

Metodología para clasificar el nivel de estrés utilizando imágenes
hiperespectrales en cultivo de aguacate

Autor:

Sebastian Carrera Carabali
Maicol Stiven Torres Cardozo

Director de trabajo de grado:
Milton Orlando Sarria-Paja

Ingeniería Electrónica
Facultad de Ingeniería
Ingeniero Electrónico
Universidad Santiago de Cali
Cali, Colombia

2023

Metodología para clasificar el nivel de estrés utilizando imágenes hiperespectrales en cultivo de aguacate.

Methodology to classify stress level using hyperspectral images in avocado crops.

Sebastián Carrera Carabali¹
sebastian.carrera00@usc.edu.co

Maicol Stiven Torres Cardozo²
maicol.torres00@usc.edu.co

Milton Orlando Sarria-Paja³
milton.sarria00@usc.edu.co

Universidad Santiago de Cali, Facultad de ingeniería, Programa de Ingeniería Electrónica (1)

Universidad Santiago de Cali, Facultad de ingeniería, Programa de Ingeniería Electrónica (2)

Universidad Santiago de Cali, Facultad de ingeniería, Programa de Ingeniería Electrónica (3)

Resumen

El presente artículo se enfoca como un estudio de factibilidad, donde la aplicación de la teledetección en base de imágenes hiperespectrales para clasificar el nivel de estrés hídrico en los cultivos de aguacate, se realicen técnicas de inteligencia artificial, la cual, permite que se explore la posibilidad de hacerlo, a través de diversos métodos de preprocesamiento basados en las bandas representativas de las imágenes hiperespectrales.

Es importante resaltar que las imágenes hiperespectrales constituyen una fuente de datos muy informativa, lo que implica la necesidad de emplear diferentes técnicas para una gestión efectiva de estos datos, permitiendo el análisis de datos imperceptibles, por lo tanto, se detalla el proceso de generación de etiquetas y la partición de la base de datos en información del vecindario de los píxeles e información localizada, así como la utilización de índices de vegetación en el contexto del análisis agrícola de cultivos de aguacate, siendo esencial para identificar áreas de interés. Además, proporcionan una visión detallada de análisis de datos, estadísticas, y rendimiento de modelos de clasificación en el contexto de la reflectancia espectral de genotipos de aguacate, señalando desafíos y áreas de mejora que impulsa la eficiencia y precisión del monitoreo de los cultivos de aguacate, reduciendo el tiempo asociado a la evaluación manual y proporcionando una herramienta para los agricultores. De este modo los resultados de predicción son variable en los 2 genotipos de aguacate que se analiza por medio de redes convolucionales y redes completamente conectadas. Esta estrategia permite alcanzar porcentajes entre el 4,4 % a 42,89 % de precisión que pueden estimar factores importantes a tener en cuenta en los análisis.

Palabras Clave: Reflectancia espectral; imágenes hiperespectrales; estrés hídrico; bandas espectrales.

Abstract

This article focuses on a feasibility study where the application of remote sensing based on hyperspectral images is used to classify the level of water stress in avocado crops using artificial intelligence techniques. This exploration involves employing various preprocessing methods based on representative bands of hyperspectral images.

It's important to highlight that hyperspectral images are highly informative data sources, necessitating the use of different techniques for effective data management, enabling the analysis of imperceptible data. Therefore, the process of label generation and partitioning the database into pixel neighborhood information and localized information is detailed. Additionally, the use of vegetation indices in the agricultural analysis of avocado crops is essential for identifying areas of interest. Furthermore, it provides a detailed insight into data analysis, statistics, and the performance of classification models in the context of spectral reflectance of avocado genotypes, pinpointing challenges and areas for improvement that enhance the efficiency and precision of avocado crop monitoring, reducing the time associated with

manual evaluation and providing a tool for farmers. Thus, the prediction results vary for the two avocado genotypes analyzed using convolutional and fully connected networks. This strategy achieves accuracy percentages ranging from 4.4% to 42.89%, estimating crucial factors to consider in the analyses.

Keywords: Spectral reflectance; hyperspectral imaging; water stress; spectral bands.

1. INTRODUCCIÓN

La agricultura de precisión (AP) es un campo de acción de gran importancia que, permite definir las propiedades y características de los cultivos y uso del suelo (Hernández, 2021). La supervisión y el monitoreo son unas de las ventajas que brinda la AP, dando apertura al uso de tecnologías avanzadas como lo es la teledetección (Jia et al., 2021). La agricultura de precisión y la teledetección posibilita la gestión de las áreas agrícolas de manera remota, utilizando drones y satélites para la supervisión desde el espacio. Además, la AP permite el enfoque de la identificación, monitoreo, gestión, y recopilación de datos en los cultivos. De este modo, la AP es destacado por permitir adquirir una comprensión exhaustiva acerca del manejo y la atención de los cultivos.

La supervisión de los cultivos mediante el uso de imágenes ha establecido procedimientos de gran relevancia en la industria agrícola, siendo una fuente de información importante para revelar el estado del cultivo monitoreado (Genangeli et al., 2023). Las imágenes RGB, bien conocidas como imagen de color real donde, en el campo del monitoreo pueden tener limitaciones en la entrega de información para un análisis profundo con respecto a las propiedades que brindan los cultivos. Además de esto, pueden resultar menos intuitivas en los colores provocando poca capacidad visible de los objetos. Por consiguiente, el uso de imágenes hiperespectrales (HSI), presentan una herramienta alternativa de detección remota, debido al desarrollo de sistemas avanzados para la adquisición de alta resolución espacial y espectral, y al desarrollo continuo de recursos informáticos más eficientes para el manejo de grandes volúmenes de datos. De este modo, el análisis de imágenes hiperespectrales ha adquirido aplicaciones significativas en la agricultura de precisión (Paoletti et al., 2019).

El estrés en cultivos que es impactado por condiciones ambientales tales como, el calor o medios físicos y químicos, provocan una interrupción en el proceso de mantener la condición interna estable llamada homeostasis, y es lo que afecta los procesos metabólicos que permiten generar energía. En este contexto, se puede evaluar el estado de salud de los cultivos en diferentes fases del proceso de producción mediante el estudio de sus características espectrales (Arias et al., 2021) la supervisión del estrés hídrico es una tarea importante en los cultivos, ya que el exceso y la poca distribución del agua es un factor relevante para las etapas de crecimiento en las plantas. El desarrollo de esta problemática motiva a generar una evaluación del estudio y revisar la factibilidad de las HSI para el cultivo de aguacate, en apoyo de técnicas como el aprendizaje profundo y la teledetección que permitan en un futuro predecir los niveles de estrés de los cultivos de aguacate.

Las condiciones ambientales que perjudican el proceso fisiológicos de las plantas se consideran afectadas por factores abióticos (De Oliveira, 2019). El estrés abiótico se posiciona como uno de los desafíos por el control de los fenómenos ambientales, dando lugar a significativas disminuciones en la producción agrícola, ocasionando una reducción en la extensión de tierra destinada al cultivo Godoy et al. (2021). Como se mencionó anteriormente, el estrés representa una enorme pérdida del rendimiento en la producción de cultivos, y a su vez impacta la producción agrícola y la disponibilidad de alimentos.

Estas tecnologías constituyen un avance en la manera en que se pueden examinar y entender los cultivos, especialmente en situaciones en las que el estrés es causado por condiciones ambientales y otros factores que afectan significativamente la producción, por lo tanto, proporcionan recursos accesibles y viables para recolectar información sobre los niveles de estrés que experimentan las plantas. Las imágenes hiperespectrales demuestran que en cultivos como el trigo Li et al.

(2022), papa Sun et al. (2022), entre otros, son algunas de las aplicaciones más comunes de estas herramientas.

El aguacate desempeña un papel económico fundamental a nivel local como global, esto se ve reflejado en el incremento de las áreas cultivadas y el crecimiento de las exportaciones, lo que ha llevado al país posicionarse entre los principales productores de aguacate a nivel mundial (Díaz Ramírez et al., 2021). Por lo tanto, se considera relevante el enfoque en este tipo de cultivo, ya que permite avanzar en la agricultura de precisión, lo que estimula avances en la literatura, especialmente en su aplicación específica al aguacate. Sin embargo, a pesar de la destacada importancia de este cultivo, la revisión de la literatura revela una ausencia de publicaciones que contribuyan a sustentaciones más sólidos en el uso de las HSI para cultivos de aguacates, particularmente en lo que representa la detección del estrés. Cabe mencionar que la literatura relacionada se enfoca en diversos tipos de cultivos y utiliza una variedad de métodos de procesamiento. Por ejemplo, algunos se enfocan en el uso de índices de vegetación Li et al. (2022), mientras que otros utilizan técnicas como la regresión parcial de mínimos cuadrados y bosque aleatorio Sun et al. (2021). Además, se exploran extracción de características basadas en modelos.

El resto de este manuscrito se distribuye de la siguiente manera: sección 2, se relaciona la literatura existente y la sección 3, presenta las bases teóricas y conceptuales que sustentan la propuesta. En la sección 4 se detalla la metodología utilizada para desarrollar los objetivos propuestos, lo que conlleva a la sección 5, donde se destacan los resultados obtenidos, y finalmente, en la sección 6 se presentan las conclusiones del trabajo de implementación.

2. ANTECEDENTES

Las imágenes hiperespectrales han evolucionado como una herramienta fundamental en diversas aplicaciones a lo largo del tiempo. Algunos ejemplos relevantes de estas aplicaciones incluyen la identificación de minerales, detección de enfermedades en los cultivos, visualización de parámetros no visibles por el ojo humano en el sector de la agricultura y zona urbana, por consiguiente, aportan una perspectiva que ha generado transformaciones significativas en una amplia gama de industrias y campos de estudio.

El propósito de este artículo es llevar a cabo una revisión exhaustiva con el objetivo de desarrollar una metodología para abordar el desafío de la clasificación del estrés en cultivos de aguacate criollo y Hass. De este modo, se aborda el problema de la detección del nivel de estrés en estos cultivos. Sin embargo, cabe mencionar que, después de realizar una búsqueda exhaustiva en bases de datos de revistas científicas, se encuentra que no hay publicaciones disponibles que aborde esta problemática en la tabla 1, por lo que es necesario explorar la literatura existente sobre cómo se ha enfrentado el problema en otros cultivos donde, permite evaluar las soluciones frente a otras aplicaciones.

Tabla 1 – Plan de búsqueda referenciado a la problemática.

Resultados de búsqueda de base de datos	Ecuación de búsqueda
	'Water stress' AND 'HSI' AND 'Avocado' AND 'Artificial Intelligence'
SpringerLink	20 resultados encontrados con la ecuación de búsqueda
Oxford Academic	0 resultados encontrados con la ecuación de búsqueda
Taylor and francis group	0 resultados encontrados con la ecuación de búsqueda
google sholar	19 resultados encontrados con la ecuación de búsqueda
ScienceDiret	0 resultados encontrados con la ecuación de búsqueda
Scopus	5 resultados encontrados con la ecuación de búsqueda
Gale	0 resultados encontrados con la ecuación de búsqueda
Dialnet	0 resultados encontrados con la ecuación de búsqueda

Fuente: Adaptación propia

En la revisión literaria se destacan las siguientes publicaciones, de las cuales presentamos un corto resumen:

'Revisión de la evaluación del estrés hídrico de los cultivos utilizando sensores remotos', es publicado el 17 de octubre del 2021. Se enfoca principalmente en la evaluación del estrés hídrico en cultivos agrícolas utilizando tecnologías de teledetección. El artículo no especifica un cultivo, no obstante, el objetivo es proporcionar una visión general de cómo pueden utilizarse para analizar el estrés hídrico en distintas variedades de cultivos. Se discute la relación entre el contenido relativo de agua (RWC), el grosor equivalente de agua (EWT), se describen varias tecnologías de teledetección tales como el análisis del estrés hídrico en el cultivo mediante la identificación de la fluorescencia de la clorofila provocada por la radiación solar sistemas de censado óptico, sistemas de censado termométrico, sistemas de temperatura de la superficie terrestre, sistemas de censado multiespectral que entran en el grupo de orbitales y aéreos, sistemas de censado hiperespectral y el sistema de censado LiDAR. Además, utilizan clasificación fenotípica de hojas por medio de imágenes digitales. El artículo revela resultados alentadores con rangos de precisión del 83 % y 99 %. La relevancia del estrés hídrico en diferentes etapas y como las técnicas de teledetección pueden proporcionar mejor información; sin estrés, estrés leve y estrés severo, Además, la metodología abarca la utilización de métodos de inteligencia artificial (ANN) que se apoyan en información recolectada por medio de sensores inalámbricos y termómetros infrarrojos Ahmad et al. (2021).

El trabajo titulado **"Monitoreo hiperespectral terrestre de estrés hídrico en trigo durante diversas etapas de crecimiento"**, publicado el 22 de septiembre de 2022 por los autores Qiang Li, Maofang Gao y Zhao-Liang Li, se enfoca en el análisis de cultivos de trigo sometidos a diferentes fases de crecimiento y condiciones de tratamiento hídrico. Para este propósito, se emplearon instrumentos de medición continua de la humedad del suelo y un dispositivo espectroscópico analítico (ASD) para obtener la curva espectral del trigo bajo diferentes tratamientos de agua. El estudio abordó una serie de parámetros, incluyendo la temperatura ambiente y del dosel, así como la humedad del suelo y el espectro del dosel del trigo. La temperatura del dosel fue registrada utilizando sensores térmicos infrarrojos (*METERSI – 400*). Durante períodos de sequía en diversas etapas de crecimiento, el trigo mostró respuestas espectrales distintivas. Los datos espectrales se obtuvieron en un rango que abarcó desde 350 hasta 2500 nm, con una resolución de 3 nm en el intervalo de 350 a 1000 nm y una resolución de 10 nm en el intervalo de 1000 a 2500 nm. Se prestaron especial atención a los parámetros relacionados con el borde rojo y los indicadores de humedad de la vegetación. De los once indicadores de humedad de la vegetación (VWI) evaluados, el índice hídrico (WI) demostró ser particularmente prometedor para identificar el estrés hídrico en el trigo a lo largo de diferentes etapas de su desarrollo Li et al. (2022).

