han desarrollado algoritmos de aprendizaje profundo de redes neuronales convolucionales para la identificación de 2 especies en específico: Guindo Santo y Quillay. En este escenario se han utilizado las capacidades de los algoritmos de aprendizaje profundo “YOLO” para incluir una mayor cantidad de clases de miel y polen a identificar y clasificar, utilizando una mayor entrada de datos que permitan dar un modelo clasificador de mayores capacidades.

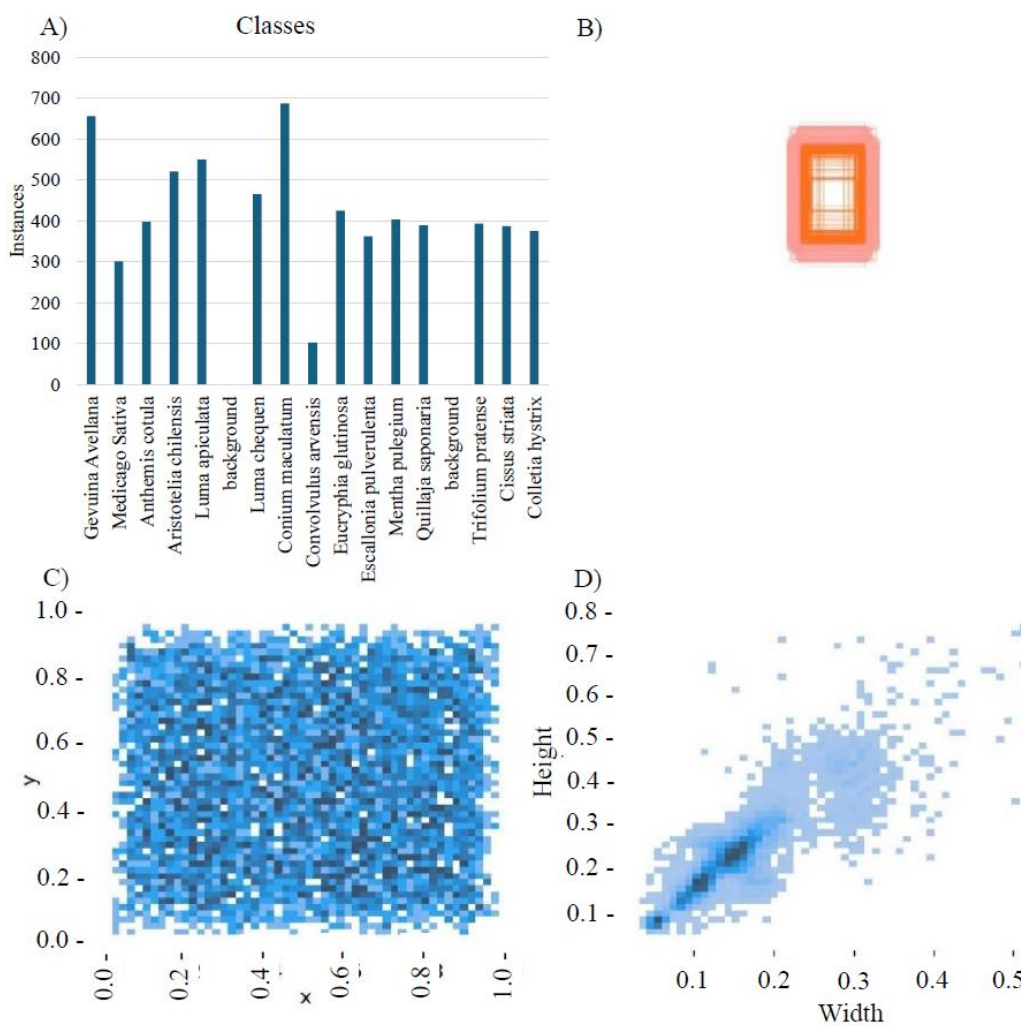


Figura 5.2.1: (a) Izquierda superior: Clases disponibles para clasificación; (b) Derecha superior: Regiones de interés en píxeles de los elementos identificados; (c, d) nivel inferior: dispersión de datos en relación con las distintas clases.

La imagen superior hace referencia a como está compuesta la base de datos, indicando las clases, la cantidad de imágenes y las regiones de interés para cada una. Esta información es la que se utiliza para entrenar los modelos, las cuales poseen características relevantes respecto a su dispersión en estas zonas de interés.

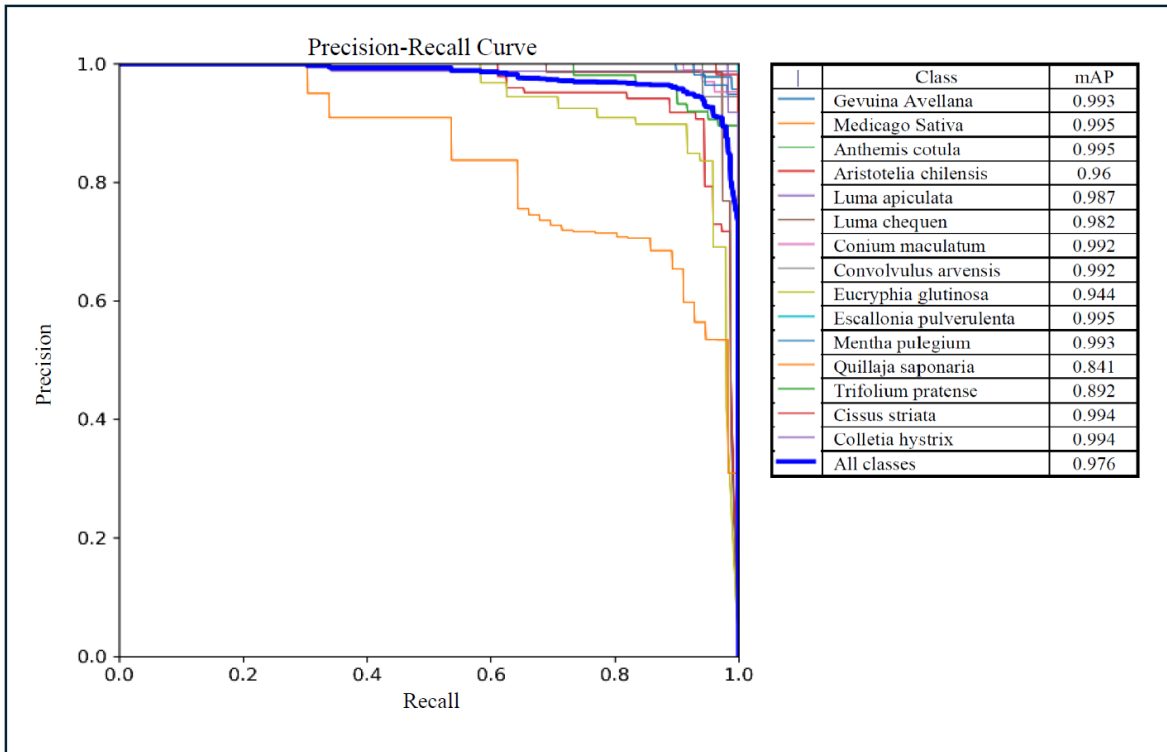


Figura 5.2.2: Curva PR del modelo clasificador YOLO.

