



UNIVERSIDAD NACIONAL DE COLOMBIA

Detección de antracnosis foliar en plantas de mango por medio de análisis de imágenes hiperespectrales

Oscar Fernando Penagos Espinel

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de ingeniería eléctrica y electrónica
Bogotá D.C., Colombia
2021

Detección de antracnosis foliar en plantas de mango por medio de análisis de imágenes hiperespectrales

Oscar Fernando Penagos Espinel

Trabajo de grado presentada como requisito parcial para optar al título de:
Magister en Ingeniería Electrónica

Director:
Ph.D Flavio Augusto Prieto Ortíz

Línea de Investigación:
Tratamiento de señales y procesamiento de imágenes
Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de ingeniería eléctrica y electrónica
Bogotá D.C., Colombia

2021

Agradecimientos

Agradezco a mi familia por el apoyo recibido en este proceso tan importante en mi vida, al director del proyecto, el profesor Flavio Augusto Prieto Ortiz Ph.D, por su mentor a y seguimiento oportuno, al estudiante de doctorado Carlos Alberto Velasquez por su ayuda y orientacion en el proyecto, y a la Universidad Nacional de Colombia por concederme la oportunidad de estudiar en tan prestigiosa institucion y aumentar mi formacion academica y personal.

Resumen

Detección de antracnosis foliar en plantas de mango por medio de análisis de imágenes hiperespectrales

La antracnosis es una enfermedad que afecta una gran cantidad de plantas, es producida principalmente por el hongo *Colletotrichum*, dada su fácil propagación por el aire, su abundante presencia en climas cálidos o en temporadas de alta temperatura como países con estaciones de verano y primavera, y su control mediante agentes químicos, genera la muerte de numerosas plantas y con ello grandes pérdidas económicas afectando principalmente a agricultores. El objetivo de este proyecto es evaluar el uso de imágenes hiperespectrales como apoyo a la detección de antracnosis, para ello se hace uso de métodos de clasificación tradicionales como LDA y KNN y enfocados en datos espectrales como SAM, además del reductor de dimensionalidad PCA con el fin de encontrar la mínima cantidad de bandas espectrales necesarias para detectar la presencia del patógeno.

Palabras clave: (Imágenes hiperespectrales, Antracnosis, Métodos de clasificación supervisada, PCA, SAM, KNN, LDA.).

Abstract

Detection of foliar anthracnose on mango plants by hyperspectral image analysis

Anthracnose is a disease that affects a large number of plants, it is mainly produced by the fungus *Colletotrichum*, given its easy propagation through the air, its abundant presence in hot climates or in seasons of high temperature such as summer and spring seasons, and its control by chemical agents, generates the death of numerous plants and with it great economic losses affecting mainly farmers. The objective of this project is to evaluate the use of hyperspectral images to support the detection of anthracnose, using traditional classification methods such as LDA and KNN and focused on spectral data such as SAM, in addition to the dimensionality reducer PCA in order to find the minimum number of spectral bands necessary to detect the presence of the pathogen.

Keywords: Hyperspectral Image, Anthracnose, Unsupervised Classification Methods, PCA, SAM, KNN, LDA.)

Contenido

Agradecimientos	IV
Resumen	V
Abstract	VI
1 Introduccion	2
1.1 Antecedentes	3
1.2 Materiales y metodos	8
1.2.1 Analisis discriminante lineal	8
1.2.2 Analisis discriminante paso a paso	9
1.2.3 Mapeador de angulo espectral (SAM)	10
2 Adquisicion de imagenes hiperespectrales en campo	11
2.1 Antecedentes	11
2.1.1 Aspectos principales para la adquisicion de las HSI	11
2.1.2 Iluminacion	12
2.2 Caracter sticas de la imagen hiperespectral	12
2.2.1 Plataforma de la camara	13
2.2.2 Calibracion	14
2.2.3 Toma de datos	15
2.3 Desarrollo del protocolo	15
3 Resultados	17
3.1 Pruebas del experimento 1	18
3.2 Clas cacion general	18
3.2.1 Clasi cacion sectorizada	19
3.2.2 Reductor de dimensionalidad	22
3.2.3 Clasi cacion por d as	25
3.3 Pruebas del experimento 2	27
3.3.1 Clasi cacion general	28
3.3.2 Clasi cacion sectorizada	28
3.3.3 Reductor de dimensionalidad	30
3.4 Resultados adicionales	33

4 Conclusiones

34

Bibliografía

37

1 Introduccion

La atracnosis es una enfermedad en las plantas causada por hongos de los generos *Colletotrichum*, *Gloesporium* y *Coniothyrium* [11], se encuentra principalmente en climas humedos y calidos o incluso temporalmente en las estaciones de verano y primavera, por lo se puede encontrar en muchos pa ses del mundo, afecta un gran numero de frutales, arboles, arbus-tivas, trepadoras, hort colas y verduras, entre otros. Adicionalmente se propaga facilmente por medio del aire y puede causar muerte del tejido vegetal de la planta. Si no se detecta a tiempo causa perdidas economicas considerables.

A pesar de que los s ntomas de la enfermedad dependen de la especie de planta afectada [15], se suele evidenciar en las hojas como manchas amarillas o marrones formando c rculos concentricos sobre el area de infeccion. Existen varias medidas preventivas para la enfermedad como lo son el cambio de ubicacion de plantas que presenten s ntomas de la enfermedad, reduccion de abonados ricos en nitrogeno, desinfeccion frecuente de herramientas de jar-diner a y cultivo, veri car las condiciones espec cas del riego que requiere cada planta, desinfeccion del suelo antes de plantar, remover plantas silvestres, no mojar la parte aerea de las plantas, entre otros. Adem as existen acciones mas limitadas que puede contribuir como medidas curativas tales como cortar y destruir las partes infectadas de plantas que presenten s ntomas y el uso de fungicidas y control qu mico [11], es importante destacar que existen diversos fungicidas para el control de la enfermedad y sus efectos var an en la especie del hongo causante de la enfermedad.

A partir de estos factores se puede resumir la problematica en cuatro factores:

- Enfermedad que se propaga facilmente por el aire y se encuentra principalmente en clima humedos y calurosos (presente en gran parte del territorio colombiano), o durante el verano y primavera en pa ses con estaciones.
- Causa destruccion del tejido vegetal si no se detecta en etapas tempranas.
- Puede generar grandes perdidas economicas a agricultores debido a los dos factores anteriormente mencionados.
- Es dependiente del control qu mico mediante el uso de fungicidas, el cual puede ser perjudicial para el mismo cultivo a tratar o para cultivos adyacentes.

Una de las soluciones desarrolladas en los últimos años para la detección oportuna de la enfermedad es el procesamiento de imágenes, especialmente en el análisis de imágenes hiperespectrales de las hojas de algunas plantas [14], [28], [1] y [23].

Una imagen hiperespectral es la representación de una escena a lo largo del espectro electromagnético, los datos son guardados a partir de múltiples imágenes en la que cada una representara un rango o una banda del espectro, se compone de tres ejes para su lectura, dos espaciales (x y y) y un eje espectral (λ).

Los rangos del espectro que pueden abarcar al igual que la resolución depende directamente de la cámara hiperespectral y sus rangos pueden ir desde los 400 nm hasta los 2000 a 2500 nm con pasos de 8 nm a 10 nm . Sin embargo el uso de datos hiperespectrales requiere un gran costo computacional por lo que es importante diseñar métodos que permitan disminuir el procesamiento necesario para el análisis, manteniendo la precisión de clasificación lo más alto posible. Sus aplicaciones principales en Colombia es el sensado de áreas mineras, agrícolas y adquisición de mapas geológicos, esto lo utilizan para determinar el tamaño, la posición y en el caso de áreas agrícolas el tipo de cultivo.

1.1. Antecedentes

A continuación se nombrarán algunos trabajos para la detección de antracnosis principalmente en cultivos de fresa, en los cuales se enfocará el sistema de clasificación utilizado, el método de análisis para la selección de bandas y la precisión.

En [14] se enfocan especialmente en la detección de la antracnosis en la fresa, su motivación en la investigación se basa en la alta velocidad de propagación de la enfermedad, justificando que detectar la enfermedad en etapas tempranas es de suma importancia. Para ello utilizaron imágenes hiperespectrales de varias plantas de fresa y las dividieron en tres grupos, infectadas pero asintomáticas, infectadas o sintomáticas, luego implementaron tres modelos de clasificación basados en análisis discriminante por pasos (SDA), análisis discriminante de Fisher (FDA) y los algoritmos k -vecinos más cercanos (kNN), Las pruebas que realizaron las dividieron en dos grupos, en interiores y en campo, obteniendo precisiones entre 78 % y 77 % en campo, mientras que en el interior la precisión varía entre 84 y 93 %.

Otro referente en la detección de antracnosis en las plantas de fresa es [28], de igual forma utilizaron un sistema de clasificación basado en un mapeador de ángulo espectral (SAM), análisis discriminante por pasos (SDA) y medida de correlación (CM), a diferencia de [14] esta investigación se enfocó en detectar la etapa de la enfermedad en plantas contagiadas y las dividieron en tres categorías, saludable, de incubación y sintomática. La investigación

concluyo con una precision en la deteccion de la etapa de la enfermedad de aproximadamente 80 % bajo los metodos nombrados, a excepcion del sistema basado en SDA el cual logro una precision del 93 %. Adicionalmente encontraron que las longitudes de onda mas adecuadas en las imagenes espectrales para la deteccion de la etapa de la enfermedad son 551, 706, 750 y 914.

En [1] utilizan analisis discriminante lineal (LDA) y SIMCA sobre imagenes hiperespectrales para la deteccion de dos tipos de hongo de difcil deteccion, el *Colletotrichum gossypii* y el cephalosporioide, adicionalmente utilizaron el modelo SPA para el tratamiento de los datos para la seleccion de variables y de esta forma aumentar la eficiencia en el sistema de clasificacion. Para este caso se encontro que la mejor longitud de onda para la deteccion del hongo esta entre 1000 y 1381 nm, ademas no se necesito tratamiento previo en las imagenes hiperespectrales. Bajo estas condiciones obtuvieron una precision del 95 %.

Los documentos [14], [28] y [1] reflejan la importancia de la deteccion en etapas tempranas de enfermedades en las plantas con una precision superior al 80 %, destacando la alta eficiencia de los sistemas de clasificacion basados en analisis discriminante lineal, principalmente el LDA (Linear Discriminant Analysis) y SDA (Stepwise Discriminant Analysis).

En [2] se enfocan en la deteccion pudricion basal del tallo (BSR) a partir del analisis de imagenes hiperespectrales, este patogeno es la principal amenaza en plantaciones de palma de aceite, las imagenes hiperespectrales utilizadas tienen una vista panoramica del campo de plantacion logrando alcanzar una cobertura de $5000m^2$, para ello se requiere elevar la camara hiperespectral a una altura de $1000m$. Para la clasificacion utilizaron el metodo SAM y la verificacion de la presencia de la enfermedad la obtuvieron mediante el uso del metodo medio selectivo de Ganoderma (GSM). La precision obtenida bajo esas consideraciones fue de 96 %, el estudio lo realizaron para plantaciones con un total de 644.325 palmas de las cuales el 1,3 % (6.789) resultaron tener la enfermedad.

Por otra parte, en [23] realizan una revision sobre los diferentes metodos para la seleccion de bandas en imagenes hiperespectrales. La seleccion de bandas tiene como objetivo seleccionar un pequeno subconjunto de bandas para eliminar la redundancia espectral y reducir los costos computacionales mientras se preserva la informacion espectral significativa, en el documento evalua el rendimiento de varios metodos de seleccion de banda, los cuales se dividen en seis categorias, metodos basados en clasificacion, en busqueda, basada en agrupacion, en dispersion, en aprendizaje incrustado y en esquema hibrido. Ellos encuentran que los metodos basados en clasificacion y agrupacion son en general mas eficientes que los demas, sin embargo los resultados varian de acuerdo a la complejidad de la escena y la dificultad de clasificacion, Los metodos basados en clasificacion tienen una baja complejidad computacional y son adecuados para conjuntos de datos hiperespectrales mas grandes, pero las

precisiones de clasificación de sus bandas seleccionadas no son satisfactorias, mientras los métodos basados en búsquedas, en clusters y en aprendizaje profundo generalmente requieren costos computacionales más altos por lo que son más adecuados para conjuntos de datos hiperespectrales más pequeños. Por otra parte, si los datos poseen bandas hiperespectrales ruidosas, los métodos MSNRPCA, OSP y la selección de banda de ruido mínimo tienen un mejor desempeño.