El artículo **"Detección de Estrés en Cultivos Agrícolas con Datos Espectrales Utilizando Métodos de Selección de Bandas"** escrito por Pablo Carmona Zuluaga y publicado en 2023 se enfoca en la identificación del estrés nutricional en tres tipos de cultivos: maíz, frijol y aguacate. El autor propone una metodología para detectar el estrés abiótico en cultivos temporales utilizando información espectroscópica, que incluye datos hiperespectrales (HSI) y firmas espectrales. Además, se emplean estrategias de selección de bandas para simplificar los datos y aplican enfoques de aprendizaje automático, utilizando técnicas de selección y reducción de bandas, así como métricas de precisión basadas en la matriz de confusión.

Las técnicas de selección de bandas utilizadas incluyen Ordenamiento de Bandas basado en la Similitud de Mahalanobis (Bombs), Descomposición en Valores Singulares (SVD) y Proyección de Subespacio Ortogonal con Extracción de Características Independiente de Clase (OCF). Para llevar a cabo la clasificación y se utilizan métodos como Máquina de Vectores de Soporte (SVM), Bosque Aleatorio (RF) y Red Neuronal (NN). La efectividad de esta metodología se evalúa en tres conjuntos de datos correspondientes a los cultivos mencionados anteriormente y que han sido sometidos a diversos tratamientos nutricionales, como deficiencias de nitrógeno y fósforo.

Las imágenes hiperespectrales (HSI) presentan desafíos debido a su gran volumen y la dificultad de obtener un número considerable de muestras. Por lo tanto, es crucial implementar técnicas de reducción de bandas para conservar la información relevante y eliminar el ruido. Por lo tanto, propone varios métodos de reducción de bandas, que incluyen Ranking, dispersión, búsqueda, agrupación, aprendizaje y enfoques híbridos.

En la revisión de la literatura, se destaca el uso de diversas técnicas, como la regresión lineal, las máquinas de soporte vectorial (SVM), el método Random Forest (RF), K-nearest neighbor y las redes neuronales artificiales (NN). Además, se menciona que los métodos de clasificación pueden ser supervisados o no supervisados, dependiendo de si los datos utilizados incluyen etiquetas asociadas. En relación con las redes neuronales, se observa una amplia variedad de arquitecturas y topologías, siendo las más comunes el perceptrón multicapa (MLP), las máquinas de aprendizaje extremo, las funciones de base radial y las redes neuronales profundas (DNN) Carmona Zuluaga (2023).

3. MATERIALES Y MÉTODOS

3.1. Imágenes hiperespectrales y teledetección

En este trabajo se aborda la aplicación de tecnologías de gran avance tal como es la teledetección en la superficie terrestre para identificar objetos de interés. La teledetección es un instrumento valioso para analizar y recopilar datos de la superficie terrestre sin entrar en contacto directo. Además, al estudiar el censo de los datos estadísticos, se explora la técnica de las imágenes hiperespectrales (HSI) como una tecnología relativamente nueva que combina datos y son característicos espacialmente en información espectral Asaari et al. (2018).

La mejor manera de comprender la tecnología hiperespectral es considerar primero la naturaleza de una imagen digital en color estándar, a diferencia de una imagen hiperespectral. En una imagen en color, las longitudes de onda de la luz están asociadas con diferentes colores: la luz azul tiene longitudes de onda de 475 nm, la verde 520 nm y la roja 650 nm Lowe et al. (2017). Una imagen en color se forma combinando tres amplias bandas de longitudes de onda: roja, verde y azul.

Una imagen en color es un ejemplo de imagen multispectral con tres bandas, cada una de las cuales registra uno de los tres colores primarios: rojo, verde y azul. Sin embargo, las imágenes multispectrales pueden contener más bandas, quizás incluso cubriendo la región infrarroja del espectro.

Las imágenes hiperespectrales (HSI) suelen cubrir cientos de bandas de longitud de onda más estrechas, dispuestas en un rango espectral. Este enfoque da como resultado un conjunto de datos denso y con un gran contenido informativo en color, con suficiente resolución espacial para incluir varios cientos de puntos de datos (píxeles) en una sola hoja.

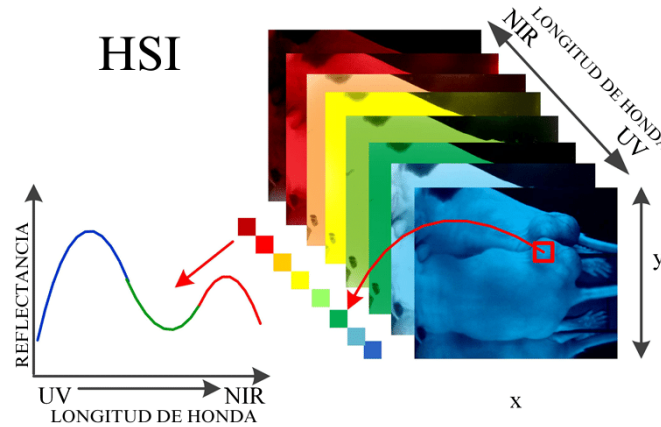
Comprender la relación de reflectancia entre las imágenes es beneficioso para extraer y sintetizar información relevante y al mismo tiempo excluir información redundante.

La capacidad de obtener datos hiperespectrales (a través de espectroscopía) con una alta resolución espectral posibilita un examen más minucioso de las características de las plantas, ofreciendo un método no invasivo para evaluar el contenido de agua Loggenberg et al. (2018), lo que convierte a las imágenes hiperespectrales en una herramienta para llevar a cabo análisis eficientes.

Las HSI capturan escenas u objetos utilizando múltiples bandas ópticas en un amplio rango espectral, creando un volumen de datos que se asemeja a un cubo de información. Cada parte de este cubo corresponde a una longitud de onda diferente, con dos dimensiones representando el área capturada y la tercera la variación en la longitud de onda (Crio-

llo Saavedra and Merino Arellano, 2023). Por otro lado, las imágenes hiperespectrales (HSI) registra la capacidad de permitir caracterizar y distinguir eficazmente fenómenos en la superficie terrestre, siendo esencial en aplicaciones de ingeniería y ciencia Pérez et al. (2021).

Figura 1 – Representación de cubo de datos hiperespectrales con firma espectral.



Fuente: Sowmya et al. (2019)

Cabe destacar que, el espectro electromagnético representa diferentes formas de energía lumínica y los sensores hiperespectrales capturan datos de más de 200 franjas, desde el ultravioleta (UV) hasta el infrarrojo de onda larga (LWIR). El término espectro refleja la variación de la intensidad de una radiación respecto a un parámetro específico, y la "firma espectral" se define como la distribución de la reflectancia o emisión de un elemento en relación con la longitud de onda, representando los valores de reflectancia durante su generación Criollo Saavedra and Merino Arellano (2023).

La tecnología de teledetección y el análisis hiperespectral combinado con el aprendizaje automático proporciona una forma conveniente de modelar el estrés Loggenberg et al. (2018). Las herramientas que proporcionan la tecnología de aprendizaje automático mejoran el rendimiento y clasifican datos espectrales mediante métodos de clasificación de categorías y datos en diferentes clases.

3.2. Aprendizaje automático

El aprendizaje automático (ML) se considera un subconjunto de la inteligencia artificial (IA) que se enfoca en crear sistemas capaces de aprender o mejorar su rendimiento a partir de los datos que utilizan. Un término amplio utilizado para describir máquinas o sistemas que imitan la inteligencia humana es "inteligencia artificial". Aunque con frecuencia se utilizan los términos aprendizaje automático e inteligencia artificial de manera intercambiable, tienen significados distintos. Es importante señalar que, aunque todo el aprendizaje automático es una forma de inteligencia artificial, no toda la inteligencia artificial involucra aprendizaje automático OCI (2022).

El aprendizaje automático aplicado a imágenes hiperespectrales se refiere a la implementación de técnicas para el análisis y procesamiento de dichas imágenes. Esto se basa en la relación entre el píxel central y su contexto, lo que puede aumentar significativamente la robustez del modelo para el estudio de los objetos presentes en la escena Nachappa et al. (2020). El aprendizaje automático se emplea en imágenes hiperespectrales para diversas tareas, como la clasificación de objetos y materiales, la detección de anomalías, la segmentación de objetos, la reconstrucción de imágenes y la mejora de la

calidad de las mismas.

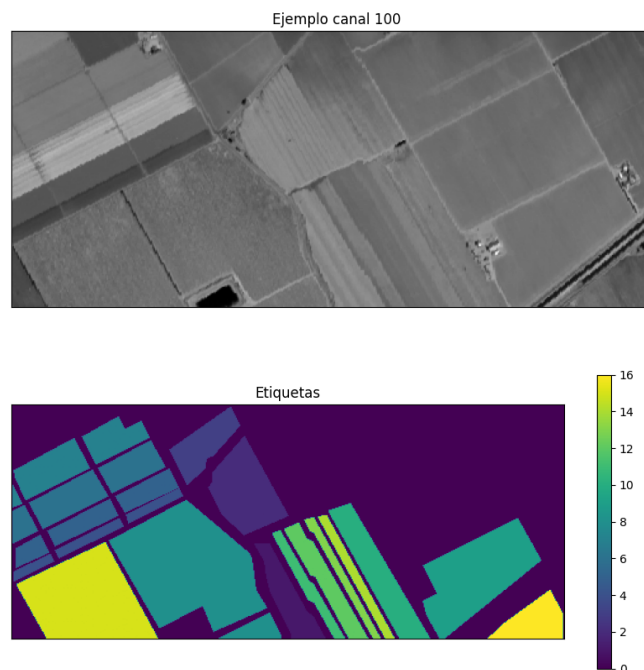
El aprendizaje automático aplicado a imágenes hiperespectrales puede involucrar técnicas como las redes neuronales convolucionales (CNN), las redes neuronales recurrentes (RNN), las máquinas de vectores de soporte (SVM), los clasificadores basados en árboles, el análisis de componentes principales (PCA), entre otros Nachappa et al. (2020). El aprendizaje supervisado es una categoría en el aprendizaje automático que hace uso de datos etiquetados para el entrenamiento. En otras palabras, cada muestra de datos está asociada con una etiqueta o clase. El objetivo es aprender una función que mapee las entradas a las salidas deseadas y luego realizar predicciones o clasificaciones en datos no etiquetados Quigley (2023).

3.3. Base de datos para evaluar algoritmos de clasificación usando HSI

Las bases de datos que son usados para evaluar algoritmos de clasificación en apoyo de imágenes hiperespectrales (HSI) es un componente esencial en la investigación en visión por computadora y procesamiento de imágenes. Estas bases de datos contienen conjuntos de imágenes hiperespectrales recopiladas en diversas condiciones y entornos, de este modo se ilustra la base de datos SALINAS.

A continuación, se demuestra un ejemplo del canal 100 en escala de gris y la imagen de abajo son las etiquetas del terreno en la figura 2:

Figura 2 – Terreno evaluado con sus respectivas divisiones de clases que representa el terreno



Fuente: Adaptación propia

SALINAS:

Es una imagen captada por el sensor AVIRIS de 224 bandas sobre el Valle de Salinas, California, y se caracteriza por una alta resolución espacial (píxeles de 3,7 metros). El área cubierta comprende -512 líneas por 217 muestras. En la escena se

descartan las 20 bandas de absorción de agua, en este caso las bandas: [108 – 112], [154 – 167], 224 esta imagen solo estaba disponible como datos de radiancia en el sensor incluye hortalizas, suelos desnudos y campos de viñedos (Jia et al., 2021).

La base de datos SALINAS en total contiene 16 clases, representado en la tabla 2

Tabla 2 – Representación de clase, base de datos SALINAS

#	Clases	Muestras
1	Malezas verdes de brócoli 1	2009
2	Malezas verdes de brócoli 2	3726
3	Barbecho	1976
4	Arado áspero en barbecho	1394
5	Barbecho suave	2678
6	Rastrojo	3959
7	Apio	3579
8	Uvas en bruto	11271
9	Desarrollo del suelo del viñedo.	6203
10	Malezas verdes senescentes del maíz	3278
11	Lechuga romana 4 semanas	1068
12	Lechuga romana 5 semanas	1927
13	Lechuga romana 6 semanas	916
14	Lechuga romana 7 semanas	1070
15	Viñedo no entrenado	7268
16	Viñedo vertical enrejado	1807

Fuente: Adaptación propia

3.4. Base de datos empleada para cultivo de aguacate

El experimento para cultivo de aguacate, se estableció a partir de una base de datos, donde se emplea datos previamente recolectados por investigadores del grupo MIRP del Instituto Tecnológico Metropolitano de Medellín, el cual, consta de un total de 64 imágenes divididas en 2 grupos, donde tenemos el grupo *G1* y *G2* de imágenes enmascarada previamente etiquetada (imágenes con solo las hojas de interés y sin fondo), donde incluye 3 clases y 2 subclases más. Se conforma por la clase de control, restricción de agua (*Restriccion_50* y *Restriccion_75*) y exceso de agua (*Exceso_125* y *Exceso_150*), es decir, Las variedades incluidas son aguacate criollo o común y aguacate Hass, lo que caracteriza la base de datos como imágenes segmentadas donde se tiene identificada la región de las hojas de la planta.

4. METODOLOGÍA

En esta investigación, se plantea la necesidad de evaluar la viabilidad de identificar niveles de estrés en los cultivos de aguacate a través del análisis de datos recopilados. Con el propósito de alcanzar este objetivo, se recurre a técnicas de análisis estadístico, detalladas a continuación, con el fin de determinar si se revelan diferencias estadísticamente significativas entre los diversos grupos al utilizar las características o métricas recomendadas en la literatura científica.

4.1. Anova, Método de análisis de la varianza

El ANOVA es una técnica estadística que se emplea para detectar diferencias significativas entre los promedios de grupos diversos, así como para identificar cuáles promedios específicos difieren entre sí. Además, permite llevar a cabo contrastes planificados antes o después del experimento para comparar los promedios, basándose en el análisis de varianza. Es importante destacar que esta técnica se utiliza al comparar las medias de tres o más grupos mediante un único factor independiente. Por lo tanto, solo puede indicar si al menos un par de estas medias es considerablemente diferente, sin especificar cuál. Para que el análisis sea válido, la variable dependiente debe mostrar normalidad y uniformidad de varianza en todos los grupos (Rencher and Schimek, 1997).

Las comparaciones que se observan en la sección 5 de resultados mediante el análisis Anova, permite gestionar interpretaciones acerca de la viabilidad del comportamiento de los datos a través de índices de vegetación, que ayudan al estudio de zonas características del cultivo. Por esta razón, esta técnica mide el factor de conformidad o inconformidad del experimento.