Predicted	0	Gevuina Avellana	1.00																	0.10	1.00	
	1	Medicago Sativa		1.00																	0.04	0.94
	2	Anthemis cotula			1.00																0.01	0.88
	3	Aristolelia chilensis				0.83															0.05	0.82
	4	Luma apiculata					1.00														0.01	0.76
	5	background																				0.71
	6	Luma chequen						0.96				0.02									0.09	0.65
	7	Conium maculatum							1.00									0.02			0.08	0.59
	8	Convolvulus arvensis								0.94												0.53
	9	Eucryphia glutinosa									0.94											0.47
	10	Escallonia pulverulenta										1.00									0.01	0.41
	11	Mentha pulegium											0.98									0.35
	12	Quillaja saponaria												0.91						0.05		0.29
	13	background																				0.24
	14	Trifolium pratense								0.06								0.97			0.08	0.18
	15	Cissus striata							0.01										0.98	0.02	0.04	0.12
	16	Colletia hystrix																		0.93	0.04	0.06
	17	background				0.17		0.03			0.06			0.09		0.30					0.04	0.00
		Gevuina Avellana																				
		Medicago Sativa																				
		Anthemis cotula																				
		Aristolelia chilensis																				
		Luma apiculata																				
		background																				
		Luma chequen																				
		Conium maculatum																				
		Convolvulus arvensis																				
		Eucryphia glutinosa																				
		Escallonia pulverulenta																				
		Mentha pulegium																				
		Quillaja saponaria																				
		background																				
		Trifolium pratense																				
		Cissus striata																				
		Colletia hystrix																				
		background																				
		True																				

Figura 5.2.3: Matriz de confusión, representa la correlación entre clases y su diferenciabilidad. Se puede visualizar que se han removido 2 clases ya que su presencia en miel es poco frecuente en mieles endémicas (Brassica Rapa, Ruda).

Luego de haber obtenido distintos modelos, optimizando hiperparámetros relevantes, se ha logrado dar con modelos que pueden describir diferencias entre las distintas clases, hasta el punto de validación con respecto a la propia base de datos; este escenario cambia sustancialmente al momento de llevarlo a pruebas reales.

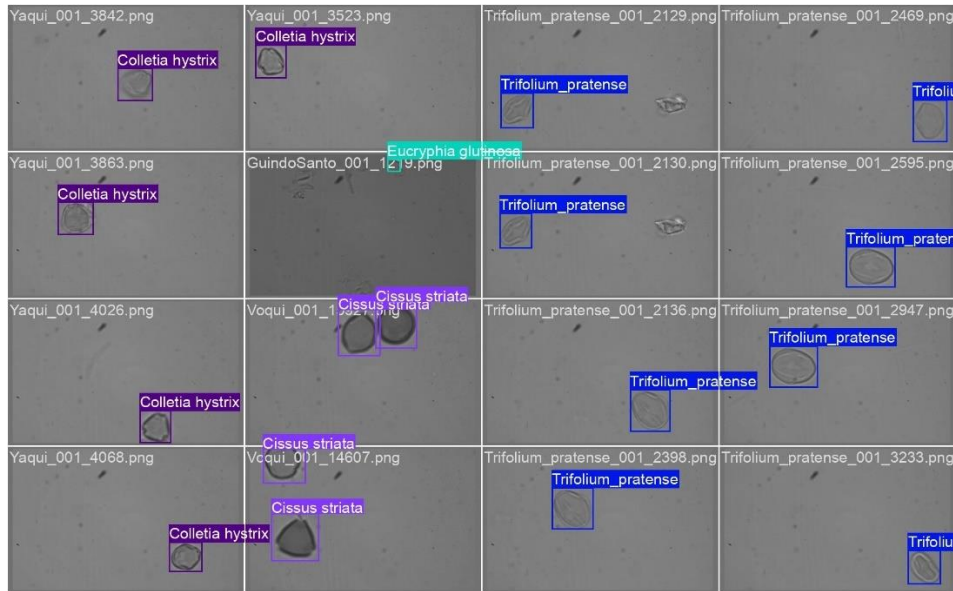


Figura 5.2.4: Imagen de referencia al proceso de entrenamiento para YOLOv8.

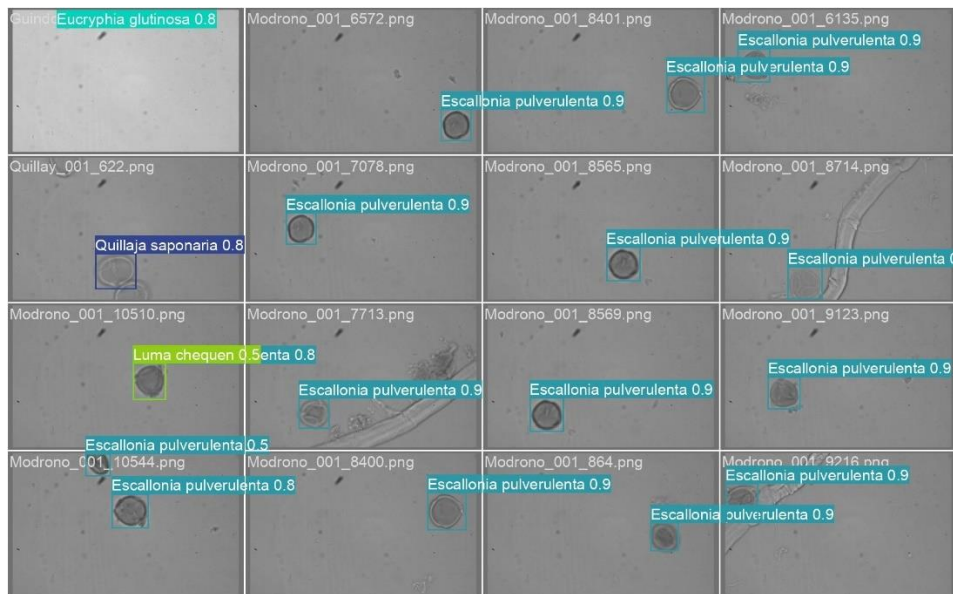


Figura 5.2.5: Imagen de referencia al proceso de validación para YOLOv8.