En [26] evalúan la clasificación de imágenes hiperespectrales a través del método SAM (Optimized Spectral Angle) para la presencia de contaminación por aflatoxinas en granos de maíz, para ello utilizaron imágenes hiperespectrales de fluorescencia junto con el método analítico VICAM para fines de cuantificación. La clasificación la dividieron en dos grupos, sintomático y asintomático. El proceso de clasificación que implementaron se basa en la siguiente expresión:

$$= \cos^{-1} \frac{X \cdot R}{\sqrt{X \cdot X} \sqrt{R \cdot R}} = \cos^{-1} \frac{\sum_{i=1}^n x_i r_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n r_i^2}}$$

donde x y r son los píxeles y los espectros de referencia respectivamente y n es el número de bandas de la imagen. La clasificación en dos clases (sintomático y asintomático) se logra mediante el uso de dos umbrales previamente establecidos por la FDA (Food and Drug Administration), los cuales son 20 y 100 ppb (partes por millón). Utilizaron un total de 504 kernels los cuales los dividieron en 252 para desarrollar los grupos de entrenamiento y los otros 252 para la validación del algoritmo. La precisión de la clasificación obtenida fue del 86;1% para el umbral de 20 ppb y del 88;1% para el umbral de 100 ppb.

En [10] hace un estado del arte sobre métodos de detección para la selección de banda sobre imágenes hiperespectrales para la clasificación de plantas, en el documento se especifica la variación de frecuencia de selección de banda de onda por diferentes factores como grupo taxonómico, estructural o funcional de la planta, además de la influencia de la forma de selección de banda al aplicarse sobre métodos de clasificación como SDA, FDA (Fisher Discriminant Analysis), LDA, SVM (Support Vector Machine) y RM (Random Forest). Algunos de los métodos principales utilizados en el estudio son PCA (Principal Component Analysis), CFS (Correlation-based Feature Selection), SAM (Spectral Angle Mapper), PLS-DA (Partial Least Squares), SPA (Successive Progressions Algorithm), entre otros. En sus conclusiones no resaltan que exista algún método en general para procesos de selección de banda debido a la gran variabilidad de los datos, y establecen que no existe un único método de selección de las mejores características, el funcionamiento de cada método varía según sea la aplicación.

El autor en [7] desarrolla un estado del arte en el que se establece la descripción matemática

de diferentes metodos de clasificacion aplicables a imagenes hiperespectrales, para ello establecieron seis categorias en las que se agrupan diferentes metodos de acuerdo a su modo de operacion.

1. Clasificacion no supervisada
 - PCA
 - ICA (Independent component analysis)
2. Clasificacion supervisada
 - Clasificador ML (Maximum Likelihood)
 - kNN (k-nearest-neighbor)
 - SAM
 - SVM
3. Random Forest
 - CART (Classification And Regression Tree)
 - BHC (Binary Hierarchy Classifier)
4. Clasificacion espacio-espectral
5. *Sparse-representation*
 - Metodologias basadas en segmentacion
6. *Deep Learning*
 - *Deep Learning* para clasificacion HSI
 - CNN (Redes Neuronales Convolucionales)

En el documento especifican que los principales inconvenientes al trabajar con este tipo de imagenes son la dimensionalidad (fenomeno de Hughes) y la falta de muestras para un adecuado proceso de entrenamiento.

Clasificación no supervisada: Son metodos basados en extraccion de caracteristicas a partir de la reduccion de dimensionalidad que no requieren ningun conocimiento previo para entrenar los datos, su funcionamiento radica en el agrupamiento automatico de los pixels con caracteristicas espectrales similares (como la media o la desviacion estandar).

Clasificación supervisada: Se realiza un entrenamiento a partir de datos etiquetados (conocimiento previo del grupo al que pertenecen), se caracteriza por aprovechar mejor la informacion espectral de las imagenes. Su funcionamiento se puede derivar en seis pasos:

- Adquisición de las imágenes hiperespectrales
- Preprocesamiento
- Selección y extracción de características espaciales espectrales
- Clasificación
- Salida y optimización
- Evaluación de rendimiento

Random Forest: Son clasificadores basados en árboles donde cada árbol se entrena con un conjunto de datos de entrenamiento. La clasificación dada por cada árbol se conoce como un voto para esa clase, el bosque elige la clase que tiene más votos. En general, los clasificadores RFC presenta dos características principales: precisión relativamente alta y velocidad de procesamiento.

Clasificación espacio-espectral: La clasificación espectral-espacial se basa en un sistema de dependencia espacial, la dependencia espacial es transportada por dos identidades llamadas píxel y etiqueta asociada. La correlación entre los píxeles relacionados espacialmente es la dependencia espacial, por lo tanto, los píxeles relacionados espacialmente se denominan píxeles vecinos. La dependencia del píxel indica la correlación de los píxeles vecinos y la dependencia de la etiqueta indica la correlación de las etiquetas de los píxeles vecinos.

Sparse-representation: Este método está orientado a aplicaciones con escasez en muestras, para ello supone que las muestras de entrenamiento se pueden representar como una combinación lineal del menor número posible de columnas. El principio de clasificación se basa en la representación de un píxel hiperespectral mediante una combinación lineal de unas pocas muestras de entrenamiento. La prueba de evaluación se guarda en un vector de dispersión el cual se recupera resolviendo un problema de optimización restringido por escasez, y puede determinar directamente la etiqueta de clase de la muestra de prueba.

Deep Learning: Representan modelos de entrenamiento jerárquico de características profundas de datos de entrada en redes neuronales, generalmente con más tres capas, la red se inicializa primero en capas a través de capacitación no supervisada y, posteriormente, se sintoniza de manera supervisada. En este esquema, las características de alto nivel se aprenden de las de bajo nivel, mientras que las características adecuadas se pueden formular para la clasificación de patrones hacia el final.

En [8] se analiza el uso de las imágenes hiperespectral para la estimación de la cantidad de fruta en bloques de huertos, específicamente que si bien el análisis de imágenes RGB para esta

tarea ha demostrado tener muy buenos resultados, el uso de imagenes hiperspectrales puede realizar esa misma funcion y adicionalmente provee la posibilidad de trabajar con tareas matizadas como la deteccion de enfermedades, la clasificacion de cultivares y la estimacion de la madurez de los frutos. El documento sustenta el uso de HSI para realizar multiples tareas con el uso de un solo sensor. El procesamiento de las imagenes inicia con la correccion de datos a partir de los calculos de la reflectancia, seguido de la segmentacion para delimitar los arboles, posteriormente la deteccion y conteo de los arboles de mango, en esta etapa se realizan procesos adicionales de filtrado, dilatacion y erosion, transformaciones de distancia e identificacion de maximos locales, por ultimo establecen parametros de optimizacion en procesos morfologicos previamente realizados, tales como la dilatacion y erosion. En el documento se especifica un alto rendimiento del algoritmo para la estimacion de la cantidad de fruta en bloques de huertos usando imagenes RGB, sin embargo, las HSI pueden desempeñar multiples aplicaciones adicionales, siendo una opcion de amplia utilidad en huertos comerciales.

1.2. Materiales y metodos

Los metodos a utilizar se basan en analisis discriminante, detectores, clasificadores y selectores de banda, estos procesos son utilizados comunmente como bases para tecnicas de seleccion de bandas espectrales y sistemas de clasificacion, a continuacion se nombraran algunos de los metodos tradicionales.

1.2.1. Analisis discriminante lineal

El analisis discriminante lineal es un metodo de clasificacion de objetos supervisado, en la cual cada objeto debe pertenecer a un solo grupo, La funcion de un analisis discriminante es poder predecir el grupo de procedencia de un objeto basado en variables o criterios cuantificables. El LDA esta basado en el teorema de Bayes, ademas existen dos enfoques en la clasificacion a partir de analisis discriminante [6]:

- Basado en la obtencion de funciones discriminantes de calculo, similar a las ecuaciones de regresion lineal multiple.
- Empleando tecnicas de correlacion canonica y de componentes principales, denominado analisis discriminante canonico.

Teorema de Bayes

Dados dos eventos A y B , la probabilidad de B dado el evento A esta dado por:

$$P[B|A] = \frac{P(AB)}{P(A)}$$

Utilizando el teorema de Bayes se puede establecer los criterios para el discriminante lineal [18]

$$P[Y = k | X = x] = \frac{P_k P(X = x | Y = k)}{\sum_{j=1}^K P_j P(X = x | Y = j)} = \frac{f_k(x)}{\sum_{j=1}^K f_j(x)}$$

donde k es la clase a comparar y x el valor observado. La clasificación para un objeto x_i será la clase i con el máximo valor obtenido a la salida.

La estimación de f_k viene dada por la siguiente ecuación:

$$f_k(X) = P(Y = k | X = x) = \frac{1}{\sigma_k} e^{-\frac{1}{2\sigma_k^2}(x - \mu_k)^2}$$

donde μ_k es la media y σ_k^2 la varianza para la clase k .

1.2.2. Analisis discriminante paso a paso

El analisis discriminante paso a paso (SDA) es un metodo en el cual una variable puede entrar o salir de un grupo en cada paso, es de gran utilidad cuando el numero de variables clasificatorias es alto. Este metodo permite clasificar las variables con mas capacidad discriminante entre un conjunto de variables mas amplio [19]. El criterio con el que un valor puede cambiar de grupo en el proceso de clasificación depende del estadístico que se utilice, el mas comun es utilizar el parametro estadístico F de la funcion Lambda de Wilks.

El SDA tiene tres implicaciones [18]:

- Las variables que se utilizaran serán las que mas separan los grupos.
- Todas las variables a utilizar deben cumplir un criterio de entrada.
- Se eliminaran las variables que cumplen un criterio de salida.

Esto significa que si el valor del estadístico F es mayor que el valor de entrada es admitido pero se elimina si es menor que el valor de salida.

Criterio de Lambda Wilks

El estadístico Lambda Wilks es el criterio mas utilizado en el SDA, mide las desviaciones de cada grupo con respecto a las desviaciones totales, se representa por un valor entre 0 y 1. Las variables con un valor Lambda Wilks cercano a cero serán las que mas diferencien a

los grupos. Si su valor es cercano a 1 los grupos estaran mezclados, por lo que las variables carecen de capacidad discriminante [19].

Otros estadísticos aplicables al analisis discriminante paso a paso [4] son:

- Varianza no explicada.
- Distancia de Mahalanobis.
- Menor razon F.
- V de Rao.

1.2.3. Mapeador de angulo espectral (SAM)

SAM es un metodo de clasificacion supervisado adaptado a datos hiperespectrales que ofrece una metrica de distancia comun [3], en la que compara un espectro de p xeles desconocido t con espectros de referencia r_i , con $i = 1; \dots; K$, para cada una de K referencias y asigna al material que tiene la distancia mas pequena:

$$\text{Class}(t) = \text{argmind}(t; r_i) \quad 1 \leq i \leq K$$

Los espectros de reflectancia de cada pixel se pueden describir como vectores en un espacio de n dimensiones, donde n corresponde al numero de bandas espectrales. La longitud del vector representa el brillo del pixel mientras que la direccion representa la caracteristica espectral del pixel, la variacion en la iluminacion afecta principalmente los cambios en la longitud del vector, mientras que la variabilidad espectral entre diferentes espectros afecta el angulo entre sus vectores correspondientes.

La clasificacion se realiza calculando los angulos espectrales entre el espectro de reflectancia del pixel objetivo y los espectros de referencia. Cada pixel se asignara a la clase de acuerdo con el valor del angulo espectral mas bajo.

2 Adquisición de imágenes hiperespectrales en campo

2.1. Antecedentes

El proceso de adquisición de datos hiperespectrales es de suma importancia en el desarrollo del proyecto, un correcto protocolo de adquisición puede mejorar los resultados en la fase de clasificación al aumentar la calidad de las imágenes hiperespectrales en cuanto a resolución, nitidez, enfoque de la cámara, cantidad y tamaño de las regiones de interés, entre otros.

En la literatura es común encontrar generalidades del proceso realizado por cada autor para la adquisición de imágenes en proyectos con tratamiento de datos hiperespectrales, los aspectos más importantes a tener en cuenta en el desarrollo del protocolo de adquisición son el tipo de vehículo para toma de datos, la plataforma de la cámara, las características de la cámara, la posición de la plataforma frente al objeto, la velocidad y altitud del vehículo, la fuente de iluminación artificial, posición y ángulo de la cámara con respecto al sol, resolución de las imágenes HSI, longitudes de onda comprendida en las imágenes HSI, cantidad de bandas espectrales, resolución espectral, pre-procesamiento y calibración de la cámara, fechas de adquisición y características del terreno, entre otros. Cada uno de los aspectos mencionados difieren en relevancia de acuerdo a las características del proyecto o el tipo de toma de datos (pueden ser tomando las imágenes HSI en laboratorio con condiciones controladas o en campo).