4.2. Manova, Método de análisis multivariante de la varianza

El MANOVA, o Análisis Multivariante de Varianza, es una herramienta estadística valiosa que posibilita el estudio de múltiples variables dependientes en relación con factores y covariables. Ofreciendo una perspectiva más abarcadora de cómo estos factores inciden en las medias de las variables dependientes en un conjunto de datos específico, al mismo tiempo que permite la contrastación de hipótesis acerca de su influencia. Las variables de factor divide la población en grupos. Utilizando este procedimiento de modelo lineal general, donde es posible contrastar hipótesis nulas sobre los efectos de las variables de factor sobre las medias de varias agrupaciones de una distribución conjunta de variables dependientes IBM (2021). De esta manera, en la sección 5 se tiene en cuenta el comportamiento de los datos a nivel de bandas espectrales, ya que contiene mucha más información en relación a cada clase.

4.3. Método de Generación de Etiquetas para identificación de la clase

En este segmento se aborda la generación de etiquetas a nivel de píxel característico, lo cual constituye un elemento fundamental en la identificación de las clases para fines de estudio. El propósito principal es identificar y diferenciar las áreas de interés en los cultivos, lo que facilita la detección de estas zonas.

El proceso implica el uso de imágenes previamente segmentadas, es decir, aquellas que representan las hojas de los cultivos. Cabe destacar que la base de datos se encuentra etiquetada con sus respectivas clases. la implementación de este método involucra la utilización de estas imágenes y se caracteriza por interpretar inicialmente cada imagen, accediendo a las dimensiones que componen el cubo espectral de cada una.

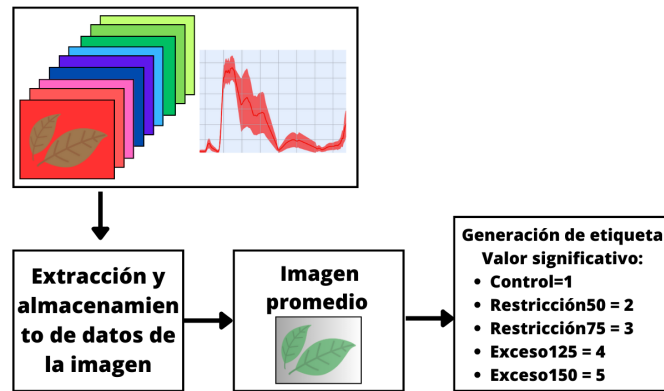
Después de obtener la información de las bandas que conforman cada imagen, es esencial procesar la abundante información proporcionada por las imágenes hiperespectrales (HSI). Por lo tanto, se calcula el promedio a través de todas sus bandas, generando como resultado una nueva imagen que permite verificar que algunos píxeles, en diferentes posiciones de los ejes, no se conviertan en 0 para ciertas bandas.

Finalmente, se crea el conjunto de etiquetas mediante la confirmación de los píxeles existentes, es decir, aquellos píxeles cuyo valor se encuentra por encima de 0 se toman en consideración y se les asigna la etiqueta correspondiente a la clase que representan en la imagen.

Estas etiquetas desempeñan un papel fundamental en la clasificación y segmentación de los datos, proporcionando información valiosa para la toma de decisiones en las implementaciones usadas, ya que la aplicación de este método contribuye

a llevar a cabo un proceso más organizado y detallado, garantizando la preservación de la información en la base de datos.

Figura 3 – Diagrama de bloques del proceso de etiquetado



Fuente: Adaptación propia

4.4. Método de partición de la base de datos en información del vecindario de los píxeles e información localizada

El proceso de generación de parches y píxeles para la clasificación se usan 2 enfoques relevantes para dichas implementaciones que se evidencian en la sección 4.7.

1. El uso de información del vecindario de los píxeles de cada una de las imágenes, se aplica la construcción de parches $K \times K$, donde K es el tamaño del kernel, lo que permite una mayor factibilidad computacional, donde cada parche contiene información espectral, y su correspondiente etiqueta original de la imagen. Esta división en parches facilita la tarea de entrenamiento y clasificación, ya que se centra en áreas específicas y captura detalles relevantes, la figura 4a se evidencia el método Jia et al. (2021).
2. El uso del segundo método se encarga de generar pares de datos y etiquetas a nivel de píxel donde, cada píxel se convierte en una muestra y se asocia con su correspondiente etiqueta de clasificación. De este modo, se tiene el uso de información localizada, es decir, manejar la información de todas las bandas a través de un solo píxel Jia et al. (2021). Cabe resaltar que, el uso de este método son esenciales para que la red neuronal que se aplique, pueda interactuar con los datos de forma más detallada, la figura 4b se evidencia el método.

Los enfoques 1, 2 también facilita la capacidad de aumentar el conjunto de datos mediante la realización de transformaciones como rotaciones y reflexiones en las imágenes que aumenta la variabilidad y evita el sobreajuste, lo que puede mejorar la capacidad de generalización del modelo de clasificación.

4.5. Análisis por bandas seleccionadas en el rojo e infrarrojo cercano de la curva de reflectancia espectral en las imágenes hiperespectrales

El análisis de plantas y vegetación, se deben tener en cuenta los rangos de longitud de onda más relevantes, ya que, incluyen la combinación de banda visible e infrarrojo cercano. Este rango de longitudes de onda puede capturar cambios en

Figura 4 – Diagramas de bloques de método por parches e información localizada (píxel a píxel)



(a) Diagrama de bloques del sistema de adquisición de datos por generación de parches

(b) Diagrama de bloques del sistema de adquisición de datos por generación de píxel a píxel

Fuente: Adaptación propia

la pigmentación de las hojas, siendo el rojo cercano [400 – 700nm] y el infrarrojo cercano [700 – 1400nm] Ahmad et al. (2021). Por ejemplo, una deshidratación grave puede afectar la estructura del mesófilo de las hojas, provocando cambios en la reflectancia del infrarrojo cercano. Sin embargo, un estrés hídrico menos grave normalmente no produce un efecto lo suficientemente significativo como para detectarlo Lowe et al. (2017). La posición del "borde rojo" se mueve a lo largo del eje de la longitud de onda dependiendo del contenido de clorofila, la biomasa y los cambios fenológicos Jia et al. (2023). Los métodos de selección de longitud de onda se basan en los valores de reflectancia obtenidos promediando todos los píxeles de la imagen de reflectancia. El análisis de correlación considera directamente la información global de la imagen Weng et al. (2023).

En el campo de la espectroscopía de luz reflejada, el objetivo principal es obtener el espectro de reflectancia, que es la propiedad fundamental buscada. La reflectancia se define como la proporción entre la intensidad de la luz reflejada por una muestra y la intensidad de la luz incidente sobre ella Quintulén et al. (2018). Dicha magnitud se altera en función de la longitud de onda en la mayoría de los materiales, ya que la energía en diferentes longitudes de onda es absorbida o dispersa en distintos grados Sánchez Bernabé (2016). Estas variaciones en la reflectancia son evidentes cuando se comparan las curvas espectrales de reflectancia en función de la longitud de onda para diferentes materiales.

En el análisis de imágenes hiperespectrales, se emplean los espectros de reflectancia para extraer datos acerca de la composición de las superficies que están siendo observadas. La reflectancia, una cantidad sin dimensiones, se representa a través de un valor numérico entre 0 y 1, o como un porcentaje en el rango de 0 % a 100 % (Sánchez Bernabé, 2016).

Las notables variaciones y la disminución marcada en las curvas espectrales de reflectancia establecen el intervalo de longitudes de onda en el cual el material elige absorber la energía entrante de manera selectiva. Estas regiones se identifican como las bandas de absorción para el análisis de información importante. Las formas de las curvas espectrales de reflectancia de las plantas exhiben un patrón distintivo influenciado por sus múltiples características. En la parte visible del espectro, la configuración de la curva está influenciada por los efectos de absorción de la clorofila y otros pigmentos presentes en las hojas, lo que a su vez determina la posición y el ángulo de la captura para el objeto de interés (Asaari et al., 2019).

La diferenciación clave entre los sistemas multispectrales y los sistemas hiperespectrales, sin importar la plataforma empleada, sigue siendo la resolución espectral, que se mide en términos de ancho de banda. Esta resolución señala tanto la cantidad como la amplitud de las regiones en el espectro de las cuales el sensor adquiere datos (Sánchez Bernabé, 2016).

4.6. Análisis por índices de vegetación (IV)

En la tabla 3, se proporciona una descripción detallada de los índices de vegetación utilizados en el análisis, asimismo las bandas espectrales asociadas y las longitudes de onda que se emplearon. Este análisis brindará una comprensión

más profunda de la relación entre estos índices y las particularidades espectrales de los datos, así como de su función en el contexto de la investigación.

Tabla 3 – Valores determinados de la longitud de onda (nm) y índices de vegetaciones

Índice	Bandas espectrales y longitud de onda (nm)				
	NIR	B. ROJO	B. AZUL	MIR	\textbf{B. VERDE}
NDVI	800 – 1100 nm	500 – 700 nm	-	-	-
SAVI	800 – 1100 nm	500 – 700 nm	-	-	-
WI	750 – 751 nm	600 – 601 nm	-	-	-
GCI	800 – 1100 nm	-	-	-	500 – 600 nm
MSI	700-1400 nm	-	-	1500 – 2300 nm	-
EVI	700 – 1400 nm	400 – 700 nm	450 – 500 nm	-	-

Fuente: Elaboración propia

Tomando como referencia el artículo presentado por Qiang Li, Maofang Gao y Zhao-Liang Li, se detallan los distintos índices de vegetación, incluyendo sus escalas de medición y sus aplicaciones tanto en el campo, en hojas como a través de satélites. Este artículo presenta varias hipótesis que facilitan un análisis más detallado sobre cómo trabajar con imágenes hiperespectrales mediante el uso de índices de vegetación. Es importante tener en cuenta que, para el cálculo de estos índices se usa la información presente en las bandas específicas de cada pixel, por lo tanto, no da información sobre la cantidad de luz reflejada en distintas longitudes de onda, sin embargo, proporcionan información relevante sobre la vegetación presente.

El Índice Hídrico (WI) es uno de los indicadores de vegetación que posibilita la evaluación de la humedad. La fórmula general para el Índice Hídrico es la siguiente:

$$WI = \frac{R - SWIR}{R + SWIR}, \tag{1}$$

Donde, R representa la reflectancia en una banda específica del espectro electromagnético, generalmente se encuentra en la zona roja del espectro y SWIR representa la reflectancia en el infrarrojo de onda corta, generalmente se encuentra en el rango en el rango cercano al infrarrojo Villatoro et al. (2021).

El Índice de Vegetación de Diferencia Normalizada (NDVI) índice de monitoreo ambiental, cuantifica la intensidad del color verde y la densidad de la vegetación en una imagen. Esta variación esta en el rango [-1, 1] Yang et al. (2022). La fórmula del Índice de Vegetación de Diferencia Normalizada es la siguiente:

$$NDVI = \frac{NIR + RED}{NIR - RED}, \tag{2}$$

Donde: NIR es la reflectancia en la región cercana al infrarrojo y RED representa la reflectancia en la banda roja del espectro Villatoro et al. (2021).

El Índice de Estrés Hídrico (MSI) se emplea para determinar el grado de estrés hídrico o la escasez de agua en la vegetación, y sus valores oscilan en una escala que va de [-3, 3]. La fórmula utilizada para calcular el Índice de Estrés Hídrico es la siguiente:

$$MSI = \frac{NIR + MIR}{NIR - MIR}, \tag{3}$$

por lo tanto: NIR es la reflectancia en el infrarrojo cercano y MIR es la reflectancia en el infrarrojo medio Yang et al. (2022).

El Índice de Diferencia de Agua Normalizada (NDWI) se utiliza con el propósito de resaltar la existencia de masas de agua, lo que contribuye a que las áreas acuáticas sean más distinguibles en la imagen de modo que este índice evalúa el contenido de humedad.

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR}, \quad (4)$$

cada variable representa lo siguiente: NIR representa la reflectancia en el infrarrojo cercano y SWIR representa la reflectancia en el infrarrojo de onda corta.

El índice de Diferencia de Agua Normalizada (NDII) ha sido empleado con el propósito de exhibir las fluctuaciones en la cantidad de agua presente en la cubierta del suelo, y La fórmula del NDII es:

$$NDII = \frac{NIR + MIR}{NIR - MIR}, \quad (5)$$

por lo tanto, NDII es el valor del Índice de Diferencia Normalizada de Infrarrojo y Aguas Abiertas, NIR es la reflectancia en el infrarrojo cercano y MIR es la reflectancia en el infrarrojo medio Quispe Almontes (2021).

La fórmula del Índice de Vegetación Mejorado (EVI) se expresa de la siguiente manera:

$$EVI = G * \frac{(NIR - RED)}{(NIR + C1 * RED - C2 * BLUE + L)}, \quad (6)$$

donde cada componente cumple un papel específico en el cálculo. NIR representa la reflectancia en el infrarrojo cercano, RED se refiere a la reflectancia en la banda roja, BLUE se relaciona con la reflectancia en la banda azul, G es el factor de ganancia, típicamente establecido en 2,5, y su función es equilibrar las contribuciones de las diferentes bandas en la fórmula. L corresponde al factor de corrección del suelo, generalmente fijado en 1, utilizado para ajustar la reflectancia del fondo del suelo. Finalmente, C1 y C2 son coeficientes de corrección que dependen de las características del sensor utilizado, siendo sus valores habituales 6 y 7,5, respectivamente (Vijith and Dodge-Wan, 2020).