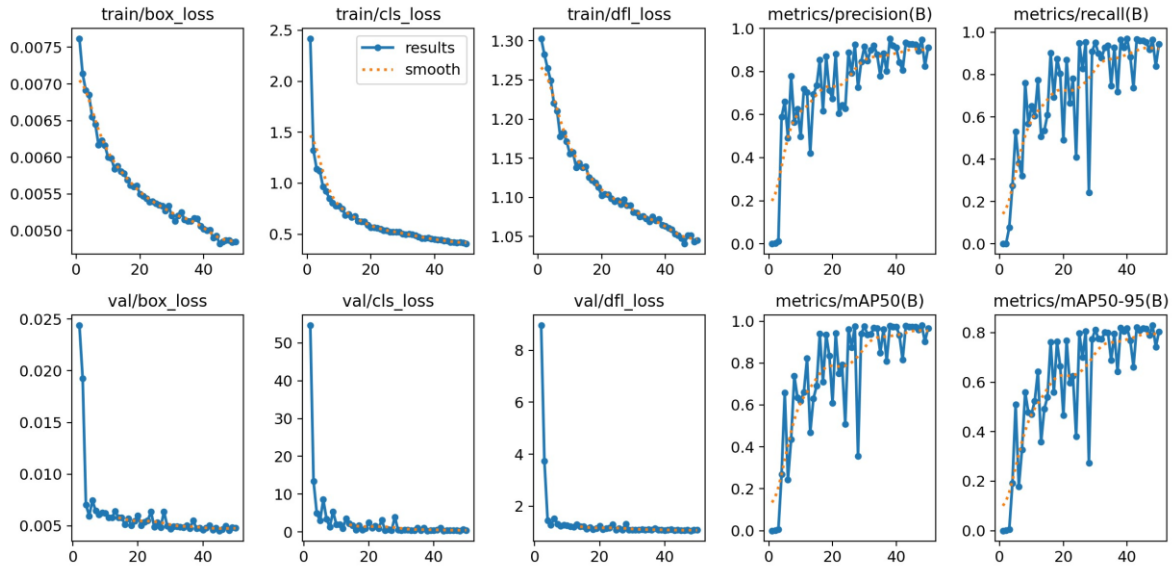


Figura 5.2.6: Métricas de evaluación para entrenamiento y validación. Entre ellas se encuentran coeficientes de pérdida y métricas principales de “Precision”, “Recall”, “mAP50”.

Los resultados obtenidos hasta ahora son satisfactorios. Se puede visualizar en los resultados de entrenamiento, específicamente en métricas de precisión y sensibilidad, una precisión promedia aproximada mAP sobre el (95~99) % en la mayoría de las clases, exceptuando la clase “Quillay” que ha obtenido algunos problemas debido a que el banco de imágenes correspondiente posee algunos errores respecto a cómo se tomaron las regiones de interés en dichas imágenes. De todas maneras, se ha logrado utilizar diversos hiperparámetros y configuraciones que han dado por resuelto parcialmente dicho problema.

Se ha implementado este modelo en *muestras sin análisis previo*, a modo de validación externa, provenientes del mismo banco de datos, pero de perfiles de miel (no sólo polen de una especie). Se han utilizado 4 muestras distintas, tabulándose los resultados de la primera muestra, a continuación:

Tabla de detecciones:	MGS_R1		
Clase	Promedio de Confianza	Cantidad de Detecciones	%
Anthemis coluta	0.51	1	0.08
Aristolelia chilensis	0.66	6	0.48
Cissus striata	0.72	53	4.27
Colletia hystrix	0.64	10	0.81
Convolvulus arvensis	0.72	8	0.64
Escallonia pulverulenta	0.7	6	0.48
Eucryphia glutinosa	0.69	732	58.94
Gevuina avellana	0.82	26	2.09
Luma apiculata	0.81	10	0.81
Luma chequen	0.66	19	1.53
Medicago sativa	0.77	43	3.46
Mentha pulegium	0.7	60	4.83
Quillaja saponaria	0.68	6	0.48
Trifolium pratense	0.77	262	21.1
Total		1242	100

Tabla 5.2.1: Conteo realizado en muestra real de “Guindo Santo”, obteniendo indicador de monofloralidad al superar el 50% de detección de polen para dicha clase.

Estos resultados han sido comparados con estudios realizados en laboratorios de Palinología y Ecología Vegetal, los cuales han determinado la presencia polínica para las mismas muestras, elaborando así un contraste para dicha tarea:

Conteo:	MGS_R1	
	N° de Granos	%
Amaranthus	38	2.63
Aristotelia chilensis	10	0.69
Colletia hystrix	374	25.85
Echium vulgare	23	1.59
Eucryphia glutinosa	732	50.59
Galega officinalis	24	1.66
Gevuina avellana	75	5.18
Lotus pedunculatus	41	2.83
Luma apiculata	22	1.52
Medicago sativa	33	2.28
Melilotus indicus	40	2.76
Mentha pulegium	15	1.04
Mutisia	0	0.00
Poaceae	4	0.28
Quillaja saponaria	16	1.11
Total	1447	100.00

Tabla 5.2.2: Conteo realizado en muestra real de “Guindo Santo”. Los resultados provienen del “Laboratorio de Palinología y Ecología Vegetal”, Campus Los Ángeles, Universidad de Concepción.

Se puede observar que se han detectado más clases de polen presentes en las muestras, e incluso de algunas inexistentes en el caso del entrenamiento del clasificador automático. Esto es de esperar, debido a que el modelo detectará polen y lo asimilará a la clase más cercana de acuerdo con los valores umbrales para cada una de estas clases.

Es posible observar en la siguiente tabla como se han desempeñado ambos métodos a modo de comparación, siendo destacable los resultados sobre 50%, ya que indicarían una miel de tipo monofloral:

COMPARACIÓN DE MÉTODOS		
Especies	Automático	Manual
Amaranthus	0.00%	2.63%
Anthemis coluta	0.08%	0.00%
Aristotelia chilensis	0.48%	0.69%
Cissus striata	4.27%	0.00%
Colletia hystrix	0.81%	25.85%
Convolvulus arvensis	0.64%	0.00%
Echium vulgare	0.00%	1.59%
Escallonia pulverulenta	0.48%	0.00%
Eucryphia glutinosa	58.94%	50.59%
Galega officinalis	0.00%	1.66%
Gevuina avellana	2.09%	5.18%
Lotus pedunculatus	0.00%	2.83%
Luma apiculata	0.81%	1.52%
Luma chequen	1.53%	0.00%
Medicago sativa	3.46%	2.28%
Melilotus indicus	0.00%	2.76%
Mentha pulegium	4.83%	1.04%
Mutisia	0.00%	0.00%
Poaceae	0.00%	0.28%
Quillaja saponaria	0.48%	1.11%
Trifolium pratense	21.10%	0.00%
TOTAL	100.00%	100.00%

Tabla 5.2.3: Comparación entre métodos automático y manual.