Para este proyecto se buscarán los aspectos más relevantes y se desarrollará un protocolo de adquisición de imágenes hiperespectrales con el uso de un UAV basado en los datos recolectados en la literatura.

2.1.1. Aspectos principales para la adquisición de las HSI

A continuación se nombrarán los detalles principales de los parámetros más significativos para el desarrollo del protocolo.

2.1.2. Iluminación

La iluminación del objeto al momento de la toma de las imágenes tienen varios propósitos, los más comunes son superar la fluctuación de la luz solar en invernaderos, reducir los efectos ocasionados por sombras de otros objetos, eliminar los efectos debidos a la reflexión especular, obtener una iluminación difusa en caso de sacar imágenes en campo con clima nublado, entre otros.

En [8] realizaron la adquisición de imágenes hiperespectrales para la estimación de extensión de campos de mango, para ello usaron un vehículo terrestre no tripulado para la adquisición, dentro del vehículo equiparon paneles de referencia de iluminación *QPcard 102* enfocándose en puntos del campo con luz solar y sombra para cubrir la mayor cantidad posible de condiciones de iluminación.

Otra fuente de iluminación artificial comúnmente utilizadas son lámparas halógenas, en [5] usaron este tipo de iluminación para superar la fluctuación solar producida en el invernadero, para ello utilizaron una lámpara de 200W y 5000lx, el objetivo de su proyecto fue identificar características de tomates de colores similares con el uso de imágenes multispectrales. Adicionalmente, en [20] también hicieron uso de una lámpara halógena para reducir el efecto de las sombras en las imágenes, para este caso la lámpara es de 150W y se posicionó a 45° con respecto al objeto de estudio. En [21] se implementó una lámpara halógena LCD Schott KL2500 para la reducción de la reflexión especular, para ello se usó un plano de polarización colocado verticalmente a la cámara.

El problema principal en el uso de lámparas en la adquisición de las imágenes es el consumo de energía que ellas necesitan, como lo especifican en [17] al hacer uso de una lámpara halógena de 21V y 150W se requirió una fuente de alimentación de voltaje regulada. Un último caso de uso de lámparas se refiere en [29] en el cual utilizan dos lámparas 3900-ER de 150W puestas a ambos lados del objeto con un ángulo de 45° cada una y una altura de 40cm.

2.2. Características de la imagen hiperespectral

Por características de la imagen hiperespectral nos referimos al tamaño de la imagen, longitudes de onda de las bandas que posee, cantidad de bandas, resolución espectral y frecuencia de toma de las imágenes.

En [8] las longitudes de onda de las imágenes adquiridas iban desde 390 hasta 890nm y su resolución espectral fue de 2nm. En [5] utilizaron una cámara AC133-60gmNIR, con la cual abarcaron longitudes de onda desde 300 hasta 1000nm con una resolución espectral de

20nm. En [27] obtuvieron imágenes con bandas entre los 350 a 1050nm con una resolución de 5nm. En [20] las longitudes de onda comprendidas van desde 393 a 1009nm.

De acuerdo a la literatura revisada las longitudes de onda de las imágenes hiperespectrales varían muy poco, en promedio van de 350 a 1000nm, mientras que la resolución espectral sí puede variar entre 2 a 20nm, el principal motivo de este cambio son las especificaciones de la aplicación y el tipo de imágenes necesarias que pueden ser hiperespectrales o multiespectrales.

Por otro lado, teniendo en cuenta la resolución de la imagen, la cantidad de bandas obtenidas y el tipo de dato utilizado se encontró que en [8] la imagen hiperespectral tiene 648 × 244 píxeles y cada dato posee 12bits. En [24] la imagen contiene 512 × 217 píxeles y 224 bandas espectrales. En [16] las dimensiones de la imagen son 614 × 512 píxeles con 224 bandas y datos de 16 bits. En [22] realizaron dos pruebas, la primera trabajaron con imágenes de 304 × 301 píxeles y bandas y la segunda con 145 × 145 píxeles y 202 bandas. En [13] utilizaron imágenes de 304 × 301 píxeles con 220 bandas y datos de 16 bits. En [25] también realizaron dos pruebas variando las dimensiones de las imágenes, en la primera la imagen era de 610 × 340 píxeles y 115 bandas espectrales, en la segunda, las imágenes fueron de 144 × 144 píxeles y 200 bandas.

En general, la resolución de las imágenes puede ser el parámetro que más varía en la literatura, se encontraron imágenes desde 144 × 144 hasta 614 × 520, los principales causantes de esta variabilidad son las condiciones del campo o laboratorio para la adquisición, la exactitud y cantidad de regiones de interés que requiere la aplicación, el tipo de cámara utilizado y el coste computacional disponible para el procesamiento de las imágenes. En cuanto a la cantidad de bandas, la mayoría de imágenes hiperespectrales trabajan entre 110 y 220 bandas. La cantidad de los datos de las imágenes son 12 o 16 bits.

2.2.1. Plataforma de la cámara

La plataforma se refiere al conjunto de elementos que acompañan la cámara en el momento de la adquisición de las imágenes, para ello se tendrá en cuenta principalmente procesos de adquisición en condiciones no controladas o en campo.

En [8] utilizaron una plataforma móvil para un vehículo terrestre que contenía una cámara hiperespectral de escaneo de línea Resonon Pika II Vis-NIR, un detector de luz 3D Velodyne HDL64E, un sistema de navegación inercial y GPS. En [27] realizaron la toma de imágenes en un cuarto oscuro con una cámara, fuente de iluminación, una mesa móvil y un sistema informático equipado con un software de adquisición de imágenes (SpectralScanner, Óptica DV).

En [12] el sistema hiperespectral, consta de tres partes principales, una etapa giratoria, dos

camaras hiperespectrales VNIR de empuje, las cuales se encuentran simetricamente en el estado giratorio, y el computador para el procesamiento. En [21] construyeron un sistema DBf que consista en la camara hiperespectral con lente macro de 25mm y filtro polarizador lineal, una fuente de luz polarizada $\lambda 600$ con un plano de polarización vertical a la camara, la distancia entre la camara y el sujeto de observación oscilaba entre 80 y 100cm.

En [17] utilizaron un sistema portátil para la adquisición de imágenes hiperespectrales, el sistema está basado en barrido de líneas con escoba de empuje, se compone de un EMCCD que se enfría termoelectricamente a -20°C con un dispositivo Peltier de doble etapa, también posee un espectrografo de imágenes ImSpector V10Em una lente de montura C Rainbow CCTV S6X11 y una fuente de iluminación de 150W. En [25] realizaron dos pruebas para la toma de datos, ambas pruebas las realizaron con un dron equipado con la camara hiperespectral y un sensor ROSIS, también utilizaron un espectrometro de imágenes infrarrojas AVIRIS con una resolución espacial de 20m.

En [29] realizaron la adquisición en condiciones controladas con una base rotatoria a una velocidad de $0,8\text{mm/s}$, dos fuentes de iluminación ubicadas a 45° del objeto, la camara hiperespectral posicionada perpendicularmente a 65cm y un computador para el procesamiento.

2.2.2. Calibración

En [5] fue necesario corregir el gris de la imagen de longitudes de onda deseadas debido a la desigualdad en la intensidad de luz de varias longitudes de onda de la lámpara y la sensibilidad desigual de la camara en varias longitudes de onda. Como referencia, colocaron una paleta D65 (color blanco) en una posición fija en el campo de visión de la camara durante la captura de la imagen del objeto con los parametros de control constantes de la camara.

En [12] utilizaron un sistema de dos camaras con una plataforma giratoria para el movimiento del objeto, para ello se debe realizar una corrección geométrica como preprocesamiento, en el eje vertical para que las líneas epipolares coincidan para que correspondan a la misma línea, esto es necesario por la posibilidad de que se produzca diferencias en el punto de vista de ambas camaras.

En [9] consideraron un ancho de banda mínimo para obtener una relación señal/ruido factible en los datos. Adicionalmente consideraron una pérdida de potencia de la lente para longitudes de onda alta (superiores a 800nm), por lo que seleccionaron cuatro canales para esas longitudes (850, 900, 950 y 1000nm).

En [14] realizaron calibraciones espectrales en el panel de referencia blanco, se llevaron a cabo periódicamente manteniendo el panel bajo la luz halógena a una proximidad similar a la de

las coronas de las plantas a la fuente de luz (Spectralon Reflectance Target, CSTM-STR-99-100; Spectra Vista Corporation, Poughkeepsie, NY, EE. UU.). La calibración se realizó cada diez a quince plantas utilizando el panel de calibración blanco.

2.2.3. Toma de datos

En [8] se utilizó un vehículo terrestre a 5km/h sin detenerse para tomar imágenes de árboles a ambos lados, se aseguró que el sol iluminara directamente los árboles desde detrás de la cámara, por lo que las adquisiciones se realizaron por la mañana desde el lado este y por la tarde desde el lado oeste. En [20] tomaron imágenes hiperespectrales del pescado en una habitación oscura para evitar interferencia de luz y obtener reflectancia espectral pura, para la adquisición realizaron un barrido de línea y escoba al igual que [17] y [29].

En [14] utilizaron una plataforma móvil sobre un riel para la toma de datos, sobre ella se encontraba la cámara y una lámpara halógena de 500W como fuente de luz artificial, la plataforma se encontraba perpendicularmente a la planta, en el trabajo realizaron la inoculación de la enfermedad a detectar sobre las plantas y recolectaron los datos a los 30, 60 y 90 días después de la inoculación.

2.3. Desarrollo del protocolo

De acuerdo a la literatura revisada y las condiciones del proyecto se estableció un protocolo de adquisición de imágenes hiperespectrales de la siguiente forma:

1. Con respecto a la hora para la toma de datos, las horas deseadas para el proceso de adquisición de las imágenes hiperespectrales son aproximadamente a las 9 am desde el lado este de la planta y a las 3 pm desde el lado oeste.
2. El ángulo de la cámara con respecto a la planta es 45° , sin embargo, la cámara no se puede encontrar en el mismo eje con el sol y la planta debido a que la sombra del mismo objeto.
3. A pesar de que la fuente de iluminación artificial es bastante utilizada en la mayoría de procesos de adquisición de imágenes hiperespectrales, sin embargo, el alto consumo de energía obliga a utilizar fuentes de alimentación externas, por lo que se hará uso de lámparas halógenas únicamente en el caso de tener clima nublado, para ello se buscará una referencia con un consumo menor o igual a 150W .
4. La distancia entre el árbol y la cámara se procurará que sea de 80 a 100cm , aunque este dato puede modificarse en campo de acuerdo a las observaciones del movimiento de las plantas debido a los motores del UAV, en caso de tener muchas perturbaciones

también se podrán hacer modificaciones del ángulo de la cámara con respecto a la planta.

5. idealmente se buscara que en el momento de la toma de datos el UAV mantenga la misma posición, en caso de requerir tomas continuas, el desplazamiento del UAV no debe ser mayor a $5km=h$.
6. El formato, resolución de las imágenes y bandas utilizables depende directamente de la cámara, en este caso se utilizará una cámara multiespectral *Parrot Sequoia+* la cual puede tomar imágenes en 4 diferentes bandas ($550nm \pm 40nm$ para el color verde, $660nm \pm 40nm$ para el color rojo, $735nm \pm 10nm$ para el borde rojo y $790nm \pm 40nm$ para el infrarrojo cercano). La resolución de las imágenes es 1280×960 para cada banda y 4608×3456 para la imagen RGB. El formato de la imagen espectral es *TIFF* por lo que los datos serán de 8 bits. Con respecto al *FOV* de la imagen espectral puede ser 62° , 49° y 74° , por lo que para calcular la altura de cobertura de la imagen se puede obtener una aproximación utilizando la siguiente ecuación:

$$h = 2 \tan\left(\frac{\theta}{2}\right) d$$

donde θ es el ángulo seleccionado de los tres posibles mencionados, h es la altura de la imagen y d es la distancia de la cámara al árbol. Por ejemplo para una distancia de $80cm$ y un ángulo de 62° , la altura de cobertura de la imagen es $96,13cm$, disminuir ese valor se puede utilizar un ángulo menor, en este caso 49° es el más pequeño de todos, por lo que al mantener la distancia constante, la altura de cobertura en este caso es $72,91cm$.

7. Durante la toma de muestras es necesario tomar una imagen hiperespectral de un panel de calibración cada 15 o 20 minutos, este tiempo estimado puede aumentar o disminuir en función de la nubosidad presentada durante la toma de imágenes.

3 Resultados

Se realizaron dos experimentos de inoculación del patógeno sobre hojas de mango en ambientes controlados. Los detalles de cada experimento se evidencian a continuación.

Experimento 1.