El Índice de Clorofila Verde (GCI) es una técnica que muestra el potencial para seguir de manera remota el estado fisiológico de los cultivos. Este índice se utiliza para estimar el contenido de clorofila en las hojas de varias especies de plantas Li et al. (2022), La fórmula del GCI se expresa como:

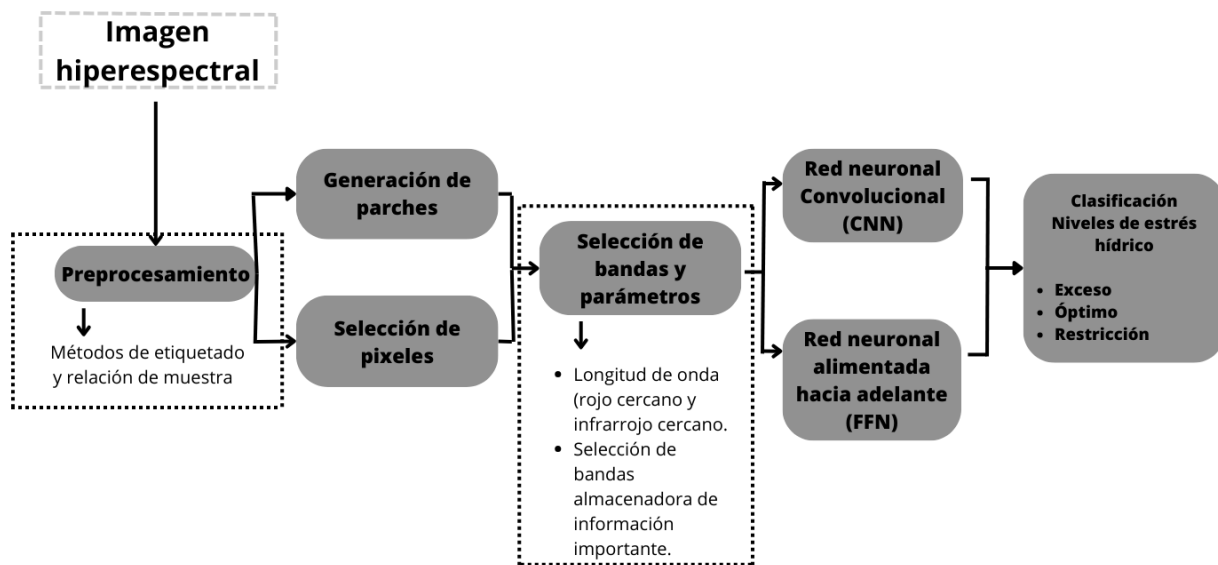
$$GCI = \frac{R_{verde}}{R_{NIR}}, \quad (7)$$

donde "Verde" representa la reflectancia en la banda de color verde y RNIR corresponde a la reflectancia en la banda del infrarrojo cercano Muñoz and Burdett (2019).

Empleando los datos disponibles, se evalúa la información contenida en cada uno de los índices mencionados usando las técnicas de análisis estadístico, como Anova o Manova. Esta evaluación permitirá determinar el potencial de cada uno de los índices para ser incorporado en un sistema de clasificación.

4.7. Técnicas de clasificación empleadas

Figura 5 – Diagrama del sistema y su proceso general de funcionamiento



Fuente: Adaptación propia

implementación 1: Sistema basado en CNN (Redes neuronales convolucionales)

El sistema 1 para imágenes hiperespectral (HSI) contiene un proceso de entrenamiento y evaluación para un modelo de red neuronal convolutiva (CNN) en el contexto de clasificación de imágenes hiperespectrales. A continuación, se explica paso a paso lo que hace cada parte de la implementación mencionada.

Creación y Configuración del Modelo CNN: Se comprende un modelo de red neuronal convolutiva (CNN) con el número de bandas espectrales, el número de clases y el tamaño del parche como argumentos. Además, se aplica una inicialización de pesos al modelo y se define la función de pérdida a utilizar siendo CrossEntropyLoss. Donde se define un optimizador SGD (Stochastic Gradient Descent) para el entrenamiento del modelo, también cabe destacar que el manejo de todos los parámetros y configuraciones varía según las necesidades del investigador y la necesidad que desea cubrir el problema.

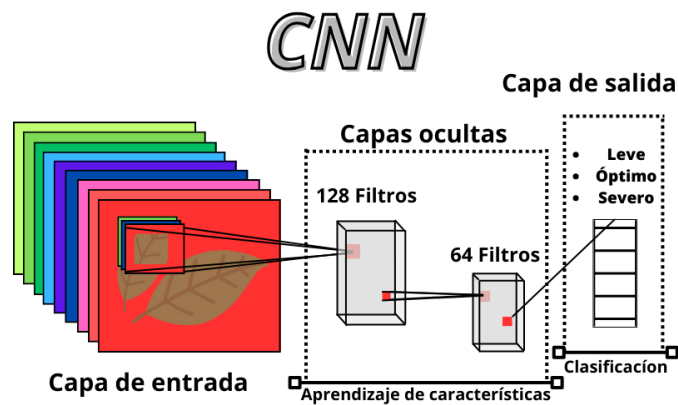
Entrenamiento y Evaluación del Modelo: se evalúa el modelo en los conjuntos de prueba y entrenamiento utilizando las funciones de prueba, la información sobre la pérdida de entrenamiento y evaluación, así como la precisión alcanzada en la evaluación es importante al momento de calificar el modelo utilizado, ya que indica que aspectos se pueden mejorar.

Para la carga de datos en la red de implementación, se resalta la realización de cada uno de los métodos que son mencionados en secciones anteriores, el cual son aplicados en este trabajo. Ya que se realiza el preprocesamiento como factibilidad de la metodología usada en dicha implementación. Además, cabe aclarar que para el entrenamiento se utilizó el grupo de imágenes enmascarado que contiene el objeto de interés sin fondo y finalmente se realiza las pruebas estadísticas que permite un análisis más profundo del modelo.

implementación 2: Sistema basado en redes completamente conectadas (Feed Forward Neural Networks)

Se conforma de un proceso de entrenamiento y evaluación para un modelo de red neuronal alimentada hacia adelante (FFN) en el contexto de clasificación de imágenes hiperespectrales. A continuación, se explica paso a paso lo que hace

Figura 6 – Diagrama de red convolucional, estructura de implementación CNN



Fuente: Adaptación propia

en cada parte de la implementación mencionada

Definición de Parámetros Iniciales: Se definen varios parámetros iniciales que controlarán el proceso de entrenamiento y la configuración del modelo. Estos parámetros incluyen el número de épocas, tasa de aprendizaje, tamaño del lote, número de trabajadores en la carga de datos y las bandas espectrales utilizadas.

Carga y Preparación de Datos: Se cargan los datos del conjunto de entrenamiento, de prueba y de validación utilizando el método de generación por píxel quien se encarga de almacenar los datos relevantes de las imágenes. Además, quien se encarga de ejecutar unos cargadores de datos para cada conjunto. El cargador de datos es una forma eficiente de cargar los datos en lotes durante el entrenamiento.

Creación y Configuración del Modelo: Se instala el modelo de red neuronal llamado FFNet_v2 con argumentos relacionados al número de bandas, número de clases y el tamaño del parche. Además se da el uso de la función de pérdida CrossEntropyLoss y el optimizador (SGD) con los parámetros previamente definidos.

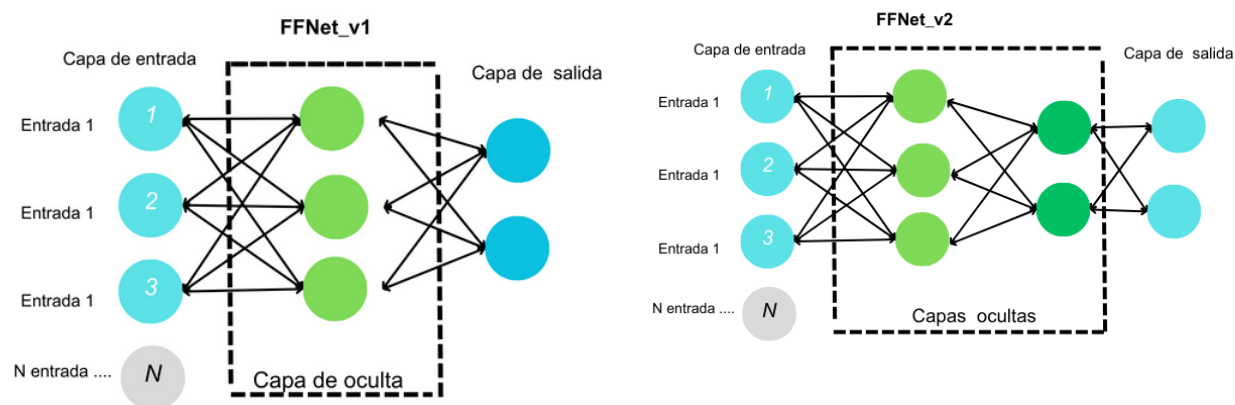
Finalmente, esta implementación ejecuta un proceso de entrenamiento y evaluación de un modelo de red neuronal para la clasificación de imágenes utilizando datos del conjunto PixelDataset. Los datos se cargan, el modelo se configura y se entrenan y evalúan durante varias épocas. Lo que termina con la evaluación y entrenamiento del modelo que se realiza con el mismo conjunto de imágenes que se le aplica al sistema de implementación por CNN.

5. RESULTADOS Y DISCUSIÓN

5.1. Base de datos SALINAS

La forma de evaluar el desempeño de un sistema de clasificación basado en redes neuronales se realizan mediante pruebas piloto con la base de datos SALINAS. Las pruebas consisten en tomar una muestra aleatoria pequeña del total de 54, 129 muestras que componen la base datos (Kang et al., 2014). Esta pequeña muestra se usa para entrenamiento y el conjunto restante para evaluar el sistema de clasificación. La estrategia es sugerida en el artículo **'Aprendizaje profundo para la clasificación de imágenes hiperespectrales con pocas muestras etiquetadas'** (Jia et al., 2021). En la figura 8 se muestra un ejemplo de los conjuntos seleccionados, donde la parte izquierda se evidencia las muestras que se seleccionan para entrenamiento (500 en total), y en el lado derecho las muestras usadas para prueba. En la tabla 4, se demuestra el comportamiento del entrenamiento en relación al número de muestras empleadas para entrenar.

Figura 7 – Representación de redes neuronales

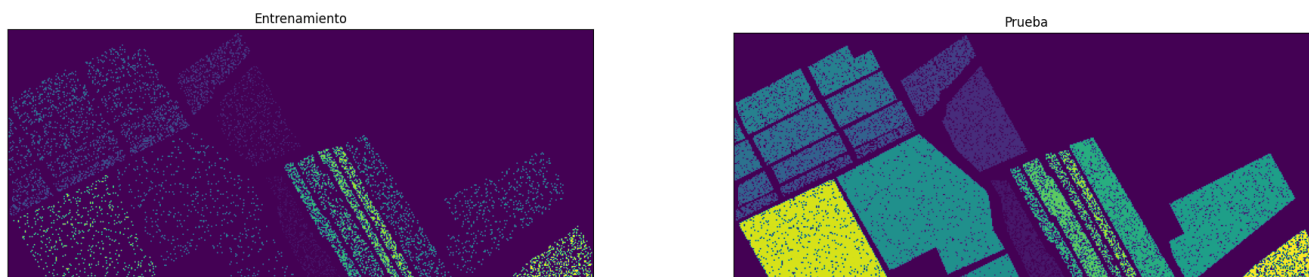


(a) Diagrama de redes completamente conectadas (Feed Forward Neural Networks), estructura FF-Versión 1
 (b) Diagrama de redes completamente conectadas (Feed Forward Neural Networks), estructura FF-Versión 2

Fuente: Adaptación propia

Además, se puede observar que el desempeño mejora en la medida que se aumenta el número de muestras para entrenar, adicionalmente se destaca el buen desempeño de la red neuronal incluso con una cantidad pequeña de muestras, comparada con el tamaño total de la base de datos. Se observa que el porcentaje de acierto supera el 90 % con tan solo 500 muestras de entrenamiento, de este modo se evalúa el algoritmo de la red neuronal variando el número de muestras de entrenamiento y utilizando parches de imágenes de 5×5 .

Figura 8 – Terreno evaluado con sus respectivas divisiones de clases que representa el terreno



Fuente: Adaptación propia

A continuación, se presentan los resultados en la tabla 4. Seguidamente se determina que a pesar de que el conjunto de prueba es mayor se nota un alto porcentaje de acierto en la medida que se aumentan las muestras de entrenamiento:

Tabla 4 – Resultados obtenidos en el uso de la base de datos SALINAS

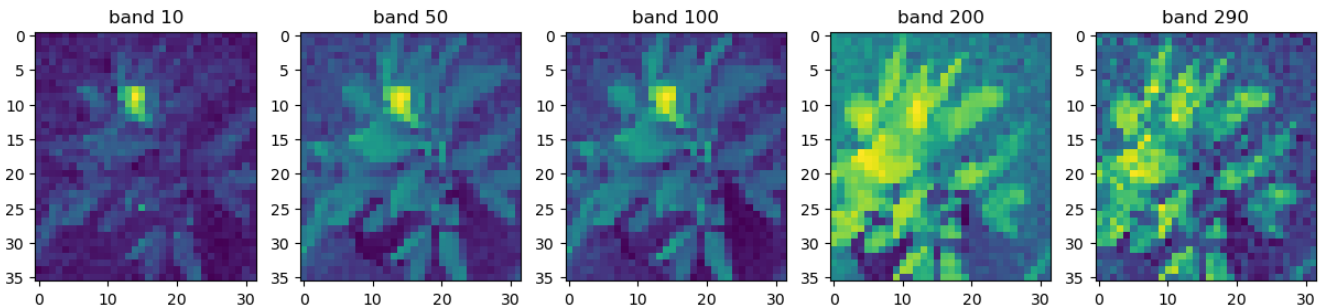
Número de muestra	Acierto(Precisión)
50	78,90 %
100	78,67 %
200	84,77 %
400	83,97 %
500	91,60 %

Fuente: Adaptación propia

Análisis de base de datos empleada para cultivo de aguacate

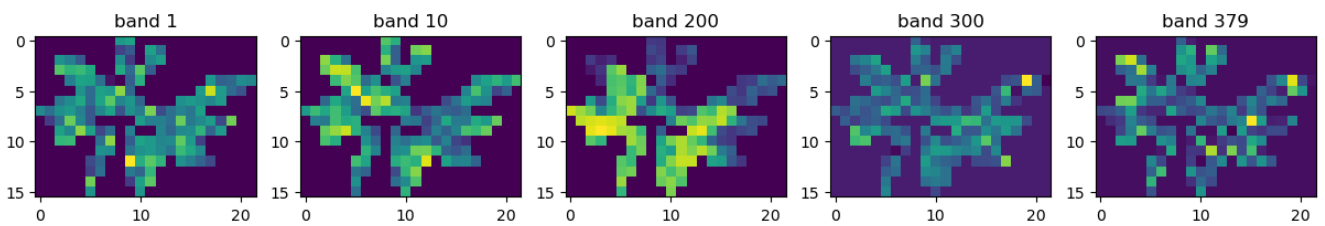
La imagen proporcionada representa una visualización de datos extraídos de la base de datos, La imagen refleja la información recopilada en cada banda, ofreciendo una visión técnica y detallada de las propiedades espectrales de la escena capturada en la figura 9 y 10.