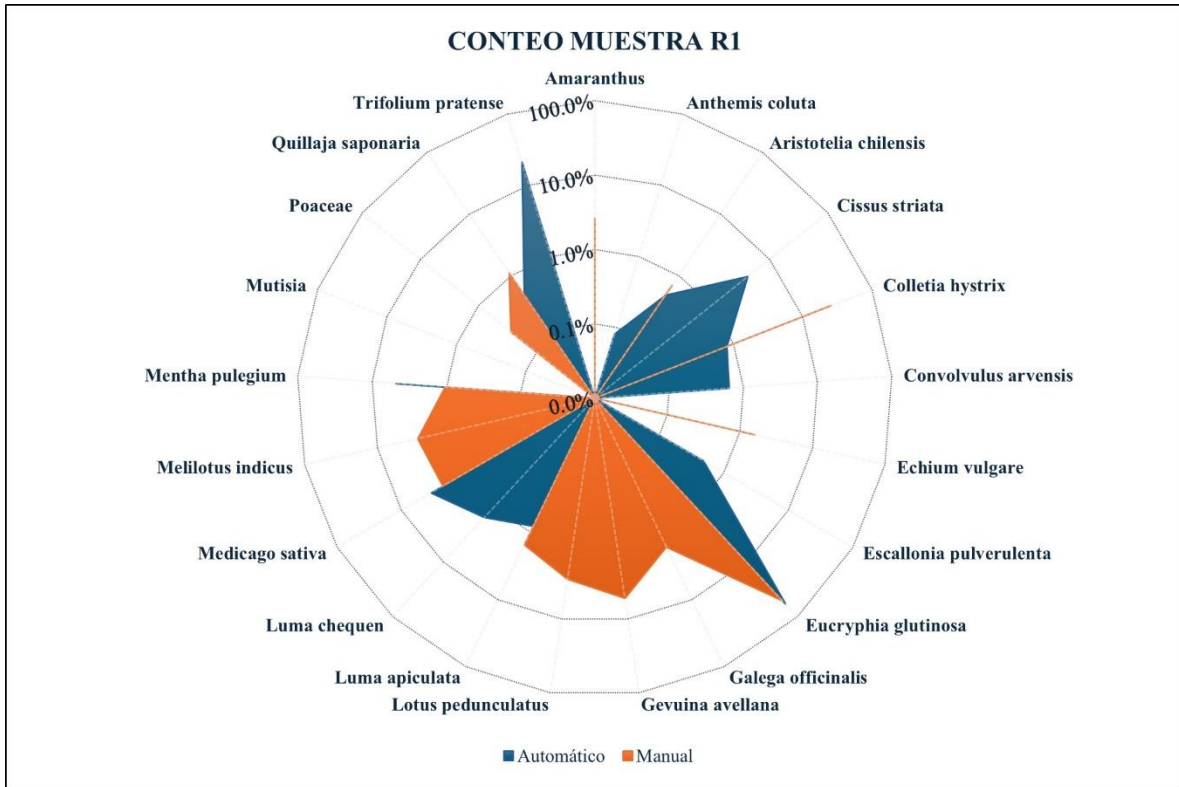


Figura 5.2.7: Firma del espectro polínico para cada método, en escala logarítmica.

En último lugar, se encuentra el gráfico radial que posee distintas clases de especies para polen de la región del Bío Bío. Dicha figura se encuentra en escala logarítmica para poder resaltar cantidades de polen que no han sido superiores al 1%.

Si bien, el clasificador automático posee la flexibilidad de ser entrenado con distintas clases de polen, la tarea se vuelve especialmente complicada al momento de etiquetar y administrar la cantidad de imágenes necesarias para cada clase. El algoritmo utilizado demanda alrededor de 1500 imágenes por clase a identificar, esto para asegurar que exista un modelo lo suficientemente detallado para clasificar correctamente. En este estudio, el número total de imágenes utilizadas está indicado en la figura 4.2.1 (a), que son alrededor de 5000 imágenes para 15 clases. Esto ya representa un problema de información que debe ser mejorado, sea por métodos de aumento de datos, que no modifican o añaden nueva información a la base de datos, o por añadimiento de datos originales, que, si logran modificar la entrada de datos a nuevas situaciones o contextos de imágenes, clasificando con una mayor eficiencia. Esto supone también, un costo computacional relevante para el entrenamiento que debe ser considerado al momento de realizar dichos estudios.

Por otro lado, el método manual llevado a cabo por los laboratorios asociados pertinentes conlleva un trabajo y tiempo considerables, esto para analizar y registrar los datos de las muestras, y dada la experiencia de los profesionales a cargo de dicha tarea, junto a la capacidad de reconocimiento de especies locales con índices de confianza relativamente altos por los métodos empleados, es que es crucial para validar esta última etapa del método propuesto para determinación de cantidad de polen en muestras de miel.

6. DISCUSIÓN DE RESULTADOS

En primer lugar, establecer 2 técnicas de análisis ópticos que difieren en tecnología y métodos para adquisición de datos resulta desafiante, esto porque desde el punto de vista de LIBS la cantidad de datos resultantes de sus mediciones puede ser de miles de características o variables, las cuales suponen un desafío para poder determinar cuáles son las más importantes al momento de comparar muestras de miel en laboratorio en un marco controlado, y que sin algoritmos de aprendizaje, semi o completamente supervisados, resultaría muy poco viable establecer un estudio a tiempo.

LIBS da resultados precisos sobre las trazas elementales que puede analizar, pero son datos que deben ser preprocesados adecuadamente para utilizarlos (corrección de ruido de fondo, corrección de desfases en longitudes de onda, entre otros problemas prácticos en la adquisición de datos). Previo a estos pasos correctivos, es que es posible procesar dicha información por algoritmos de aprendizaje automático, que es donde se han encontrado diferencias relevantes entre clases, que puede mejorar considerando distintos enfoques en la limpieza de los datos desde LIBS, como también, la cantidad de registros o disparos.

En segundo lugar, la visión artificial no posee la capacidad de LIBS para analizar múltiples longitudes de onda, ya que el tamaño de datos sería excesivo en volumen y la orientación de los sensores está enfocada en otro objetivo: en la cantidad de píxeles distribuidos espacialmente a través de tramos específicos de longitudes de onda digitales. Esto permite utilizar las herramientas de microscopía para visualizar el polen detectado en muestras normalizadas de intensidad.

Los resultados obtenidos en esta etapa presentan un rendimiento destacable para algunas clases, las que han sido mejor representadas en las bases de datos ingresadas al algoritmo de aprendizaje profundo “YOLOv8”, pero que, para otras clases, como “*Trifolium Pratense*” o “*Mentha Pulegium*”, resulta difícil de establecer. Esto porque en los rangos de medición establecidos para visión artificial las diferencias entre algunas clases son muy sutiles (tamaño, forma, si el polen está hidratado o seco, técnicas de desenfoque al momento de identificar la clase de polen...), y dado estos hechos, es que existen resultados que poseen un rendimiento menor para identificarlas como tal, estableciendo inferencias con un relativo bajo nivel de confianza. Esto puede ser resuelto a través de una mayor cantidad de imágenes y etiquetas por cada clase, abarcando distintas perspectivas de desenfoque y replicando de mejor manera las mediciones hechas por expertos palinólogos, como también haciendo uso de otros algoritmos, tales como estimadores de posición, para preprocesar la información antes de ingresar en etapas de aprendizaje profundo.