- **Tipo de mango:** Mango Tommy Atkins.
- **Días de muestro:** 6 días, el muestreo empezó a partir del tercer día de la inoculación.
- **Periodo de muestreo:** Diario.
- **Método de inoculación:** Localizado y por el envés de la hoja.
- **Cantidad de muestras:** 10 hojas, 5 de control y 5 inoculadas.
- **Cantidad de puntos de inoculación:** 12 puntos de inoculación por cada hoja, para un total de 60 puntos.

Experimento 2.

- **Tipo de mango:** Mango keitt.
- **Días de muestro:** 11 días, el muestreo empezó a partir del tercer día de la inoculación.
- **Periodo de muestreo:** Diario.
- **Método de inoculación:** Localizado y por el haz de la hoja.
- **Cantidad de muestras:** 20 hojas, 10 de control y 10 inoculadas..
- **Cantidad de puntos de inoculación:** 12 puntos de inoculación por cada hoja, para un total de 120 puntos.

3.1. Pruebas del experimento 1

3.2. Clasificación general

El primer experimento llevado a cabo consiste en la detección del patógeno con muestras obtenidas sobre las hojas de mango tipo Tommy Atkins, para ello se utilizaron un total de 927 muestras, 399 de ellas con presencia del patógeno y 528 libre del mismo. Para el entrenamiento de los clasificadores se utilizaron el 50% de las muestras de cada grupo y para la validación el 50% restante. La variable utilizada como medida de discriminación fue la reflectancia, para ello se adquirieron muestras a nivel de píxel y se obtuvieron los vectores de 160 componentes (correspondiente al número de bandas). En la figura 3-1 se observa una gráfica con el promedio de reflectancia para los grupos mencionados, muestras con presencia del patógeno (para el cual se le denominará en adelante como muestras enfermas) y muestras sin presencia del patógeno (muestras sanas).

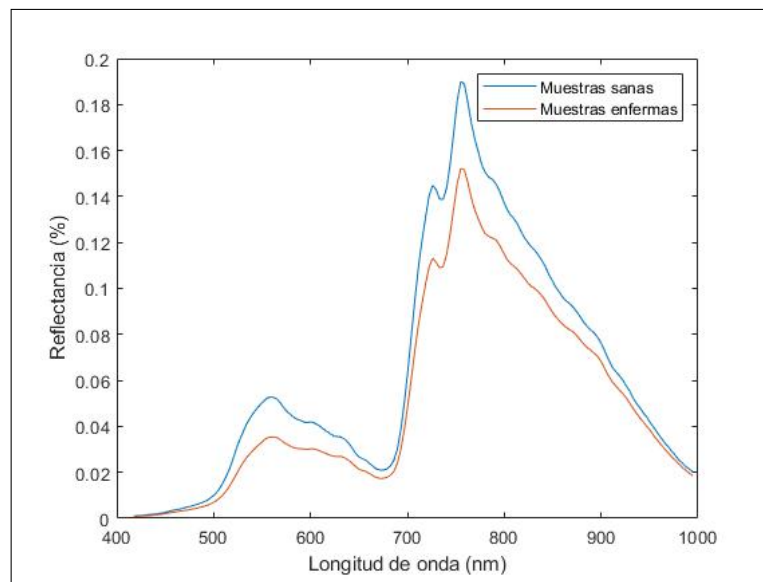


Figura 3-1: Promedio de curvas espectrales para los grupos de clasificación (muestras enfermas vs muestras sanas).

Los clasificadores implementados fueron LDA, SAM y KNN, para KNN se utilizó tres diferentes valores de K (3, 5 y 9), los valores fueron seleccionados impares debido a que se tienen únicamente dos grupos para la clasificación y los valores no fueron mayor a 10 por la baja cantidad de muestras obtenidas para algunos casos en concreto.

Clasificador	Matriz de confusión	
	LDA	158
42		246
SAM	168	117
	31	147
KNN (K=3)	145	47
	54	217
KNN (K=5)	139	45
	60	219
KNN (K=9)	133	43
	66	221

Tabla 3-1: Matrices de confusión para cada clasificador

Clasificador	Recall	Precisión	F-Score	Exactitud	Selectividad
LDA	0.8977	0.7900	0.8404	0.8707	0.8542
SAM	0.5895	0.8442	0.6942	0.6803	0.8258
KNN (K=3)	0.7552	0.7286	0.7469	0.7819	0.8097
KNN (K=5)	0.7554	0.6985	0.7258	0.7732	0.7849
KNN (K=9)	0.7557	0.6683	0.7093	0.7646	0.7700

Tabla 3-2: Métricas de clasificación.

De las tablas **3-1** y **3-2** se puede destacar un mejor desempeño del clasificador LDA frente a los demás, obteniendo métricas sobre 84 % a excepción de la precisión, la cual se encuentra en 79 %, caso contrario obtenido con el clasificador SAM, obteniendo una precisión mayor frente a las demás métricas, esto indica un alto desempeño en la clasificación de muestras enfermas pero rendimiento inferior a los demás métodos implementados en la clasificación de muestras sanas. Por último el clasificador KNN obtuvo resultados similares para los diferentes valores de K, en este se observan exactitudes entre 76,5 % y 78 %, al igual que lo ocurrido en el clasificador SAM, la clasificación de muestras enfermas logra un mejor resultado para todas las métricas, sin embargo, su diferencia es mucho menor, esto se refleja en datos como la precisión y el F-score el cual se mantienen entre 70 % y 75 %, también cabe destacar que al aumentar el valor de K, la exactitud disminuye.

3.2.1. Clasificación sectorizada

A partir de los resultados obtenidos en la prueba anterior, se decidió separar las muestras en categorías de acuerdo al lugar de procedencia sobre la hoja (haz y envés). La cantidad de

muestras obtenida en cada categoría para los mismos grupos utilizados son las siguientes:

Cantidad de muestras por el envés de la hoja con presencia del patógeno: 97

Cantidad de muestras por el haz de la hoja con presencia del patógeno: 103

Cantidad de muestras por el envés de la hoja sin presencia del patógeno: 120

Cantidad de muestras por el haz de la hoja sin presencia del patógeno: 144

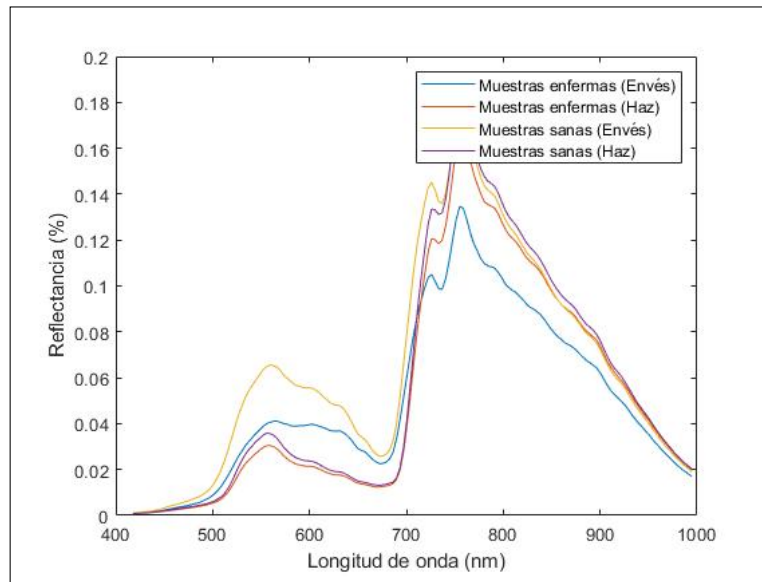


Figura 3-2: Promedio de curvas espectrales para los grupos de clasificación divididos por el haz y el envés de la hoja.

Clasificador	Muestras por el haz		Muestras por el envés	
LDA	93	9	97	0
	10	135	0	119
SAM	45	40	85	3
	58	104	11	117
KNN (K=3)	64	38	84	8
	39	106	12	112
KNN (K=5)	57	41	83	4
	46	103	13	116
KNN (K=9)	50	38	84	7
	53	106	12	113

Tabla 3-3: Matrices de confusión clasificación por el haz y envés de la hoja.

En la figura 3-2 se observa el promedio de las formas espectrales para las muestras sanas y enfermas en las dos categorías creadas (muestras por el haz y por el envés), se puede destacar un mayor nivel de reflectancia para muestras obtenidas por el envés, adicionalmente, la diferencia de reflectancia entre las muestras sanas y enfermas obtenidas por el envés alcanza a ser hasta 5 % mayor que la diferencia entre las muestras obtenidas por el haz.

Clasificación por el haz

En las tablas 3-3 y 3-4 se observan las matrices de confusión y las métricas para las muestras obtenidas por el haz, se evidencia una exactitud más baja frente a la prueba anterior, a excepción del clasificador LDA, el cual logró un valor de 92 %. Adicionalmente, se observó a partir de los resultados de selectividad, recall y precisión la misma tendencia en la cual la clasificación de muestras sanas presenta mejores resultados frente a la clasificación de muestras enfermas, de igual manera, en este caso al aumentar el valor de K para el clasificador KNN la exactitud disminuye y el clasificador SAM a pesar de estar enfocado en el análisis de datos hiperespectrales obtuvo los resultados más bajos en cada una de las métricas calculadas.

Clasificador	Recall	Precisión	F-Score	Exactitud	Selectividad
LDA	0.9118	0.9029	0.9073	0.9231	0.9310
SAM	0.5294	0.4669	0.4787	0.6032	0.6420
KNN (K=3)	0.6274	0.6214	0.6244	0.6883	0.7310
KNN (K=5)	0.5816	0.5534	0.5672	0.6478	0.6913
KNN (K=9)	0.5682	0.4854	0.5236	0.6316	0.6667

Tabla 3-4: Métricas en la clasificación con muestras por el haz de la hoja

Clasificación por el envés

Clasificador	Recall	Precisión	F-Score	Exactitud	Selectividad
LDA	1.0000	1.0000	1.0000	1.0000	1.0000
SAM	0.9659	0.8854	0.9239	0.9352	0.9141
KNN (K=3)	0.9130	0.8750	0.8936	0.9074	0.9032
KNN (K=5)	0.9540	0.8646	0.9071	0.9213	0.8992
KNN (K=9)	0.9231	0.8750	0.8984	0.9120	0.9040

Tabla 3-5: Métricas en la clasificación con muestras por el envés de la hoja.

En las tablas 3-3 y 3-5 se observan los resultados para la clasificación con muestras obtenidas por el envés de la hoja, se evidencia una exactitud para todos los clasificadores superior al 90 %, el clasificador LDA obtuvo una clasificación correcta del 100 % en ambos grupos, el

clasi cador SAM en este caso genero mejores resultados que el clasi cador KNN con una exactitud del 93 %, ademas se observo que SAM y KNN lograron una distribucion mas pareja en la clasi cacion de muestras enfermas con respecto a la clasi cacion con muestras sanas debido a que la diferencia entre selectividad y recall var a entre 1 % y 5 %. En general la clasi cacion para muestras por el envase generan mejores resultados con exactitudes hasta un 30 % mas altas en comparacion a la clasi cacion con muestras por el haz.

3.2.2. Reductor de dimensionalidad

Posteriormente, se utilizo un reductor de dimensionalidad, su objetivo es calcular la minima cantidad de componentes o bandas necesarias para detectar correctamente la presencia del patogeno sin bajar en gran medida el desempeño obtenido en las pruebas realizadas anteriormente. Para ello se implemento el metodo PCA, en este caso se planea reducir la cantidad de bandas utilizadas en los datos, se plantea verificar el comportamiento de los clasi cadores a medida que aumenta la cantidad de componentes utilizados, La cantidad de componentes graficados en cada prueba varian de 1 a 40, a excepcion de KNN.

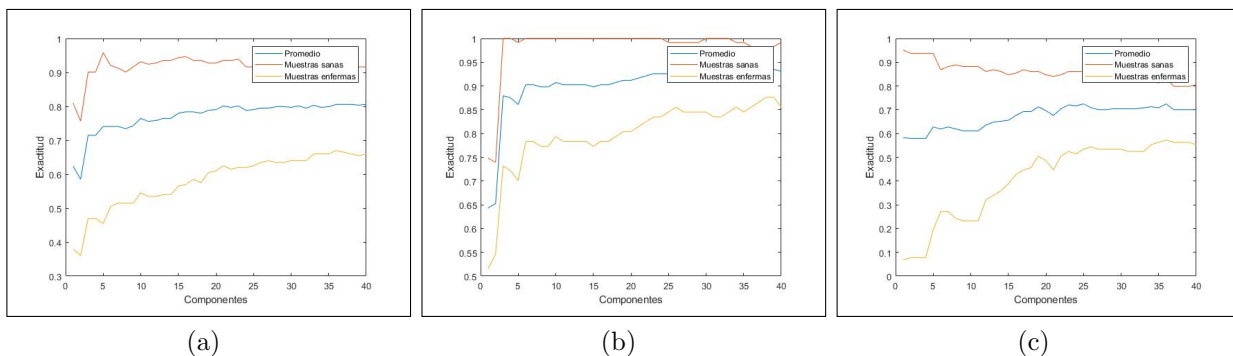


Figura 3-3: Implementacion de PCA con el clasi cador LDA (a) Muestras en general (b) Muestras por el envase (c) Muestras por el haz.