Figura 9 – Imagen representativa de la base de datos respecto al cultivo de aguacate, una clase al azar



Fuente: Adaptación propia

Figura 10 – Imagen representativa de la base de datos respecto al cultivo de aguacate, una clase al azar



Fuente: Adaptación propia

A continuación, se observará las siguientes subsecciones que se realizan para la base de datos enfocada al cultivo de aguacate:

5.2. Curva de reflectancia espectral para el genotipo 1(G_1) y 2(G_2)

La tabla 5, se ilustra las representaciones de los grupos para cada uno de los genotipos, que ayudaran a tener conocimiento de las iniciales en el resto de la sección:

En la figura 11, contiene información importante, sobre el comportamiento espectral (curva de reflectancia espectral) que demuestra cada una de las clases y a su vez el comparativo entre las mismas.

Primeramente, la figura 11, para el genotipo G_1 se observa los parámetros de reflectancia con respecto a la longitud de onda que denomina a cada grupo, a partir de valores significativos que representa el rojo cercano é infrarrojo cercano, son quienes determinan características diferenciales de información relevante para el análisis que podría distinguir entre las clases.

El grupo $G_1_Control$ indica un pico de reflectancia un poco mayor a 0,12, $T1_G1$ se encuentra casi llegando a 0,12, $T2_G1$ está en el valor de 0,11 en un punto medio de 0,10 y 0,12, $T3_G1$ muestra que esta se encuentra cerca de 0,12 y la última clase $T4_G1$ muy cerca de 0,11, el cual hace referencia al valor más alto de reflectancia (reflejo de luz captada por los sensores que se captan en las tomas de la cámara) que tiene las clases, cuando se le realiza el cálculo de la media referente

Tabla 5 – Representación de genotipos 1 y 2

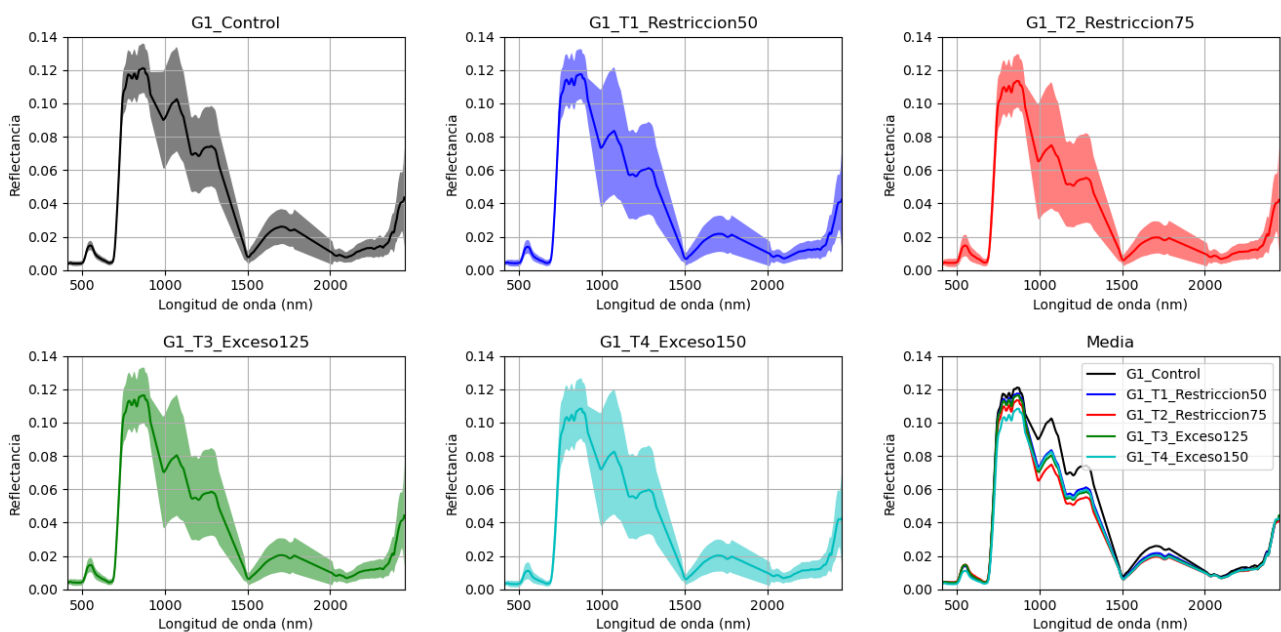
Genotipo 1	Genotipo 2
G1_Control	G2_Control
T1_G1 = Restricción50	T1_G2 = Restricción50
T2_G1 = Restricción75	T2_G2 = Restricción75
T3_G1 = Exceso125	T3_G2 = Exceso125
T4_G1 = Exceso150	T4_G2 = Exceso150

Fuente: Adaptación propia

a todas las imágenes del mismo directorio del grupo.

La última gráfica se observa todas las clases de manera conjunta, donde el análisis de dicha gráfica es ver, cual es la diferencia que hay entre las bandas, ya que el punto de partida, es la longitud de onda donde se puede verificar el rango al que pertenece el rojo cercano e infrarrojo cercano, donde las curvas espectrales se nota una diferencia según la información que puede brindar las gráficas, desde los [700 – 900nm] y también comenzando en [1000 – 1400nm] por tanto, se muestra una pequeña diferencia de información que se tralapan, demostrando similitud de los píxeles a través de las bandas.

Figura 11 – Curva de reflectancia espectral por clase, grupo G1



Fuente: Adaptación propia

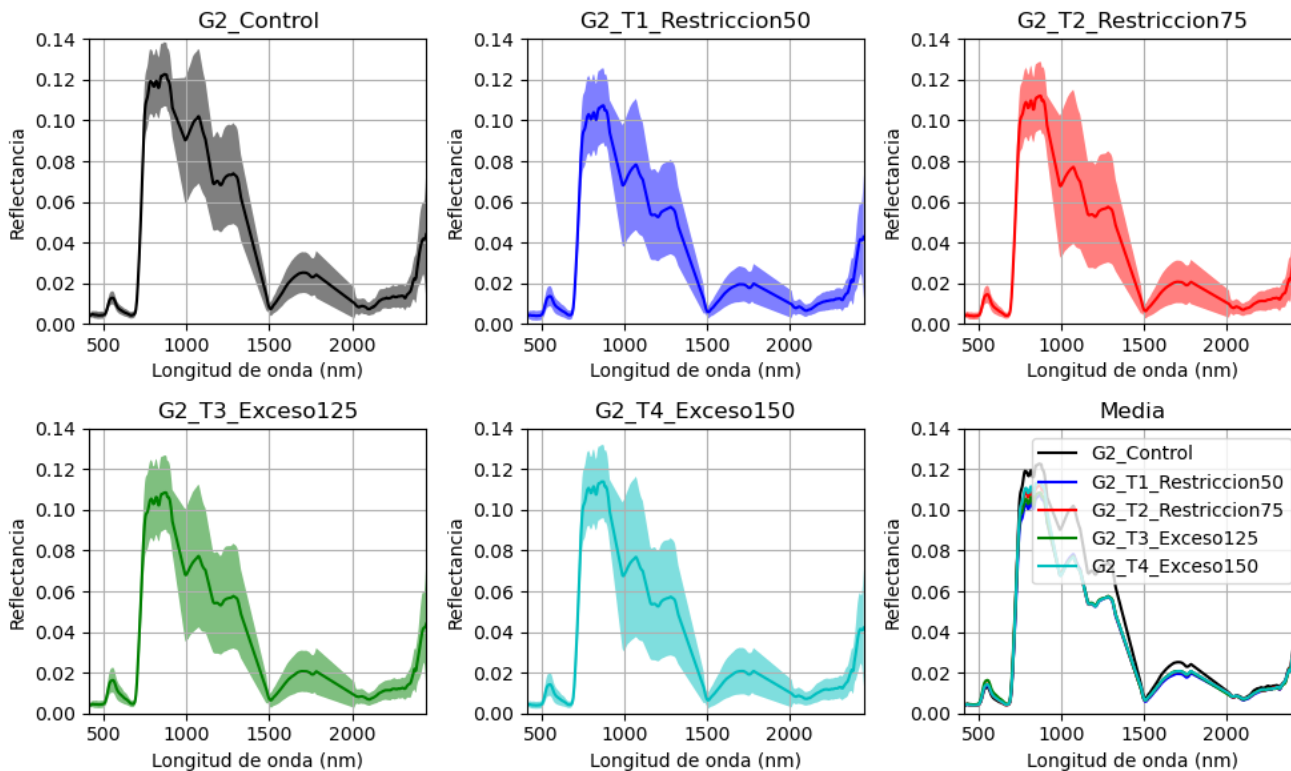
La figura 12 del grupo G2, se observa los parámetros de reflectancia con respecto a la longitud de onda que denomina a cada grupo a partir de valores significativos que representa el rojo cercano e infrarrojo cercano quienes determinan características diferenciales de información relevante para el análisis que podría diferenciar entre las clases.

El grupo G2_Control indica un pico de reflectancia un poco mayor a 0,12, T1_G2 se encuentra casi llegando a 0,11, T2_G2 está en el valor de 0,11 ubicado en un punto medio de 0,10 y 0,12, T3_G2 muestra que esta se encuentra cerca de 0,11 y la última clase T4_G2 un poco elevado de 0,11, el cual hace referencia al valor más alto de reflectancia (reflejo de luz captada por los sensores que se captan en las tomas de la cámara) que tiene las clases cuando se le realiza el cálculo de la media referente a todas las imágenes del mismo directorio del grupo.

La última gráfica se observa todas las clases de manera conjunto, donde el análisis de dicha gráfica es ver, cual es la diferencia

que hay entre las bandas, ya que el punto de partida, es la longitud de onda donde se puede verificar el rango al que pertenece el rojo cercano e infrarrojo cercano, donde las curvas espectrales se nota una diferencia según la información que puede brindar las gráficas, se encuentra también comenzando en [800 – 1400nm], se muestra información claramente similar a la figura anterior, del cual se asemeja al análisis.

Figura 12 – Curva de reflectancia espectral por clase, grupo G2



Fuente: Adaptación propia

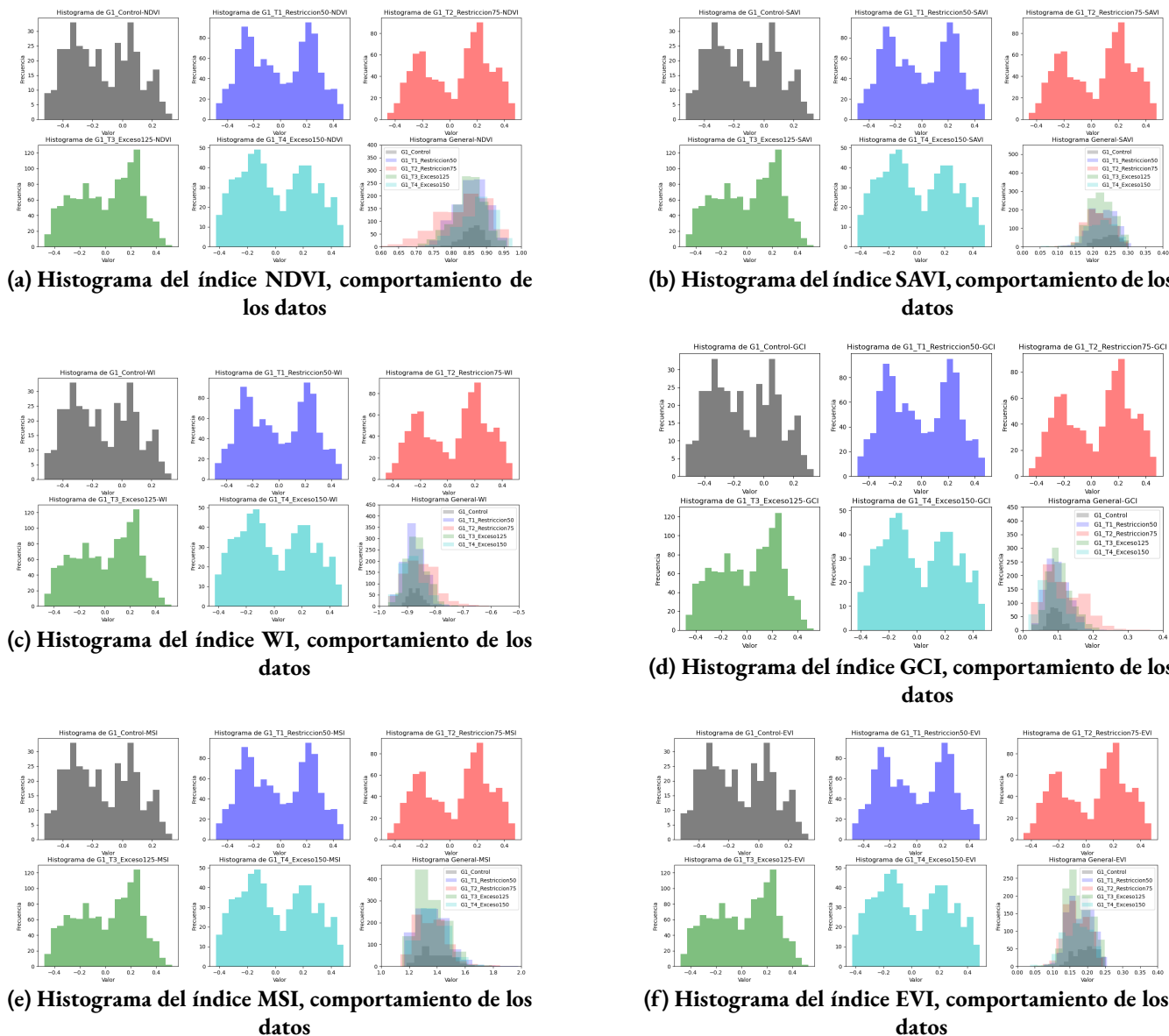
Cálculo de los índices de vegetación (IV)

Los índices de vegetación son características importantes a graficar y calcular, donde se encuentran de manera descrita en la sección 4.6. El observar el comportamiento de los píxeles que se reflejan por la luz que toma los sensores en la captura de las imágenes, el cual las imágenes de todas las figuras se detalla que tienen 6 gráficas, donde se conforma por 5 primeras gráficas que permiten un análisis detallado, ya que son histogramas de cada una de las clases de forma individual, facilitando un análisis de los datos que contiene cada clase al momento de realizar el cálculo del índice por píxel como parámetro de pre procesamiento implementado, es decir, primeramente se toma todos los valores necesarios de las longitudes de onda en el infrarrojo cercano (NIR), banda roja, banda verde, infrarrojo medio (MIR) y banda azul que determine cada índice de vegetación (IV) que son referentes a datos que contienen las bandas que estructuran las imágenes de las base de datos, por consiguiente se realiza el promedio de estas longitudes de ondas (bandas espectrales), donde finalmente se saca el cálculo del índice pixel por pixel, denotando la generación del histograma.