De esta manera, los resultados tabulados de cada método aplicado, y que contienen todo el estudio respecto a sus análisis elementales por LIBS, como estadísticas de origen botánico por visión artificial, responden al cumplimiento del último objetivo específico [OE-6], donde es posible visualizar las características y sus indicadores de calidad, autenticidad y origen creados para los modelos clasificadores estudiados.

Haciendo uso de los clasificadores desarrollados en base a LIBS, es posible afirmar que existe una clasificación clara entre distintas adulteraciones, que, si bien unos modelos poseen mejores rendimientos que sus símiles, en el caso de Fructosa y Glucosa especialmente, que es donde se han obtenido resultados con mayor cantidad de componentes diferenciables, y particularmente especial en Quillay con Fructosa. Contando con esta herramienta desarrollada, es que es posible hablar de presencia o ausencia de adulteración en una muestra.

Por otro lado, la identificación de origen botánico establecido por visión artificial también ha tenido mejores desarrollos de clasificación para algunas clases, de las cuales cabe recalcar a “*Eucryphia Glutinosa (Guindo Santo)*”, ya que su clasificación ha sido la que mayor desempeño posee. Dado este hecho, es que es posible realizar procesos estadísticos relativos a la concentración polínica en muestras de miel, y que si bien mejoraría con un modelo ampliamente documentado para cada clase (imágenes de cada clase), da un indicador respecto a la cantidad de polen que debiera estar presente en una muestra de miel. Esto, al momento de ser comparado con expertos palinólogos, ha resultado en desempeños similares.

El punto de cruce entre ambas tecnologías radica en que una concentración adulterada de polen en una muestra de miel, que puede ser generada por adulteraciones en su composición, es distinta y va en decremento de calidad y cantidad de polen, en comparación a lo observado para una muestra pura. Dicha estimación y conteo permiten en teoría, corresponder estadísticamente al porcentaje de adulteración establecido para cada una de las muestras a analizar, realizando en conjunto al análisis LIBS, la decisión sobre si una muestra de miel está adulterada o no, y si lo está, determinar los posibles compuestos presentes que sean reconocibles.

7. TRABAJO FUTURO

Se ha establecido que la efectividad de los modelos de aprendizaje automático mejora significativamente con el incremento en la cantidad y calidad de los datos. Sin embargo, es esencial priorizar la calibración adecuada de instrumentos como los sistemas LIBS y de microscopía. Esto incluye la selección cuidadosa de longitudes de onda preliminarmente y el preprocesamiento de datos para minimizar ruidos indeseados. Una fuente de datos bien calibrada y estructurada es crucial para la mejora en la construcción de clasificadores precisos.

Con el avance en los algoritmos de aprendizaje profundo, se están desarrollando modelos más eficaces para tecnologías de visión artificial. Una mejora sustancial en estos sistemas consiste en considerar las variaciones espaciales de posición y orientación para una identificación más precisa de objetos. La aplicación de estos algoritmos avanzados permite a los modelos diferenciar con mayor exactitud entre distintos objetos, como las diferentes clases de polen en este estudio.

Además, es importante reconocer las técnicas utilizadas por expertos palinólogos, que a menudo incluyen métodos de enfoque dinámico en el análisis de muestras. Esto sugiere que la metodología de análisis puede diferir del enfoque automatizado, el cual realiza barridos completos de imágenes y puede generar falsos positivos en caso de insuficiencia de datos para el entrenamiento, afectando la confiabilidad de la detección.

La colaboración con expertos en palinología es valiosa para incorporar conocimientos biológicos y ambientales en los modelos, permitiendo que estos se basen no solo en datos visuales sino también en un entendimiento contextual de las muestras. Esto puede mejorar significativamente la precisión en la clasificación e identificación del polen.

Todo lo expuesto anteriormente, también propone analizar estadísticamente el conteo sobre la presencia de polen de cada especie sobre una muestra, comprendiendo que palinológicamente existen cargas o densidades polínicas mayores en muestras de miel pura, las cuales, en teoría, debieran disminuir en presencia ante adulteraciones. Para esto es necesario tomar muestras adulteradas controladamente y realizar el conteo polínico, que estadísticamente puede contener diferencias relevantes contra el patrón puro de miel, que, ante el análisis por algoritmos de aprendizaje automático, haría posible el entrenamiento de un clasificador de visión artificial capaz de detectar adulteración mediante análisis de densidad de polen.

Finalmente, es crucial explorar la aplicación de estos sistemas avanzados en contextos reales, como la industria de la apicultura o estudios de biodiversidad. Esto permitiría evaluar su eficacia y realizar ajustes basados en aplicaciones prácticas. Dicha validación en entornos reales no solo confirma la utilidad de la tecnología, sino que también puede abrir nuevas áreas de investigación y oportunidades para mejorar los algoritmos y técnicas de procesamiento de datos.

8. CONCLUSIONES

Se demostró que, es posible utilizar algoritmos de aprendizaje automático para resolver desafíos de identificación de elementos químicos a través de LIBS, y dar con indicadores de autenticidad, detectando casos de fraude por tramos de adulteración, donde elementos presentes en azúcares pueden ser caracterizados y visualizados a través de esta técnica y procesamiento propuestos, siendo utilizados así en el proceso de creación de estos índices, acercando un método tecnológico que permitiría cumplir la legislación vigente, que exige normar la actividad apícola chilena.

El resultado preliminar de LIBS es examinado y evaluado en paralelo a la implementación de modelos de visión artificial, los cuales entrenados bajo imágenes de microscopía, aportan una perspectiva de origen botánico, reconociendo distintas especies o clases mediante el conteo de polen detectado, realizando una clasificación morfológica de las distintas muestras, donde los casos claros de adulteración no reflejan una cantidad mínima de polen o concentración polínica suficiente corroborados a través de validaciones externas de laboratorios de palinología y ecología vegetal, que son números rígidos para superar el umbral de clase. Este modelo propuesto, otorga posibilidades de optimizar la tarea de clasificación utilizando estimadores de pose antes de clasificar las muestras como tal, y así buscar elevar el rendimiento de clasificación en clases que no han obtenido un desempeño mínimo deseado, debido a su baja capacidad de inferencia para dichas clases.