En la figura 3-3 se observa la implementacion de PCA con el uso del clasi cador LDA, se realizo las pruebas para tres categorias, muestras en general y por el haz y envase de la hoja. Para las muestras en general, la mayor exactitud promedio obtenida fue de 81 %, la clasi cacion para muestras sanas tuvo un maximo de 92 % y para muestras enfermas un maximo de 66 %, luego del uso de 4 componentes en el reductor de dimensionalidad, la clasi cacion para muestras sanas se establecio sobre 90 % y para las muestras enfermas permanecio aproximadamente en 50 %. Para las muestras obtenidas por el envase se observo un pico de clasi cacion correcta de muestras enfermas del 100 % y se mantiene luego de 4 componentes, la exactitud promedio alcanza un 90 % luego de 7 muestras. Con respecto a la clasi cacion para muestras por el haz, la exactitud alcanza un maximo de 70 % el cual se consigue luego del uso de 20 componentes y la clasi cacion de muestras sanas decrece de un

95 % inicialmente obtenido hasta un 80 %.

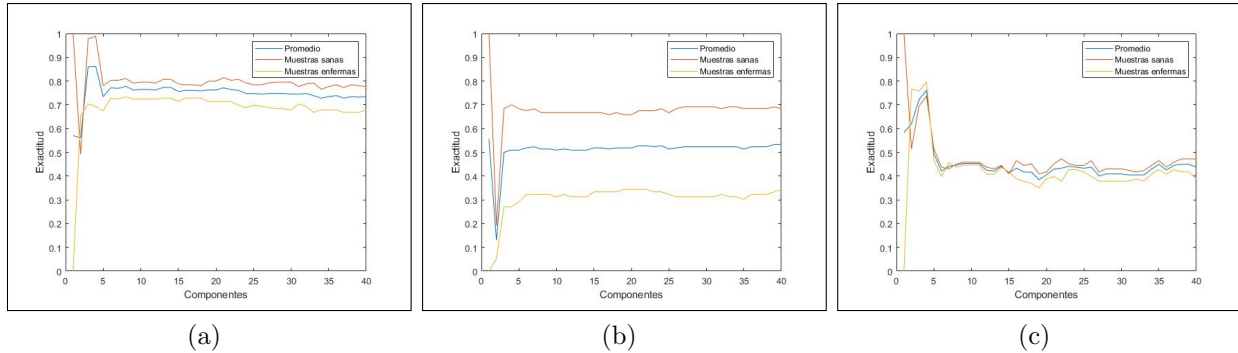


Figura 3-4: Implementación de PCA con el clasificador SAM (a) Muestras en general (b) Muestras por el envés (c) Muestras por el haz.

En la figura 3-4 se ilustra el comportamiento del clasificador SAM al añadir el reductor de dimensionalidad, se realizó las pruebas para las mismas tres categorías mencionadas. Para las muestras en general. La mayor exactitud promedio obtenida fue de 77 %, la clasificación para muestras sanas tuvo un máximo de 80 % y para muestras enfermas 72 %. Luego del uso de 6 componentes en el reductor de dimensionalidad la clasificación para muestras sanas se estableció en 80 % con tolerancia del 2 % y para muestras enfermas en 70 % con tolerancia del 3 %.

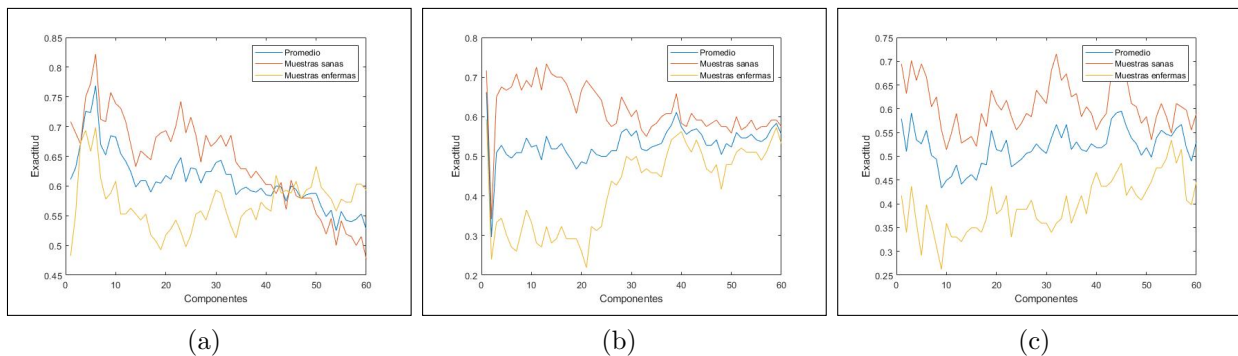


Figura 3-5: Implementación de PCA con el clasificador KNN ($K=3$) (a) Muestras en general (b) Muestras por el envés (c) Muestras por el haz.

La aplicación del reductor de dimensionalidad con el clasificador KNN (con $K=3$) genera resultados caóticos frente a los otros clasificadores, los porcentajes de clasificación en promedio pueden oscilar entre 47 % hasta 83 % al variar un solo componente, tiende a estabilizarse luego de 50 componentes, en el cual logra clasificaciones de 60 % para muestras sanas y 56 % de muestras enfermas pero con tolerancia de 5 % en ambos casos.

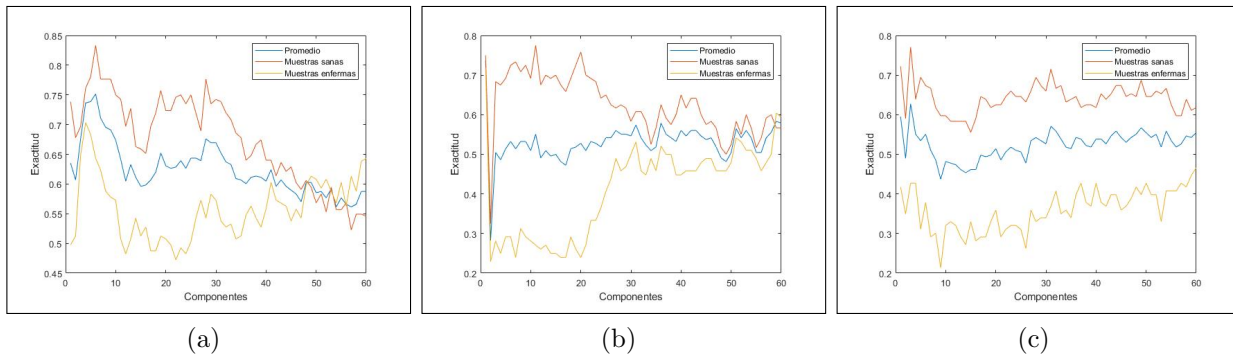


Figura 3-6: Implementacion de PCA con el clasificador KNN ($K=5$) (a) Muestras en general (b) Muestras por el envase (c) Muestras por el haz.

Para $K=5$, en el grupo de muestras generales, los porcentajes de exactitud para muestras sanas y enfermas varían entre 27 % y 75 %, sin embargo, el promedio es más estable logrando una tolerancia menor al 15 % con una tendencia final sobre el 50 % para los tres grupos. Para la categoría de muestras por el envase, la tolerancia es mucho menor (alrededor de 5 %), además los porcentajes de exactitud convergen a 58 % para todos los grupos. Por otro lado, para las muestras por el haz, los grupos son más estables, las muestras sanas alcanzan una exactitud de aproximadamente 65 % mientras las muestras enfermas un 40 %, para un promedio total del 53 %.

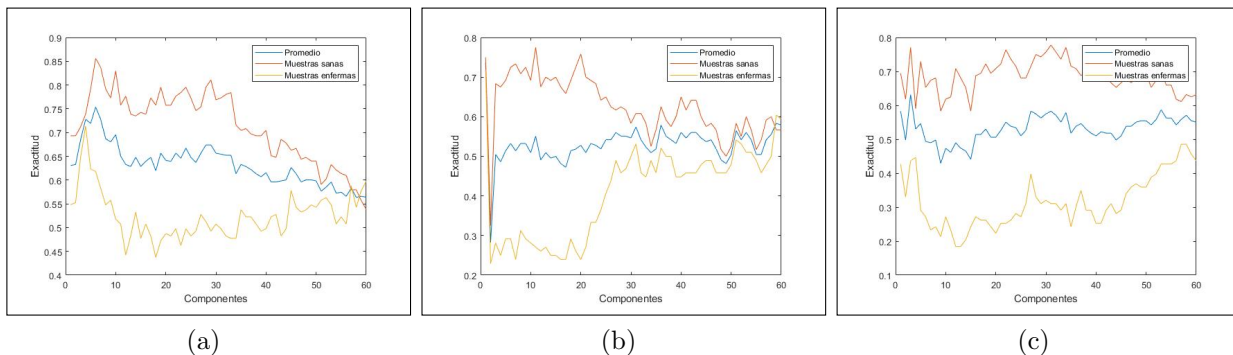


Figura 3-7: Implementacion de PCA con el clasificador KNN ($K=9$) (a) Muestras en general (b) Muestras por el envase (c) Muestras por el haz.

Por último, para $K=9$, en el grupo de muestras generales, la exactitud en la clasificación de muestras sanas inicia en 80 % y baja a medida que se aumenta la cantidad de componentes utilizados hasta 54 %, para muestras enfermas se observa lo contrario, inicio desde 45 % y subió hasta 59 %, es importante resaltar que con el uso de menos de 10 componentes, se obtienen porcentajes de clasificación superiores a 70 %. Para el grupo de muestras por el envase, con el uso de 5 a 25 componentes, las muestras sanas obtienen exactitudes entre 75 % y 85 %. Por otro lado, los resultados para muestras enfermas varían entre 25 % y 30 %, luego

del uso de 30 o más componentes, la clasificación para muestras sanas baja con resultados entre 52% y 65%, pero para las muestras enfermas la exactitud aumenta hasta obtener resultados entre 45% y 60%, el promedio general se mantiene entre 50% y 60%. Para el grupo de muestras por el haz los resultados son irregulares, la clasificación para muestras sanas varía entre 5% y 85%, y para muestras enfermas entre 80% y 20%.

3.2.3. Clasificación por días

La última prueba realizada para el experimento uno corresponde a la clasificación separando los grupos de acuerdo al día de obtención de las muestras, esto permite conocer la mínima cantidad de días necesarios para poder detectar la presencia del patógeno de forma adecuada, para ello se utilizaron las métricas de selectividad y sensibilidad para medir el desempeño del algoritmo en cuanto a la clasificación de muestras enfermas y muestras sanas respectivamente. En la figura 3-8 se observa el promedio del valor de reflectancia para las muestras enfermas separadas por el día de obtención de cada una, de ahí se puede observar como los valores de reflectancia disminuyen a medida que aumenta el día de muestreo, esto se observa principalmente en los rangos de 550 a 600 nm y 750 a 900 nm.

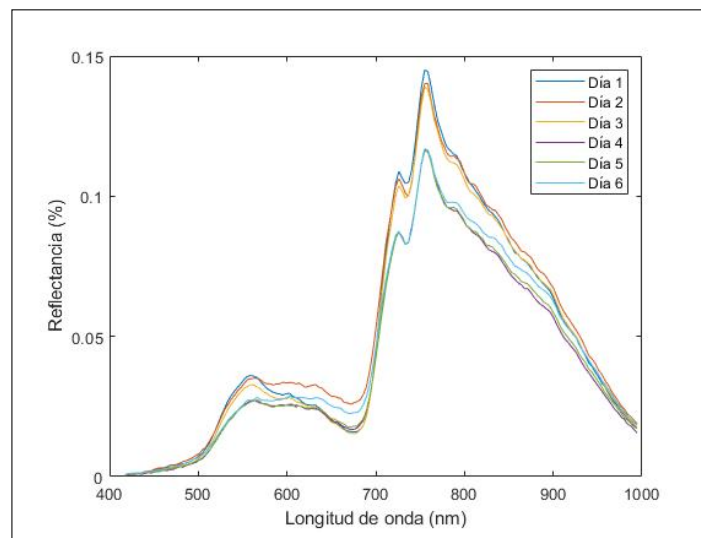


Figura 3-8: Promedio de curvas espectrales por días.