Los histogramas en general es una herramienta que puede ayudar a entender la distribución de los datos, e incluso a identificar algunas diferencias que puedan presentarse entre grupos, ya que esto puede servir para el análisis de diversas aplicaciones que permita un buen procesamiento de los datos a tratar, de forma que se forje un análisis exploratorio significativo para determinar un estudio satisfactorio.

A continuación, se observa que en las 5 primeras gráficas (*Gris = Control*, *Azul = T1*, *Rosa = T2*, *Verde = T3*, *A.celeste = T4*) de cada una de las figuras 13a-13f, se muestra el histograma de cada clase, donde se puede ver que los datos no tienen una distribución normal, ya que no se forma una figura tipo campana simétrica, asimismo los datos muestran un comportamiento no congruente, lo que significa no agruparse alrededor de toda la media generando variabilidad en los valores. Además de determinar estas cualidades, también se distingue diferentes picos máximos que caracterizan diferentes variaciones en la frecuencia de los datos a lo largo de la distribución del eje x, permitiendo una clara tendencia. Los histogramas generales (gráfica multicolor) que se presenta de manera agrupada todas las clases, es claro evidenciar que las clases no tienen una visible diferencia entre las misma, ya que se traslapan y por ende se muestra como pixeles muy similares al momento de variar cualquier tipo de cálculo de se relacione con los diferente IV llegando a características no informativas para una buena clasificación.

Figura 13 – Comportamiento de los datos a nivel de índices de vegetación, grupo G1

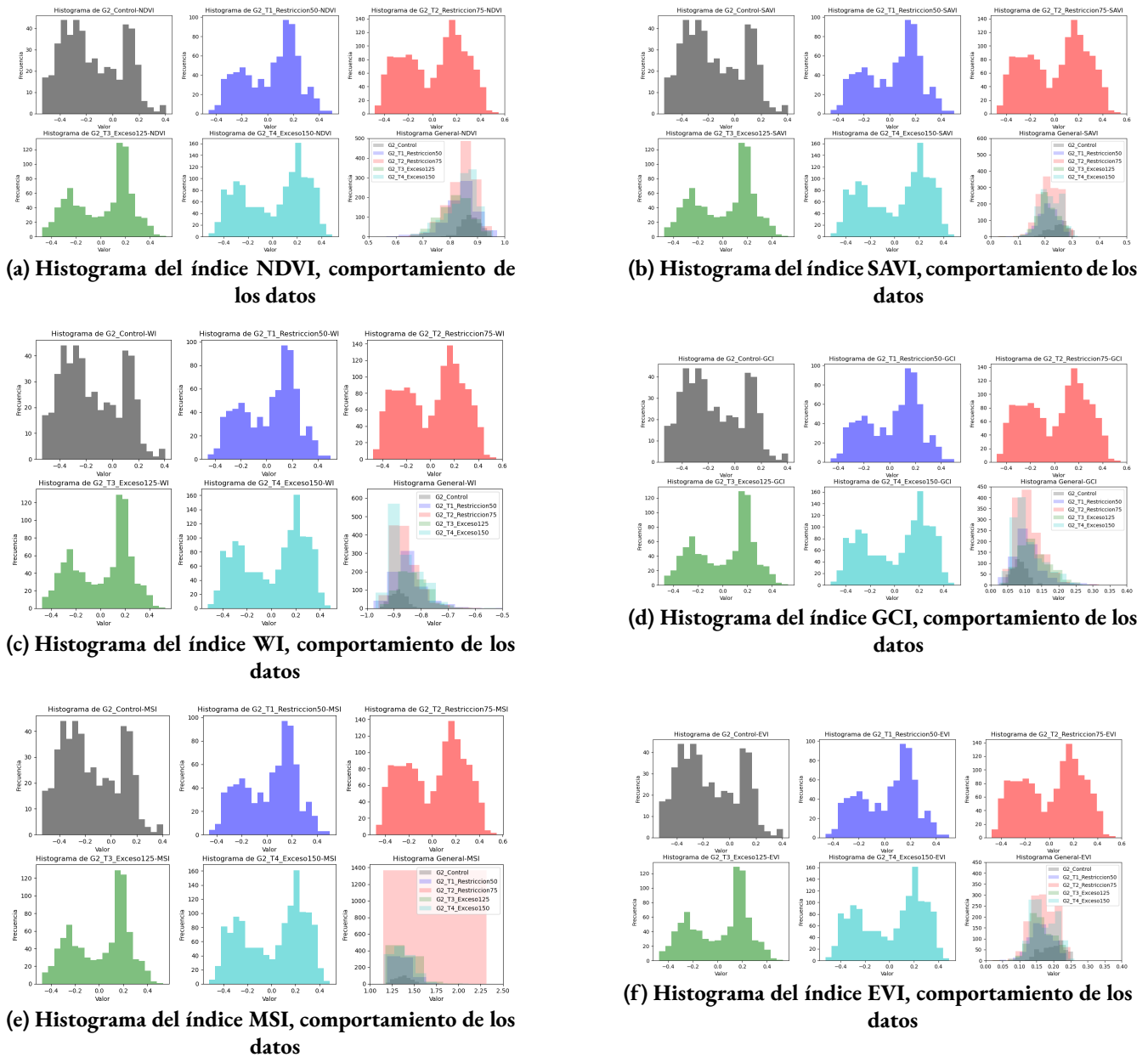


Fuente: Adaptación propia

Además, se destaca un comportamiento muy diferente entre el grupo G1 y G2, ya que al observar las gráficas generales de todas figuras, no se puede detallar una clara tendencia sobre las clases, generando una discrepancia no profunda para determinar que clase predomina sobre el comportamiento de los pixeles que identifica cada clase, siendo una particularidad

resaltante en cada momento del cultivo representado en las imágenes. A continuación, se observaran los comportamientos representativos del grupo *G2* en relación con los índices que se calcula, donde se busca una mejor interpretación de las imágenes y distribución de los datos en las figuras 14b-14f

Figura 14 – Comportamiento de los datos a nivel de índices de vegetación, grupo G2



Fuente: Adaptación propia

5.3. Anova (Análisis de la varianza)

Las siguientes figuras del grupo *G1* determinan el análisis Anova (análisis de varianza), es una técnica estadística que permite diferenciar entre las medias de una o más grupos. A continuación, se evidencia las siguientes interpretaciones:

La Figura 15a, del índice NDVI se evidencia 2 gráficas donde, primeramente se tiene las comparaciones entre cla-

ses y en segundo lugar contiene un diagrama de cajas, vemos que las comparaciones de medias, se observa que las clases se denota una sobre posición, es decir en este caso cuando las clases se sobreponen de acuerdo a las medidas en el rango de la gráfica, conlleva a que las clases tenga igualdad entre medias, por lo tanto no habría una diferencia que las caracterice entre ellas, pero cabe recalcar que no todas las clases presentan una media igual y que por este tipo de característica también haya una posibilidad de que al menos 1 de las clases sea diferente del estudio, en este caso podría ser la clase de $T1_G1$ y $T2_G1$, en la tabla 6, se tiene 2 datos estadísticos, donde La estadística-F es una medida de cuánto varían las medias entre los grupos en comparación con cuánto varían las medias dentro de los grupos y el valor-p es la probabilidad de observar una estadística F tan extrema como la que se ha calculado si la hipótesis nula es cierta, partiendo de si el $Valor - p < 5\%$ (hay por lo menos 1 media que es diferente de las demás - $M1 \neq M2 \neq M3$) y si $Valor - p > 5\%$ (no se puede descartar la igualdad, $M1 = M2 = M3$). La tabla 3, cabe resaltar que los parámetros se siguen teniendo en cuenta los valores de las longitudes de ondas establecidas por la tabla.

En el diagrama de cajas se podrá ver la distribución de los datos, es decir, como se comportan los datos. Se observa unos puntos o asteriscos, el cual se le llama valores atípicos (mayores o menores), ya que son valores que se salen del rango de la media de la clase, demostrando que en la gráfica existe valores atípicos menores que son valores muy alejados de la media y tiende a 0. La media de todas las clases se entiende que se encuentra en rango similar, a comparación de la caja morada ($T4_G1$) quien es la que marca diferencia. Además de otra característica que brinda los diagramas de cajas, son la longitud de las cajas y bigotes (líneas extremas tipo T), determinan que si son largas más disperso son los datos, encontrando que las clases tienen diferentes distribuciones de datos y no son simétricos entre clases.

El índice SAVI de la figura 15b, la comparación de medias es similar, detallando que la clase de control termina siendo diferente de las demás clases, por consiguiente, es visible el diagrama de cajas, donde la caja azul perteneciente de la clase control, se muestra por tener una media que se diferencia de las demás, además también hay presencia de los valores atípicos menores.

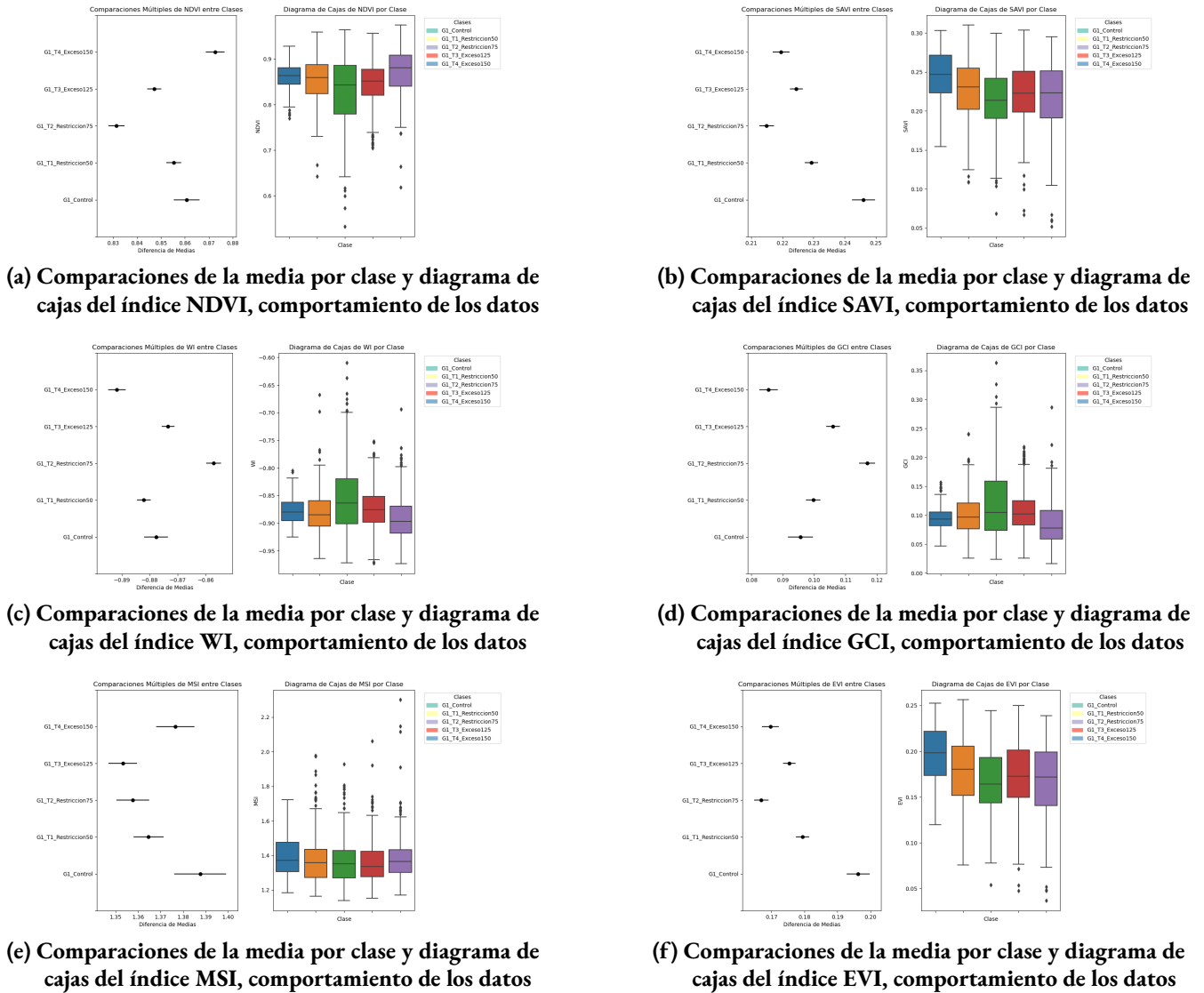
El índice WI, figura 15c y GCI, figura 15d, la comparación de medias es similar, evidenciando que existe 2 posibles medias diferenciales, la clase de $T4_G1$ y $T2_G1$ son las clases que tienden a ser diferentes de las demás clases, que a su vez se puede ver que, en el diagrama de cajas, la caja verde y morada referente a $T2_G1$ y $T4_G1$, se muestra por tener una media que se diferencia de las demás, además también hay presencia de los valores atípicos mayores.

El índice MSI, figura 15e, la comparación de medias se demuestra que las medias son iguales, ya que si se verifica se sobreponen entre sí mismas, el cual, se puede distinguir que, en el diagrama de cajas, todas las clases presentan medias muy similares y distribuciones simétrica, también se caracteriza por tener presencia de valores atípicos mayores en casi todas las clases a diferencia de la de control.

El índice EVI, la figura 15f, la comparación de medias es similar, detallando que la clase de control se caracteriza por ser diferente de las demás clases, por lo tanto, el diagrama de cajas, la caja azul apropiada de la clase control, se muestra por tener una media que se diferencia de las demás, además también hay presencia de los valores atípicos menores.

Continuando con el análisis de Anova, esta sección indica figuras del análisis de la clase $G2$ con los mismos índices, breve mente aquí también se puede ver casos donde persiste la igualdad de las medias siendo aún más marcada la sobreposición de las clases y otros tipos de comportamiento de los datos que se distribuyen en los diagramas de cajas, la figura 16a, es el índice NDVI, figura 16b, SAVI, figura 16c, WI, figura 16d, GCI, figura 16e, MSI y la figura 16f, es EVI, el cual se encuentra la representación, además de la tabla 7, están los valores estadísticos de Anova.