REFERENCIAS

- [1]. Visciano, P. and Schirone, M. (2021) **“Food frauds: Global incidents and misleading situations”**, Trends in Food Science & Technology, 114, pp. 424–442. doi:10.1016/j.tifs.2021.06.010.
- [2]. Vinothkanna, A. et al. (2024) **“Advanced detection tools in food fraud: A systematic review for holistic and rational detection method based on research and patents”**, Food Chemistry, 446, p. 138893. doi:10.1016/j.foodchem.2024.138893.
- [3]. Molina, C. et al. (2023) **“Temporada apícola 2022 – 23”**, Oficina de Estudios y Políticas Agrarias, Ministerio de Agricultura, Chile.
- [4]. Shakoori, Z. et al. (2023) **“Assessing the quality of bee honey on the basis of Melissopalynology as well as chemical analysis”**, PLOS ONE, 18(8). doi:10.1371/journal.pone.0289702.
- [5]. Olsson, O. et al. (2021) **“Efficient, automated and robust pollen analysis using deep learning”**, Methods in Ecology and Evolution, 12(5), pp. 850–862. doi:10.1111/2041-210x.13575.
- [6]. Diehn, S., Zimmermann, B., Tafintseva, V. et al. (2020) **“Discrimination of grass pollen of different species by FTIR spectroscopy of individual pollen grains”**. Anal Bioanal Chem 412, 6459–6474. doi.org/10.1007/s00216-020-02628-2.
- [7]. Ren, J., Zhao, Y. and Yu, K. (2022) **“Libs in agriculture: A review focusing on revealing nutritional and toxic elements in soil, water, and crops”**, Computers and Electronics in Agriculture, 197, p. 106986. doi:10.1016/j.compag.2022.106986.
- [8]. Mureșan, C.I. et al. (2022) **“Honey Botanical Origin and honey-specific protein pattern: Characterization of some European honeys”**, LWT, 154, p. 112883. doi:10.1016/j.lwt.2021.112883.
- [9]. **Ley 21.489**, DE PROMOCIÓN, PROTECCIÓN Y FOMENTO DE LA ACTIVIDAD APÍCOLA, MINISTERIO DE AGRICULTURA DE CHILE, OCTUBRE 2022.
- [10]. **“Calidad de la miel producida en la región de Los Lagos”**, Iris Lobos, Investigadora INIA Remehue; Mariela Silva, Ing. en Alimentos INIA Remehue, INSTITUTO DE INVESTIGACIONES AGROPECUARIAS - INFORMATIVO N° 214 - AÑO 2019.
- [11]. **“Producción y caracterización de mieles monoflorales: una forma de diversificar los productos apícolas chilenos”**, M. Neira, J. Sánchez, L. Ciampi, C. Soto, Universidad Austral de Chile – Fundación de Innovación Agraria, 2004.
- [12]. **“Diversificación de la cadena apícola, a través de la valorización de pólenes apícolas producidos en Chile y la evaluación de sus propiedades biológicas específicas”**, G. Montenegro, X. Ortega, Pontificia Universidad Católica de Chile - Fundación de Innovación Agraria, 2009.
- [13]. D. A. Ruiz Yupangui, G. A. Insuasti Castelo, C. Pilamunga Capus, y E. V. Fonseca Navarrete, **“Afectación térmica de la miel de abeja en una metodología alternativa, monitoreando hidroximetilfurfural, número de diastasa y vitamina C”**, CCD, vol. 3, n.º 2.1, pp. 41-63, may 2020, doi: <https://doi.org/10.33262/concienciadigital.v3i2.1.1219>.

- [14]. A. Atiogbe Huzortey et al., "**445 nm Laser-Induced Fluorescence monitoring of 5-Hydroxymethylfurfural in Honey**," 2022 Conference on Lasers and Electro-Optics (CLEO), 2022, pp. 1-2.
- [15]. Bázár, G., Romvári, R., Szabó, A., Somogyi, T., Éles, V., Tsenkova, R. 2016. "**NIR detection of honey adulteration reveals differences in water spectral pattern**". *Food Chem*, 194, 873–880.
- [16]. Escuredo, O. et al. (2014) "**Contribution of botanical origin and sugar composition of honeys on the crystallization phenomenon**", *Food Chemistry*, 149, pp. 84–90. doi:10.1016/j.foodchem.2013.10.097.
- [17]. Al-Rubaie, Wissam K., Dhia F. Al-Fekaiki, Alaa Kareem Niamah, Deepak Kumar Verma, Smita Singh, and Ami R. Patel. 2024. "**Current Trends and Technological Advancements in the Study of Honey Bee-Derived Peptides with an Emphasis on State-of-the-Art Approaches: A Review**" *Separations* 11, no. 6: 166. <https://doi.org/10.3390/separations11060166>.
- [18]. Zúñiga-Díaz, D., Cetzal-Ix, W., López-Castilla, H. *et al.* "**A review of the melliferous flora of Yucatan peninsula, Mexico, on the basis for the honey production cycle**". *J Ethnobiology Ethnomedicine* **20**, 40 (2024). <https://doi.org/10.1186/s13002-024-00681-0>.
- [19]. Węglińska, M., Szostak, R., Kita, A., Nemś, A., & Mazurek, S. (2020). "**Determination of nutritional parameters of bee pollen by Raman and infrared spectroscopy**". *Talanta*, 212, 120790. <https://doi.org/10.1016/j.talanta.2020.120790>.
- [20]. Astolfi, G., Gonçalves, A., Menezes, G., Borges, F., Astolfi, A., Matsubara, E., Alvarez, M. and Pistori, H., "**POLLEN73S: An image dataset for pollen grains classification**", 2022, doi: <https://doi.org/10.1016/j.ecoinf.2020.101165>.
- [21]. S. Battiato, A. Ortis, F. Trenta, L. Ascari, M. Politi and C. Siniscalco, "**POLLEN13K: A Large-Scale Microscope Pollen Grain Image Dataset**," 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 2456-2460, doi: 10.1109/ICIP40778.2020.9190776.
- [22]. M. Boldeanu, H. Cucu, C. Burileanu and L. Mărmureanu, "**Automatic Pollen Classification Using Convolutional Neural Networks**" 2021 44th International Conference on Telecommunications and Signal Processing (TSP), 2021, pp. 130-133, doi: 10.1109/TSP52935.2021.9522626.
- [23]. J. C. da Silva Soares et al., "**Classification of pollen grain images with MobileNet**," 2021 XLVII Latin American Computing Conference (CLEI), 2021, pp. 1-10, doi: 10.1109/CLEI53233.2021.9639998.
- [24]. M. Boldeanu, M. González-Alonso, H. Cucu, C. Burileanu, J. M. Maya-Manzano and J. T. M. Buters, "**Automatic Pollen Classification and Segmentation Using U-Nets and Synthetic Data**" in *IEEE Access*, vol. 10, pp. 73675-73684, 2022, doi: 10.1109/ACCESS.2022.3189012.
- [25]. Handalage, Upulie & Kuganandamurthy, Lakshini. (2021). "**Real-Time Object Detection Using YOLO: A Review**". 10.13140/RG.2.2.24367.66723.
- [26]. N. Yang, V. Joos, A. -L. Jacquemart, C. Buyens and C. De Vleeschouwer, "**Using Pure Pollen Species When Training a CNN to Segment Pollen Mixtures**," 2022 IEEE/CVF Conference on Computer Vision

