

DETECCIÓN DE ESTRÉS EN CULTIVOS AGRÍCOLAS CON DATOS ESPECTRALES UTILIZANDO MÉTODOS DE SELECCIÓN DE BANDAS

PABLO CARMONA ZULUAGA

MAESTRÍA EN AUTOMATIZACIÓN Y CONTROL INDUSTRIAL

DIRECTORA:

MARÍA CONSTANZA TORRES MADROÑERO



Institución Universitaria

INSTITUTO TECNOLÓGICO METROPOLITANO
FACULTAD DE INGENIERÍAS
MEDELLÍN, 2023

DETECCIÓN DE ESTRÉS EN CULTIVOS AGRÍCOLAS CON DATOS ESPECTRALES UTILIZANDO MÉTODOS DE SELECCIÓN DE BANDAS

Autor: Pablo Carmona Zuluaga

Directoras: María Constanza Torres Madroñero

Facultad: Facultad de Ingenierías

Titulación: Maestría en Automatización y Control Industrial

Palabras clave: estrés abiótico, selección de bandas, firma espectral, hiperespectral, aprendizaje de máquina, sensado remoto, espectrómetro, agricultura de precisión

Resumen

Los sensores espectrales permiten obtener información sobre las propiedades del material a través de la luz reflejada. A partir de este tipo de datos se estudia el estrés abiótico nutricional en los cultivos de maíz, frijol y aguacate. Para este fin, se propone una metodología que utiliza técnicas de reducción de dimensiones basada en selección de bandas y discrimina diferentes tratamientos nutricionales en las plantas. Los métodos elegidos son Bombs, SVD y OCF para selección de bandas y SVM, RF y NN para clasificación y detección del nivel nutricional. La metodología se prueba con tres bases de datos sometidas a diferentes tratamientos nutricionales: cultivo de maíz sometido a deficiencias de nitrógeno, cultivo de frijol sometido a deficiencias por fósforo, y aguacate sometido a deficiencias por fósforo. Los resultados arrojan que para diferencia nutricional de 50 % se detecta el estrés por nitrógeno con precisión de hasta 98 % en maíz y para estrés por fósforo con precisión de hasta 66 % en frijol y 77 % en aguacate. Sin embargo, estos valores bajan al tener diferencia nutricional de 25 % entre clases, obteniendo máxima precisión de 74 % para maíz y 38 % para aguacate. Entonces, se concluye que la detección de niveles de estrés de 50 % es factible para cualquiera de las bases de datos con la metodología propuesta. Sin embargo, actualmente solo es aplicable a la detección de diferencias nutricionales de 25 % en maíz. Es posible que estos resultados mejoren con una mayor cantidad de datos y se considera un importante trabajo a futuro comprobarlo.

CROP STRESS DETECTION USING ESPECTRAL DATA AND BAND SELECTION

Author: Pablo Carmona Zuluaga

Supervisor: María Constanza Torres Madroñero

Department: Faculty of Engineering

Degree: Master's Degree in Automation and Industrial Control

Keywords: abiotic stress, band selection, spectral signature, hyperspectral, machine learning, remote sensing, spectrometer, precision agriculture

Abstract

Spectral sensors obtain information about the material properties through reflected light. Using this type of data, abiotic nutritional stress in corn, bean, and avocado crops is studied. A methodology is proposed that employs dimensionality reduction through band selection and classifies different stress levels present in plants. The chosen methods are Bombs, SVD, and OCF for band selection and SVM, RF, and NN for classification and nutritional level detection. The methodology is tested with three databases subjected to different nutritional treatments: maize crop subjected to nitrogen deficiencies, bean crop subjected to phosphorus deficiencies, and avocado subjected to phosphorus deficiencies. Results show that for a 50% nutritional difference, nitrogen stress detection is up to 98% in corn, and phosphorus stress detection is up to 66% in beans and 77% in avocado. However, these values decrease when there are 25% nutritional differences between classes, with a maximum accuracy of 74% for corn and 38% for avocado. It is concluded that stress level detection of 50% is feasible for any of the databases with the proposed methodology, but it is currently only applicable to the detection of 25% nutritional differences in corn. It is possible that these results will improve with a greater amount of data and it is considered an important future work to verify it.

Agradecimientos

En primer lugar, quiero expresar mi más sincera gratitud a mi directora de tesis, la Doctora Maria Constanza Torres Madroñero. Su dedicación, conocimientos y paciencia fueron fundamentales para el desarrollo de este proyecto. Agradezco sinceramente sus valiosas sugerencias, sus comentarios constructivos y su guía constante a lo largo de esta investigación.

Deseo agradecer a los miembros de mi comité de evaluación, Miguel A. Goenaga, Ariolfo Velasco y Andres Eduardo Castro. Su experiencia y conocimientos en el área de estudio fueron invaluable para mejorar la calidad de mi trabajo. Agradezco sus valiosas observaciones y sugerencias, las cuales me permitieron fortalecer mis argumentos y enfoques.

Ofrezco mi gratitud a Agrosavia por su trabajo en la agronomía y preparación de los experimentos. Quiero agradecer especialmente a Manuel Mauricio Goes, Camilo Acevedo y a todos los compañeros que trabajaron conmigo en la toma de datos, ya que sin su esfuerzo, no hubiera sido posible llegar a este punto.

Agradezco de corazón a mi madre, Teresita Zuluaga, y a mi novia, Valentina Gutierrez, por su amor incondicional y paciencia. Me brindaron un apoyo constante, motivación y comprensión durante los momentos desafiantes. Además, a mis amigos Nicolás Molina, Brayhan Zuluaga y Mateo Velásquez, su colaboración y ánimo fueron un motor fundamental para alcanzar este objetivo.

Por último, quiero reconocer al Instituto Tecnológico Metropolitano y a Min-ciencias, que hicieron posible la realización de este trabajo de grado a través del programa RC 475-2020 "Sistemas inteligentes para el monitoreo de cultivos permanentes y transitorios". Su apoyo financiero y recursos fueron de gran importancia para llevar a cabo esta investigación.

Acrónimos

HSI	Hiperspectral Images - Imágenes Hiperespectrales
NN	Neural Network - Red Neuronal
RF	Random Forest
SVM	Support Vector Machines
SVD	Singular Value Decomposition
OCF	Optimal Clustering Framework
MLP	Multi layer perceptron - perceptrón multicapa
DNN	Deep Neural Networks - Redes Neuronales Profundas
MI	Mutual Information, Información Mutua
CNN	Convolutional Neural Networks - Redes Neuronales Convolucionales
E-FDPC	Enhanced Fast Density-Peak-based Clustering
BOMBS	multi-objective optimization model - modelo de optimización multiobjetivo

Tabla de Contenido

Resumen	II
Abstract	III
Agradecimientos	IV
Acrónimos	V
1 Introducción	1
1.1 Justificación	1
1.2 Problema de Investigación	3
1.3 Objetivos	5
1.3.1 General	5
1.3.2 Específicos	5
1.4 Estructura del documento	5
1.5 Contribuciones	5
2 Estado del Arte y Marco Teórico	7
2.1 Agricultura de precisión	8
2.2 Percepción Remota	8
2.3 Procesamiento de datos espectrales para monitoreo de vegetación	9
2.4 Métodos de selección de bandas	10
2.4.1 Ranking	11
2.4.2 Dispersión	11

2.4.3	Búsqueda	11
2.4.4	Agrupación	12
2.4.5	Aprendizaje	12
2.4.6	Híbridos	13
2.5	Métodos seleccionados	14
2.5.1	Singular value decomposition band subset selection (SVD)	14
2.5.2	Optimal Clustering Framework	15
2.5.3	Multi-objective optimization module	16
3	Metodología	17
3.1	Toma de datos	18
3.1.1	Maíz: deficiencia de nitrógeno	19
3.1.2	Aguacate: Deficiencia de fósforo	20
3.1.3	Fríjol: Deficiencia de fósforo	21
3.2	Preprocesamiento	22
3.3	Técnicas de clasificación	24
3.4	Metodología para detección de estrés usando selección de bandas	26
4	Resultados y Discusión	28
4.1	Maíz	29
4.2	Aguacate	37
4.3	Fríjol	47
4.4	Discusión	52
	Conclusiones y Trabajo Futuro	54
	A Bandas seleccionadas	55
	B Parámetros en búsqueda de grilla	58
	Bibliografía	66

Índice de figuras

3.1	Fotografías experimento maíz	19
3.2	Fotografías experimento Aguacate	20
3.3	Fotografía plantación Frijol	21
3.4	Firma espectral sin procesar	22
3.5	Firma espectral cortada para eliminar bandas de mayor ruido.	22
3.6	Firma espectral luego de filtro de mediana deslizante	23
3.7	Esquema metodología detección de estrés abiótico usando selección de bandas	26
4.1	Firma promedio para cada clase	29
4.2	Conjunto de 20 bandas seleccionadas con SVD	34
4.3	Matriz de confusión clasificación SVM: 20 bandas SVD	35
4.4	Conjunto de 28 bandas seleccionadas con BOMBS	36
4.5	Matriz de confusión clasificación SVM: BOMBS 28 bandas	37
4.6	Firma promedio para cada clase	38
4.7	Conjunto de 32 bandas seleccionadas con BOMBS	43
4.8	Matriz de confusión clasificación SVM: BOMBS 32 bandas	44
4.9	Conjunto de 32 bandas seleccionadas con OCF	45
4.10	Matriz de confusión clasificación NN: 32 bandas OCF	46
4.11	Firma promedio para cada clase	47
4.12	Conjunto de 32 bandas seleccionadas con SVD	51
4.13	Matriz de confusión clasificación SVM: 32 bandas SVD	52

Índice de Tablas

3.1	Bases de datos: tamaño y rango espectral	23
4.1	Distribución de muestras de maíz entre clases	29
4.2	Precisión con RF para dos experimentos en Maíz.	30
4.3	Precisión con NN para dos experimentos en Maíz.	32
4.4	Precisión con SVM para dos experimentos en Maíz.	33
4.5	Métricas derivadas de clasificación SVM: SVD 20 bandas	34
4.6	Métricas derivadas de clasificación SVM: BOMBS 28 bandas	36
4.7	Distribución de muestras de aguacate entre clases	38
4.8	Precisión con RF para dos experimentos en aguacate.	39
4.9	Precisión con NN para dos experimentos en aguacate.	40
4.10	Precisión con SVM para dos experimentos en aguacate.	41
4.11	Métricas derivadas de clasificación SVM: BOMBS 32 bandas	44
4.12	Métricas derivadas de clasificación NN: OCF 32 bandas	46
4.13	Distribución de muestras de fríjol entre clases	47
4.14	Precisión con RF en fríjol	48
4.15	Precisión con NN en fríjol	49
4.16	Precisión con SVM en fríjol	50
4.17	Métricas derivadas de clasificación SVM: SVD 32 bandas	51
A.1	Selección 20 bandas algoritmo SVD dataset Maíz	55
A.2	Selección 28 bandas algoritmo Bombs dataset Maíz	56
A.3	Selección 32 bandas algoritmo Bombs dataset Aguacate	56

A.4	Selección 32 bandas algoritmo OCF dataset Aguacate	57
A.5	Selección 32 bandas algoritmo SVD dataset Frijol	57
B.1	Parametros usados en la busqueda de grilla para el algoritmo RF .	58
B.2	Parametros usados en la busqueda de grilla para el algoritmo SVM	58
B.3	Parametros usados en la busqueda de grilla para el algoritmo NN .	59

Capítulo 1

Introducción

Contenido

1.1 Justificación	1
1.2 Problema de Investigación	3
1.3 Objetivos	5
1.3.1 General	5
1.3.2 Específicos	5
1.4 Estructura del documento	5
1.5 Contribuciones	5

1.1. Justificación

La Misión de Sabios Colombia–2019 estableció el foco temático Tecnologías Convergentes e Industrias 4.0 para el avance en el desarrollo y competitividad del país a través de actividades de ciencias, tecnología e innovación. La Industria 4.0 o también conocida como la cuarta revolución industrial está relacionada con diversas tecnologías habilitadoras, entre las que se encuentran el Internet de las Cosas (IoT), sistemas ciber-físicos, y las tecnologías de la información y la comunicación (TICs) [1]. La Industria 4.0 tiene varias aplicaciones, sin embargo, es de interés para el país su aplicación en la agroindustria, campo en el cual se encuentran diversos trabajos previos [2–5]. Las Industrias 4.0 aplicadas a la agricultura buscan el desarrollo de nuevos sistemas más baratos de sensado, procesadores de bajo costo, comunicación móvil, sistemas TICs en la nube, y analítica de datos. En trabajos anteriores, este campo de aplicación se denomina

Agricultura 4.0 o agricultura digital, la cual se considera un área dentro de la agricultura de precisión. A través de este campo tecnológico se busca aumentar la competitividad del uso del suelo en cultivos permanentes y transitorios como parte de la transformación social y técnica, y a través de la convergencia de varias tecnologías [6].

Por otra parte, el Plan Estratégico de Ciencia, Tecnología e Innovación del Sector Agropecuario – PECTIA 2017-2027 [7] destaca las oportunidades de investigación y desarrollo que existen desde diversos campos de la ingeniería para la solución de problemas del agro colombiano. Entre el Ministerio de TICs y el Ministerio de Agricultura y Desarrollo Rural se establecieron ejes temáticos para el desarrollo de las actividades de ciencia, tecnología e innovación, enmarcados en el enfoque denominado agricultura electrónica. En este eje temático se destacan el desarrollo e implementación de soluciones basadas en agricultura de precisión, soluciones TICs para la captura y procesamiento de información en campo, gestión de información, gestión del conocimiento y mejora de la asistencia técnica, gestión de información climática y Big-Data.

Adicionalmente, la Agenda Regional de I+D+i “Sociedades y territorios sostenibles e inclusivos” construida por el grupo G8 de Antioquia (UPB, CES, EAFIT, UNAL, UdeA, UdeM, La Salle, e ITM) establece como campo estratégico la “Naturaleza y Recursos” que incluye la línea estratégica “Tecnificación del Agro”, la cual busca impulsar el desarrollo e implementación de “herramientas y procedimientos con un alto grado de tecnificación en el aprovechamiento óptimo de los recursos naturales, con el objetivo de contribuir en la transformación del sector agroindustrial colombiano hacia un sector más productivo y competitivo” [8].

Para lograr ese desarrollo productivo y competitivo del sector agroindustrial se deben enfrentar diferentes retos. Por ejemplo, el rendimiento de los cultivos es afectado por factores bióticos como enfermedades y plagas, y por factores abióticos como el cambio climático, los niveles de humedad y la fertilidad del suelo. El agua es considerada el segundo factor más limitante en un cultivo [9]. Este elemento es esencial para los diferentes procesos químicos y fisiológicos de la planta. La deficiencia de agua en el suelo reduce la productividad de un cultivo, afectando la absorción de nutrientes. Otros factores abióticos que afectan los cultivos son la alta temperatura, aplicación indiscriminada de agroquímicos y las enfermedades causadas por patógenos de la raíz y del brote [10]. Estos factores se agravan más con los efectos del cambio climático, ya que se espera un incremento en plaga y enfermedades, reduciendo los rendimientos del cultivo. El estudio presentado en [11], reveló una amenaza seria para la seguridad alimentaria: por cada grado Celsius que aumente la temperatura del planeta, se perderá entre el 10 % y 25 % de la productividad de arroz, maíz y trigo, es decir,

cada año se dejarían de producir cerca de 213 millones de toneladas.

Monitorear las variables que afectan la productividad y calidad de los cultivos, como humedad y nutrientes, presenta varias dificultades. Esto debido a la gran variedad de factores que inciden en los cultivos, tales como: la topología de cultivo, sequías, inundaciones, corrientes de aire, ciclo del cultivo, entre otros, generando una alta variabilidad temporal y espacial dentro del mismo [12]. Para tener un nivel de control sobre estas variables es necesario tener datos actuales. Sin embargo, el procedimiento regular de análisis de suelos conlleva un alto costo laboral, requiriendo la toma de muestras y análisis en laboratorio. Dado estos costos, se realizan pocas muestras espaciadas dentro de la extensión del cultivo, perdiendo nivel de detalle del comportamiento del estrés del cultivo [13]. Por esta razón, es necesario la exploración de nuevas tecnologías que permitan realizar evaluaciones periódicas de los cultivos más efectivas y rápidas.

Una tecnología con gran potencial para el monitoreo de cultivos es la percepción remota basada en sensores multiespectrales e hiperespectrales. Los sistemas espectrales miden la radiancia reflejada y emitida por una superficie a lo largo del espectro electromagnético [14]. Hoy en día existen diversos sistemas espectrales que pueden ser empleados para monitoreo de cultivos. Estos sistemas incluyen satélites (Landsat, Sentinel, Planet Scope) y cámaras para drones (Micasense, Hyspex, SOC). En el mercado se encuentran también soluciones desarrolladas directamente para esta actividad, como el dron P4 Multiespectral de DJI. El impacto de estas tecnologías, depende de los factores que afectan en la respuesta espectral de las plantas, tales como: condiciones geográficas, condiciones ambientales, fenológicas, de suelo, condiciones de estrés biótico o abiótico, así como la variabilidad vegetal. Sin embargo, hacen falta estudios que permitan entender la relación entre cultivos y condiciones de estrés específicas.