Figura 15 – Comportamiento de los datos a nivel de índices de vegetación aplicando Anova, grupo G1



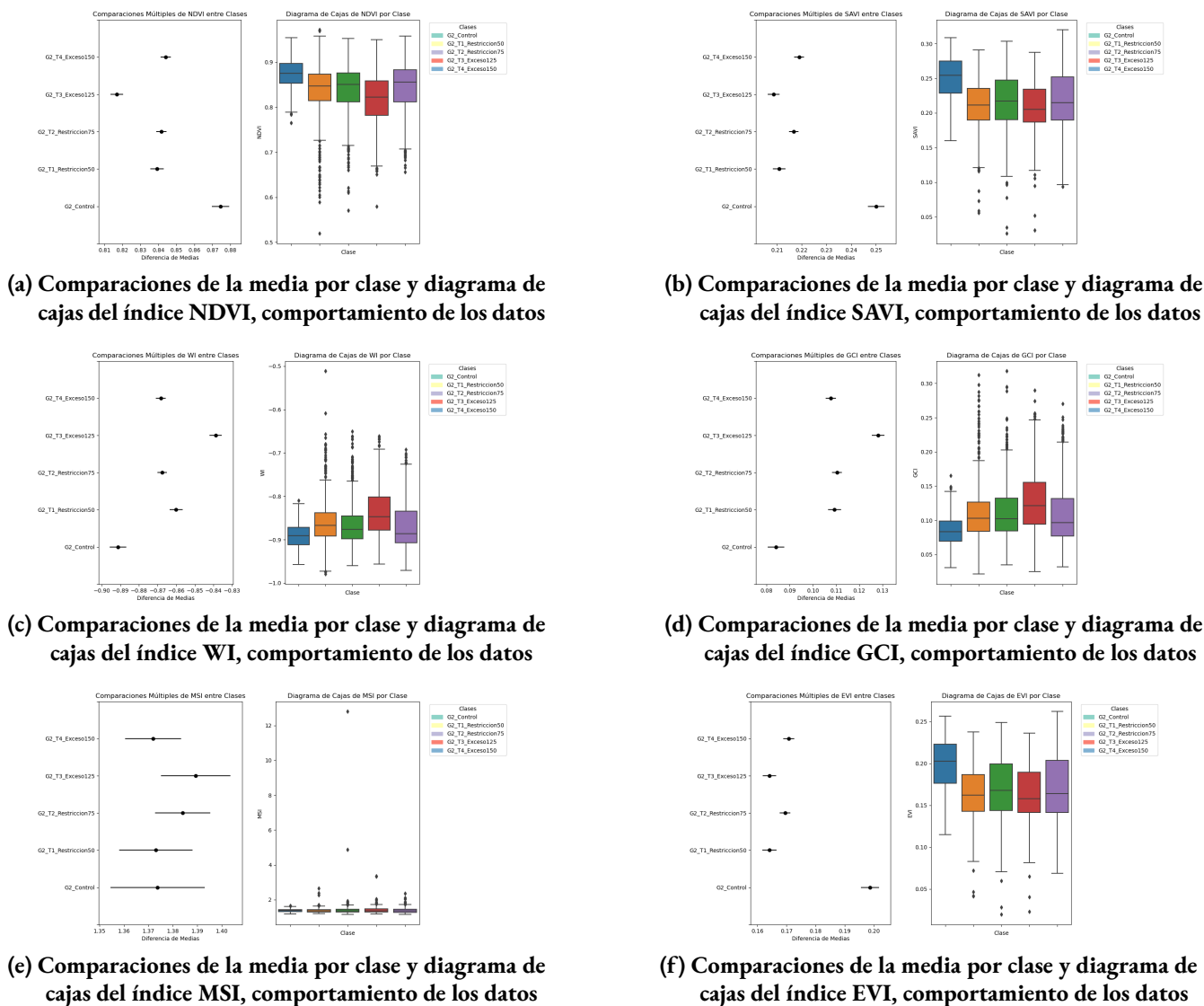
Fuente: Adaptación propia

Tabla 6 – Datos estadístico de Anova respecto a los índices, grupo G1

Índice	Estadística Anova	
	Clase G1	
	Estadística-F	Valor-P
NDVI	65,538	8,76E – 40
SAVI	57,951	1,26E – 31
WI	76,866	6,54E – 49
GCI	68,490	3,59E – 40
MSI	9,802	6,87E + 00
EVI	57,503	2,93E – 32

Fuente: Elaboración propia

Figura 16 – Comportamiento de los datos a nivel de índices de vegetación aplicando Anova, grupo G2



Fuente: Adaptación propia

Tabla 7 – Datos estadístico de Anova respecto a los índices, grupo G2

Índice	Estadística Anova Clase G2	
	Estadística-F	Valor-P
NDVI	92,007	1,86E – 59
SAVI	112,944	4,10E – 76
WI	96,054	1,07E – 62
GCI	93,428	1,35E – 61
MSI	1,275	0,277190
EVI	93,362	1,53E – 60

Fuente: Elaboración propia

5.4. Manova (Análisis multivariante de varianza)

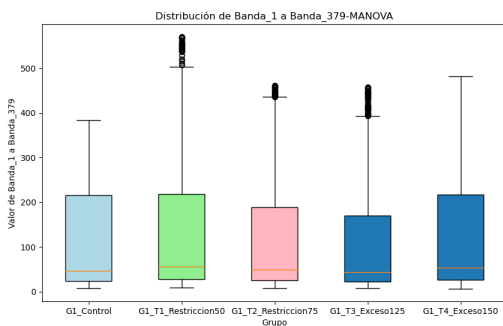
El Análisis multivariante de varianza (MANOVA) proporciona un análisis de regresión y análisis de varianza para variables dependientes múltiples por una o más variables de factor.

Las figuras 17a,17b,17c,17d representadas por gráficas de cajas y histogramas, donde se puede observar que se distribuye en 2 representaciones significativas para el análisis Manova, el cual pertenece al grupo $G1$ y se genera 2 presentaciones donde, se tiene gráficas teniendo en cuenta todas las 379 bandas de las imágenes y como segunda representación se realiza la elección de un rango de bandas que se estima de la banda 80 a la 130, ya que pertenecen al rojo cercano y el infrarrojo cercano, siendo bandas importantes en el campo de las imágenes hiperespectrales para determinar objetos de interés que se necesite identificar en el caso de estudio. Además, para el grupo $G2$ también se ilustra las misma gráficas que permite analizar el comportamiento de los datos a través de las bandas por cada uno de los grupos.

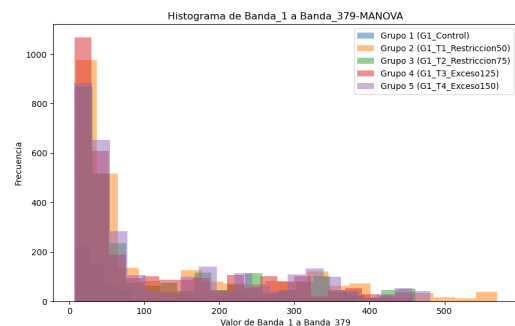
El grupo $G1$, se puede evidenciar que en el diagrama de cajas de la figura 17a, en representación de todas las bandas, es claro que la mediana de todas las clases no tienen una clara diferencia entre las misma, lo que lleva a determinar que no hay una diferencia significativa entre clases, de este modo, se demuestra lo complejo que se distribuyen los datos en las imágenes, ya que la longitud de las cajas son similares pero también en la clase $T2_G1$, $T2_G1$ y $T3_G1$ existen valores atípicos mayores, es decir, valores fuera de lo común del conjunto de datos de las imagen. La figura 17b, del histograma de todas las bandas representativas, no se puede tener una visible información, ya que, es claro el comportamiento de las clases se mantienen de forma conjunta en los mismos valores de frecuencia, demostrando que ninguna de las clases tiene una diferencia significativa, que permita estandarizar una clasificación.

En la selección del rango de bandas [80 – 130] en la figura 17c y 17d, igualmente las medianas y la longitud de cajas tienen un comportamiento similar a las figuras anteriores 17a y 17b, a diferencia de que no existen valores atípicos.

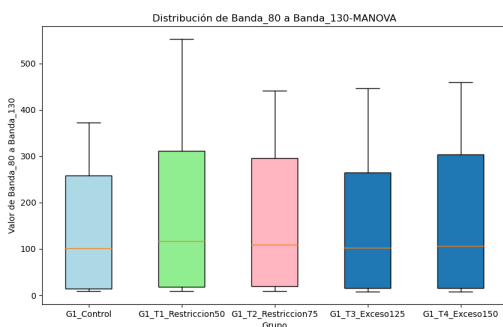
Figura 17 – Comportamiento de los datos a nivel de bandas espectrales aplicando Manova, grupo $G1$



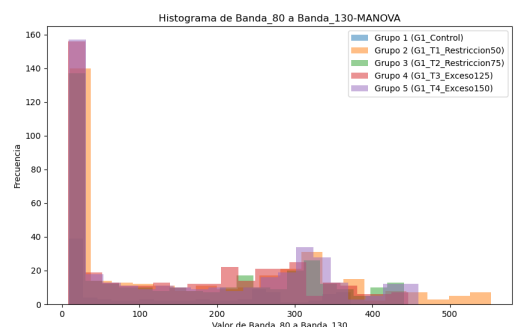
(a) Histograma de las 379 bandas del análisis Manova. Comportamiento de los datos grupo $G1$



(b) Diagrama de cajas de las 379 bandas del análisis Manova. Comportamiento de los datos grupo $G1$



(c) Diagrama de cajas por selección de bandas del análisis Manova. Comportamiento de los datos grupo $G1$



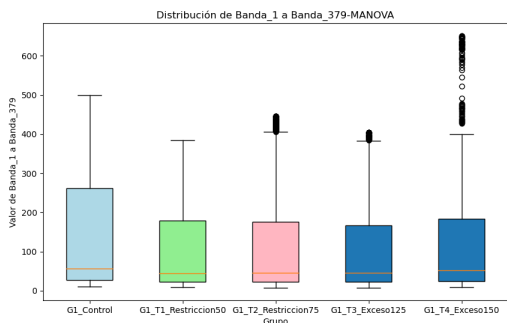
(d) Histograma por selección de bandas del análisis Manova. Comportamiento de los datos grupo $G1$

Fuente: Adaptación propia

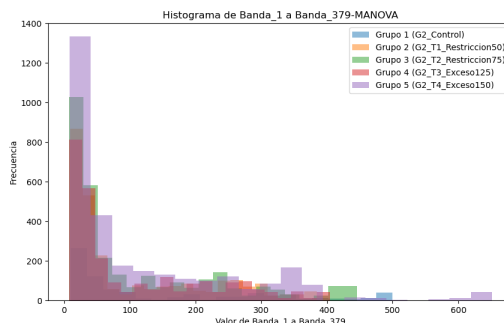
Para el grupo $G2$, no es mucha la diferencia del comportamiento de los datos, ya que muestra similitud al grupo $G1$. A continuación se observa las figuras 18a y 18b referente a todas las bandas y las figuras 18c y 18d, donde se selecciona

las bandas importantes en la banda del rojo cercano y el infrarrojo cercano.

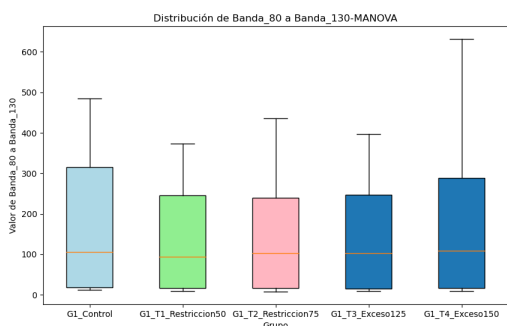
Figura 18 – Comportamiento de los datos a nivel de bandas espectrales aplicando Manova, grupo G2



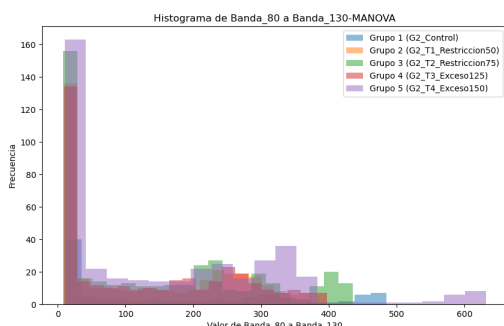
(a) Histograma de las 379 bandas del análisis Manova. Comportamiento de los datos grupo G2



(b) Diagrama de cajas de las 379 bandas del análisis Manova. Comportamiento de los datos grupo G2



(c) Diagrama de cajas por selección de bandas del análisis Manova. Comportamiento de los datos grupo G2



(d) Histograma por selección de bandas del análisis Manova. Comportamiento de los datos grupo G2

Fuente: Adaptación propia

5.5. Resultados de entrenamiento

5.5.1. Grupo G1

A través de la implementación de dos modelos de entrenamiento, las redes completamente conectadas (Feed Forward Neural Networks) y la Red Neuronal Convolutiva (CNN), se presentan los resultados del proceso de entrenamiento, centrándonos específicamente en el primer genotipo de aguacate, denominado *CNN_G1*.

Los resultados derivados del proceso de entrenamiento del modelo *CNN_G1* ponen de manifiesto un rendimiento que no satisface las expectativas en términos de métricas fundamentales, como la precisión (accuracy), la exhaustividad (recall) y la puntuación F1 (F1-score). Es fundamental destacar que, a pesar de este bajo rendimiento global, se observa una situación diferente en lo que respecta a la Clase 2. Esta categoría presenta una precisión del 21,01 %, una exhaustividad del 100 %, y una puntuación F1 del 34,72 % a partir de la segunda época de entrenamiento y en las épocas posteriores. En relación con las otras categorías, todas las métricas arrojan resultados nulos.

Luego de realizar un análisis exhaustivo de las métricas registradas en cada etapa de entrenamiento, se ha identificado que el mejor rendimiento se alcanza en la época 2. Las métricas correspondientes a esta fase se detallan a continuación en la tabla 8.

Tabla 8 – Datos de entrenamiento, implementación CNN grupo G1

Métrica	Resultado
Exactitud (Accuracy)	20,96 %
Precisión	4,4 %
Exhaustividad (Recall)	20,96 %
Puntuación F1 (F1-Score)	7,28 %

Fuente: Elaboración propia

Es relevante destacar que, aunque estas métricas siguen siendo modestas en términos absolutos, representan una ligera mejoría en comparación con los valores registrados en etapas previas.