- and Pattern Recognition Workshops (CVPRW), 2022, pp. 1694-1703, doi: 10.1109/CVPRW56347.2022.00176.
- [27]. M. Boldeanu, H. Cucu, C. Burileanu and L. Marmureanu, “**Automatic Pollen Classification Using Convolutional Neural Networks**”, 2021 44th International Conference on Telecommunications and Signal Processing (TSP), 2021, pp. 130-133, doi: 10.1109/TSP52935.2021.9522626.
- [28]. N. Yang, V. Joos, A.-L. Jacquemart, C. Buyens and C. De Vleeschouwer, "Using Pure Pollen Species When Training a CNN to Segment Pollen Mixtures," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2022, pp. 1694-1703, doi: 10.1109/CVPRW56347.2022.00176.
- [29]. M. Boldeanu, C. Marin, D. Ene, L. Marmureanu, H. Cucu and C. Burileanu, "MARS: the First Romanian Pollen Dataset using a Rapid-E Particle Analyzer," 2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), 2021, pp. 145-150, doi: 10.1109/SpeD53181.2021.9587447.
- [30]. M. D’Orazio et al., "Electro-Optical Classification of Pollen Grains via Microfluidics and Machine Learning," in IEEE Transactions on Biomedical Engineering, vol. 69, no. 2, pp. 921-931, Feb. 2022, doi: 10.1109/TBME.2021.3109384.
- [31]. A. Di Menno di Bucchianico et al., “**Combined effects of air pollution and allergens in the city of Rome**,” Urban For. Urban Green., vol. 37, pp. 13–23, 2019, doi: 10.1016/j.ufug.2018.04.001.
- [32]. A. Mishra, Z. M. Jalaluddin and C. V. Mahamuni, "Air Quality Analysis and Smog Detection in Smart Cities for Safer Transport using Machine Learning (ML) Regression Models," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022, pp. 200-206, doi: 10.1109/CSNT54456.2022.9787618.
- [33]. N. V. R. Masna, S. D. Paul, C. Chen, S. Mandal, and S. Bhunia, “**Eat, but verify: Low-cost portable devices for food safety analysis**,” IEEE Consum. Electron. Mag., vol. 8, no. 1, pp. 12–18, Jan. 2019.
- [34]. Ali, H., Rafique, K., Ullah, R. et al. **Classification of Sidr honey and detection of sugar adulteration using right angle fluorescence spectroscopy and chemometrics**. Eur Food Res Technol 248, 1823–1829 (2022), doi: 10.1007/s00217-022-04008-9.
- [35]. Sahlan M, Karwita S, Gozan M, Hermansyah H, Yohda M, Yoo YJ, Pratami DK. **Identification and classification of honey's authenticity by attenuated total reflectance Fourier-transform infrared spectroscopy and chemometric method**. Vet World. 2019 Aug;12(8):1304-1310. doi: 10.14202/vetworld.2019.1304-1310. Epub 2019 Aug 23. PMID: 31641312; PMCID: PMC6755387.
- [36]. X. Zhang, S. Bao, C. Yang, S. Mad and J. Yang, "An External Contour Extraction Method for Pollen Images", 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 2021, pp. 691-695, doi: 10.1109/ICAICE54393.2021.00135.
- [37]. Bleha, R.; Shevtsova, T.V.; Živčáková, M.; Korbářová, A.; Ježková, M.; Saloň, I.; Brindza, J.; Synytsya, A. **Spectroscopic Discrimination of Bee Pollen by Composition, Color, and Botanical Origin**. Foods 2021, 10, 1682, doi: <https://doi.org/10.3390/foods10081682>

- [38]. Franco Rivas, Francisco Pérez, Claudio Sandoval, Ignacio Sanhueza, Benjamín Sepúlveda, Jorge Yañez, and Sergio Torres, "**Copper concentrate dual-band joint classification using reflectance hyperspectral images in the VIS-NIR and SWIR bands**," *Appl. Opt.* 62, 2970-2977 (2023), doi: <https://doi.org/10.1364/AO.477193>.
- [39]. A. H. Saputro and C. Aprichilia, "**Classification System of Honey Floral Origin based on Visual Near-Infrared Imaging**," 2019 International Conference on Sustainable Information Engineering and Technology (SIET), Lombok, Indonesia, 2019, pp. 125-129, doi: 10.1109/SIET48054.2019.8986024.
- [40]. Dimitrios Stefas et.al. (2020) "**Laser induced breakdown spectroscopy for elemental analysis and discrimination of Honey Samples**", *Spectrochimica Acta Part B: Atomic Spectroscopy*, doi: <https://doi.org/10.1016/j.sab.2020.105969>.
- [41]. C. Fabre, "**Advances in Laser-Induced Breakdown Spectroscopy analysis for geology: A critical review**", *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2020, doi: <https://doi.org/10.1016/j.sab.2020.105799>.
- [42]. Nespeca, M.G. et al. "**Detection and quantification of adulterants in honey by Libs**", *Food Chemistry*, 311, p. 125886. 2020. doi:10.1016/j.foodchem.2019.125886.
- [43]. Stefas, D. et al. "**Honey discrimination based on the bee feeding by laser induced breakdown spectroscopy**", *Food Control*, 134, p. 108770. 2022. doi:10.1016/j.foodcont.2021.108770.
- [44]. Valiente, J.M. et al. "**Automatic pollen recognition using convolutional neural networks: The case of the main pollens present in Spanish citrus and Rosemary Honey**", *Journal of Food Composition and Analysis*, 123, p. 105605. 2023. doi:10.1016/j.jfca.2023.105605.
- [45]. S. Richiero, C. Sandoval, C. Oberlin, A. Schmitt, J.-C. Lefevre, A. Bensalah-Ledoux, D. Prigent, C. Coquidé, A. Valois, F. Giletti, F. Pelascini, L. Duponchel, P. Dugourd, C. Comby-Zerbino, and V. Motto-Ros, "**Archaeological Mortar Characterization Using Laser-Induced Breakdown Spectroscopy (LIBS) Imaging Microscopy**," *Applied Spectroscopy*, vol. 76, no. 8, pp. 978–987, 2022. DOI: 10.1177/00037028211071141.
- [46]. T.A. Alrebdi, A. Fayyaz, H. Asghar, A. Zaman, M. Asghar, F.H. Alkallas, A. Hussain, J. Iqbal, and W. Khan, "**Quantification of Aluminum Gallium Arsenide (AlGaAs) Wafer Plasma Using Calibration-Free Laser-Induced Breakdown Spectroscopy (CF-LIBS)**," *Molecules*, vol. 27, p. 3754, 2022. DOI: 10.3390/molecules27123754.
- [47]. N. Kumar et al., "**Discrimination of Various pure Honey samples and its Adulterants using FTIR Spectroscopy Coupled with Chemometrics**," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 808-811, doi: 10.1109/ICACCS51430.2021.9441971.
- [48]. S. Das, D. K. Agarwal, A. Patnaik, V. R. Rao and T. Kundu, "**Study of Gold nanoparticles binding dynamics on functionalized surface by ATR-evanescent wave absorption method**," 2022 Workshop on Recent Advances in Photonics (WRAP), Mumbai, India, 2022, pp. 1-2, doi: 10.1109/WRAP54064.2022.9758288.