1.2. Problema de Investigación

La detección de estrés en plantas tradicionalmente se realiza tomando muestras físicas en cada individuo, lo cual representa una gran inversión en tiempo y mano de obra. La realización de este análisis por medio de detección remota presenta ventajas sobre los métodos tradicionales, ya que permite obtener datos de un espacio muestral más extenso y desde la distancia [15]. Dentro de las posibles herramientas de detección remota se opta por las imágenes multiespectrales e hiperespectrales. Las cámaras espectrales son sistemas pasivos que emplean la luz solar como fuente de energía para medir propiedades físicas de una superficie a lo largo de diferentes longitudes de onda del espectro electromagnético. La

respuesta espectral de una superficie forma una firma, cuya forma depende de los materiales y sus características físicas y químicas. De esta forma, existen en la literatura diversos estudios que emplean la firma espectral de vegetación para el monitoreo de los cultivos [16–19]. Las imágenes hiperespectrales (HSI) suelen tener cientos de bandas, sin embargo, el método más utilizado para realizar la detección de estrés en cultivos (los índices de vegetación) toman un subconjunto de tres bandas [20–22], lo que es una pérdida de información útil. Como alternativa se tiene la implementación de algoritmos de aprendizaje de máquina que permitan detectar y analizar las secciones más importantes de la imagen para la clasificación de estrés en plantas.

Para hacer uso de sistemas multiespectrales e hiperespectrales para el monitoreo y detección de estrés en cultivos se debe establecer las resoluciones espaciales y espectrales, lo cual afecta directamente el costo de los sistemas. Para establecer la resolución y rangos espectrales que permitan detectar el estrés se pueden emplear métodos automáticos de selección de bandas, estos métodos permiten determinar las longitudes de ondas discriminantes para un problema específico [23]. Adicionalmente, los métodos de selección de bandas permiten reducir la redundancia de información en los datos espectrales, disminuyendo el efecto de la “maldición de la dimensionalidad” en los sistemas automáticos de aprendizaje. Para esto se encuentran variadas técnicas de selección de bandas en la literatura [23]. Sin embargo, específicamente para el problema de detección de estrés, y hasta lo que conocemos, no se han aplicado técnicas automáticas de selección de bandas relevantes. La mayoría de estudios se centran en el uso de los índices de vegetación, limitándose a aprovechar dos o tres bandas del espectro. En este sentido se considera que para solucionar la clasificación de HSI de vegetación, hace falta investigación en lo relacionado a metodologías que permitan abordar técnicas de aprendizaje de máquina como reemplazo de los índices de vegetación y que incluyan la selección de bandas para reducción de dimensionalidad y redundancia.

Lo anterior permitió plantear la pregunta de investigación: ¿Es posible plantear una metodología que permita la detección de estrés abiótico en cultivos transitorios, como el maíz y el frijol, a partir de datos espectrales y utilizando algoritmos de aprendizaje de máquina y selección de bandas para reducción de dimensiones?

1.3. Objetivos

1.3.1. General

Desarrollar una metodología para la detección de estrés abiótico en cultivos transitorios, como el maíz y el frijol, empleando datos espectrales (HSI y firmas espectrales), técnicas de selección de bandas para reducción de dimensiones y técnicas de aprendizaje de máquina.

1.3.2. Específicos

- Comparar técnicas de reducción de dimensiones basadas en selección de bandas de tipo agrupamiento o aprendizaje para el procesamiento de datos espectrales capturados sobre cultivos transitorios.
- Proponer una metodología basada en técnicas de selección de bandas y que utilice técnicas de aprendizaje de máquina para la detección de estrés abiótico en cultivos transitorios utilizando datos espectrales.
- Evaluar la metodología para detección de estrés abiótico en cultivos transitorios empleando datos espectrales reales usando métricas de precisión derivadas de la matriz de confusión (precisión, especificidad y sensibilidad)

1.4. Estructura del documento

Una vez finalizada esta introducción se abarca, en el Capítulo 2, la revisión del estado del arte y las bases teóricas que fundamentan el presente trabajo. Posteriormente en el Capítulo 3, se detallan las bases de datos y la configuración experimental que se llevó a cabo para obtener los resultados que se describen en el Capítulo 4. Por último, el Capítulo 5, se describen las conclusiones de la tesis y se sugieren trabajos futuros.

1.5. Contribuciones

- Ponencia “Abiotic Maize Stress Detection Using Hyperspectral Signatures and Band Selection”. Third International Conference, SmartTech-IC 2022, Quito, Ecuador, November 16-18, 2022

-
- Artículo “Spectral Library of Maize Leaves under Nitrogen Deficiency Stress”. Data 2023, 8, 2. <https://doi.org/10.3390/data8010002>
 - Artículo aceptado “Abiotic Maize Stress Detection Using Hyperspectral Signatures and Band Selection”. SmartTech-IC 2022 proceedings, Communications in Computer and Information Science (CCIS) series Springer.

Capítulo 2

Estado del Arte y Marco Teórico

Contenido

2.1 Agricultura de precisión	8
2.2 Percepción Remota	8
2.3 Procesamiento de datos espectrales para monitoreo de vegetación	9
2.4 Métodos de selección de bandas	10
2.4.1 Ranking	11
2.4.2 Dispersión	11
2.4.3 Búsqueda	11
2.4.4 Agrupación	12
2.4.5 Aprendizaje	12
2.4.6 Híbridos	13
2.5 Métodos seleccionados	14
2.5.1 Singular value decomposition band subset selection (SVD)	14
2.5.2 Optimal Clustering Framework	15
2.5.3 Multi-objective optimization module	16

Sinopsis

Este proyecto de tesis de maestría busca explorar técnicas de aprendizaje de máquinas y reconocimiento de patrones para la detección de condiciones de estrés abiótico en cultivos agrícolas. A continuación, se presenta, un breve análisis

del estado del arte, incluyendo las áreas de agricultura de precisión, percepción remota, procesamiento de imágenes espectrales, y enfoques de detección de estrés desde datos espectrales. A partir de la revisión de literatura se establece la necesidad de nuevas metodologías de procesamiento que aprovechen técnicas modernas de reconocimiento de patrones para el análisis de datos espectrales en aplicaciones de agricultura de precisión.

2.1. Agricultura de precisión

Actualmente la agricultura se encuentra pasando por una cuarta revolución, principalmente gracias a las nuevas tecnologías que están siendo aplicadas al monitoreo y recolección de datos sobre las plantas. A esto se le llama agricultura de precisión y puede definirse como las estrategias que utilizan herramientas avanzadas de información, comunicación y análisis para el proceso de toma de decisiones, lo que ayuda a mejorar el rendimiento general de los cultivos, a reducir la pérdida de agua y nutrientes, o a disminuir los impactos ambientales [24]. Algunas de las variables agrícolas sobre las que se considera que puede tenerse control son la aplicación de pesticidas, fertilizante, semillas, mano de obra o la cantidad de combustible usado [25]. Las tecnologías emergentes de mayor relevancia para la agricultura de precisión incluyen la inteligencia artificial, Big Data, internet de las cosas, los sistemas de información geográfica, los sistemas de posicionamiento global y la percepción remota [26].

2.2. Percepción Remota

La percepción remota se refiere a la obtención de datos sobre un evento u objeto desde la distancia. Esto se hace mediante el uso de sensores que permitan la medición de la radiación electromagnética que emite o refleja el objeto medido [15]. Una de las formas más utilizadas de la percepción remota es el monitoreo satelital de los recursos de la tierra, aunque las observaciones de campo se han hecho más comunes con el tiempo y son utilizadas para obtener mayor precisión sobre el fenómeno estudiado. La respuesta típica en el espectro electromagnético (usualmente visible, infrarrojo o microondas) para diferentes materiales varía, lo que permite aplicar la percepción remota a la diferenciación de vegetación, supervisión de crecimiento de plantas, monitoreo hídrico y agricultura de precisión por mencionar algunas aplicaciones [27]. Cuando la percepción remota satelital comenzó, la mayor parte de las investigaciones se enfocaban en el uso de los datos obtenidos para clasificar la superficie de la tierra, aquellos interesados en

agricultura tenían énfasis especial en la clasificación de los diferentes cultivos. En los últimos años, este análisis se ha enfocado en las características biofísicas de las plantas [27].

Entre los sistemas de percepción remota se encuentran los sistemas multiespectrales e hiperespectrales. Estos sistemas pasivos miden la energía reflejada o emitida por una superficie a lo largo de decenas o cientos de bandas espectrales. Las imágenes capturadas por estos sensores pueden verse como un cubo: con dos dimensiones espaciales y una tercera dimensión que almacena la información espectral. La forma y caracterización de la respuesta espectral permite la identificación de materiales presentes en el campo de visión del sensor [14].

La firma espectral de vegetación ha sido ampliamente estudiada [16–19], permitiendo determinar efectos puntuales del estrés en la reflectancia. Por ejemplo, se ha demostrado que cuando una planta está bajo estrés hídrico, la reflectancia entre las bandas espectrales del rojo y verde incrementa debido a la reducción de concentración de clorofila [19, 28]. Si hay estrés por nitrógeno, la reflectancia en las bandas rojo y borde-rojo decrece, dado que los niveles de clorofila disminuyen, cambiando el color de la hoja de verde a un tono amarilloso [19].

2.3. Procesamiento de datos espectrales para monitoreo de vegetación

Para el monitoreo del estrés en las plantas a partir de imágenes espectrales, se encuentran en la literatura métodos basados en índices de vegetación, modelos biofísicos y métodos de aprendizaje de máquina.

Los índices de vegetación se han estudiado para predecir numerosas características como área foliar, contenido de clorofila, contenido de agua en la planta, biomasa, rendimiento, entre otros [20–22]. A pesar del amplio uso de los índices de vegetación en agricultura, estos se limitan a emplear pocas bandas de la información proporcionada por los sistemas espectrales. Aunque son deseables por su simplicidad y baja complejidad computacional, no explotan toda la capacidad de los datos. Adicionalmente, no es claro aún cual es la mejor combinación de bandas para obtener una mejor caracterización de los cultivos [29].

Una alternativa a los índices de vegetación son los modelos biofísicos, los cuales exploran la interacción de la luz con las plantas en términos de sus características biofísicas [30]. Los modelos biofísicos describen la variación espectral como una función de la cobertura, la hoja y las características del suelo [31]. La estimación de las características biofísicas depende de un proceso de in-

versión, el cual consiste en minimizar la diferencia entre datos reales y simulados. Trabajos previos han demostrado que la inversión de este tipo de modelos de transferencia desde datos capturados por sensores remotos es (1) costosa computacionalmente si se incluyen las características complejas de las plantas, (2) varios de los parámetros presentan una interrelación que dificulta su estimación, y (3) la resolución espacial de las cámaras limita la información disponible para la inversión [31].

Por último, el aprendizaje de máquina es una familia de algoritmos que identifican patrones en conjuntos de datos, para el estrés abiótico en plantas se ha usado mayormente en la literatura un acercamiento supervisado. Lo que significa que cada muestra debe estar etiquetada y que el desempeño y complejidad de técnicas usables depende de la cantidad de datos a la mano. La reducción de precio en sensores espectrales ha permitido aumentar las bases de datos existentes en los últimos diez años, a su vez permitiendo la implementación de algoritmos de mayor complejidad. Barbedo [32], presenta un análisis de esta tendencia desde el año 2010, donde se destaca el uso de regresión lineal, máquinas de soporte vectorial (SVM), random forest (RF), K-nearest neighbor y de redes neuronales artificiales (NN). Las NN presentan una gran cantidad de arquitecturas y topologías, las más usadas son: multi layer perceptron (MLP), extreme learning machines, radial basis function y redes neuronales profundas (DNN) [32]. Las redes convolucionales son las de mayor importancia en la literatura dentro de las DNN, dado su buen desempeño general para clasificación de imágenes, pero tienen el problema de requerir bases de datos de gran tamaño [33, 34].

2.4. Métodos de selección de bandas

Las técnicas basadas en aprendizaje de máquina pueden aplicarse directamente sobre los datos originales o sobre nuevos espacios de transformación. En el caso de las HSI es de interés los métodos de reducción de dimensiones. Las HSI presentan considerables complicaciones al momento de analizarlas. Esto se debe a que estos datos son de un gran volumen computacional, lo que complica obtener un gran número de muestras [35]. Además, presentan redundancia entre canales adyacentes de la imagen y puede que una parte del rango espectral del sensor no sea relevante para una aplicación específica. Por tanto, la reducción de dimensionalidad se convierte en un proceso de vital importancia, en este se quiere eliminar la redundancia en sus componentes mientras se mantiene la información vital [36]. Se tienen dos acercamientos posibles que cumplen con este objetivo. Primero la extracción de características, en esta se representan los datos originales en un espacio de características diferente, modificando los datos

iniciales lo que podría complicar su interpretación final [37]. Segundo, se puede implementar métodos de selección de bandas, el objetivo de estas técnicas es elegir el conjunto de bandas que contenga la información relevante para el problema en cuestión e ignorar las que presentan información redundante [38]. Este último acercamiento presenta ventajas para la interpretación por lo que se revisará con mayor profundidad. La selección de bandas puede clasificarse actualmente en seis grupos diferentes considerando los métodos utilizados [23], estos son: ranking, dispersión, búsqueda, agrupación, aprendizaje e híbridos, y se presentan con mayor detalle a continuación:

2.4.1. Ranking

Los métodos basados en clasificación cuantifican la importancia de cada banda espectral dependiendo de un criterio predefinido de selección y posteriormente elige las que obtuvieron mejor puntaje en una secuencia ordenada. Estos métodos pueden ser supervisados o no supervisados, dependiendo de si los datos utilizados poseen etiquetas. Las técnicas no supervisadas consideran el principio de la disimilitud o correlación de bandas. En cuanto a las técnicas supervisadas, estas implican conocimiento previo sobre los datos espectrales para poder realizar la construcción del criterio de clasificación [39, 40].

2.4.2. Dispersión

Los métodos basados en la dispersión hacen uso de la regresión para revelar estructuras subyacentes de las HSI. Las bandas representativas pueden entonces ser encontradas resolviendo un problema de optimización considerando restricciones de dispersión. Los métodos actuales pueden ser categorizados en "sparse nonnegative matrix factorization", representación de dispersión y los basados en regresión de dispersión [23, 41, 42].

2.4.3. Búsqueda

Los métodos basados en búsqueda convierten la selección de bandas en un problema de optimización en el cual se debe considerar dos factores principalmente 1) función de criterio 2) estrategia de búsqueda, basado en estos se realiza una búsqueda de las bandas que den como resultado una solución óptima. La función de criterio puede ser de dos tipos principales, basada en similitud o basada en información. Dentro de las funciones basadas en similitud se encuentra la distancia Euclidiana, la distancia Bhattacharyya, la distancia Jeffries–Matusita,

el mapeo del ángulo espectral o la medición del índice de similitud estructural. Por otro lado, dentro de las funciones basadas en información se encuentra la divergencia de información espectral, divergencia transformada y información mutua (MI) [43]. En cuanto a la estrategia de búsqueda se encuentran agrupados en tres categorías, búsqueda incremental, búsqueda actualizada y búsqueda de eliminación. Con estas se quiere determinar la mejor manera de encontrar la solución óptima para la función de criterio [23, 44].

2.4.4. Agrupación

Los métodos basados en agrupación unen las bandas originales en conjuntos similares y seleccionan las más representativas de cada conjunto para construir la selección final de bandas. Los primeros trabajos en esta área tratan la agrupación jerárquica de bandas, en este intentan minimizar la varianza dentro de cada conjunto seleccionado y al mismo tiempo maximizar la varianza entre los diferentes conjuntos. Las bandas representativas pueden ser seleccionadas usando mediciones de información, como la MI [43]. desde entonces muchos algoritmos diferentes han sido propuestos, sin embargo, la mayoría de estos se basan en tres técnicas, propagación de afinidad, graph clustering y k-medias [23, 45, 46].

