

CENTRO DE INVESTIGACIÓN EN
CIENCIAS DE INFORMACIÓN GEOESPACIAL, A.C.

CentroGeo

Centro Público de Investigación SECIHTI

Monitoreo espaciotemporal del grado de afectación
causado por Huanglongbing en cultivos de naranjo
(*Citrus sinensis*)

TESIS

Que para obtener el grado de
Maestra en Ciencias de Información Geoespacial

Presenta

Brenda Soto Rivera

Directora de Tesis

Dra. Daniela Alejandra Moctezuma Ochoa

CENTRO DE INVESTIGACIÓN EN
CIENCIAS DE INFORMACIÓN GEOESPACIAL, A.C.

CentroGeo

Centro Público de Investigación SECIHTI

Monitoreo espaciotemporal del grado de afectación
causado por Huanglongbing en cultivos de naranjo
(*Citrus sinensis*)

TESIS

Que para obtener el grado de
Maestra en Ciencias de Información Geoespacial

Presenta

Brenda Soto Rivera

Directora de Tesis

Dra. Daniela Alejandra Moctezuma Ochoa

Sinodales

Dra. Nohemí Álvarez Jarquín

Dra. Daniela Juanita López Araujo

Dr. Hugo Carlos Martínez

Ciudad de México, marzo 2025

Resumen

El monitoreo efectivo del Huanglongbing (HLB) en cultivos de naranjo es esencial para apoyar la toma de decisiones informadas en el sector citrícola. Este proyecto se centra en el uso de tecnologías de teledetección, como los satélites Sentinel-2 y los Vehículos Aéreos No Tripulados (VANTs), para recopilar datos espaciotemporales que permitan una clasificación precisa del grado de afectación por HLB. La investigación integra el algoritmo de aprendizaje automático K-Vecinos más cercanos (KNN), y el modelo estadístico autorregresivo integrado de media móvil estacional (SARIMA), para clasificar y pronosticar la evolución de la enfermedad en el tiempo y el espacio. Se desarrolló un sistema de monitoreo y pronóstico basado en series de tiempo del índice de vegetación de diferencia normalizada (NDVI) y ortomosaicos de alta resolución generados a partir de imágenes de VANTs, facilitando la identificación de áreas afectadas.

El estudio se lleva a cabo en una huerta del municipio de Ciudad Fernández, San Luis Potosí, donde la prevalencia de HLB ha afectado significativamente los cultivos de naranjo. La hipótesis central de la investigación plantea que la combinación de datos geoespaciales, aprendizaje automático y modelos estadísticos mejora la precisión y eficiencia del monitoreo de HLB a gran escala. Los resultados muestran que el modelo KNN alcanzó una precisión global del 86.08 % en la clasificación del grado de afectación, mientras que el modelo SARIMA $(1, 0, 1) \times (1, 1, 0, 12)$ demostró un rendimiento adecuado en el pronóstico de la evolución temporal del HLB, con proyecciones para los próximos 12 meses.

Además de los aspectos técnicos, este enfoque innovador resalta la importancia de la accesibilidad y el uso de software libre para facilitar su implementación y replicabilidad en otras regiones productoras de cítricos. La integración de estas herramientas representa un avance significativo frente a los métodos tradicionales, ofreciendo un sistema más eficiente y accesible para la detección y gestión del HLB en cultivos de naranjo, con el potencial de contribuir a la creación de estrategias de intervención más efectivas y sostenibles.

Palabras clave: Huanglongbing, Teledetección, NDVI, Aprendizaje automático, Modelos estadísticos.

Abstract

Effective monitoring of Huanglongbing (HLB) in orange orchards is essential for supporting informed decision-making in the citrus sector. This project focuses on the use of remote sensing technologies, such as Sentinel-2 satellites and Unmanned Aerial Vehicles (UAVs), to gather spatiotemporal data that enables accurate classification of HLB severity. The research integrates the K-Nearest Neighbors (KNN) machine learning algorithm and the Seasonal Autoregressive Integrated Moving Average (SARIMA) statistical model, to classify and forecast the disease's progression over time and space. A monitoring and forecasting system based on time series of the Normalized Difference Vegetation Index (NDVI) and high-resolution UAV-generated orthomosaics was developed, facilitating the identification of affected areas.