- [49]. Bertelli, D., Lolli, M., Papotti, G., Bortolotti, L., Serra, G., Plessi, M. 2010. “**Detection of Honey Adulteration by Sugar Syrups Using One-Dimensional and Two-Dimensional High-Resolution Nuclear Magnetic Resonance**”. *J. Agric. Food Chem*, 58, 8495–8501.
- [50]. Rosalie A. Multari, David A. Cremers, Jo Anne M. Dupre, and John E. Gustafson, 2013, “**Detection of Biological Contaminants on Foods and Food Surfaces Using Laser-Induced Breakdown Spectroscopy (LIBS)**”, *Journal of Agricultural and Food Chemistry* 61 (36), 8687-8694, doi: 10.1021/jf4029317.
- [51]. Megherbi, M., Herbreteau, R., Faure, R., Dessalces, G., Grenier-Loustalot, M.F. 2008. “**Solid Phase Extraction of Oligo- and Polysaccharides: Application to Maltodextrins and Honey Qualitative Analysis**”. *J Liq Chrom & Rel Tech*, 31,7, 1033-1046
- [52]. Spiteri, M., Jamin, E., Thomas, F., Rebours, A., Lees, M., Rogers, K.M., Rutledge, D.N. 2015. “**Fast and global authenticity screening of honey using 1H-NMR profiling**”. *Food Chem*, 189, 60–66.
- [53]. Wang, S., Guo, Q., Wang, L., Lin, L., Shi, H., Cao, H., Cao, B. 2015. “**Detection of honey adulteration with starch syrup by high performance liquid chromatography**”. *Food Chem*, 172, 669-674.
- [54]. Skaff, W, El Hajj, R, Hanna-Wakim, L, Estephan, N. “**Detection of adulteration in honey by infrared spectroscopy and chemometrics: Effect on human health. Journal of Food Processing and Preservation**”, 2022, 46, e15438, doi: <https://doi.org/10.1111/jfpp.15438>.
- [55]. P. Mekha, N. Teeyasuksaet, T. Sompowloy and K. Osathanunkul, "Honeybee Sound Classification Using Spectrogram Image Features," 2022 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), 2022, pp. 205-209, doi: 10.1109/ECTIDAMTNCON53731.2022.9720352.
- [56]. Li, X., Prebble, J., de Lange, P., Raine, J. and Newstrom-Lloyd, L., 2022. “**Discrimination of pollen of New Zealand mānuka (*Leptospermum scoparium* agg.) and kānuka (*Kunzea* spp.) (Myrtaceae)**”, doi: <https://doi.org/10.1371/journal.pone.0269361>.
- [57]. V. Stojnić, V. Risojević and R. Pilipović, "Detection of pollen bearing honeybees in hive entrance images," 2018 17th International Symposium INFOTEH-JAHORINA (INFOTEH), 2018, pp. 1-4, doi: 10.1109/INFOTEH.2018.8345546.
- [58]. Huang, Z.; Liu, L.; Li, G.; Li, H.; Ye, D.; Li, X. “**Nondestructive Determination of Diastase Activity of Honey Based on Visible and Near-Infrared Spectroscopy.**” *Molecules* 2019, 24, 1244, doi: <https://doi.org/10.3390/molecules24071244>.
- [59]. Hudz N, Leontiev D, Wiczorek PP (2019) “**Spectral characteristics of 5-hydroxymethylfurfural as a related substance in medicinal products containing glucose**”. *Pharmacia* 66(3): 121-125, doi: 10.3897/pharmacia.66.e35969.
- [60]. Bunaciu, A.A. and Aboul-Enein, H.Y. (2022) “**Honey discrimination using Fourier transform-infrared spectroscopy**, doi: <https://doi.org/10.3390/chemistry4030060>.

- [61]. NIR Spectroscopy “**A guide to near-infrared spectroscopic analysis of industrial manufacturing processes**”, Metrohm industries, 8.108.5026EN – 2013-02.
- [62]. M. A. S. Rünzel, E. E. Hassler, R. E. L. Rogers, G. Formato and J. A. Cazier, “**Designing a Smart Honey Supply Chain for Sustainable Development**,” in IEEE Consumer Electronics Magazine, vol. 10, no. 4, pp. 69-78, 1 July 2021, doi: 10.1109/MCE.2021.3059955.
- [63]. Global Alliance of Chronic Respiratory Diseases, “**COVID-19 pandemic alert: Time to focus on lung health-Beijing call to action for lung health promotion**,” J. Thorac. Dis., vol. 12, no. 6, pp. 3238–3241, 2020, doi: 10.21037/jtd-gard-20-002.
- [64]. von Chamier, L., Laine, R.F., Jukkala, J. et al. **Democratising deep learning for microscopy with ZeroCostDL4Mic**. Nat Commun 12, 2276 (2021), doi: 10.1038/s41467-021-22518-0.
- [65]. Monteiro, F.C., Pinto, C.M., Rufino, J. (2021). “**Deep Learning Recognition of a Large Number of Pollen Grain Types**”. In: Pereira, A.I., et al. Optimization, Learning Algorithms and Applications. OL2A 2021. Communications in Computer and Information Science, vol 1488. Springer, Cham. https://doi.org/10.1007/978-3-030-91885-9_28.
- [66]. Kubera, E., Kubik-Komar, A., Kurasiński, P., Piotrowska-Weryszko, K. and Skrzypiec, M., 2022. “**Detection and Recognition of Pollen Grains in Multilabel Microscopic Images**”, doi: <https://doi.org/10.3390/s22072690>.
- [67]. Nadine, S.M., Hanna, M., Paul, S, et al. (2021) “**A systematic review of threshold values of pollen concentrations for symptoms of allergy. Aerobiologia.**”, 37(3): 395-424.
- [68]. M.N. Khan, Q. Wang, B.S. Idrees, W. Xiangli, G. Teng, X. Cui, Z. Zhao, K. Wei, and M. Abrar, “**A Review on Laser-Induced Breakdown Spectroscopy in Different Cancers Diagnosis and Classification**,” Frontiers in Physics, vol. 10, p. 821057, 2022. DOI: 10.3389/fphy.2022.821057.
- [69]. Trenta, F., Ortis, A. and Battiato, S., 2022. “**Fine-Grained Image Classification for Pollen Grain Microscope Images**”, doi: https://doi.org/10.1007/978-3-030-89128-2_33.
- [70]. X. Hao, S. Shoucheng and J. Yujie, “**Pollen texture migration deblurring method based on self-coding generator**,” 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2022, pp. 1269-1273, doi: 10.1109/ITAIC54216.2022.9836728.