La primera prueba para la clasificación por días se realizó para el grupo de muestras generales (tabla 3-6), de ahí se puede observar que el clasificador LDA obtuvo una exactitud del 100% para todos los días, por otro lado KNN logra obtener 70% de clasificación para muestras enfermas luego del 2 día y 80% a partir del tercero, mientras para muestras sanas, se logran exactitudes superiores al 75% hasta el inicio del cuarto día. Para el clasificador SAM, la exactitud para muestras enfermas logra un 70% para el día 5 y 82% para el día 6, para las muestras sanas se logró una exactitud del 72% desde el día 3, la cual aumentó hasta un 84%

para el ultimo día.

Clas.	Día 1		Día 2		Día 3		Día 4		Día 5		Día 6	
	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
LDA	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SAM	0.36	0.53	0.36	0.63	0.53	0.72	0.53	0.68	0.71	0.75	0.82	0.84
KNN (K=3)	0.58	0.39	0.72	0.59	0.78	0.69	0.84	0.53	0.74	0.60	0.69	0.74
KNN (K=5)	0.53	0.54	0.81	0.48	0.84	0.62	0.75	0.61	0.77	0.57	0.63	0.82
KNN (K=9)	0.53	0.36	0.69	0.48	0.88	0.55	0.75	0.64	0.74	0.57	0.41	0.84

Tabla 3-6: Porcentaje general de clasificación por días.

Para el grupo de muestras por el haz se obtuvieron los resultados reflejados en la tabla 3-7, de ahí se puede observar nuevamente que LDA obtiene exactitudes del 100 % para todos los días. Por otro lado, la clasificación de muestras sanas para los clasificadores SAM y KNN inicia entre 0 % y 36 % y culmina entre 70 % y 93 %, para muestras enfermas, desde el día uno se obtienen exactitudes desde 65 % y aumenta hasta un máximo de 95 % para el día 4, luego disminuye hasta un 60 % para el día 6.

Clas.	Día 1		Día 2		Día 3		Día 4		Día 5		Día 6	
	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
LDA	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SAM	0.64	0.36	0.67	0.23	0.90	0.23	0.80	0.46	0.75	0.58	0.64	0.93
KNN (K=3)	0.68	0.07	0.58	0.23	0.60	0.30	0.90	0.23	0.75	0.47	0.57	0.70
KNN (K=5)	0.68	0.14	0.83	0.15	0.75	0.15	0.95	0.23	0.62	0.42	0.61	0.73
KNN (K=9)	0.75	0.00	0.75	0.16	0.75	0.15	0.90	0.38	0.67	0.32	0.57	0.73

Tabla 3-7: Porcentaje de clasificación por días por el haz.

Los resultados de clasificación por días en el grupo de muestras por el envés se observan en la tabla 3-8, el clasificador LDA logra una exactitud del 100 % en todos los días, SAM obtuvo un desempeño similar a LDA, salvo por los primeros 5 días para muestras sanas, los cuales son superiores a 75 %. Para el clasificador KNN, la clasificación de muestras sanas inician desde un 21 % para el día uno y terminan con resultados entre 91 % y 95 % para el día 6, en

el caso de las muestras enfermas se observan resultados hasta de 80% para el primer día y alcanza porcentajes de 100% luego del día 2.

Clas.	Día 1		Día 2		Día 3		Día 4		Día 5		Día 6	
	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
LDA	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SAM	1.00	0.81	1.00	0.93	1.00	0.73	1.00	0.80	1.00	0.75	1.00	1.00
KNN (K=3)	0.55	0.62	1.00	0.79	0.78	0.69	1.00	0.80	0.90	0.85	0.85	0.95
KNN (K=5)	0.80	0.28	0.90	0.71	1.00	0.87	0.90	0.80	0.80	0.81	0.90	0.95
KNN (K=9)	0.75	0.21	0.95	0.71	0.95	0.73	0.95	0.80	0.80	0.87	0.80	0.91

Tabla 3-8: Porcentaje de clasificación por días por el envés.

3.3. Pruebas del experimento 2

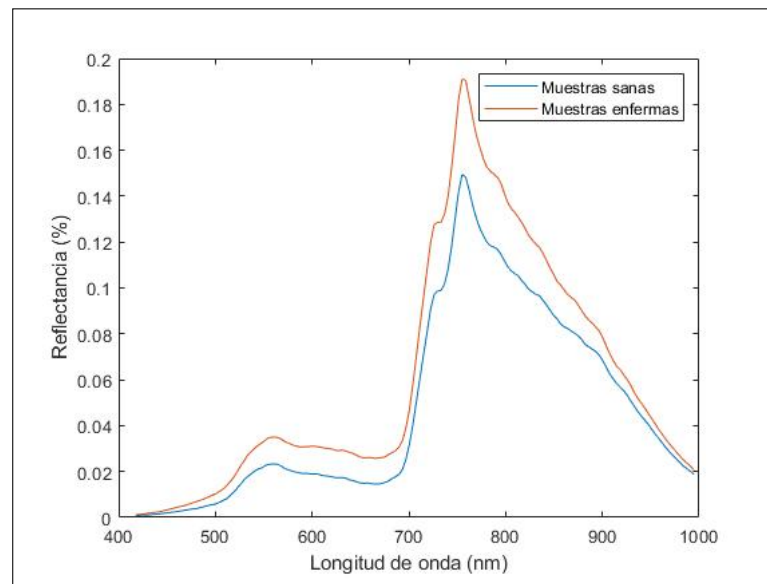


Figura 3-9: Promedio de curvas espectrales para los grupos de clasificación (muestras enfermas vs muestras sanas).

Para el experimento 2 se realizaron las mismas pruebas (a excepción de la clasificación por días debido a la baja cantidad de muestras), utilizando los clasificadores mencionados, para este caso la cantidad de muestras enfermas y muestras sanas es de 58 para cada uno. En la figura 3-9 se observa el promedio de reflectancia para el grupo de muestras enfermas y

muestras sanas, de ahí se puede observar que la re ectancia para este experimento es mayor en el grupo de muestras enfermas con porcentajes cercanos al 20 %, superando hasta en 4 % la re ectancia, con respecto al grupo de muestras sanas.

3.3.1. Clasi cacion general

La primera prueba se realizo con el grupo de muestras generales, los resultados se observan en las tablas 3-9 y 3-10, de ahí se puede observar que el clasi cador LDA obtuvo una exactitud del 100 %, el clasi cador SAM logro una cexactitud del 65 % y KNN porcentajes entre 63 % y 57 %, los cuales disminu an a medida que se aumentaba el valor de K.

Clasificador	Matríz de confusión	
	LDA	29
	0	29
SAM	24	15
	5	14
KNN (K=3)	22	14
	7	15
KNN (K=5)	20	14
	9	15
KNN (K=9)	17	13
	12	16

Tabla 3-9: Matriz de confusion para cada clasi cador

Clasificador	Recall	Precisión	F-Score	Exactitud	Selectividad
LDA	1.0000	1.0000	1.0000	1.0000	1.0000
SAM	0.6153	0.8276	0.7059	0.6552	0.7368
KNN (K=3)	0.6111	0.7586	0.6769	0.6379	0.6818
KNN (K=5)	0.5882	0.6896	0.6349	0.6034	0.6250
KNN (K=9)	0.5667	0.5862	0.5763	0.5690	0.5714

Tabla 3-10: Exantitud en la clasi cacion general

3.3.2. Clasi cacion sectorizada

Posteriormente se realizo la clasi cacion sectorizada, en la cual se separan las muestras en dos categor as, muestras obtenidas por el envés de la hoja y muestras obtenidas por el haz. Para este experimento se obtuvieron 12 muestras por el envés de la hoja, de las cuales 6

corresponden a muestras con presencia del patógeno y 6 sin presencia del patógeno, así mismo, se tomaron 46 muestras por el haz de la hoja, 23 correspondientes a muestras con presencia del patógeno y 23 muestras sin presencia del patógeno. Para este experimento la cantidad de muestras por el envés de la hoja es mucho más baja, esto se debe a que únicamente se realizó el muestreo por ese lado de la hoja durante los últimos dos días del experimento, debido al bajo desarrollo de la enfermedad.

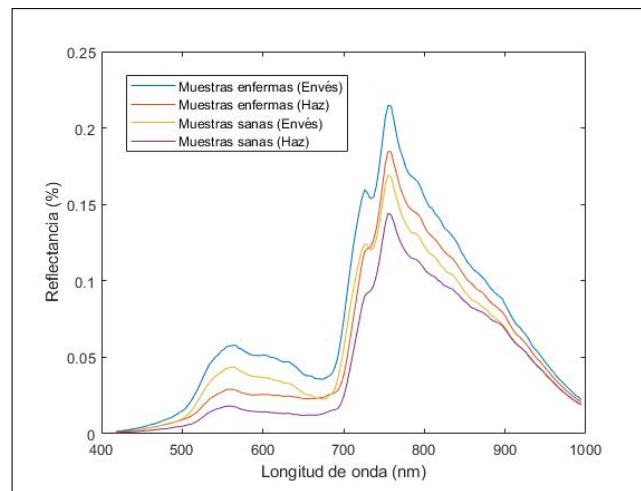


Figura 3-10: Promedio de curvas espectrales para los grupos de clasificación divididos por el haz y el envés de la hoja.

En la figura 3-10 se observa el promedio de reflectancia para cada grupo, de ahí se puede resaltar que en cada categoría, la reflectancia es mayor para el grupo de muestras enfermas, además las muestras obtenidas por el envés reflejan un mayor nivel de reflectancia alcanzando porcentajes superiores a 20% para una longitud de onda de 780 nm.

Clasificador	Muestras por el haz		Muestras por el envés	
	Enfermas	Sanas	Enfermas	Sanas
LDA	23	0	6	0
	0	23	0	6
SAM	12	5	2	1
	11	18	4	5
KNN (K=3)	14	7	3	2
	9	16	3	4
KNN (K=5)	14	7	2	1
	9	16	4	5
KNN (K=9)	8	2	2	1
	15	21	4	5

Tabla 3-11: Matrices de confusión clasificación por el haz y envés de la hoja

Clasificador	Recall	Precisión	F-Score	Exactitud	Selectividad
LDA	1.0000	1.0000	1.0000	1.0000	1.0000
SAM	0.7059	0.5217	0.6000	0.6522	0.6207
KNN (K=3)	0.6667	0.6087	0.6364	0.6521	0.6400
KNN (K=5)	0.6667	0.6087	0.6364	0.6521	0.6400
KNN (K=9)	0.8000	0.3478	0.4849	0.6304	0.5833

Tabla 3-12: Metricas en la clasificación con muestras por el haz de la hoja

En las tablas 3-11 y 3-12 se observan los resultados de clasificación para muestras obtenidas por el haz, se puede resaltar que LDA logro resultados de 100 % en cada metrica por otro lado, la precision de SAM es de 52 % y su exactitud es de 65 %, indicando que para este experimento la clasificación de muestras enfermas reaja un mejor desempeño frente a la clasificación de muestras sanas. Para el clasificador KNN la precision, la exactitud, la selectividad y el F-Score disminuyen a medida que aumenta el valor de K, a pesar de que la clasificación entre muestras enfermas y muestras sanas varían continuamente entre ellas, la exactitud se mantiene con una tolerancia del 2.5 %.

Clasificador	Recall	Precisión	F-Score	Exactitud	Selectividad
LDA	1.0000	1.0000	1.0000	1.0000	1.0000
SAM	0.6667	0.3333	0.4444	0.5833	0.5556
KNN (K=3)	0.6000	0.5000	0.5454	0.5833	0.5714
KNN (K=5)	0.6667	0.3333	0.4444	0.5833	0.5556
KNN (K=9)	0.6667	0.3333	0.4444	0.5833	0.5556

Tabla 3-13: Metricas en la clasificación con muestras por el envés de la hoja.

Los resultados en las pruebas de clasificación para las muestras por el envés se observan en las tablas 3-11 y 3-13, el clasificador LDA obtuvo 100 % de clasificación correcta en ambos grupos, SAM y KNN lograron desempeños similares, en ambos clasificadores se observa que la selectividad es mayor a la precisión, demostrando al igual que la prueba anterior, que la clasificación para muestras enfermas genera mejores resultados que la clasificación de muestras sanas para este experimento. La exactitud se mantiene en 58.33 % en los dos clasificadores, la poca variación en los datos son resultados de la baja cantidad de muestras.

3.3.3. Reductor de dimensionalidad

La última prueba realizada para este experimento corresponde a la implementación de PCA como reductor de dimensionalidad, en esta ocasión la cantidad de componentes máximos para muestras por el haz es de 40 y para muestras por el envés es 12, esto se debe a que

PCA no permite seleccionar una cantidad de componentes mayor al numero de muestras de la etapa de entrenamiento.