2.4.5. Aprendizaje

Los métodos encontrados en esta sección requieren de datos etiquetados como base para su funcionamiento, en la percepción remota usando HSI existe una considerable escasez de este tipo de bases de datos. Esto se debe principalmente a la complejidad de la tarea de etiquetar imágenes satelitales píxel por píxel. Sin embargo, las bases de datos libres que se encuentran en esta área son suficientes para la prueba de los algoritmos, pero en aplicaciones específicas poder usarlos depende de la capacidad de etiquetar los datos. Son pocas las técnicas encontradas en esta área que traten las HSI para la selección de bandas. Se tiene por ejemplo el uso de redes neuronales convolucionales (CNN) en [47, 48], estas se aplican de una manera similar a la utilizada en el procesamiento de imágenes no espectrales. Aunque se obtiene precisión considerable, se deja abierto el campo para otros tipos de redes que representen mejor el volumen de las HSI. Se plantea entonces el uso de CNN 3D en [49] y de redes neuronales basadas en atención en [50, 51], todas plantean el uso de técnicas de aprendizaje de máquina de estado de arte, consiguiendo un alto desempeño frente a las diferentes métricas de precisión medidas en cada artículo.

2.4.6. Híbridos

Estos métodos consideran una combinación de dos o más de los esquemas presentados anteriormente para la selección de bandas. Dentro de lo más usual en la literatura se ha visto la combinación de métodos basados en agrupamiento con los basados en ranking para formar un algoritmo híbrido [52, 53].

A pesar de los avances en el área de percepción remota basada en datos espectrales para la detección de estrés, aún no se encuentran metodologías completamente confiables que permitan detectar las deficiencias nutricionales o hídricas de un cultivo. Por lo cual es necesario de más estudios, que permitan analizar la información espectral de los datos.

2.5. Métodos seleccionados

2.5.1. Singular value decomposition band subset selection (SVD)

Descripción

El método SVD [54] para la selección de bandas espectrales es de tipo no supervisado y es basado en algebra lineal. Se basa en la búsqueda de un subconjunto de columnas independientes que revelan descomposiciones. Hace parte de los métodos de ranking descritos en la sección anterior, ya que el subconjunto hallado contiene las bandas ordenadas de mayor a menor importancia. El algoritmo de selección se resume a continuación.

Algoritmo 1 SVD

- 1: Construya una representación matricial del cubo hiperespectral de la siguiente manera. Sea $X = [x_1, x_2, \dots, x_N]$, donde N es el número de firmas espectrales. Observe que cada fila corresponde a una banda del cubo y cada columna corresponde a una firma.
- 2: Construya la matriz normalizada $Z = [z_1, z_2, \dots, z_N]$ que se obtiene restando la media de cada píxel y normalizando cada banda a la varianza unitaria de la siguiente manera:

$$z = \frac{D_1}{2(x - \mu x)}$$

donde $D = \text{diag}\{\sigma_{x_{21}}, \sigma_{x_{22}}, \dots, \sigma_{x_{2d}}\}$ y $\sigma_{x_{2i}}$ es la varianza de la i -ésima banda.

- 3: Calcule la descomposición en valores singulares de Z y determine su rango d (se puede determinar el d de esta forma o usar un valor a priori).
 - 4: Calcule la factorización QR con pivote de la matriz V_1^T donde V_1 está formado por los primeros d vectores singulares de X . P es la matriz resultante de la factorización.
 - 5: Calcule $Y = PX$.
 - 6: **Fin:** Tome las primeras d filas de Y como las bandas seleccionadas.
-

Justificación elección método

Este método es de uso común en la literatura y su acercamiento matemático le permite desempeñarse bien en datos de diferente distribución. Además, po-

see gran velocidad de procesamiento. Estas características son las razones por las que ha sido escogido para los experimentos que serán presentados en las siguientes secciones de este documento.

2.5.2. Optimal Clustering Framework

Descripción

Optimal Clustering Framework [55] (OCF) está diseñado para extraer el valor de la contribución de cada banda hacia las diferentes clases de datos en la imagen. El framework implica una estrategia de ranking en agrupación, donde el método de agrupamiento deseado, como K-Means, separa la imagen de entrada en un número arbitrario de clases. Luego, se clasifican la contribución de las bandas hacia cada clase, utilizando una función objetivo seleccionada para determinar las bandas con mejor desempeño.

Para este experimento, se utilizó el Enhanced fast density-peak-based clustering (E-FDPC) [56] como método de ranking para evaluar el aporte de cada banda a las diferentes clases. Primero, determinando la densidad local, y segundo, evaluando que la cercanía a bandas con mayor densidad local. El método de agrupación utilizado fue "Self-tuning spectral clustering" [57], tomando "corte de rango superior" (Top Rank Cut - TRC) como función objetivo. Esta sección cuenta con tres pasos principales, primero el proceso de clustering espectral se realiza sobre la base de datos. Segundo, se elige dentro de cada grupo las bandas con mejor ranking obtenido por E-FDPC y se calcula la similitud entre ellas. Luego, se minimiza el valor de similitud obtenida por cluster.

Justificación elección método

Los autores de este algoritmo obtuvieron precisión que superaba el estado del arte al momento de publicación y que actualmente sigue siendo competitiva, pero de mayor importancia para esta investigación es la capacidad de adaptarse a diferentes distribuciones de datos. Los resultados presentados en [55], se realizaron en cuatro bases de datos libres diferentes: Indian Pines, Kennedy Space Center, Salinas y Pavia University. La consistencia de clasificación presentada por los autores es entonces el principal motivo de elegir el método OCF en la metodología propuesta.

2.5.3. Multi-objective optimization module

Descripción

Zhang, Gong y Chan propusieron un modelo de optimización multiobjetivo [58] (BOMBS), que explora las relaciones entre la redundancia y la cantidad de información en las bandas espectrales. Una función objetivo es utilizada para medir cada una de estas dos características. La entropía de formación promedio (Average formation entropy) se utilizó para medir la cantidad de información en una banda específica. Esto generalmente se define usando variables aleatorias continuas, sin embargo, también puede ser un problema discreto con cada banda del conjunto de datos como una variable aleatoria discreta X . La distribución de probabilidad $p(X)$:

$$p(X) = \frac{h(X)}{Num} \quad (2.1)$$

Donde $h(X)$ es el histograma de una banda y Num es la totalidad de píxeles en $h(X)$. Este proceso otorga una puntuación alta a las imágenes claras y una puntuación más baja a las bandas ruidosas.

El segundo aspecto que debe medirse es la correlación entre bandas. La entropía cruzada calcula la distancia relativa entre las distribuciones de probabilidad de las bandas A y B, dada por la ecuación:

$$H(A, B) = E_A(-\log(B)) + E_B(-\log(A)) \quad (2.2)$$

Donde E_x es la expectativa que obedece a la distribución de probabilidad de x . Para encontrar una solución óptima y compensar estas dos funciones objetivo, los autores propusieron algoritmo multiobjetivo inspirado en el sistema inmunitario.

Justificación elección método

Se presenta en [55] que BOMBS obtiene clasificación superior al estado del arte en el momento de publicación, además de ser similar en su desempeño a OCF. Los resultados presentados que presentaron los autores, se realizaron en tres bases de datos libres diferentes: Indian Pines, Salinas y Pavia University. La consistencia de clasificación presentada es ideal para las pruebas a realizar como parte de a metodología presentada.

Capítulo 3

Metodología

Contenido

3.1 Toma de datos	18
3.1.1 Maíz: deficiencia de nitrógeno	19
3.1.2 Aguacate: Deficiencia de fósforo	20
3.1.3 Frijol: Deficiencia de fósforo	21
3.2 Preprocesamiento	22
3.3 Técnicas de clasificación	24
3.4 Metodología para detección de estrés usando selección de bandas	26

Sinopsis

En este Capítulo se definen los datos y los experimentos realizados. Inicialmente se define la base de datos y la información extraída de esta a partir de experimentos previos y pruebas adicionales. Posteriormente se describe cómo se realiza la agrupación de los datos espectrales y cómo se lleva a cabo la selección de características relevantes. Por último se definen las pautas para realizar evaluación de la predicción.

3.1. Toma de datos

Se establecieron experimentos con aguacate (cultivo permanente), maíz y frijol (cultivos transitorios) en los centros de investigación de AGROSAVIA “La Selva” y “Turipaná”. La Selva está ubicado en el sector Llanogrande, municipio de Rionegro, Antioquia Colombia. Esta zona presenta una temperatura media anual de 17 °C, con precipitaciones de 1917 mm, humedad relativa del 78 %, luminosidad solar de 1726 h/año y evapotranspiración de 1202 mm. Se encuentra en la zona de vida ecológica Bosque Húmedo Montano Bajo (bh-MB), en el suelo de la unidad cartográfica Asociación Rionegro, en una terraza aluvial baja del río Rionegro, plana, con pendiente menor al 3 %, pobremente agotado. Por su parte, el Centro de Investigación Turipaná se encuentra ubicado en el Valle del Sinú (Cordoba), con una temperatura media anual de 28 °C, precipitaciones de 1200 mm y humedad relativa del 87 % [59].

Para recolectar el conjunto de datos espectrales se usó un espectrómetro modelo FLAME S VIS NIR de Ocean Insight, fibra óptica QP600-2-VIR-NIR recubierta en metal, generador de luz HL-2000-LL y una pinza de fabricación específica para el proyecto en el extremo final para aislar la muestra de la luz externa. La pinza permite fijar la muestra, la fuente de luz con un ángulo de 45 grados, y la fibra del espectrómetro de forma perpendicular a la superficie. El rango del espectrómetro está entre 350 y 1000 nm, con una resolución de 1,33 nm; en total, se capturan 2049 bandas a lo largo del espectro electromagnético. Para el proceso de captura se utilizó el software OceanView 2.0, ofrecido por el fabricante, donde se configura que cada firma es el resultado del promedio de 10 muestras.

Las firmas espectrales requieren de calibración respecto a un nivel de luz conocido, por lo que se requiere acompañar cada sesión con la toma de un estándar blanco y negro. La corrección a reflectancia se realiza mediante la ecuación (3.1). El software Ocean View 2.0 requiere los estándares de color al inicio de toma de muestras, pero calcula la calibración a reflectancia de forma automática.

$$\frac{(FirmaEspectral - Negro)}{(Blanco - Negro)} \quad (3.1)$$

3.1.1. Maíz: deficiencia de nitrógeno

Siembra 5 de abril de 2022

Lugar: Centro de investigación La Selva

Toma de datos: 23 de junio de 2022

Al momento de tomar las muestras las plantas se encontraban entre las etapas V10 a V12 (final de crecimiento e inicio de floración). Se aplicaron cuatro tratamientos: 100 %, 75 %, 50 %, 25 % del requerimiento de nitrógeno. La dosis recomendada de nitrógeno (100 %) fue calculada a partir de un análisis previo de suelo. Cada tratamiento contó con treinta surcos de 1.4 m, espaciados a 0.2 m entre plantas y 0.8 m entre surcos [60]. En la Figura 3.1 se presentan dos vistas de la plantación, la vista (a) presenta todo el experimento desde una vista aérea y la vista (b) desde el suelo mirando las secciones de 50 % y 25 % de nutrición.



(a) Vista aérea



(b) Vista en campo

Figura 3.1: Fotografías experimento maíz

3.1.2. Aguacate: Deficiencia de fósforo

Edad: 18 meses

Lugar: Centro de investigación La Selva

Toma de Datos: 9 de mayo de 2022

Se aplicaron cuatro tratamientos basados en distintas dosis de fósforo: 100 %, 75 %, 50 %, 25 % del requerimiento óptimo. La dosis recomendada de fósforo 100 % fue calculada a partir del nivel disponible en el suelo del experimento, obtenido por un análisis de suelo. El diseño de experimento constó de un bloque completos al azar, con arreglo de tratamientos en parcelas divididas, con tres repeticiones, para un total de 96 plantas dentro de invernadero. En la Figura 3.2 se presentan dos vistas de la plantación, la vista (a) el experimento fuera del invernadero desde una vista aérea y la vista (b) desde el suelo en una etapa avanzada de deficiencia nutricional.



(a) Vista aérea



(b) Vista en campo

Figura 3.2: Fotografías experimento Aguacate

3.1.3. Fríjol: Deficiencia de fósforo

Siembra: 20 de agosto de 2021

Lugar: Centro de investigación Turipaná

Toma de datos: 03 de noviembre de 2021

Se aplicaron dos tratamientos basados en distintas dosis de fósforo: 100 % y 50 % de la dosis óptima. La dosis recomendada 100 % fue calculada a partir del fósforo disponible en el suelo del experimento, obtenido por un previo análisis de suelo. En la Figura 3.3 se presenta una vista de la plantación desde el suelo, el mismo día de la toma de datos.



Figura 3.3: Fotografía plantación Fríjol

3.2. Preprocesamiento

Las bases de datos obtenidas presentan ruido en las secciones extremas del rango espectral, lo que impide el análisis de los datos. La Figura 3.4 presenta un ejemplo aleatorio de firma espectral de maíz sin modificaciones.

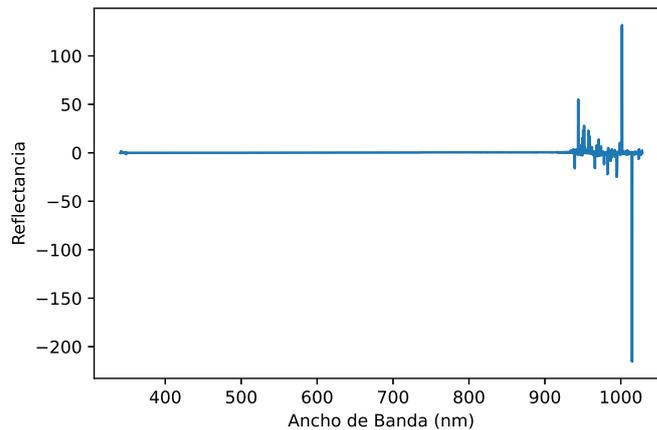


Figura 3.4: Firma espectral sin procesar

Para eliminar este problema se corta la firma intentando conservar la mayor cantidad del rango espectral inicial. La cantidad de información preservada varía entre cada base de datos. un ejemplo de firma recortada se ve en la Figura 3.5.

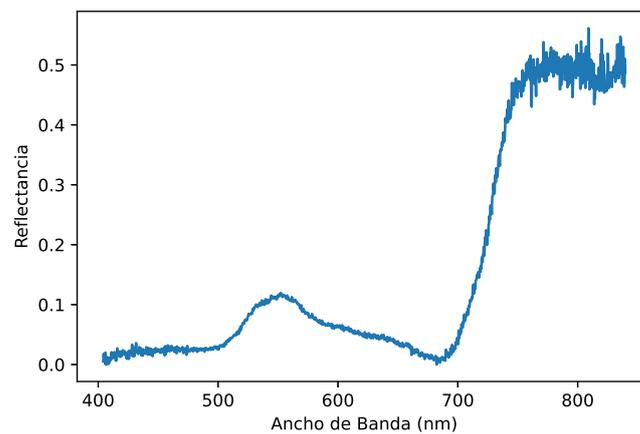


Figura 3.5: Firma espectral cortada para eliminar bandas de mayor ruido.

El segundo paso del preprocesamiento aplica a cada firma espectral un filtro de mediana deslizante con una ventana de diez puntos, el resultado de este proceso puede verse en la Figura 3.6.

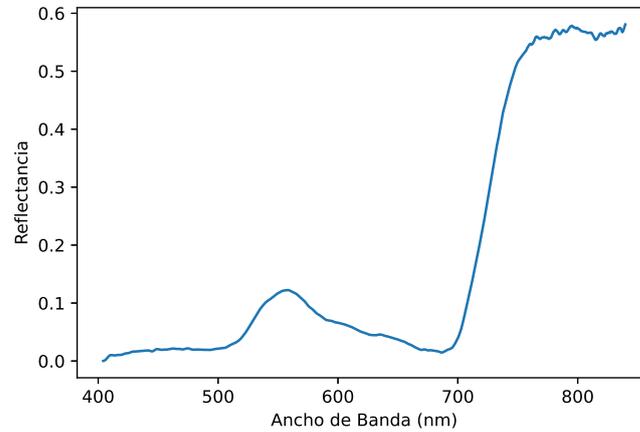


Figura 3.6: Firma espectral luego de filtro de mediana deslizante

Por último, se calcula la mediana y la desviación estándar para cada una de las clases consideradas. Se eliminan las firmas que quedan fuera de ± 2 desviación estándar de la mediana. Se presentan los tamaños inicial y final para cada base de datos, además del rango espectral resultado de los procesos descritos en la Tabla 3.1:

	Maíz	Aguacate	Fríjol
Firmas Iniciales	2390	445	1451
Firmas filtradas	2033	376	1274
Rango Espectral (nm)	450 - 950	420 - 930	450 - 950

Tabla 3.1: Bases de datos: tamaño y rango espectral

3.3. Técnicas de clasificación

Se considera en esta sección la elección de tres algoritmos de clasificación, cuyo principal objetivo es comprobar la posibilidad de detección de estrés a partir de los dataset etiquetados. Se considera como parámetros de selección que los algoritmos funcionaran bien en las siguientes condiciones: espacios de alta dimensionalidad, no lineales y sin necesidad de gran cantidad de datos.