Los resultados del modelo de las redes completamente conectadas (Feed Forward Neural Networks) para el genotipo G1 destacan su mejor desempeño en la época 19, especialmente en términos del F1-Score. A continuación, se presentan detalladamente las métricas correspondientes a la época 19 de manera visual en la tabla 9

Tabla 9 – Datos de entrenamiento, implementación FF grupo G1

Métrica	Resultado
Exactitud	25,39 %
Precisión (Clase 1)	25,39 %
Exhaustividad (Clase 1)	100 %
Puntuación F1 (Clase 1)	40,50 %
Precisión ponderada	6,45 %
Exhaustividad ponderada	25,39 %
Puntuación F1 ponderada	10,28 %

Fuente: Elaboración propia

5.5.2. Grupo G2

Los resultados de la Red Neuronal Convolutiva (CNN) para el genotipo G2 indican que la elección de la época 2 se respalda en su máximo valor de Exactitud (Accuracy) en comparación con las demás épocas. En consecuencia, en la época 2, y la Clase 4 muestra un mejor rendimiento en términos de F1-Score, aproximadamente con un valor de 43,43 %. Esto implica que la Clase 4 logra un equilibrio superior entre precisión y exhaustividad en contraposición a las demás clases durante este periodo de entrenamiento visto en la tabla 10.

Tabla 10 – Datos de entrenamiento, implementación CNN grupo G2

Métricas	Resultados
Exactitud (Accuracy)	27,32 %
Precisión	42,89 %
Exhaustividad (Recall)	43,99 %
Puntuación F1 (F1-Score)	43,43 %

Fuente: Elaboración propia

A pesar de ello, es relevante destacar que la mejor época de entrenamiento en términos de Exactitud (Accuracy) continúa siendo la época 2, con un valor de 27,32 %.

A continuación, en la tabla 11 se detallan las métricas de Precisión, Recall y F1-Score para la época 2:

Tabla 11 – Datos de entrenamiento, implementación CNN grupo G2

Métricas	Resultados
Precisión	18 %
Exhaustividad (Recall)	25, 89 %
Puntuación F1 (F1-Score)	20, 46 %

Fuente: Elaboración propia

Estos valores representan las cifras más elevadas en comparación con las otras épocas, lo que subraya que la época 2 lidera en términos de rendimiento en estas métricas. Los resultados para la FF en el genotipo G2 mostraron que, considerando las métricas proporcionadas, se hace patente que la época 4 se distingue por encima de las épocas previas. En este momento, la Clase 4 logra una precisión del 27,9 %, un recall del 100 %, y una puntuación F1 de 0,4361, lo que representa el mejor rendimiento documentado hasta este punto para esta categoría.

En resumen, el óptimo desempeño del modelo se materializa en la época 4, dado que en este instante se alcanzan las métricas más elevadas para la Clase 4 en términos de precisión, recall y F1-Score.

Desde una perspectiva global, el rendimiento destacado del modelo se encuentra en la época 2, donde se obtienen las métricas más sobresalientes. Las métricas presentadas se centran específicamente en la Clase 4, y es fundamental destacar que las métricas correspondientes a las demás categorías permanecen en cero, lo que refleja un rendimiento considerablemente deficiente en estas clases. Las métricas de la tabla 12 son las siguientes:

Tabla 12 – Datos de entrenamiento, implementación FF grupo G2

Métricas	Resultados
Precisión	18 %
Exhaustividad (Recall)	25, 89 %
Puntuación F1 (F1-Score)	20, 46 %

Fuente: Elaboración propia

6. CONCLUSIONES

En el presente trabajo se abordó el problema de clasificación de los niveles de estrés en el cultivo de aguacate. Se evalúa distintos métodos que permiten el procesamiento de los datos representativos de las imágenes hiperespectrales que conforman la base de datos del cultivo de aguacate.

Inicialmente se realiza la evaluación del sistema de clasificación donde se está empleando una base de datos ampliamente conocida, lo cual permitió implementar y dar validez a la metodología de clasificación a utilizar en el problema abordado.

Posteriormente, al emplear la base de datos del cultivo de aguacate, se exploran diferentes estrategias que ayudan a identificar posibles características que pueden aportar y a su vez ser empleadas en un sistema de clasificación. De este modo, diferentes métodos como la partición de la base de datos en información del vecindario de los píxeles e información localizada que permitió el manejo de las imágenes sin ninguna pérdida de información. También fue empleado el método que permite analizar características (píxeles) puntuales de las imágenes de las diferentes clases, donde el sistema de clasificación

pudo trabajar de forma factible.

De acuerdo con los resultados, y contrario a lo esperado, al momento de analizar las curvas de reflectancia espectral se observa que hay una mayor variabilidad en las bandas correspondientes al rojo e infrarrojo, pero existe un alto traslape entre las diferentes clases y, la variabilidad presente en dichas bandas no parece indicar que se pueda establecer claras diferencias en el problema abordado, ya que, los análisis para las imágenes hiperespectrales donde se evidencia en cada una de las gráficas, por cada uno de los índices de vegetación es claro las tendencias de los pixeles referente a cada clase, el cual determina que son muy similares y no encontrar diferencias significativas.

Para confirmar lo dicho anteriormente, se aplica una serie de análisis estadísticos como es Anova y Manova sobre los índices que típicamente recomienda la literatura. Al realizar el análisis sobre los diferentes índices evaluados se observa que hay, por lo menos cálculos estadísticos que demuestran que al menos alguna de las clases tuviera una diferencia significativa respecto a las demás clases, pero no suficiente para generar una buena clasificación. Ya que en las gráficas de cada una de estos métodos estadísticos, siempre se observó que la distribución y comportamiento de los datos no fueron concluyentes para caracterizar una ventaja en la mejora del rendimiento de los modelos empleados. De este modo, esto indica que por lo menos en este experimento, no es factible evaluar el nivel de estrés en cultivos de aguacate empleando imágenes HSI.

Finalmente se contextualiza por los resultados obtenidos por el diseño CNN (Red neuronal convolucional) para los grupos $G1$ y $G2$, dado que en cuestión de métricas del modelo tiene valores nulos para la mayoría de las categorías, provocando que el modelo no logra aprender de manera efectiva la clasificación durante el proceso de entrenamiento, a pesar de que se realizaron cambios de parámetros que exige la red neuronal, concluyendo un rendimiento deficiente, sin embargo, se observa que la métrica para una de las clases tiene un mínimo progreso, siendo perteneciente a niveles bajos para la no clasificación.

Además, el sistema basado en redes completamente conectadas (Feed Forward Neural Networks) en los grupos $G1$ y $G2$, es importante tener en cuenta que, a pesar de realizar otra arquitectura diferente al CNN y de tener diferentes características de análisis para el uso de imágenes, se determina que no es posible ultimar resultados satisfactorios y concluyentes siendo no óptimo para evaluar el nivel de estrés en cultivos de aguacates, ya que presenta porcentajes de precisión bajos para determinar un resultado de buen desempeño.

La evaluación de la base de datos SALINAS que contiene diferentes clases y que en su mayoría se denota por tener varios entornos y diferentes tipos de vegetación, permite demostrar que el algoritmo usado por este estudio, lo caracteriza por ser funcional de manera óptima en la tarea de clasificación de las HSI, dando soporte de cada una de las medidas que se tuvieron en cuenta para el estudio y afirmar la viabilidad de este, es decir, se realizó la prueba en otro tipo de entorno que estuviera ligado a las imágenes hiperespectrales. Esta estrategia permite llegar a niveles de acierto del 90 % a medida que se aumenta la cantidad de muestras, el cual, aumenta mas información al sistema de clasificación.

Referencias

- Ahmad, U., Alvino, A., and Marino, S. (2021). A review of crop water stress assessment using remote sensing. *Remote Sensing*, 13(20):4155.
- Arias, F., Zambrano, M., Broce, K., Medina, C., Pacheco, H., and Nunez, Y. (2021). Hyperspectral imaging for rice cultivation: Applications, methods and challenges. *AIMS Agriculture and Food*, 6(1):273–307.
- Asaari, M. S. M., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., and Scheunders, P. (2019). Analysis of hyperspectral images

- for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Computers and Electronics in Agriculture*, 162:749–758.
- Asaari, M. S. M., Mishra, P., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., and Scheunders, P. (2018). Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform. *ISPRS journal of photogrammetry and remote sensing*, 138:121–138.
- Carmona Zuluaga, P. (2023). Detección de estrés en cultivos agrícolas con datos espectrales utilizando métodos de selección de bandas.
- Criollo Saavedra, A. C. and Merino Arellano, T. A. (2023). Espectroscopía con imágenes hiperespectrales y detección de bandas principales para estimar parámetros de control de calidad en la harina de pescado con modelos de machine learning.
- De Oliveira, A. (2019). *Abiotic and Biotic Stress in Plants*. BoD–Books on Demand.
- Díaz Ramírez, L., García Botina, M. J., Jäger, M., and Hurtado, J. J. (2021). Plan de investigación y desarrollo de la cadena productiva del aguacate hass en el valle del cauca a partir de sus principales brechas tecnológicas.
- Genangeli, A., Avola, G., Bindi, M., Cantini, C., Cellini, F., Grillo, S., Petrozza, A., Riggi, E., Ruggiero, A., Summerer, S., et al. (2023). Low-cost hyperspectral imaging to detect drought stress in high-throughput phenotyping. *Plants*, 12(8):1730.
- Godoy, F., Olivos-Hernández, K., Stange, C., and Handford, M. (2021). Abiotic stress in crop species: improving tolerance by applying plant metabolites. *Plants*, 10(2):186.
- Hernández, R. R. (2021). La agricultura de precisión. una necesidad actual. *Ingeniería Agrícola*, 11(1).
- IBM (2021). Análisis multivariante de varianza (MANOVA) - Documentación de IBM.
- Jia, S., Jiang, S., Lin, Z., Li, N., Xu, M., and Yu, S. (2021). A survey: Deep learning for hyperspectral image classification with few labeled samples. *Neurocomputing*, 448:179–204.
- Jia, Y., Shi, Y., Luo, J., and Sun, H. (2023). Y-net: Identification of typical diseases of corn leaves using a 3d–2d hybrid cnn model combined with a hyperspectral image band selection module. *Sensors*, 23(3):1494.
- Kang, X., Li, S., Fang, L., Li, M., and Benediktsson, J. A. (2014). Extended random walker-based classification of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1):144–153.
- Li, Q., Gao, M., and Li, Z.-L. (2022). Ground hyper-spectral remote-sensing monitoring of wheat water stress during different growing stages. *Agronomy*, 12(10):2267.
- Loggenberg, K., Strever, A., Greyling, B., and Poona, N. (2018). Modelling water stress in a shiraz vineyard using hyperspectral imaging and machine learning. *Remote Sensing*, 10(2):202.
- Lowe, A., Harrison, N., and French, A. P. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant methods*, 13(1):80.
- Muñoz, P. S. R. and Burdett, E. (2019). Evolución de la cobertura forestal en los alcornocales próximos al estrecho de gibraltar a través del índice de vegetación evi. *Ecosistemas*, 28(3):73–80.
- Nachappa, T. G., Ghorbanzadeh, O., Gholamnia, K., and Blaschke, T. (2020). Multi-hazard exposure mapping using machine learning for the state of salzburg, austria. *Remote Sensing*, 12(17):2757.
- OCI (2022). ¿Qué es el aprendizaje automático? | Oracle Colombia.
- Paoletti, M. E., Haut, J. M., Plaza, J., and Plaza, A. (2019). Estudio comparativo de técnicas de clasificación de imágenes hiperespectrales. *Revista Iberoamericana de Automática e Informática Industrial*, 16(2):129–137.
- Pérez, J. L. et al. (2021). Imágenes hiperespectrales y sus aplicaciones en estudios de suelos, cultivos y bosques, en la era de la cuarta revolución industrial. *UD y la geomática*, (16).

Quigley, E. (2023). Qué es el aprendizaje combinado.

Quintulén, V., Andrés, S., and PIDERIT MORENO, M. B. (2018). Reflectancia de las envolventes verticales y su influencia sobre disponibilidad de luz natural en el cañón urbano de la ciudad de concepción. *Revista hábitat sustentable*, 8(1):6–15.

Quispe Almontes, J. G. (2021). Eficiencia de los índices de vegetación espectral para identificar procesos de escorrentías dominantes, en la quebrada arapato, distrito de tamburco, provincia de abancay, región de apurímac.

Rencher, A. C. and Schimek, M. (1997). Methods of multivariate analysis. *Computational Statistics*, 12(4):422–422.

Sánchez Bernabé, E. (2016). Procesado de imágenes hiperespectrales.

Sowmya, V., Soman, K., and Hassaballah, M. (2019). Hyperspectral image: Fundamentals and advances. *Recent Advances in Computer Vision: Theories and Applications*, pages 401–424.

Sun, C., Zhou, J., Ma, Y., Xu, Y., Pan, B., and Zhang, Z. (2022). A review of remote sensing for potato traits characterization in precision agriculture. *Frontiers in Plant Science*, 13:871859.

Sun, J., Yang, W., Zhang, M., Feng, M., Xiao, L., and Ding, G. (2021). Estimation of water content in corn leaves using hyperspectral data based on fractional order savitzky-golay derivation coupled with wavelength selection. *Computers and Electronics in Agriculture*, 182:105989.

Vijith, H. and Dodge-Wan, D. (2020). Applicability of modis land cover and enhanced vegetation index (evi) for the assessment of spatial and temporal changes in strength of vegetation in tropical rainforest region of borneo. *Remote Sensing Applications: Society and Environment*, 18:100311.

Villatoro, B., Barrera, C., and Tagua, K. (2021). Uso De Índice De Vegetación Ndvi Como Herramienta Para Realizar. pages 313–331.

Weng, S., Ma, J., Tao, W., Tan, Y., Pan, M., Zhang, Z., Huang, L., Zheng, L., and Zhao, J. (2023). Drought stress identification of tomato plant using multi-features of hyperspectral imaging and subsample fusion. *Frontiers in Plant Science*, 14:1073530.

Yang, G., Huang, K., Sun, W., Meng, X., Mao, D., and Ge, Y. (2022). Enhanced mangrove vegetation index based on hyperspectral images for mapping mangrove. *ISPRS Journal of Photogrammetry and Remote Sensing*, 189:236–254.