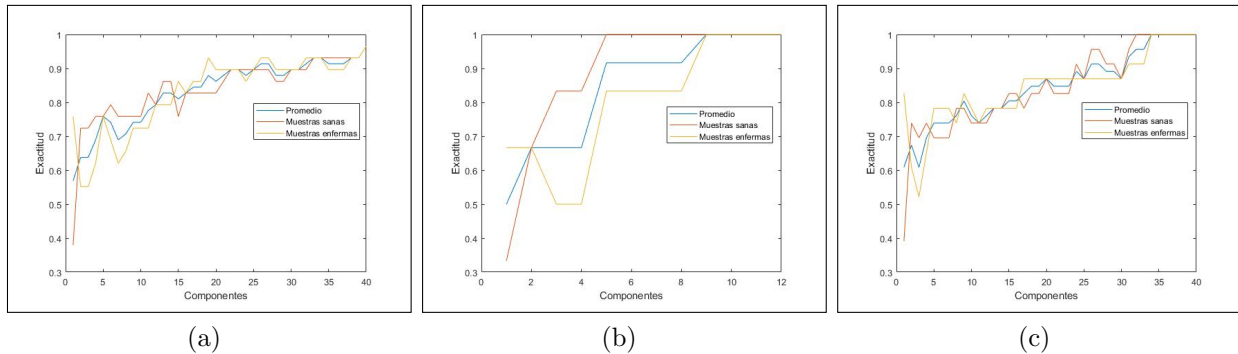


Figura 3-11: Implementacion de PCA con el clasificador LDA (a) Muestras en general (b) Muestras por el enves (c) Muestras por el haz.

En la figura 3-11 se observan los resultados de clasificacion para el clasificador LDA, para el grupo de muestras generales y el grupo de muestras por el haz inician con una exactitud promedio de 60% e incrementan linealmente hasta el 100% de clasificacion, luego de 32 componentes en ambos casos. Las pruebas con el grupo de muestras por el enves inicia con un desempeño promedio del 50% y aumenta hasta un 100%, luego de 9 componentes. Los resultados de muestras enfermas y muestras sanas obtienen resultados similares por lo que estan muy cercanos al promedio. Sin embargo, a pesar de ser muy baja la diferencia, la clasificacion de muestras enfermas tiene mejores resultados frente a la clasificacion de muestras sanas.

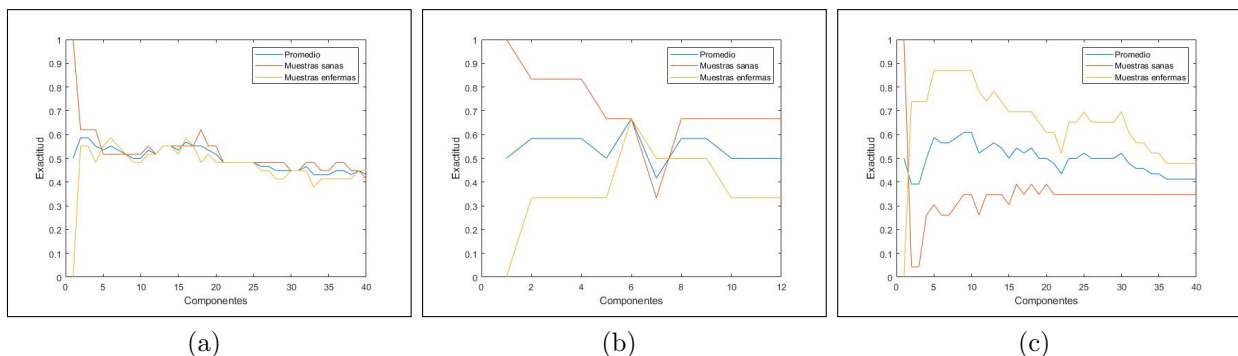


Figura 3-12: Implementacion de PCA con el clasificador SAM (a) Muestras en general (b) Muestras por el enves (c) Muestras por el haz.

Los resultados de la implementacion de PCA junto al clasificador SAM se observan en la figura 3-12, para el grupo de muestras generales los resultados de clasificacion inicia en 60% y disminuye hasta un 40% luego de 40 componentes. Los porcentajes de muestras sanas y

enfermas son similares a lo largo de la gráfica, para el grupo de las muestras por el envés, la clasificación de muestras sanas inician con un 100% y disminuye hasta un mínimo de 30% con el uso de 7 componentes, para las muestras enfermas inicia en 0% de exactitud y aumenta hasta 65% luego de 6 componentes, ahí se consiguió el mejor desempeño con una exactitud promedio del 68%. Para las muestras obtenidas por el haz se observa una gran diferencia entre la clasificación de muestras enfermas y muestras sanas, para las muestras enfermas la exactitud inicia en 88% y baja hasta un 50% y para muestras sanas inician en 4% y suben hasta un máximo de 35% luego del uso de 20 componentes. El promedio se mantiene entre 60% y 50%, y luego de 30 componentes baja hasta el 40%.

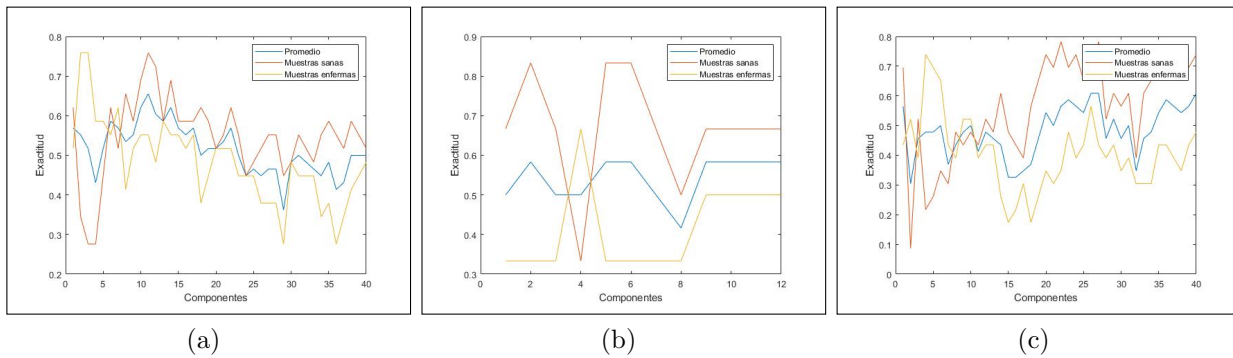


Figura 3-13: Implementación de PCA con el clasificador KNN ($K=3$) (a) Muestras en general (b) Muestras por el envés (c) Muestras por el haz.

Para el clasificador KNN con $K = 3$, en el grupo de muestras generales la exactitud varía constantemente, la exactitud de clasificación para muestras sanas y enfermas fluctúa entre 78 y 30%, el promedio se mantiene entre 65% y 35% pero no logra estabilizarse en algún punto, para los casos de muestras por el haz y envés se observa el mismo comportamiento, con exactitudes entre 32% y 85% para el envés y 10% a 79% para el haz.

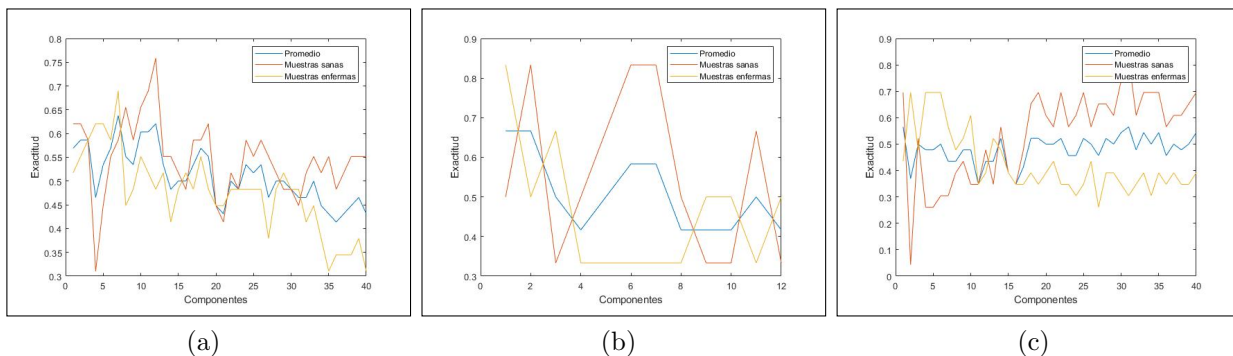


Figura 3-14: Implementación de PCA con el clasificador KNN ($K=5$) (a) Muestras en general (b) Muestras por el envés (c) Muestras por el haz.

Para las pruebas del clasificador KNN con $K = 5$ (tabla 3-14) y $K = 9$ (tabla 3-15), los resultados son los mismos, la exactitud cambia continuamente y los porcentajes pueden variar entre 5% y 90% en una sola prueba. Esto indica un muy bajo desempeño en la implementación de reductor de dimensionalidad PCA junto al clasificador KNN.

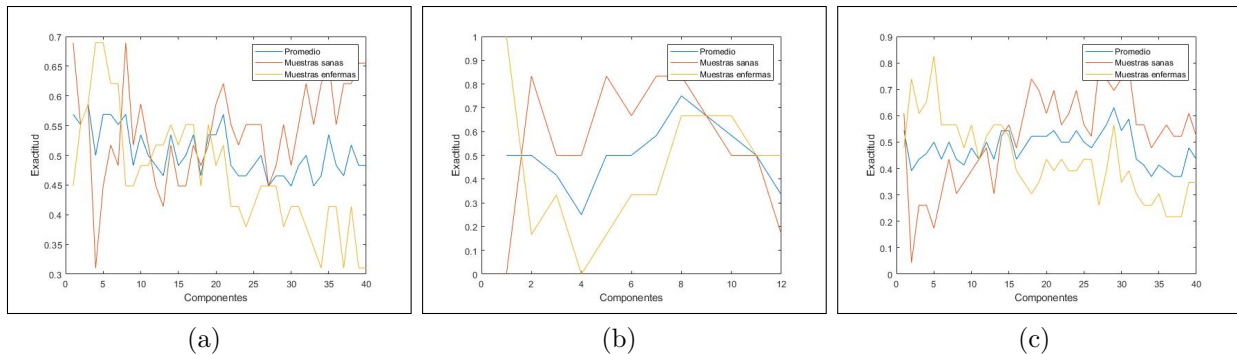


Figura 3-15: Implementación de PCA con el clasificador KNN ($K=9$) (a) Muestras en general (b) Muestras por el envés (c) Muestras por el haz.

3.4. Resultados adicionales

Como resultados adicionales y relacionados con el desarrollo del presente trabajo se tiene:

1. Divulgación académica: Penagos, Oscar; Velasquez, Carlos; Prieto, Flavio; *Detection of anthracnose in mango leaves by hyperspectral image analysis*, IEEE Xplore, Simposio STSIVA, 2021. (Aprobado el 31 de agosto de 2021).
2. Participación en el proyecto de investigación "Detección en campo de antracnosis a nivel foliar en árboles de mango mediante el análisis de imágenes espectrales y técnicas de aprendizaje de máquina", código 48334 y financiado por la División de Investigación (sede Bogotá D.C.) de la Universidad Nacional de Colombia.

4 Conclusiones

De acuerdo a los resultados obtenidos en el experimento 1, la clasificación con muestras obtenidas por el envés refleja una clara superioridad en la exactitud con respecto a la clasificación con muestras por el haz, la clasificación por el envés, el recall, la exactitud y la selectividad son superiores a 90 % indicando una alta precisión en cuanto a la clasificación correcta tanto de muestras enfermas como de muestras sanas. Adicionalmente, el clasificador LDA obtuvo un rendimiento perfecto, logrando una exactitud del 100 % en este caso. Para las muestras por el haz, la exactitud obtenida con los clasificadores SAM y KNN oscilan entre 60 y 69 %, también se observa que la selectividad es más alta que la precisión y el F-Score, indicando que los clasificadores predicen mejor la presencia del patógeno que la ausencia del mismo, el clasificador LDA logró un mejor desempeño comparado con los otros dos clasificadores con una exactitud general del 92 %, asimismo, la selectividad en este clasificador fue de 93 %, por lo que nuevamente la clasificación de muestras enfermas genera mejores resultados que la clasificación de muestras sanas.