Support Vector Machines (SVM)

Es un algoritmo robusto que funciona mejor en datasets de alta dimensión y puede ser usado en problemas no lineales, además es uno de los más usados en la literatura [32].

La búsqueda de parametros se hizo mediante la estrategia de búsqueda de grilla, se puede ver el espacio de esta búsqueda en la Tabla B.2. Scikit-learn SVC se usó para esta implementación y se obtuvieron los siguientes parámetros: $C = 2000$, $\text{kernel} = \text{'poly'}$, $\text{degree} = 5$, $\text{gamma} = 1$, $\text{random_state} = 0$.

Random Forest (RF)

Es un método muy usado en ML donde muchos árboles de decisión se usan juntos, es tomado en muchas ocasiones como modelo de referencia. Esto se debe a que trabaja bien en problemas no lineales con múltiples variables y que es poco afectado por el overfitting o varianza en el entrenamiento. No se encontraron muchos casos de uso en detección o clasificación de estrés, sin embargo, [32] presenta que se ha encontrado que tiene rendimiento similar a SVM en esta aplicación.

La búsqueda de parámetros se hizo mediante la estrategia de búsqueda de grilla, se puede ver el espacio de esta búsqueda en la Tabla B.1. Scikit-learn RandomForestClassifier fue usado para esta implementación y se obtuvieron los siguientes parámetros: $\text{max_depth} = 60$, $\text{min_samples_leaf} = 3$, $\text{min_samples_split} = 6$, $\text{n_estimators} = 400$, $\text{random_state} = 0$.

Fully connected Neural Networks (NN)

Esta familia de algoritmos es muy usada para problemas no lineales, pero que pueden llegar a tener problemas en datasets de alta dimensionalidad si no se

cuenta con gran cantidad de datos. Dentro de las NN hay muchas arquitecturas para problemas específicos o de uso general. Se consideró que con la cantidad de datos recolectada no es posible tener en cuenta las DNN. Por otro lado, la morfología de los datos hace que deban analizarse uno a uno, lo que hace inviable tener en cuenta las CNN, muy usadas para imágenes hiperespectrales. Se seleccionó entonces que la red sería multi layer perceptron (MLP), la más usada dentro de la literatura considerando la exclusión de DNN y CNN [32–34]. El MLP es entonces un tipo de NN que consta de una serie de capas donde cada nodo (neurona) está conectado a cada nodo de la siguiente capa, lo que la hace una red de propósito general.

Se utilizó Scikit-learn MLPClassifier para esta implementación con los siguientes parámetros: `hidden_layer_sizes=(500,300,100,50)`, `activation='relu'`, `alpha=0.000001`, `solver='adam'`, `random_state=0`. Los anteriores parámetros fueron encontrados mediante la estrategia de búsqueda de grilla, se puede ver el espacio de esta búsqueda en la Tabla B.3.

3.4. Metodología para detección de estrés usando selección de bandas

En términos generales la metodología presenta un esquema de trabajo por bloques que pueden alterarse dependiendo de los detalles de los datos sobre los cuales se va implementar, un esquema general se presenta en la Figura 3.7. Constando de tres procesos principales:

Preprocesamiento: El objetivo principal de este bloque es la reducción de ruido y eliminación de muestras atípicas. Se vio en etapas tempranas del desarrollo de esta metodología, que diferentes métodos no supervisados de selección de bandas tienden a centrarse en regiones con alta variabilidad. Para el caso de las firmas espectrales recolectadas esta variabilidad era considerada ruido y se presentaba en las secciones extremas. Ver sección 3.2 para detalles de implementación.

Selección de bandas: Esta etapa ha sido diseñada para aplicar varios métodos, dado que dependiendo del dataset objetivo las técnicas varían en precisión. En la sección 2.5 se presentan los métodos implementados para los experimentos de esta investigación. Sin embargo, la metodología propuesta tiene la flexibilidad de añadir nuevos algoritmos dentro del esquema de funcionamiento.

Clasificación: Tiene la función principal de comprobar la capacidad de la metodología para detectar estrés abiótico a partir de muestras etiquetadas. Se considera flexible en cuanto a la posibilidad de añadir o modificar métodos para futuras iteraciones de esta metodología. en la sección 3.4 se presentan los métodos escogidos y sus parámetros.

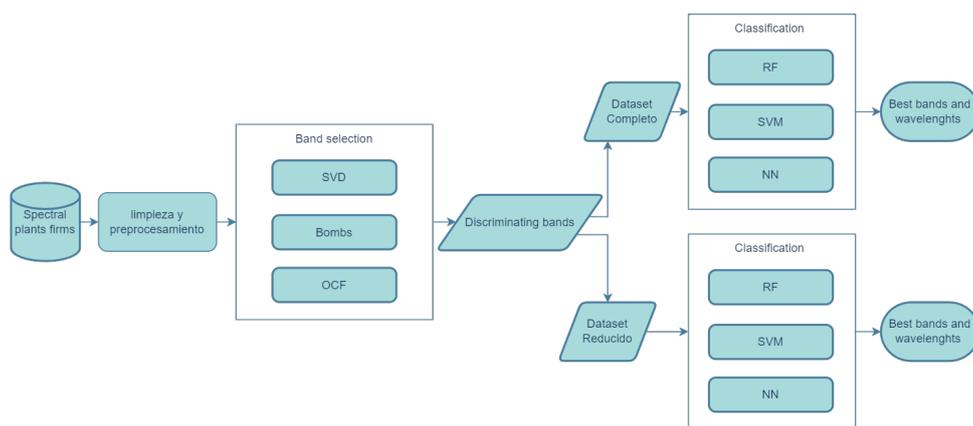


Figura 3.7: Esquema metodología detección de estrés abiótico usando selección de bandas

Se realizan dos experimentos dentro de esta investigación, en la Figura 3.7 se presentan en la división luego de obtener las bandas seleccionadas. El primer experimento (Ex1) analiza la base de datos completa, donde se clasifican todos los niveles de restricción para cada uno de los datasets. El segundo experimento (Ex2), se clasifica únicamente el valor óptimo de nutrición o 100% y el de restricción de 50% de nutrición. A través de esto se pretende corroborar si la diferencia de estrés entre las diferentes clases afecta la posible detección. Se debe considerar que la base de datos de frijol consta de un solo experimento dado que la plantación solo contó con dos niveles diferentes de nutrición.

Capítulo 4

Resultados y Discusión

Contenido

4.1 Maíz	29
4.2 Aguacate	37
4.3 Fríjol	47
4.4 Discusión	52

Sinopsis

Esta sección presenta el desempeño general de la metodología para detectar estrés en maíz, fríjol y aguacate usando reducción de dimensionalidad con tres métodos de selección de bandas. Se realizaron dos experimentos, Ex1 considera todas las clases para comprobar la viabilidad de detección de estrés considerando una diferencia de nutrición del 25%. El segundo Ex2 toma la clase de óptima nutrición y la que tuvo estrés del 50%. En cualquiera de los experimentos se utiliza una división de 85/15, 85% para etapa de entrenamiento y 15% para prueba. Para cada dataset se realiza la selección de 4, 8, 12, 16, 20, 24, 28 y 32 bandas relevantes mediante los métodos OCF, SVD y BOMBS, presentados en la sección 2.5. La clasificación se realizó utilizando tres algoritmos de aprendizaje de máquina: Random Forest (RF), red neuronal (NN) y máquinas de soporte vectorial (SVM), cuyas descripciones y parámetros utilizados se presentaron en la sección 3.4. El análisis inicial presenta la precisión de clasificación para determinar las bandas más relevantes de entre las seleccionadas. A estas últimas se les genera una matriz de confusión y se analiza a partir de las métricas exactitud, precisión y sensibilidad.

4.1. Maíz

En la Figura 4.1, se grafica la media de todas las muestras pertenecientes a cada una de las clases de restricción de nitrógeno, donde #N se refiere al porcentaje de nitrógeno respecto a la cantidad óptima de este nutriente. La Tabla 4.1 presenta cuál es el número total de muestras perteneciente a cada clase y el número utilizado para etapa de entrenamiento.

Desde el valor promedio de las firmas de la Figura 4.1 es difícil evidenciar diferencias espectrales para cada nivel de estrés. Sin embargo, en la región visible, se puede observar alrededor de los 550 nm una diferencia entre la firma de las plantas con mayor estrés (25 %) y los demás datos; por otra parte, la región del infrarrojo cercano muestra una diferencia entre las firmas sin deficiencia de nitrógeno (100 %) y las demás plantas. En las siguientes secciones se presentan los resultados de Ex1 y Ex2.

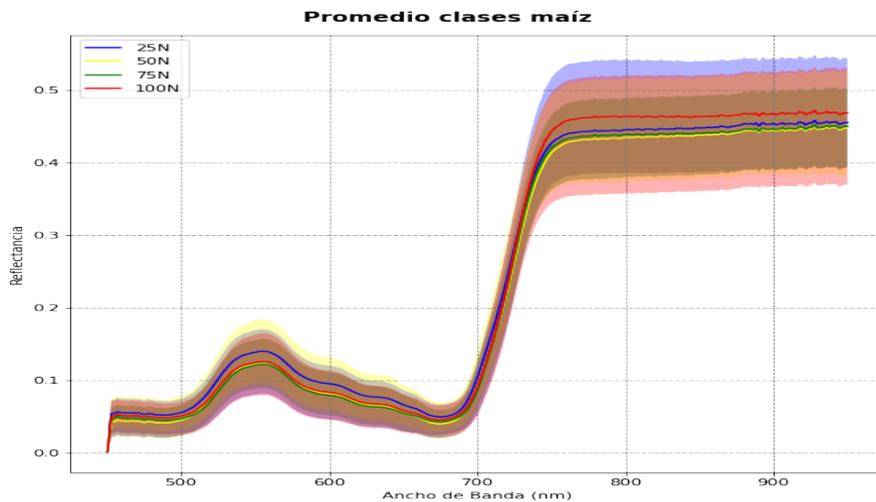


Figura 4.1: Firma promedio para cada clase

Clase	Total	Entrenamiento
25N	506	430
50N	501	426
75N	516	439
100N	510	434

Tabla 4.1: Distribución de muestras de maíz entre clases

Clasificación Random Forest

La Tabla 4.2 presenta la precisión obtenida con el clasificador RF empleando todas las bandas y el conjunto de bandas seleccionadas por los métodos BOMBS, SVD y OCF, estas varían entre 4 a 32. En esta tabla se consideran ambos experimentos y se señala como mejor resultado el porcentaje más alto, escogiendo la menor cantidad de bandas si se obtuvo el mismo resultado más de una vez.

		General (1485 bandas)	Ex1: 0.603		
			Ex2: 0.816		
Bandas	Experimento	BOMBS	SVD	OCF	
4	Ex 1	0.498	0.538	0.472	
	Ex 2	0.824	0.784	0.810	
8	Ex 1	0.511	0.567	0.538	
	Ex 2	0.843	0.850	0.817	
12	Ex 1	0.521	0.590	0.574	
	Ex 2	0.869	0.856	0.824	
16	Ex 1	0.593	0.570	0.554	
	Ex 2	0.869	0.856	0.850	
20	Ex 1	0.590	0.577	0.567	
	Ex 2	0.863	0.850	0.843	
24	Ex 1	0.590	0.577	0.541	
	Ex 2	0.856	0.856	0.843	
28	Ex 1	0.587	0.561	0.554	
	Ex 2	0.856	0.863	0.843	
32	Ex 1	0.570	0.564	0.548	
	Ex 2	0.863	0.850	0.837	

Tabla 4.2: Precisión con RF para dos experimentos en Maíz.

El experimento 1 (Ex1) considera todos los niveles de deficiencia de nitrógeno (25 %, 50 %, 75 % y 100 %). Para este experimento se obtuvo un 60.3 % de precisión de referencia empleando todas las bandas. Para las bandas seleccionadas por SVD, el mejor resultado se obtuvo con 12 bandas con una precisión de 59 %. Por su parte, con el método OCF, la mejor precisión fue del 57.4 % con 12 bandas. El mejor resultado fue obtenido con el método de selección de bandas BOMBS, con una precisión del 59.3 %. Aunque este resultado es menor al de referencia por 1.66 %, se logra una reducción significativa de los datos, de 1485 bandas a 16.

El experimento 2 (Ex2) considera solo dos niveles de nitrógeno (50 % y 100 %). En este caso, la precisión general de clasificación incrementa a 81.6 % empleando todas las bandas. Para el método de selección de bandas SVD, el mejor resultado fue el obtenido con 28 bandas, con una precisión del 86.3 %. Por su parte, OCF obtiene un 85 % de precisión con 16 bandas. El mejor resultado, similar al experimento anterior, fue obtenido con el método BOMBS, empleando 12 o 16 bandas se logra una precisión del 86.9 %, teniendo un resultado 6.49 % mayor a la referencia. En este caso se logra con los tres métodos de selección de bandas mejorar la precisión general de clasificación, como también reducir significativamente el número de bandas en los datos espectrales.

Clasificación Neural Network

La Tabla 4.3 presenta la precisión obtenida con el clasificado NN empleando todas las bandas y el conjunto de bandas seleccionadas por los métodos BOMBS, SVD y OCF, estas varían entre 4 a 32. En esta tabla se consideran ambos experimentos y se señala como mejor resultado el porcentaje más alto, escogiendo la menor cantidad de bandas si se obtuvo el mismo resultado más de una vez.

Para el experimento 1 (Ex1), se obtiene una referencia de precisión del 48.1 % empleando todas las bandas de las firmas. En este caso, los tres métodos de selección de bandas logran incrementar este porcentaje de precisión. En el caso de SVD se logra un 55.7 % de precisión empleando 16 bandas, y empleando OCG se obtiene un 51.1 % empleando 20 bandas. El mejor resultado nuevamente es obtenido por el método BOMBS, que supera la referencia en un 21.1 % obteniendo una precisión de 58.4 % con 16 bandas.

Para el experimento 2 (Ex2) se tiene una precisión general de clasificación del 81.6 % con todas las bandas. En el caso de SVD, se logra un 84.2 % con solo 12 bandas, y con OCF se obtiene una precisión del 80.9 % con 28 bandas. El mejor resultado de precisión es obtenido con la selección de bandas de BOMBS, que logra una precisión del 86.2 % con 20 bandas, superando el porcentaje obtenido con todo el dataset por 5.64 %.

Clasificación Support Vector Machines

La Tabla 4.4 presenta la precisión obtenida con el clasificado SVD empleando todas las bandas y el conjunto de bandas seleccionadas por los métodos BOMBS, SVD y OCF, estas varían entre 4 a 32. En esta tabla se consideran ambos experimentos y se señala como mejor resultado el porcentaje más alto,

		General (1485 bandas)	Ex1: 0.481		
			Ex2: 0.816		
Bandas	Experimento	BOMBS	SVD	OCF	
4	Ex 1	0.459	0.482	0.475	
	Ex 2	0.664	0.783	0.664	
8	Ex 1	0.436	0.515	0.505	
	Ex 2	0.730	0.829	0.697	
12	Ex 1	0.462	0.528	0.498	
	Ex 2	0.724	0.842	0.737	
16	Ex 1	0.584	0.557	0.502	
	Ex 2	0.757	0.796	0.763	
20	Ex 1	0.466	0.475	0.511	
	Ex 2	0.862	0.651	0.730	
24	Ex 1	0.430	0.495	0.446	
	Ex 2	0.678	0.750	0.796	
28	Ex 1	0.554	0.538	0.508	
	Ex 2	0.855	0.730	0.809	
32	Ex 1	0.525	0.423	0.508	
	Ex 2	0.737	0.638	0.770	

Tabla 4.3: Precisión con NN para dos experimentos en Maíz.

escogiendo la menor cantidad de bandas si se obtuvo el mismo resultado más de una vez.

El experimento1 (Ex1) tiene un valor de referencia de exactitud de 70.8% empleando todas las bandas. Con el método OCF, el mejor resultado de precisión fue de 65.9% empleando 28 y 32 bandas. Por su parte, con el método BOMBS, se logra una clasificación del 69.5% con 24 bandas. En este caso, el mejor resultado fue el obtenido con las bandas seleccionadas con SVD con 74.4% de precisión y 20 bandas, siendo este el único que supera la referencia, por un 5.08%.

Para el experimento 2 (Ex2), los resultados con SVM muestran un alto desempeño empleando todas las bandas con un 98.7% de precisión general. Para las bandas seleccionadas con SVD, el mejor resultado es de 96.7% y para OCF el mejor porcentaje de clasificación es del 93.4%, ambos con 32 bandas. Nuevamente, se nota que el método BOMBS logra los mejores resultados, con una precisión igual a la de referencia del 98.7% empleando 28 bandas.