Los resultados de la clasificación separando las muestras del experimento 1 brindan una aproximación de la mínima cantidad de días necesarios luego de la inoculación del patógeno adecuadamente la enfermedad. Para el caso de la clasificación con muestras obtenidas por el envés, se observa que a partir del segundo día de muestreo, la clasificación de muestras sanas mostrada a través del Recall genera porcentajes superiores al 90 % para KNN y 100 % para LDA y SAM. Para las muestras enfermas se lograron resultados similares a partir del cuarto día de muestreo, en el cual los clasificadores SAM y KNN lograron una selectividad de 0.8. Para las muestras obtenidas por el haz de la hoja, únicamente durante el cuarto día de muestreo, la clasificación de muestras sanas superó el 80 % en todos los clasificadores, luego decae nuevamente este valor. Por otro lado la clasificación de muestras sanas alcanzó su máximo en el día 6 con porcentajes de 70 %, en los demás días los resultados son inferiores a 50 % e incluso en el primer día, para KNN, estos porcentajes son inferiores a 15 %. Sin embargo, LDA en cada día sin importar el lugar de procedencia de la muestra, logró resultados de 100 % de exactitud en todos los casos.

Para el experimento 2, la cantidad de muestras obtenidas se redujo abruptamente, esto indica que la inoculación por el haz de la hoja disminuye considerablemente el desarrollo de la enfermedad. Adicionalmente, se observó que las muestras enfermas generan un mayor nivel de re-ectancia que las muestras sanas, esta condición se evidenció tanto para muestras por

el haz como para muestras por el envés. En cuanto a los resultados de la clasificación, la exactitud de la clasificación de muestras por el envés fue de 58 % para KNN y SAM, mientras la clasificación con muestras por el haz de la hoja, logro exactitudes entre 63 y 65 % para los mismos clasificadores. También se evidencio que en este experimento, la selectividad es menor que el Recall, precisión y F-Score, indicando que la clasificación de muestras enfermas tiene mejores resultados que la clasificación de muestras sanas, esto puede ser provocado por el nivel de reflectancia superior en las muestras enfermas. En este experimento el clasificador LDA obtuvo un desempeño de 100 % de clasificación en todas las pruebas realizadas.

Por otra parte, el uso del reductor de dimensionalidad PCA, en el experimento 1, indico que para el clasificador LDA es necesario el uso de al menos 20 componentes para obtener una exactitud promedio del 75 %, en caso de utilizar solo muestras por el envés, con 7 componentes se logra una exactitud del 90 % y con 20 componentes sube hasta 95 %. Para el clasificador SAM, hacen falta unicamente 5 componentes para obtener una exactitud promedio de 80 % y 50 % si solo se toman en cuenta muestras por el envés, este fue el unico caso en el cual el grupo de muestras generales (haz y envés) logra un mejor desempeño que la clasificación solamente tomando muestras por el envés. Por ultimo para KNN no se observo un resultado adecuado con el reductor de dimensionalidad, la exactitud no se estabiliza en ningun en ningun numero concreto de componentes, la exactitud varía en todos los casos entre 50 y 75 %. En el experimento 2, LDA obtiene una exactitud de 75 % con 5 componentes en el reductor PCA, y aumenta linealmente hasta un maximo de 95 % luego de 25 componentes, para muestras por el envés, se obtiene una exactitud de 100 % a partir de 9 componentes, con el clasificador SAM, la exactitud en las 3 pruebas realizadas (muestras generales, por el haz y por el envés), la exactitud varía entre 45 y 60 %, por ultimo para KNN, los resultados son los mismos obtenidos en el experimento 1, sin importar el valor de K en el clasificador, la exactitud no se estabiliza o tiene a un valor en concreto y puede variar entre 30 y 70 %.

En general, la clasificación con muestras obtenidas por el envés de la hoja refleja un mejor desempeño comparado con la clasificación de muestras sanas, demostrado a través del experimento 1 en el cual todos los clasificadores alcanzaron porcentajes de clasificación superiores a 90 %, además LDA logro exactitudes de 100 % en la mayoría de las pruebas y 90 % en su menor desempeño. Al momento de utilizar el reductor de dimensionalidad PCA, si se busca priorizar el uso de la mínima cantidad de componentes a utilizar, se recomienda el uso de SAM, el cual con el uso de 5 componentes se obtiene una exactitud general de 90 % con muestras por el envés, pero si el objetivo es buscar maximizar la exactitud, se recomienda el clasificador LDA, el cual alcanza una exactitud de 95 % luego de 20 componentes. Al comparar los resultados del experimento 1 y 2, se observa que la enfermedad se desarrolla con mayor facilidad si el hongo causante de la enfermedad se ubica sobre el envés de la hoja, asimismo, es más fácil detectar la enfermedad si se obtienen muestras por el envés de la hoja.

Como trabajo futuro se plantea reemplazar la cámara hiperespectral para la toma de muestras a una multispectral, para ello se pretende analizar los resultados obtenidos con PCA como reductor de dimensionalidad para seleccionar las longitudes de onda más significantes en la detección de la presencia del patógeno. Adicionalmente, se propone extender los experimentos con una mayor cantidad de muestras e implementar otros reductores de dimensionalidad enfocados en el análisis de datos espectrales, como lo pueden ser OSP (Orthogonal Subspace Projection) o técnicas de deep learning.

Bibliografía

- [1] AIRES, Priscila ; FERNANDES GAMBARRA NETO, Francisco ; COUTINHO, Wirton ; ARAÚJO, Alderi ; SILVA, Gilvan ; GOUVEIA, Josivanda ; MEDEIROS, Everaldo: Near infrared hyperspectral images and pattern recognition techniques used to identify etiological agents of cotton anthracnose and ramulosis. En: *Journal of Spectral Imaging 7* (Abril, 2018)
- [2] ALIAS, M. S. ; ISMAIL ADNAN, A. M. ; JUGAH, K. ; ISHAQ, I. ; FIZREE, Z. A.: Detection of Basal Stem Rot (BSR) disease at oil palm plantation using hyperspectral imaging. En: *2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2013, p. 1{4
- [3] BERTELS, LUC ; DERONDE, Bart ; KEMPENEERS, Pieter ; DEBRUYN, Walter ; PROVOOST, Sam: Optimized Spectral Angle Mapper classification of spatially heterogeneous dynamic dune vegetation, a case study along the Belgian coastline. En: *9th International Symposium on Physical Measurements and Signatures in Remote Sensing (ISPMSRS). Beijing, China* (Noviembre,2002)
- [4] CENTE, IBM K. *Análisis discriminante: Método de inclusión por pasos*. https://www.ibm.com/support/knowledgecenter/es/SSLVMB_subs/statistics_mai nhel p_ddi ta/spss/base/i dh_di sc_met. html , (15-05-2020). 2019
- [5] FENG, Q. ; CHEN, J. ; LI, X. ; LI, C. ; WANG, X.: Multi-spectral Image Fusion Method for Identifying Similar-colored Tomato Organs. En: *2019 IEEE International Conference on Unmanned Systems and Artificial Intelligence (ICUSAI)*, 2019, p. 142{145
- [6] DE LA FUENTE FERNÁNDEZ, Santiago. *Análisis discriminante*. http://www.estadistica.net/Master-Econometria/Analisi s_Di scri mi nante. pdf, (06-05-2020). 2011
- [7] GOGINENI, Rajesh ; CHATURVEDI, Ashvini: *Hyperspectral Image Classification*. Diciembre, 2019
- [8] GUTIÉRREZ, Salvador ; WENDEL, Alexander ; UNDERWOOD, James: Ground based hyperspectral imaging for extensive mango yield estimation. En: *Computers and Electronics in Agriculture* 157 (Febrero, 2019), p. 126{135

- [9] HEMMLEB, Matthias ; A, F. ; B, A. ; C, A. ; MAIERHOFER, Christiane ; WG, Commission: Multi-spectral data acquisition and processing techniques for damage detection on building surfaces, Enero, 2006
- [10] HENNESSY, Andrew ; CLARKE, Kenneth ; LEWIS, Megan: Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability. En: *Remote Sensing* 12 (Enero, 2020), p. 113
- [11] IBÉRICA, Decco. >Que es la antracnosis y como afecta a los cultivos? <https://www.deccoiberica.es/que-es-la-antracnosis-y-como-afecta-a-los-cultivos/>,(15-05-2020). 2018
- [12] KARACA, A. C. ; ÇEŞMECI, D. ; ERTÜRK, A. ; GÜLLÜ, M. K. ; ERTÜRK, S.: Hyperspectral change detection with stereo disparity information enhancement. En: *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2014, p. 1{4
- [13] KAVZOGLU, Taskin ; TONBUL, Hasan ; COLKESEN, Ismail: Dimensionality reduction for hyperspectral images to improve object-based image classification using feature selection and principal component analysis, Octubre, 2018
- [14] LU, Jinzhu ; EHSANI, Reza ; SHI, Yeyin ; ABDULRIDHA, Jaafar ; DE CASTRO, Ana ; XU, Yunjun: Field detection of anthracnose crown rot in strawberry using spectroscopy technology. En: *Computers and Electronics in Agriculture* 135 (Abril, 2017), p. 289{299
- [15] MASSÓ-GARDEN. *Plantas y enfermedades, Antracnosis*. <http://www.massogarden.com/es/plagas-y-enfermedades-es/antracnosis>,(15-05-2020). 2015.
- [16] QIANWEN, L. ; BINGLIANG, H.: The Hyper-Spectral Image Compression System Based on DSP. En: *2008 International Workshop on Education Technology and Training 2008 International Workshop on Geoscience and Remote Sensing Vol. 2*, 2008, p. 171{174
- [17] QIN, Jianwei ; BURKS, Thomas ; KIM, moon s. ; CHAO, Kuanglin ; RITENOUR, Mark: Citrus canker detection using hyperspectral reflectance imaging and PCA-based image classification method. En: *Sensing and Instrumentation for Food Quality and Safety 2* (Septiembre, 2008), p. 168{177
- [18] RODRIGO, Joaquín A. *Análisis discriminante lineal (LDA) y análisis discriminante cuadrático (QDA)*. https://www.cienciadedatos.net/documentos/28_Linear_discriminant_analysis_lda_y_quadrat_discriminant_analysis_qda, (15-05-2020). 2016
- [19] RODRIGUEZ, María J.: *Estadística informaticacasos y ejemplos con el SPSS*. <http://www.digitaliapublishing.com/a/639> : Alicante: Universidad de Alicante, 2001

- [20] SABERIOON, M. ; CÍSAŘ™, P. ; LABBÉ, L.: In VIVO Fish Diet Discrimination using Selected Hyperspectral Image Classification Methods. En: *2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2018, p. 1{5
- [21] STAMATAS, Georgios ; BALAS, Costas ; KOLLIAS, Nikiforos: Hyperspectral Image Acquisition and Analysis of Skin. En: *Proc SPIE* 4959 (Julio, 2003)
- [22] SU, Hongjun ; SHENG, Yehua ; DU, Peijun ; CHEN, Chen ; LIU, Kui: Hyperspectral image classification based on volumetric texture and dimensionality reduction. (Noviembre, 2014)
- [23] SUN, W. ; DU, Q.: Hyperspectral Band Selection: A Review. En: *IEEE Geoscience and Remote Sensing Magazine* 7 (2019), Nr. 2, p. 118{139
- [24] WANG, A. S. ; PELC, N. J.: Sufficient Statistics as a Generalization of Binning in Spectral X-ray Imaging. En: *IEEE Transactions on Medical Imaging* 30 (2011), Nr. 1, p. 84{93
- [25] YANG, Jing-Hui ; WANG, Li-Guo ; QIAN, Jin-Xi: Hyperspectral image classification based on spatial and spectral features and sparse representation. En: *Applied Geophysics* 11 (Diciembre, 2015), p. 489{499
- [26] YAO, Haibo ; HRUSKA, Zuzana ; KINCAID, Russell ; ONONYE, Ambrose ; BROWN, Robert ; CLEVELAND, Thomas: Spectral Angle Mapper classification of fluorescence hyperspectral image for aflatoxin contaminated corn, Julio, 2010, p. 1{4
- [27] YE, X. ; IINO, K. ; ZHANG, S. ; OSHITA, S.: Nondestructive monitoring of chicken meat freshness using hyperspectral imaging technology. En: *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2015, p. 1{4
- [28] YEH, Yu-Hui ; CHUNG, Wei-Chang ; LIAO, Jui-Yu ; CHUNG, Chia-Lin ; KUO, Yan-Fu ; LIN, Ta-Te: Strawberry foliar anthracnose assessment by hyperspectral imaging. En: *Computers and Electronics in Agriculture* 122 (Marzo, 2016), p. 1{9
- [29] ZHANG, Baohua ; FAN, Shuxiang ; LI, Jiangbo ; HUANG, Wenqian ; ZHAO, Chunjiang ; QIAN, Man ; ZHENG, Ling: Detection of Early Rottenness on Apples by Using Hyperspectral Imaging Combined with Spectral Analysis and Image Processing. En: *Food Analytical Methods* 8 (Enero, 2015)