		General (1485 bandas)	Ex1: 0.708		
			Ex2: 0.987		
Bandas	Experimento	BOMBS	SVD	OCF	
4	Ex 1	0.426	0.456	0.410	
	Ex 2	0.750	0.737	0.704	
8	Ex 1	0.616	0.649	0.574	
	Ex 2	0.862	0.901	0.849	
12	Ex 1	0.616	0.685	0.570	
	Ex 2	0.928	0.954	0.803	
16	Ex 1	0.636	0.721	0.580	
	Ex 2	0.961	0.974	0.822	
20	Ex 1	0.643	0.744	0.652	
	Ex 2	0.947	0.961	0.908	
24	Ex 1	0.695	0.738	0.656	
	Ex 2	0.954	0.961	0.921	
28	Ex 1	0.682	0.725	0.659	
	Ex 2	0.987	0.961	0.914	
32	Ex 1	0.689	0.692	0.659	
	Ex 2	0.967	0.967	0.934	

Tabla 4.4: Precisión con SVM para dos experimentos en Maíz.

Experimento 1: Mejores Bandas

Comparando las Tablas 4.2 a 4.4, se evidencia que el mejor resultado de clasificación considerando todos los niveles de fertilización es el obtenido con el SVM y 20 bandas, con una precisión del 74.4%. Estas bandas son representadas de forma gráfica por una línea punteada en la Figura 4.2, que se superponen al promedio de las clases como referencia. La lista completa de bandas seleccionadas y su correspondiente ancho de banda en el espectro electromagnético se presentan en el Apéndice, Tabla A.1. Se observa que un mayor porcentaje de bandas se encuentra por encima de 670 nm, correspondiendo a las secciones de borde rojo e infrarrojo cercano en el espectro electromagnético.

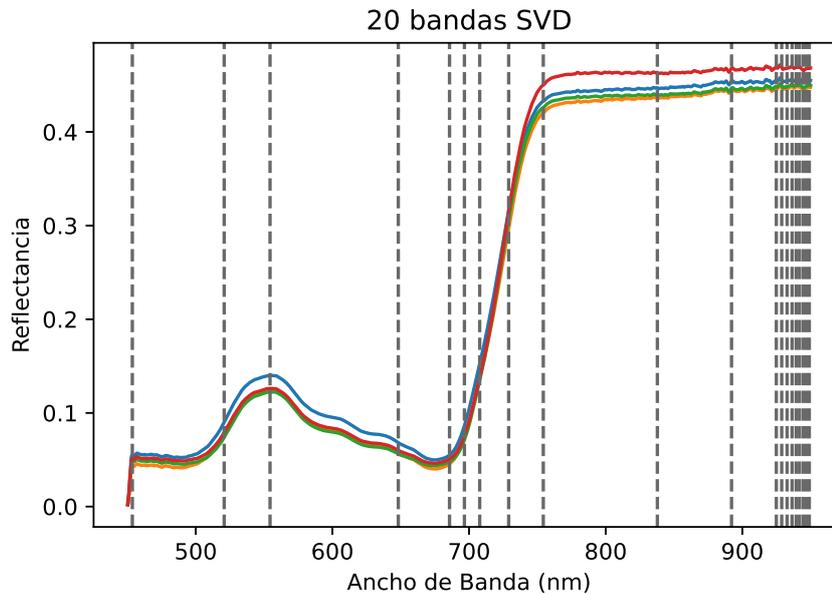


Figura 4.2: Conjunto de 20 bandas seleccionadas con SVD

	Precisión	Sensibilidad
25N	0.68	0.64
50N	0.7	0.68
75N	0.69	0.75
100N	0.91	0.90
General	0.74	0.74

Tabla 4.5: Métricas derivadas de clasificación SVM: SVD 20 bandas

Para analizar los resultados de clasificación, se presenta la matriz de confusión en la Figura 4.3, junto con la Tabla 4.5 que muestra las métricas de precisión y sensibilidad derivadas de la matriz. A partir de estos se puede ver como la clase de óptima nutrición 100N tiene falsos positivos (FP) y falsos negativos (FN) menores a 2%, diferenciándose bien de las otras clases y obteniendo precisión y sensibilidad superiores al 90%. Por el otro lado entre las clases 25N, 50N y 75N se presenta FN y FP entre 2% y 5%, bajando la precisión y sensibilidad de estas clases y afectando la precisión general.

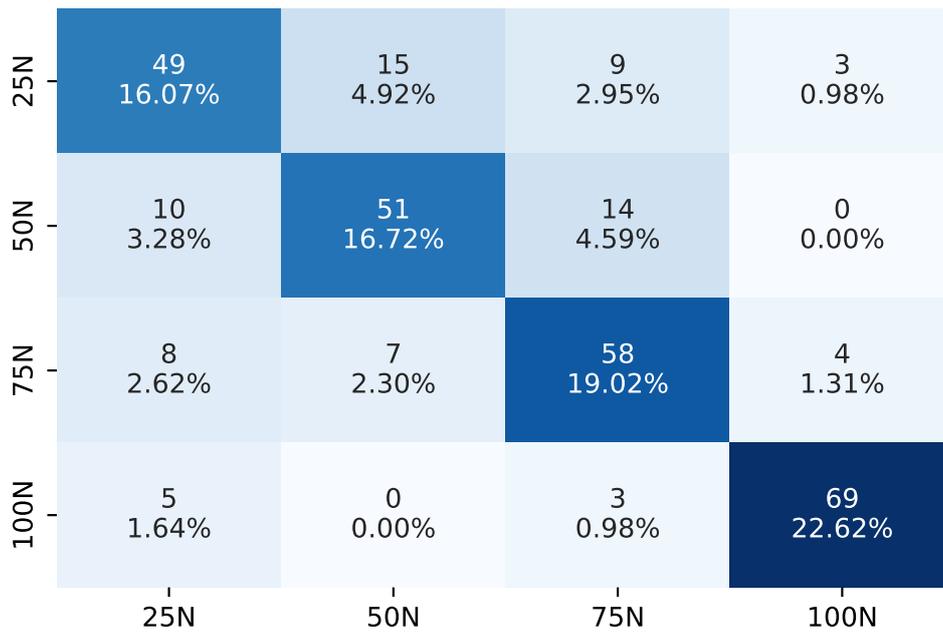


Figura 4.3: Matriz de confusión clasificación SVM: 20 bandas SVD

Experimento 2: Mejores Bandas

Considerando solo el 50% y 100% de nutrición, se observa en las Tablas 4.2 a 4.4 que el mejor resultado de clasificación es el obtenido por BOMBS y 28 bandas, con una precisión del 98.7%. Estas bandas son representadas de forma gráfica por una línea punteada en la Figura 4.2, que se superponen al promedio de las clases como referencia. La lista completa de bandas seleccionadas y su correspondiente ancho de banda en el espectro electromagnético se presentan en el Apéndice, Tabla A.2.

Se puede ver que una concentración similar de bandas en la región visible, sobre todo cerca de 550 nm, y en la región del infrarrojo cercano. Eso tiene sentido considerando que en el promedio de las clases visualmente se evidencia diferencia en estas regiones.

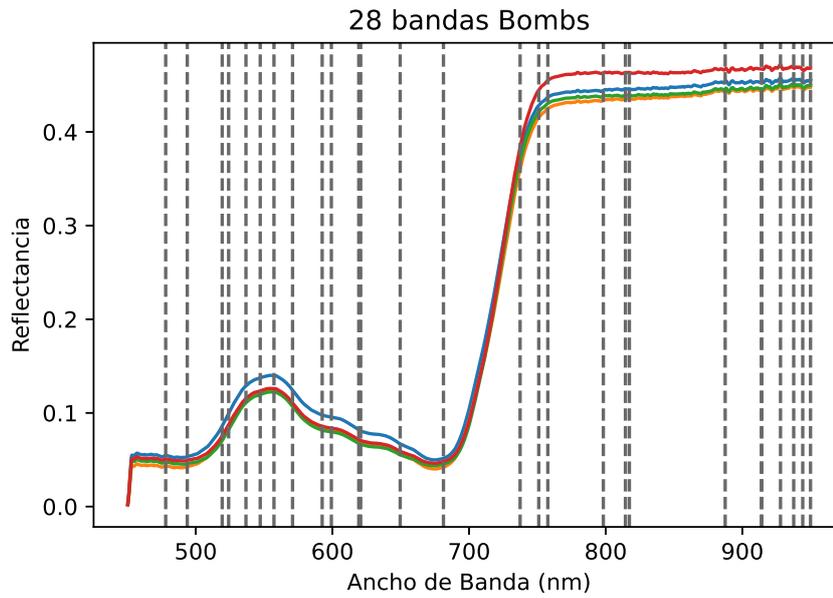


Figura 4.4: Conjunto de 28 bandas seleccionadas con BOMBS

	Precisión	Sensibilidad
50N	1.00	0.97
100N	0.97	1.00
General	0.99	0.99

Tabla 4.6: Métricas derivadas de clasificación SVM: BOMBS 28 bandas

Para analizar los resultados de clasificación, se presenta la matriz de confusión en la Figura 4.5, junto con la Tabla 4.6 que muestra las métricas de precisión y sensibilidad derivadas de la matriz. A partir de estos se puede ver que fue posible diferenciar el nivel de restricción nutricional entre las clases, se presenta error por FP en 2 de las 79 muestras de la clase 100N. Esto se refleja en la Tabla 4.6, donde se encuentra resultado de 100 % para la precisión de 50N y para la sensibilidad de 100N.

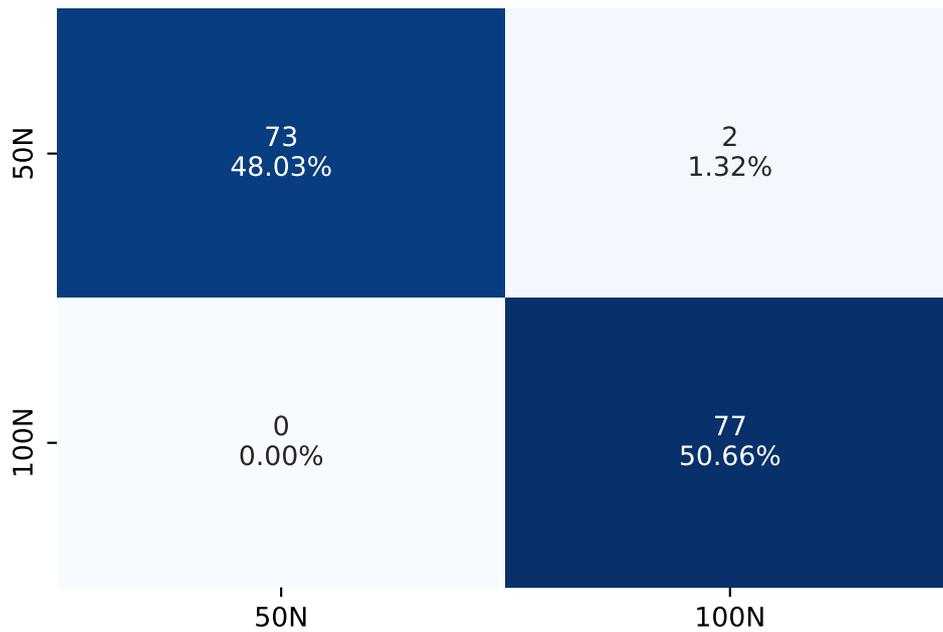


Figura 4.5: Matriz de confusión clasificación SVM: BOMBS 28 bandas

4.2. Aguacate

En la Figura 4.6 se presenta la media de todas las muestras pertenecientes a cada una de las clases de restricción de fósforo, #F se refiere al porcentaje de fósforo respecto al ideal (100%). Consta de cuatro clases, en la Tabla 4.7 se presenta el total de muestras pertenecientes a cada una y el total de datos para entrenamiento.

Observando el valor promedio de las firmas presentado en la Figura 4.6, se observa una diferencia marcada entre las clases en el sector infrarrojo, superior a 730 nm aproximadamente. Sin embargo, las clases no se encuentran en orden que permita identificar un patrón, ya que en orden de mayor a menor amplitud en este sector se tiene 25F, 50F, 100F y 75F. Por otro lado, en el rango visible del espectro se encuentra que 50F y 100 F pueden diferenciarse de las otras clases, pero 25F y 75F están superpuestas, dificultando su clasificación y la búsqueda de algún patrón en los datos. En las siguientes secciones se presentan los resultados de Ex1 y Ex2.

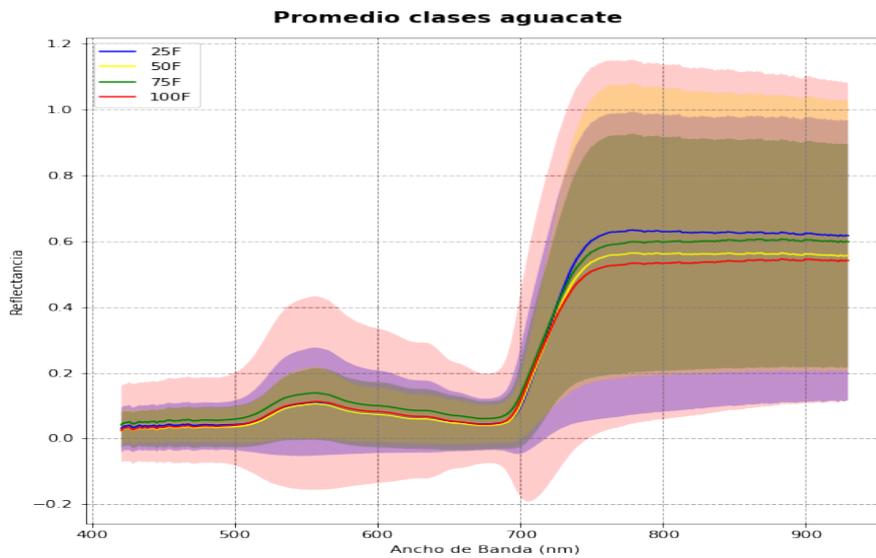


Figura 4.6: Firma promedio para cada clase

Clase	Total	Entrenamiento
25F	99	84
50F	83	70
75F	105	89
100F	89	76

Tabla 4.7: Distribución de muestras de aguacate entre clases

Clasificación Random Forest

La Tabla 4.8 presenta la precisión obtenida con el clasificador RF considerando tres clasificadores, tres algoritmos de reducción de dimensionalidad y dos experimentos. Se señala como mejor resultado el porcentaje más alto, escogiendo la menor cantidad de bandas si se obtuvo el mismo resultado más de una vez.

El experimento 1 (Ex1) considera todos los niveles de deficiencia de fósforo (25%, 50%, 75% y 100%). Se obtuvo un 35.1% de exactitud de referencia empleando todas las bandas. Para las bandas seleccionadas por SVD, el mejor resultado se obtuvo con 12 y 20 bandas con una precisión de 36.8%. Por su parte, con el método OCF la mejor precisión fue del 36.8% con 4 y 28 bandas. Por último, la selección de 12 bandas con BOMBS obtuvo una precisión del 26.3%. La selección con SVD y OCF logran superar la referencia y reducir considerable-

		General (1498 bandas)	Ex1: 0.351	Ex2: 0.615
Bandas	Experimento	BOMBS	SVD	OCF
4	Ex 1	0.246	0.351	0.368
	Ex 2	0.500	0.654	0.692
8	Ex 1	0.228	0.333	0.298
	Ex 2	0.500	0.692	0.654
12	Ex 1	0.263	0.368	0.281
	Ex 2	0.538	0.654	0.692
16	Ex 1	0.246	0.333	0.298
	Ex 2	0.500	0.654	0.692
20	Ex 1	0.246	0.368	0.333
	Ex 2	0.577	0.577	0.654
24	Ex 1	0.228	0.351	0.351
	Ex 2	0.615	0.500	0.692
28	Ex 1	0.211	0.316	0.368
	Ex 2	0.577	0.654	0.654
32	Ex 1	0.246	0.316	0.298
	Ex 2	0.577	0.538	0.654

Tabla 4.8: Precisión con RF para dos experimentos en aguacate.

mente la cantidad de información, sin embargo, el bajo porcentaje obtenido de la clasificación no permite diferenciar las cuatro clases.

El experimento 2 (Ex2) considera solo dos niveles de estrés nutricional en fósforo (50% y 100%). En este caso, la precisión general de clasificación incrementa a 61.5% empleando todas las bandas. Para el método de selección de bandas SVD, el mejor resultado fue el obtenido con 8 bandas, con 69.2%. Por su parte, OCF también obtiene un 69.2% para los sets de 4, 12, 16 y 24 bandas discriminantes. Con el método BOMBS, empleando 24 bandas se logra una precisión de 61.5%. El mejor resultado, similar al experimento anterior, fue obtenido con dos métodos, SVD y OCF. Ambos superan el margen de clasificación por 12.5%, permitiendo la reducción del dataset a entre 4 y 24 bandas.

Clasificación Neural Networks

La Tabla 4.9 presenta la precisión obtenida con el clasificador NN considerando tres clasificadores, tres algoritmos de reducción de dimensionalidad y dos

experimentos. Se señala como mejor resultado el porcentaje más alto, escogiendo la menor cantidad de bandas si se obtuvo el mismo resultado más de una vez.

		General (1498 bandas)	Ex1: 0.386		
			Ex2: 0.538		
Bandas	Experimento	BOMBS	SVD	OCF	
4	Ex 1	0.193	0.263	0.246	
	Ex 2	0.538	0.654	0.654	
8	Ex 1	0.246	0.281	0.281	
	Ex 2	0.692	0.692	0.692	
12	Ex 1	0.228	0.298	0.281	
	Ex 2	0.538	0.615	0.692	
16	Ex 1	0.228	0.246	0.281	
	Ex 2	0.654	0.692	0.692	
20	Ex 1	0.298	0.298	0.263	
	Ex 2	0.577	0.692	0.769	
24	Ex 1	0.316	0.281	0.263	
	Ex 2	0.654	0.692	0.731	
28	Ex 1	0.281	0.316	0.298	
	Ex 2	0.692	0.731	0.692	
32	Ex 1	0.263	0.298	0.298	
	Ex 2	0.577	0.615	0.769	

Tabla 4.9: Precisión con NN para dos experimentos en aguacate.

El experimento 1 (Ex1) considera todos los niveles de deficiencia de fósforo, y se obtuvo un 38.6 % de exactitud de referencia empleando todas las bandas. Para las bandas seleccionadas por SVD, el mejor resultado se obtuvo con 28 bandas para una exactitud de 31.6%. Por su parte, con el método OCF la mejor precisión fue del 29.8 % con 32 bandas. Por último, la selección de 24 bandas con BOMBS, obtiene una precisión del 31.6%. En esta ocasión ningún algoritmo supera el marco de referencia y se considera que no es posible diferenciar las clases.

El experimento 2 (Ex2) considera solo dos niveles de estrés nutricional en fósforo. En este caso, la precisión general de clasificación incrementa a 53.8 % empleando todas las bandas. Para el método de selección de bandas SVD, el mejor resultado fue el obtenido con 28 bandas para 73.1 % de exactitud. Con el método BOMBS, empleando 8 y 28 bandas se logra una precisión de 69.2 %. El

mejor resultado entre los métodos de selección de bandas fue obtenido por OCF, con un resultado de 76.9% para los sets de 20 y 32 bandas discriminantes. En este experimento cabe destacar que todos los métodos obtuvieron un desempeño mejor a la referencia, destacándose OCF al obtener exactitud 42.9% mayor a la referencia y reducir de 1498 bandas a 20.

Clasificación Support Vector Machines

La Tabla 4.10 presenta la precisión obtenida con el clasificador SVM considerando tres clasificadores, tres algoritmos de reducción de dimensionalidad y dos experimentos. Se señala como mejor resultado el porcentaje más alto, escogiendo la menor cantidad de bandas si se obtuvo el mismo resultado más de una vez.

		General (1498 bandas)	Ex1: 0.351	Ex2: 0.423	
Bandas	Experimento	BOMBS	SVD	OCF	
4	Ex 1	0.281	0.281	0.316	
	Ex 2	0.500	0.538	0.538	
8	Ex 1	0.228	0.298	0.333	
	Ex 2	0.577	0.654	0.577	
12	Ex 1	0.246	0.351	0.351	
	Ex 2	0.577	0.654	0.615	
16	Ex 1	0.333	0.333	0.316	
	Ex 2	0.538	0.500	0.654	
20	Ex 1	0.228	0.351	0.368	
	Ex 2	0.385	0.577	0.615	
24	Ex 1	0.228	0.316	0.368	
	Ex 2	0.500	0.577	0.654	
28	Ex 1	0.298	0.211	0.368	
	Ex 2	0.423	0.731	0.654	
32	Ex 1	0.386	0.246	0.298	
	Ex 2	0.423	0.500	0.577	

Tabla 4.10: Precisión con SVM para dos experimentos en aguacate.

El experimento 1 (Ex1) considera todos los niveles de deficiencia de fósforo, y se obtuvo un 35.1% de exactitud de referencia empleando todas las bandas. Para las bandas seleccionadas por SVD, el mejor resultado se obtuvo con 12 y 20

bandas para una exactitud de 35.1 %. Por su parte, con el método OCF la mejor precisión fue del 36.8 % usando 20, 24 y 28 bandas. Por último, la selección de 32 bandas con BOMBS, obtiene una precisión del 38.6 %. El algoritmo BOMBS supera el marco de referencia, pero se considera que no es posible diferenciar las clases dado el bajo porcentaje de precisión del experimento.

El experimento 2 (Ex2) considera solo dos niveles de estrés nutricional en fósforo, 50 % y 100 %. En este caso, la precisión general de clasificación es de 42.3 % empleando todas las bandas. Con el método BOMBS, empleando 8 y 12 bandas se logra una un resultado de 57.7 %. El método de selección de bandas OCF, obtuvo un resultado de 65.4 % para los sets de 16, 24 y 28 bandas discriminantes. El método SVD obtuvo los mejores resultados, considerando una exactitud de 73.1 % para 28 bandas. Los tres métodos de reducción superaron la referencia, siendo SVD el de mayor diferencia, con una exactitud 72.8 % mayor a la obtenida con todas las bandas.

Experimento 1: Mejores Bandas

Al comparar las Tablas 4.8 a 4.10 en los valores referentes a todos los niveles de nutrición (Ex1), se destaca el resultado obtenido por BOMBS con 32 bandas y algoritmo de clasificación SVM, con una exactitud de 38.6%. Estas bandas se presentan de gráficamente por una línea punteada en la Figura 4.7, se superponen al promedio de las clases como referencia. La lista completa de bandas seleccionadas y su correspondiente ancho de banda en el espectro electromagnético se presentan en el Apéndice, Tabla A.3.

Se observa que las bandas se están seleccionando mayormente en las regiones extremas del rango, correspondiendo a infrarrojo para bandas superiores a 730 nm y por el borde inferior seleccionando la mayoría cercano al color violeta 400nm.

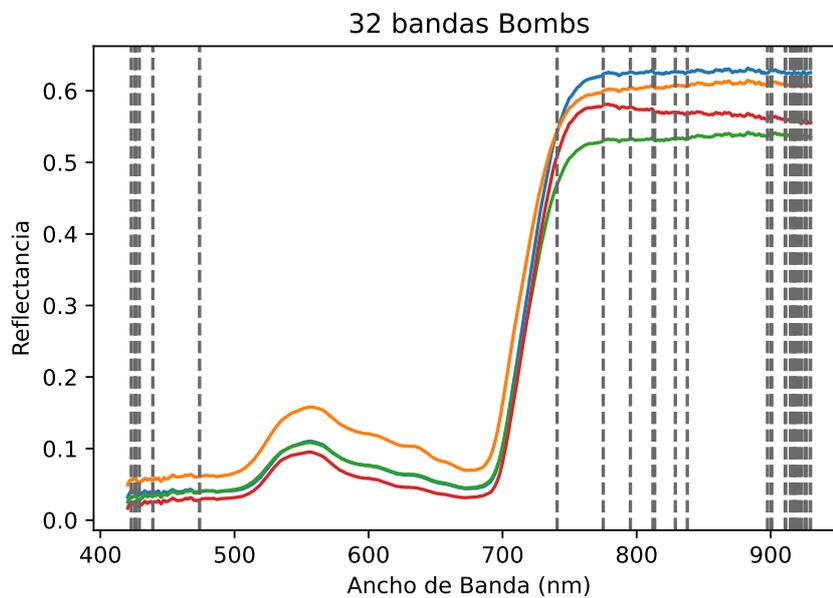


Figura 4.7: Conjunto de 32 bandas seleccionadas con BOMBS

	Precisión	Sensibilidad
25F	0.39	0.60
50F	0.25	0.15
75F	0.54	0.44
100F	0.31	0.31
General	0.37	0.37

Tabla 4.11: Métricas derivadas de clasificación SVM: BOMBS 32 bandas

Para analizar los resultados de clasificación, se presenta la matriz de confusión en la Figura 4.8, junto con la Tabla 4.11 que muestra las métricas de precisión y sensibilidad derivadas de la matriz. Para cualquier clase de el dataset se observa que hay mayor número de errores que aciertos. Esto tiene sentido considerando la exactitud total alcanzada de 38.6 %. 25F y 75F presentan mayor diferenciabilidad, considerando un 60 % de sensibilidad y un 54 % de precisión, respectivamente. Sin embargo, se considera que este experimento no es posible detectar los niveles propuestos del nutriente fósforo.

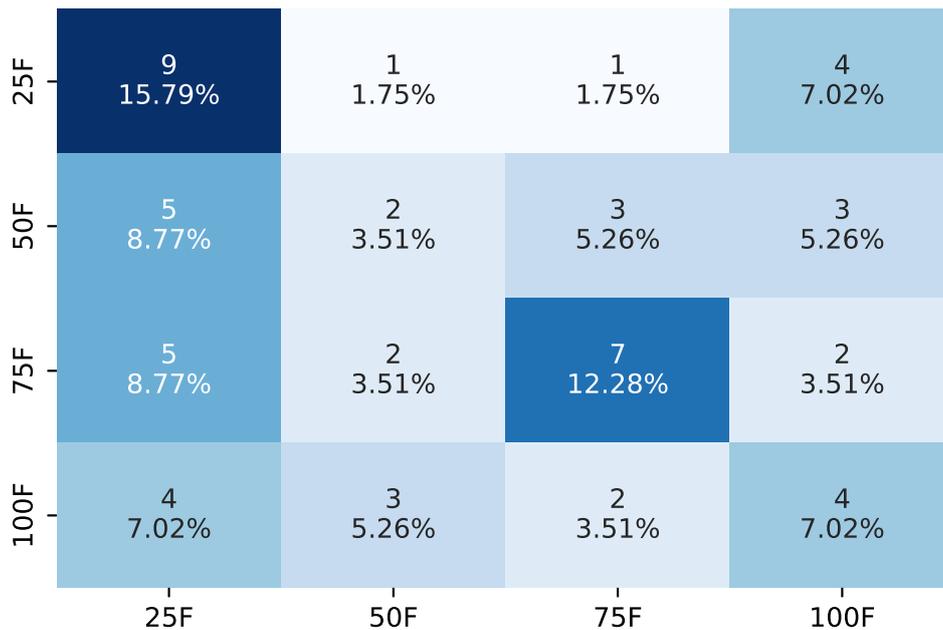


Figura 4.8: Matriz de confusión clasificación SVM: BOMBS 32 bandas

Experimento 2: Mejores Bandas

Para este experimento del dataset de aguacate se consideran dos niveles de estrés nutricional 50 % y 100 %, es decir Ex2 en las Tablas 4.8 a 4.10. Se mejoran considerablemente los resultados comparados con el experimento anterior, obteniendo una exactitud de 76.9% mediante 32 bandas seleccionadas con el método OCF y el algoritmo NN. Estas bandas son representadas de forma gráfica por una línea punteada en la Figura 4.9, que se superponen al promedio de las clases como referencia. La lista completa de bandas seleccionadas y su correspondiente ancho de banda en el espectro electromagnético se presentan en el Apéndice, Tabla A.4.

Se observa que las bandas seleccionadas se encuentran distribuidas a lo largo de la longitud de onda con mayor concentración alrededor del borde rojo (670nm - 730nm) y alrededor de el pico verde (500nm - 600nm). Se encuentra bandas en otras secciones del espectro, pero con menor abundancia. Esto es opuesto a la selección con mayor exactitud del Ex1 para esta misma base de datos.

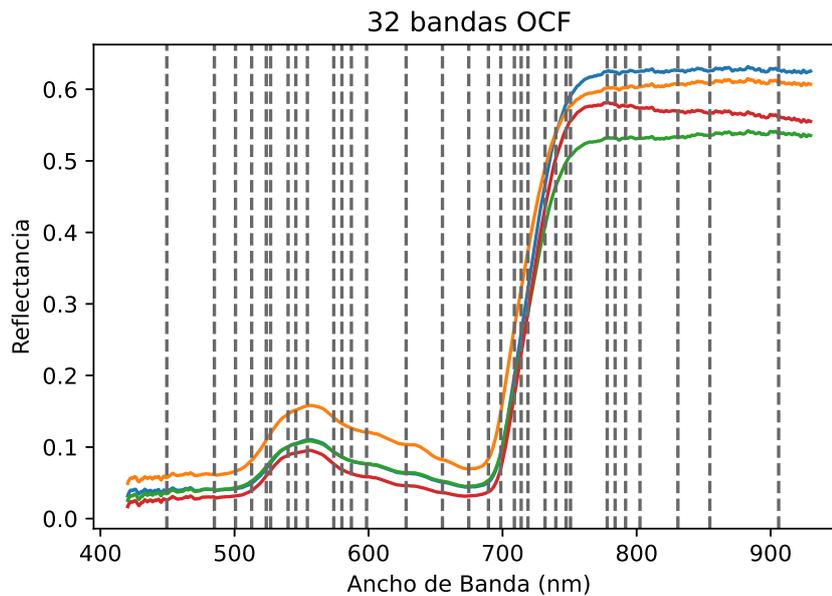


Figura 4.9: Conjunto de 32 bandas seleccionadas con OCF

	Precisión	Sensibilidad
50F	0.73	0.85
100F	0.82	0.69
General	0.78	0.77

Tabla 4.12: Métricas derivadas de clasificación NN: OCF 32 bandas

Para analizar los resultados de clasificación, se presenta la matriz de confusión en la Figura 4.10, junto con la Tabla 4.12 que muestra las métricas de precisión y sensibilidad derivadas de la matriz. Para la clase 50F se observa que se una precisión de 73 % considerando que 4 de las 15 muestras pertenecientes a esta fueron erróneamente clasificadas y una sensibilidad de 85 %. La precisión de la clase 100F es un poco mayor alcanzando el 82 % y una sensibilidad del 69 %.

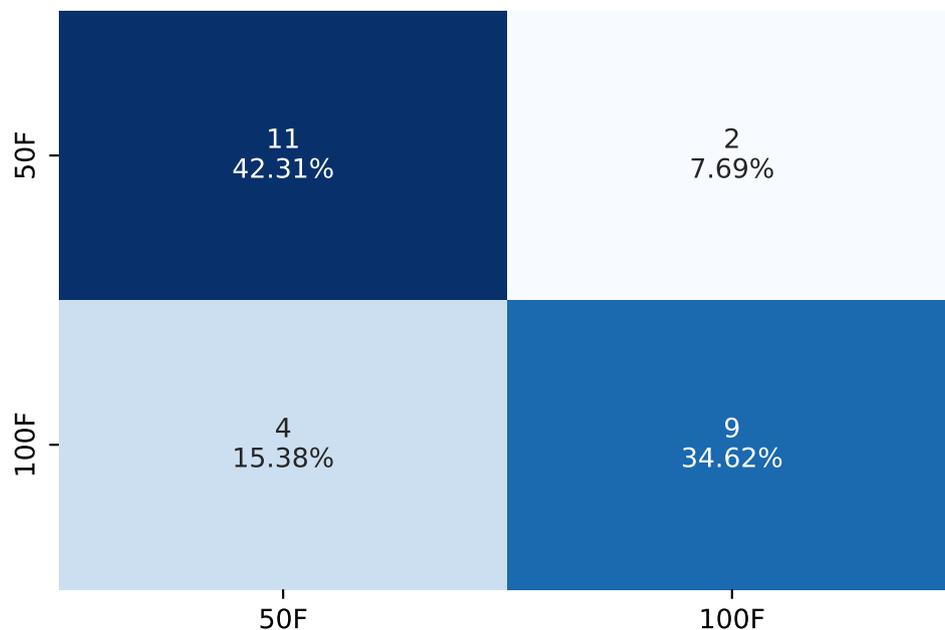


Figura 4.10: Matriz de confusión clasificación NN: 32 bandas OCF

4.3. Frijol

En Figura 4.11, se presenta la firma media de todas las muestras pertenecientes a las dos clases de restricción, #F se refiere al porcentaje de fósforo respecto a la cantidad óptima de este nutriente. La Tabla 4.1 presenta cual es el número total de muestras perteneciente a cada clase y el número utilizado para etapa de entrenamiento. Considerando que esta base de datos solo tiene dos clases, se realiza solo un experimento teniendo en cuenta ambas.

De las firmas espectrales promedio mostradas en la Figura 4.11, se evidencia que ambas clases tienen un comportamiento muy similar, superponiéndose en algunos sectores. Sin embargo, se nota una mayor amplitud en las regiones del pico verde y el infrarrojo para la clase 50F con respecto a 100F. Sin embargo, esta diferencia es poco notable lo que indica que en la base de datos hay firmas que pueden confundirse entre las dos clases, reduciendo la posibilidad de detectar el estrés.

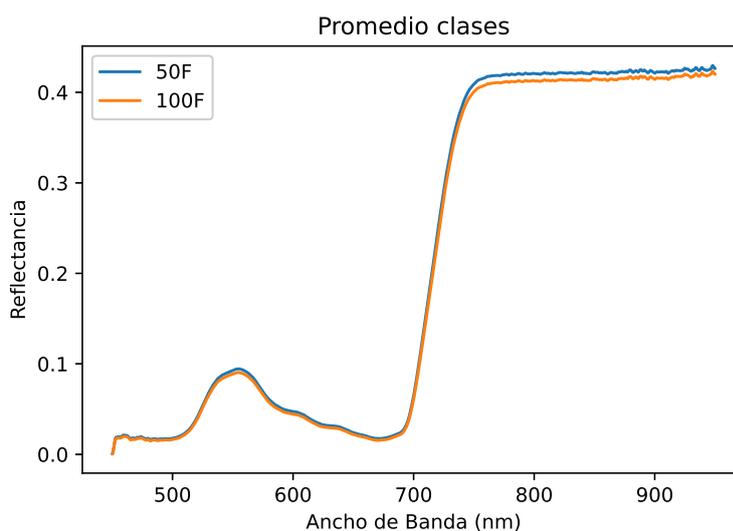


Figura 4.11: Firma promedio para cada clase

Clase	Total	Entrenamiento
50F	649	551
100F	625	531

Tabla 4.13: Distribución de muestras de frijol entre clases

Clasificación Random Forest

La Tabla 4.14 presenta la precisión obtenida con el clasificador RF. Se presenta que el resultado de referencia, es decir, clasificando con todas las bandas, obtiene 54.7% de exactitud. El método de reducción SVD se consigue la mayor precisión con 32 bandas, al obtener 58.3%. El mejor resultado obtenido con el algoritmo OCF empleando 12 bandas, con una exactitud del 57.8%. Por otro lado, BOMBS con las 8 bandas seleccionadas obtuvo mayor precisión de entre los algoritmos de reducción, con un 58.9%. Aunque en general los tres métodos logran superar el punto de referencia, BOMBS permite una reducción de 1485 bandas a las 8 más discriminantes y supera el punto de referencia por un 6.58%.

General: 0.547			
Bandas	BOMBS	SVD	OCF
4	0.562	0.568	0.547
8	0.589	0.526	0.536
12	0.516	0.536	0.578
16	0.562	0.562	0.557
20	0.552	0.552	0.531
24	0.542	0.562	0.526
28	0.583	0.568	0.542
32	0.531	0.583	0.542

Tabla 4.14: Precisión con RF en frijol

Clasificación Neural networks

La Tabla 4.15 presenta la precisión obtenida con el clasificador NN. La precisión al clasificar con todas las bandas es 44.8 % y se toma como punto de referencia para los resultados con dataset reducidos. El método SVD se consigue mayor precisión con 28 bandas, obteniendo 52.6 %. El mejor resultado obtenido con el algoritmo OCF es 52.1 % y se consigue con sets de 4, 8 y 24 bandas discriminantes. BOMBS por su parte, con las 12 bandas seleccionadas obtuvo mayor precisión de entre los algoritmos de reducción, con un 56.8 %. Aunque en general los tres métodos logran superar el punto de referencia, El método BOMBS obtiene una mejora de 26.9 % respecto a la referencia, mientras reduce a 12 bandas la base de datos.

General: 0.448			
Bandas	BOMBS	SVD	OCF
4	0.531	0.521	0.521
8	0.474	0.464	0.521
12	0.568	0.500	0.510
16	0.500	0.500	0.510
20	0.547	0.484	0.500
24	0.510	0.505	0.521
28	0.505	0.526	0.500
32	0.510	0.505	0.469

Tabla 4.15: Precisión con NN en frijol

Clasificación Support Vector Machines

La Tabla 4.16 presenta la precisión obtenida con el clasificador SVM. Se toma de referencia la exactitud obtenida con todas las bandas de la base de datos, consiguiendo un 44.8%. El primer método de reducción a analizar es BOMBS, cuyo mayor resultado de precisión es 65.1% y se logra con 24 bandas. El algoritmo OCF obtiene 56.2% de precisión con 20 bandas discriminantes. Por último, SVD supera a los dos métodos anteriores con una precisión de 66.7% utilizando las 32 bandas discriminantes. Este resultado supera la referencia por 10.4%.

General: 0.604			
Bandas	BOMBS	SVD	OCF
4	0.536	0.500	0.531
8	0.635	0.573	0.542
12	0.562	0.646	0.531
16	0.609	0.661	0.552
20	0.625	0.651	0.562
24	0.651	0.651	0.542
28	0.625	0.661	0.526
32	0.635	0.667	0.505

Tabla 4.16: Precisión con SVM en frijol

Mejores Bandas

En Frijol, al analizar las Tablas 4.14 a 4.16 se observa que la mayor exactitud (66.7 %) se obtuvo mediante 32 bandas seleccionadas con el método SVD y el algoritmo SVM. Estas bandas son representadas de forma gráfica por una línea punteada en la Figura 4.12, que se superponen al promedio de las clases como referencia. La lista completa de bandas seleccionadas y su correspondiente ancho de banda en el espectro electromagnético se presentan en el Apéndice, Tabla A.5.

Se observa en la Figura 4.12 que un mayor porcentaje de bandas se encuentra por encima de 670 nm, correspondiendo a las secciones de borde rojo e infrarrojo cercano, concentrándose especialmente en el extremo superior a 900 nm.

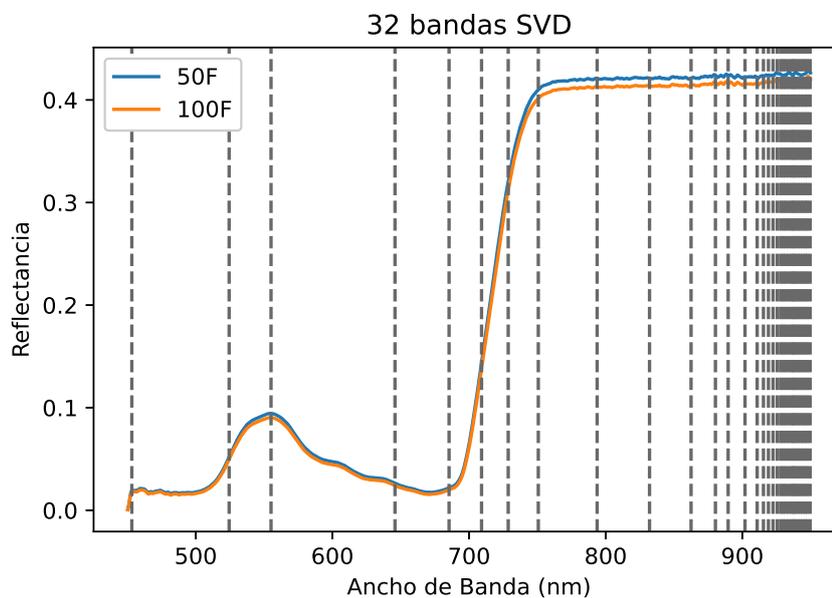


Figura 4.12: Conjunto de 32 bandas seleccionadas con SVD

	Precisión	Sensibilidad
50F	0.67	0.69
100F	0.67	0.64
General	0.67	0.67

Tabla 4.17: Métricas derivadas de clasificación SVM: SVD 32 bandas

Para analizar los resultados de clasificación, se presenta la matriz de confusión en la Figura 4.13, junto con la Tabla 4.17 que muestra las métricas de precisión y sensibilidad derivadas de la matriz. Se encuentran una precisión igual para ambas clases 67%, y una sensibilidad similar entre clases con un 5% de diferencia.

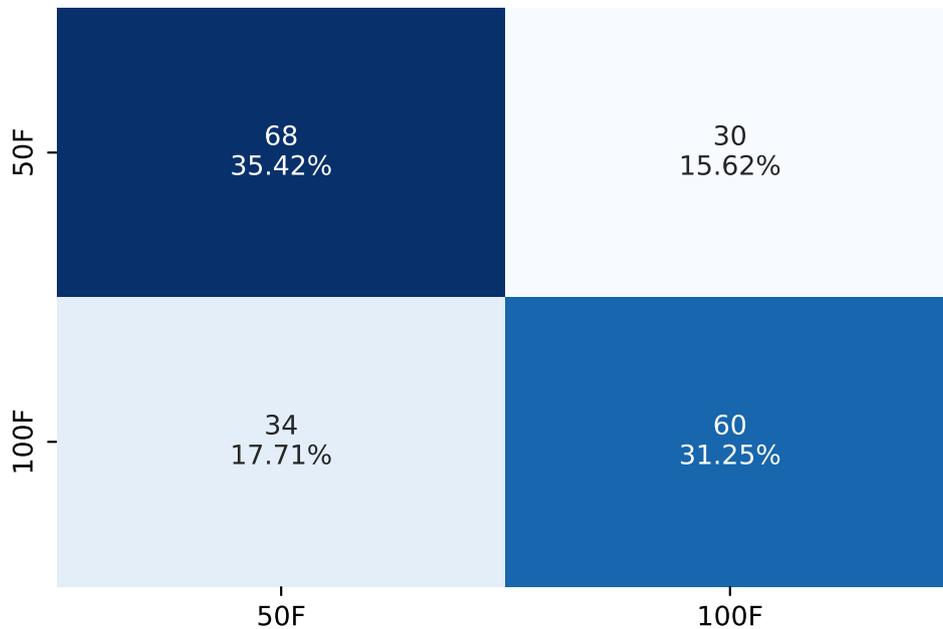


Figura 4.13: Matriz de confusión clasificación SVM: 32 bandas SVD

4.4. Discusión

Según nuestro conocimiento actual, no existen estudios que utilicen técnicas de selección de bandas como método para la reducción de dimensiones de datos espectrales para la detección de estrés. La mayoría de los estudios en la literatura emplean índices de vegetación, los cuales se calculan a partir de 2 o 3 bandas. Por ejemplo, en [61] se evalúan enfoques basados en índices de vegetación para estudiar el estado del agua y el nitrógeno, mientras que [62] determina la deficiencia de nitrógeno en maíz dulce. Además, [63] utilizó medidas de reflectancia para estudiar las respuestas espectrales de cuatro cultivares comunes de frijol sometidos a estrés hídrico en la etapa de floración. Todos los estudios mencionados utilizan índices de variedades basados en bandas NIR, RED-Edge,

RED o Green. El rango espectral utilizado para los índices está relacionado con las longitudes de onda seleccionadas en estos artículos.

El enfoque de reducción de dimensionalidad presentado en este documento difiere de los trabajos anteriores dado que se centra en encontrar secciones críticas para el estrés abiótico en lugar de utilizar solo un conjunto preestablecido de bandas. Para esto, se seleccionan regiones con baja redundancia y alto valor informativo, lo que permite enfocarse en características específicas de maíz, aguacate y frijoles que son importantes para comprender cómo responden al estrés abiótico.

Se destaca de los resultados obtenidos la alta dependencia de la precisión con el cultivo específico. Por ejemplo, con el maíz se lograron porcentajes de detección de estrés del 98.7 % usando la selección de bandas BOMBS and SVM; por su parte, para el aguacate esta precisión fue del 76.9% usando OCF y NN; finalmente, para el frijol fue del 66.7 % empleando SVD y SVM. Esto evidencia la necesidad de estudiar las metodologías de detección de estrés de acuerdo con las condiciones propias de cada cultivo; lo cual, a su vez, representa un reto para el uso de datos espectrales en agricultura de precisión, dada la dificultad para la generalización de los datos.

Por otro lado, se hace evidente que el problema de cuantificar el estrés (cuatro clases – experimento 1) es más complejo que el problema de detectarlo (dos clases – experimento 2). Con los datos disponibles del maíz y frijol, se evidencian precisiones inferiores (del 74.4 % SVD y SVM para maíz y 38.6 % BOMBS y SVM para aguacate) en la cuantificación del estrés. Aquí se encuentra una primera limitación del estudio, la cual resulta de la cantidad de datos disponibles por tratamiento, y se dificulta la generalización de los modelos: entre mayor es el número de tratamientos nutricionales aplicados a las plantas, se requiere un mayor número de muestras para el entrenamiento de los datos, sin embargo, el trabajo de campo es costoso.

Finalmente, las bandas seleccionadas como relevantes para cada uno de los cultivos sirven de base para la construcción de nuevos estudios y herramientas para el monitoreo del estrés. En el caso del maíz, los mejores resultados se obtuvieron con las bandas seleccionadas por BOMBS, con un total de 28 bandas; en aguacate, el método OCF permitió seleccionar 20 bandas; y en frijol, 32 bandas con SVM. Las bandas seleccionadas pueden ser la base para la construcción de dispositivos espectrales (espectrómetros o cámaras) con un menor número de bandas, lo cual permitiría la reducción de los costos de estos dispositivos.

Conclusiones y Trabajo Futuro

La metodología propuesta permite determinar si el estrés abiótico en plantas es diferenciable y detectarlo en caso positivo. Esto se presenta con la caracterización no supervisada del estrés mediante la reducción de bandas, que en todos los casos ha logrado superar el porcentaje de exactitud en clasificación de referencia utilizando todas las longitudes de onda. Por tanto, si con el mismo preprocesamiento y algoritmos de clasificación, se obtienen mejores métricas al utilizar la reducción de dimensionalidad propuesta, pero el porcentaje alcanzado es inferior al 60 %, se puede decir que los niveles presentes de estrés son muy similares entre ellos para diferenciarlos. Se puede concluir entonces que la detección de estrés abiótico nutricional considerando diferencias del 50 % (Experimento 2), es factible para las tres bases de datos a través de la metodología propuesta. Sin embargo, actualmente solo es aplicable a la detección de diferencias nutricionales de 25 % (Experimento 1), en la base de datos de maíz.

La revisión de técnicas de reducción de dimensiones basadas en selección de bandas, permitió mostrar la relevancia de emplear varias técnicas para lograr la caracterización del estrés en un espacio menor. Esto puede decirse dado que se ha mostrado que no se tiene una técnica que supere a las demás, para datasets con objetivo diferente, ya sea tipo de planta o tipo de estrés, la mejor selección de bandas varía entre diferentes métodos. Lo que plantea como trabajo a futuro, la posibilidad de adicionar técnicas de selección a la metodología de detección de estrés abiótico.

Uno de los mayores desafíos en la ejecución del proyecto fue la dificultad en obtención de datos precisos, repetibles y reproducibles. Se considera que aumentar la dimensión de las bases de datos e incluir tomas de diferentes momentos de la vida y evolución del estrés en las plantas, podrían mejorar la capacidad de detección presentado por esta metodología.

Apéndice A

Bandas seleccionadas

Banda	Longitud de onda
9	453.586
192	520.796
285	554.377
551	648.226
660	685.717
692	696.615
725	707.802
788	729.010
864	754.334
1121	837.818
1294	892.101
1401	924.885
1414	928.826
1427	932.758
1439	936.380
1449	939.392
1457	941.798
1466	944.500
1474	946.899
1481	948.994

Tabla A.1: Selección 20 bandas algoritmo SVD dataset Maíz

Banda	Longitud de onda	Banda	Longitud de onda
75	477.997	647	681.276
118	493.797	813	737.372
188	519.343	854	751.018
201	524.063	874	757.645
236	536.733	998	798.281
265	547.189	1048	814.447
293	557.247	1057	817.343
331	570.841	1279	887.457
392	592.524	1365	913.923
411	599.243	1366	914.228
468	619.297	1411	927.917
470	619.998	1443	937.585
472	620.699	1465	944.200
555	649.611	1484	949.892

Tabla A.2: Selección 28 bandas algoritmo Bombs dataset Maíz

Banda	Longitud de onda	Banda	Longitud de onda
7	422.993	1402	900.739
14	425.616	1435	910.866
15	425.990	1436	911.171
16	426.365	1448	914.839
23	428.985	1449	915.144
50	439.072	1454	916.670
144	473.942	1457	917.584
903	740.708	1458	917.889
1007	775.108	1459	918.194
1069	795.358	1464	919.717
1121	812.191	1469	921.238
1125	813.480	1474	922.758
1173	828.886	1483	925.491
1201	837.818	1485	926.098
1392	897.659	1487	926.705
1401	900.431	1497	929.734

Tabla A.3: Selección 32 bandas algoritmo Bombs dataset Aguacate

Banda	Longitud de onda	Banda	Longitud de onda
78	449.499	751	689.469
174	484.988	778	698.653
217	500.753	808	708.816
250	512.795	823	713.881
280	523.700	838	718.935
289	526.964	876	731.689
325	539.982	900	739.708
341	545.749	923	747.365
365	554.377	933	750.687
420	574.051	1016	778.060
437	580.104	1034	783.951
457	587.208	1058	791.779
489	598.536	1091	802.497
573	628.045	1179	830.804
651	655.147	1254	854.614
708	674.770	1419	905.963

Tabla A.4: Selección 32 bandas algoritmo OCF dataset Aguacate

Banda	Longitud de onda	Banda	Longitud de onda
8	453.215	1382	919.108
202	524.426	1393	922.455
287	555.095	1403	925.491
544	645.798	1411	927.917
659	685.376	1417	929.734
729	709.154	1423	931.549
787	728.675	1430	933.664
853	750.687	1437	935.777
984	793.732	1442	937.284
1103	832.081	1447	938.790
1199	862.486	1453	940.595
1256	880.312	1459	942.399
1286	889.626	1465	944.200
1326	901.969	1471	946.000
1355	910.866	1476	947.498
1370	915.449	1483	949.593

Tabla A.5: Selección 32 bandas algoritmo SVD dataset Fríjol

Apéndice B

Parámetros en búsqueda de grilla

Estimadores #	Profundidad máxima	Mínimo divisiones	Mínimo hojas
10	3	2	2
25	6	3	3
50	12	4	5
100	20	5	7
200	30	6	9
400	60	8	10
325	100	10	

Tabla B.1: Parametros usados en la búsqueda de grilla para el algoritmo RF

C	grado	kernel
50	2	rbf
75	3	poly
150	4	
300	5	
500	6	
1000	7	
2000	8	
3000	9	
4000	10	

Tabla B.2: Parametros usados en la búsqueda de grilla para el algoritmo SVM

Layers sizes	alpha	activation
(50)	0.00000001	relu
(50, 50)	0.0000001	tanh
(100)	0.000001	
(100,50)	0.00001	
(200,100,50)	0.0001	
(500,300,100,50)	0.00000001	
(1000,500,300,100,50)		
(2000,500,50)		
(2000,500,250,100,50,20)		

Tabla B.3: Parametros usados en la busqueda de grilla para el algoritmo NN

Bibliografía

- [1] Yang Lu. Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, 6:1–10, 6 2017.
- [2] I. Kovács and I. Husti. The role of digitalization in the agricultural 4.0 – how to connect the industry 4.0 to agriculture? *Hungarian Agricultural Engineering*, pages 38–42, 2018.
- [3] Olumide Olajide Ojo, Satya Shah, Alec Coutroubis, Mercedes Torres Jimenez, and Yolanda Munoz Ocana. Potential impact of industry 4.0 in sustainable food supply chain environment. *2018 IEEE International Conference on Technology Management, Operations and Decisions, ICTMOD 2018*, pages 172–177, 7 2018.
- [4] Dieisson Pivoto, Paulo Dabdab Waquil, Edson Talamini, Caroline Pauletto Spanhol Finocchio, Vitor Francisco Dalla Corte, and Giana de Vargas Moraes. Scientific development of smart farming technologies and their application in brazil. *Information Processing in Agriculture*, 5:21–32, 3 2018.
- [5] Giorgia Bucci, Deborah Bentivoglio, and Adele Finco. Precision agriculture as a driver for sustainable farming systems: State of art in literature and research biogas production from waste from local food, wood and sugar cane industries for increasing self-sufficiency of energy in sancti spiritus, cuba view project. *Romanian Society for Quality Assurance*, 2018.
- [6] Iván Montenegro-Trujillo, Aleidys Hernandez, Diego Chavarro, Maria Isabel Vélez, Galo Tovar, Angela Milena Niño, and Alejandro Olaya. Macrotendencias hacia el 2030 el mundo y américa latina. *Colciencias – Subdirección General – Unidad de Diseño y Evaluación de Políticas*, Doc. Trabajo N0. 02, 11 2018.
- [7] Colciencias. Plan estratégico de ciencia y tecnología agropecuaria 2017 - 2027. 2016.

- [8] G8 Antioquia. Agenda regional de investigación y desarrollo tecnológico hacia la innovación. 2020.
- [9] Hayat Ullah, Raquel Santiago-Arenas, Zannatul Ferdous, Ahmed Attia, and Avishek Datta. Improving water use efficiency, nitrogen use efficiency, and radiation use efficiency in field crops under drought stress: A review. *Advances in Agronomy*, 156:109–157, 1 2019.
- [10] Aslenis Melo R., Pedro Ariza, Sofía Lissbrant, and Adriana Tofiño. Evaluation of agrochemicals and bioinputs for sustainable bean management on the caribbean coast of colombia. *Agronomía Colombiana*, 33:203–211, 9 2015.
- [11] Curtis A. Deutsch, Joshua J. Tewksbury, Michelle Tigchelaar, David S. Battisti, Scott C. Merrill, Raymond B. Huey, and Rosamond L. Naylor. Increase in crop losses to insect pests in a warming climate. *Science*, 361:916–919, 8 2018.
- [12] Joseph A. Santanello, Christa D. Peters-Lidard, Matthew E. Garcia, David M. Mocko, Michael A. Tischler, M. Susan Moran, and D. P. Thoma. Using remotely-sensed estimates of soil moisture to infer soil texture and hydraulic properties across a semi-arid watershed. *Remote Sensing of Environment*, 110:79–97, 9 2007.
- [13] Mutlu Ozdogan, Yang Yang, George Allez, and Chelsea Cervantes. Remote sensing of irrigated agriculture: Opportunities and challenges. *Remote Sensing 2010, Vol. 2, Pages 2274-2304*, 2:2274–2304, 9 2010.
- [14] Robert A. Schowengerdt. Remote sensing : models and methods for image processing. page 515, 2007.
- [15] M. Weiss, F. Jacob, and G. Duveiller. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236:111402, 1 2020.
- [16] K. Steddom, M. W. Bredehoeft, M. Khan, and C. M. Rush. Comparison of visual and multispectral radiometric disease evaluations of cercospora leaf spot of sugar beet. <https://doi.org/10.1094/PD-89-0153>, 89:153–158, 2 2007.
- [17] Stephanie Delalieux, Jan van Aardt, Wannes Keulemans, Eddie Schrevens, and Pol Coppin. Detection of biotic stress (*venturia inaequalis*) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications. *European Journal of Agronomy*, 27:130–143, 7 2007.

- [18] A. K. Mahlein, U. Steiner, H. W. Dehne, and E. C. Oerke. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. *Precision Agriculture* 2010 11:4, 11:413–431, 6 2010.
- [19] Nikolaos Katsoulas, Angeliki Elvanidi, Konstantinos P. Ferentinos, Murat Kacira, Thomas Bartzanas, and Constantinos Kittas. Crop reflectance monitoring as a tool for water stress detection in greenhouses: A review. *Biosystems Engineering*, 151:374–398, 11 2016.
- [20] Susan L. Ustin, A. A. Gitelson, Stéphane Jacquemoud, Michael Schaepman, Gregory P. Asner, John A. Gamon, and Pablo Zarco-Tejada. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment*, 113:S67–S77, 9 2009.
- [21] Jinru Xue and Baofeng Su. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, 2017, 2017.
- [22] K. R. Thorp, M. A. Gore, P. Andrade-Sanchez, A. E. Carmo-Silva, S. M. Welch, J. W. White, and A. N. French. Proximal hyperspectral sensing and data analysis approaches for field-based plant phenomics. *Computers and Electronics in Agriculture*, 118:225–236, 10 2015.
- [23] Weiwei Sun and Qian Du. Hyperspectral band selection: A review. *IEEE Geoscience and Remote Sensing Magazine*, 7:118–139, 6 2019.
- [24] Francis J. Pierce and Peter Nowak. Aspects of precision agriculture. *Advances in Agronomy*, 67:1–85, 1 1999.
- [25] Athanasios Balafoutis, Bert Beck, Spyros Fountas, Jurgen Vangeyte, Tamme Van Der Wal, Iria Soto, Manuel Gómez-Barbero, Andrew Barnes, and Vera Eory. Precision agriculture technologies positively contributing to ghg emissions mitigation, farm productivity and economics. *Sustainability* 2017, Vol. 9, Page 1339, 9:1339, 7 2017.
- [26] Rajendra P. Sishodia, Ram L. Ray, and Sudhir K. Singh. Applications of remote sensing in precision agriculture: A review. *Remote Sensing* 2020, Vol. 12, Page 3136, 12:3136, 9 2020.
- [27] P. Shanmugapriya, S. Rathika, T. Ramesh, and P. Janaki. Applications of remote sensing in agriculture - a review. *International Journal of Current Microbiology and Applied Sciences*, 8:2270–2283, 1 2019.

- [28] Bappa Das, Rabi N. Sahoo, Sourabh Pargal, Gopal Krishna, Rakesh Verma, Viswanathan Chinnusamy, Vinay K. Sehgal, and Vinod K. Gupta. Comparison of different uni- and multi-variate techniques for monitoring leaf water status as an indicator of water-deficit stress in wheat through spectroscopy. *Biosystems Engineering*, 160:69–83, 8 2017.
- [29] Sha Huang, Lina Tang, Joseph P. Hupy, Yang Wang, and Guofan Shao. A commentary review on the use of normalized difference vegetation index (ndvi) in the era of popular remote sensing. *Journal of Forestry Research*, 32:1–6, 2 2021.
- [30] Stéphane Jacquemoud, Wout Verhoef, Frédéric Baret, Cédric Bacour, Pablo J. Zarco-Tejada, Gregory P. Asner, Christophe François, and Susan L. Ustin. Prospect + sail models: A review of use for vegetation characterization. *Remote Sensing of Environment*, 113:S56–S66, 9 2009.
- [31] S. Jacquemoud, F. Baret, B. Andrieu, F. M. Danson, and K. Jaggard. Extraction of vegetation biophysical parameters by inversion of the prospect + sail models on sugar beet canopy reflectance data. application to tm and aviris sensors. *Remote Sensing of Environment*, 52:163–172, 6 1995.
- [32] Jayme Garcia Arnal Barbedo. Detection of nutrition deficiencies in plants using proximal images and machine learning: A review. *Computers and Electronics in Agriculture*, 162:482–492, 7 2019.
- [33] Zongmei Gao, Zhongwei Luo, Wen Zhang, Zhenzhen Lv, and Yanlei Xu. Deep learning application in plant stress imaging: A review. *AgriEngineering 2020, Vol. 2, Pages 430-446*, 2:430–446, 7 2020.
- [34] Serosh Karim Noon, Muhammad Amjad, Muhammad Ali Qureshi, and Abdul Mannan. Use of deep learning techniques for identification of plant leaf stresses: A review. *Sustainable Computing: Informatics and Systems*, 28:100443, 12 2020.
- [35] Simon J. Hook. Nasa 2014 the hyperspectral infrared imager (hyspiri)-science impact of deploying instruments on separate platforms. *HyspIRI Group*, 2014.
- [36] Dalton Lunga, Saurabh Prasad, Melba M. Crawford, and Okan Ersoy. Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning. *IEEE Signal Processing Magazine*, 31:55–66, 2014.

- [37] Weiwei Sun, Avner Halevy, John J. Benedetto, Wojciech Czaja, Chun Liu, Hangbin Wu, Beiqi Shi, and Weiyue Li. Ul-isomap based nonlinear dimensionality reduction for hyperspectral imagery classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 89:25–36, 3 2014.
- [38] He Yang, Qian Du, and Genshe Chen. Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5:544–554, 2012.
- [39] Chein I. Chang and Keng Hao Liu. Progressive band selection of spectral unmixing for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52:2002–2017, 2014.
- [40] Chein I. Chang and Su Wang. Constrained band selection for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44:1575–1585, 6 2006.
- [41] Weiwei Sun, Weiyue Li, Jialin Li, and Yenming Mark Lai. Band selection using sparse nonnegative matrix factorization with the thresholded earth's mover distance for hyperspectral imagery classification. *Earth Science Informatics 2015 8:4*, 8:907–918, 1 2015.
- [42] Zhouxiao Guo, Haichuan Yang, Xiao Bai, Zhihong Zhang, and Jun Zhou. Semi-supervised hyperspectral band selection via sparse linear regression and hypergraph models. *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1474–1477, 2013.
- [43] M. Kamandar and H. Ghassemian. Maximum relevance, minimum redundancy band selection for hyperspectral images. *19th Iranian Conference on Electrical Engineering*, 2011.
- [44] Lilian Chaves B. dos Santos, Silvio Jamil F. Guimarães, and Jefersson A. dos Santos. Efficient unsupervised band selection through spectral rhythms. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):1016–1025, 2015.
- [45] T. Imbiriba, J. C.M. Bermudez, C. Richard, and J. Y. Tourneret. Band selection in rkhs for fast nonlinear unmixing of hyperspectral images. *2015 23rd European Signal Processing Conference, EUSIPCO 2015*, pages 1651–1655, 12 2015.

- [46] Yuan Yuan, Jianzhe Lin, and Qi Wang. Dual-clustering-based hyperspectral band selection by contextual analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 54:1431–1445, 3 2016.
- [47] Rui Cai, Yuan Yuan, and Xiaoqiang Lu. Hyperspectral band selection with convolutional neural network. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11259 LNCS:396–408, 11 2018.
- [48] Ying Zhan, Dan Hu, Haihua Xing, and Xianchuan Yu. Hyperspectral band selection based on deep convolutional neural network and distance density. *IEEE Geoscience and Remote Sensing Letters*, 14:2365–2369, 12 2017.
- [49] Akrem Sellami, Mohamed Farah, Imed Riadh Farah, and Basel Solaiman. Hyperspectral imagery classification based on semi-supervised 3-d deep neural network and adaptive band selection. *Expert Systems with Applications*, 129:246–259, 9 2019.
- [50] Jing Wang, Jun Zhou, and Weiqing Huang. Attend in bands: Hyperspectral band weighting and selection for image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12:4712–4727, 12 2019.
- [51] Pablo Ribalta Lorenzo, Lukasz Tulczyjew, Michal Marcinkiewicz, and Jakub Nalepa. Hyperspectral band selection using attention-based convolutional neural networks. *IEEE Access*, 8:42384–42403, 2020.
- [52] Jihao Yin, Yisong Wang, and Zhanjie Zhao. Optimal band selection for hyperspectral image classification based on inter-class separability. *2010 Symposium on Photonics and Optoelectronic, SOPO 2010 - Proceedings*, 2010.
- [53] Alope Datta, Susmita Ghosh, and Ashish Ghosh. Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8:2814–2823, 6 2015.
- [54] Emmanuel Arzuaga-Cruz, Luis O. Jimenez-Rodriguez, and Miguel Velez-Reyes. Unsupervised feature extraction and band subset selection techniques based on relative entropy criteria for hyperspectral data analysis. *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, 5093:462, 2003.

- [55] Qi Wang, Fahong Zhang, and Xuelong Li. Optimal clustering framework for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 56:5910–5922, 2018.
- [56] Sen Jia, Guihua Tang, Jiasong Zhu, and Qingquan Li. A novel ranking-based clustering approach for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 54:88–102, 1 2016.
- [57] Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. *NIPS*, 2004.
- [58] Mingyang Zhang, Maoguo Gong, and Yongqiang Chan. Hyperspectral band selection based on multi-objective optimization with high information and low redundancy. *Applied Soft Computing Journal*, 70:604–621, 2018.
- [59] Emiro Suárez, Sony Reza, Eliecer Díaz, Fredy García, Iván Pastrana, Hugo Cuadrado, and Manuel Espinosa. Efectos de las condiciones ambientales sobre el comportamiento ingestivo en bovinos de carne en un sistema intensivo en el valle del sinú. *Ciencia y Tecnología Agropecuaria*, 13:207–212, 1 2013.
- [60] Maria C. Torres-Madronero, Manuel Goez, Manuel A. Guzman, Tatiana Rondon, Pablo Carmona, Camilo Acevedo-Correa, Santiago Gomez-Ortega, Mariana Durango-Flórez, Smith V. López, July Galeano, and Maria Casamitjana. Spectral library of maize leaves under nitrogen deficiency stress. *Data 2023, Vol. 8, Page 2*, 8:2, 12 2022.
- [61] Milica Colovic, Kang Yu, Mladen Todorovic, Vito Cantore, Mohamad Hamze, Rossella Albrizio, and Anna Maria Stellacci. Hyperspectral vegetation indices to assess water and nitrogen status of sweet maize crop. *Agronomy 2022, Vol. 12, Page 2181*, 12:2181, 9 2022.
- [62] Brayden W. Burns, V. Steven Green, Ahmed A. Hashem, Joseph H. Massey, Aaron M. Shew, M. Arlene A. Adviento-Borbe, and Mohamed Milad. Determining nitrogen deficiencies for maize using various remote sensing indices. *Precision Agriculture*, 23:791–811, 6 2022.
- [63] Daniel Soares Alves, Pablo Ricardo Nitsche, Vania Moda Cirino, Maria Brigida dos Santos Scholz, José dos Santos Neto, and Paulo Henrique Caramori. Remote sensing allows to estimate water stress and yield losses of bean cultivars. *IRRIGA*, 26:1–12, 3 2021.