The study was conducted in an orchard located in Ciudad Fernández, San Luis Potosí, where HLB prevalence has significantly impacted orange crops. The central hypothesis of the research suggests that combining geospatial data, machine learning, and statistical models enhances the accuracy and efficiency of large-scale HLB monitoring. The results show that the KNN model achieved an overall accuracy of 86.08% in classifying HLB severity, while the SARIMA (1, 0, 1) x (1, 1, 0, 12) model demonstrated satisfactory performance in forecasting the temporal evolution of HLB, with projections for the next 12 months.

In addition to the technical aspects, this innovative approach emphasizes the importance of accessibility and the use of open-source software to facilitate its implementation and replicability in other citrus-producing regions. The integration of these tools represents a significant advancement over traditional methods, offering a more efficient and accessible system for detecting and managing HLB in orange orchards, with the potential to contribute to the development of more effective and sustainable intervention strategies.

Key words: Huanglongbing, Remote Sensing, NDVI, Machine Learning, Statistical Models.

"Proteger las plantas, proteger la vida"
(Food and Agriculture Organization [FAO], [2021](#)).

"I love you sweet leaf, though you can't hear"
(Black Sabbath, [1971](#))

Agradecimientos

A CONACYT, posteriormente CONAHCYT y, actualmente, SECIHTI, por la oportunidad de acceder a una Beca Nacional de Posgrado con número 1199333 del programa de Maestría en Ciencias de Información Geoespacial en el Centro de Investigación en Ciencias de Información Geoespacial con sede en la Ciudad de México, desde el 1 de septiembre de 2022 hasta el 31 de agosto de 2024. Agradezco profundamente el respaldo y la confianza brindada, que me permitieron enriquecer mis conocimientos y avanzar en mi formación académica, personal y profesional.

Quiero agradecer a la vida y a la interconexión subyacente en el plano de existencia. Por no permitirme renunciar, por continuar incluso cuando perdí a quien fue mi guía en este proceso; la crisis personal y emocional pusieron a prueba mi determinación. Por elegir reconstruirme, enfrentando la desesperanza con trabajo, la confusión con disciplina y la soledad con fuerza interior. Agradezco por no detenerme, aunque el vacío y la incertidumbre pesaran sobre mí. Por transformar el dolor en fuerza y la pérdida en aprendizaje. Porque aún en la oscuridad, seguí caminando, sabiendo que la luz no es un destino, sino el proceso mismo hacia la iluminación.

A mi familia. Gracias por ser mi apoyo incondicional, por entender los días en los que estuve aislada y ausente, inmersa en este proceso, por creer en mí y brindarme paciencia y aliento. Sé que también les cuesta expresar emociones, y que nuestra forma de demostrar afecto puede ser poco común, pero eso no hace menos genuino el cariño, el respeto y la admiración que siento por ustedes. Valoro cada uno de los sacrificios y esfuerzos que han hecho. Esto es tan suyo como mío.

Quiero dedicar un agradecimiento especial al Dr. Jesús Antonio Sosa Herrera, quien, aunque ya no esté físicamente, dejó una huella imborrable. Más allá de haber sido mi director de tesis y profesor en la maestría, fue un mentor excepcional y una fuente inagotable de inspiración. Su dedicación, sabiduría y apoyo incondicional fueron fundamentales en mi formación durante esta etapa. El legado que dejó, tanto en el ámbito académico como en la vida de quienes aprendimos de él, perdurará por siempre.

A la Dra. Daniela Alejandra Moctezuma Ochoa, quien fungió como mi directora de tesis y miembro de mi comité supervisor, por su paciencia, guía y compromiso inquebrantable. Su orientación constante y activa fueron fundamentales, brindándome el impulso necesario para concluir. Su experiencia y aportes fueron clave para el enriquecimiento de este trabajo.

Al Dr. Hugo Carlos Martínez, quien además de haber sido parte de mi comité supervisor brindando invaluable comentarios de mejora, también formó parte de mi jurado de tesis, junto con la Dra. Nohemí Álvarez Jarquín y la Dra. Daniela Juanita López Araujo, a quienes también agradezco profundamente por sus observaciones y retroalimentación.

Al Dr. Moisés Roberto Vallejo Pérez, Coordinador para la Innovación y Aplicación de la Ciencia y Tecnología de la Universidad Autónoma de San Luis Potosí, por su disposición y compromiso para establecer el contacto con el Comité Estatal de Sanidad Vegetal de San Luis Potosí (CESAVESLP), lo que permitió acceder a recursos y conocimientos clave para la investigación. Al Lic. Felipe Monsiváis Pinal, presidente del CESAVESLP, y al Ing. José René Moreno López, Coordinador de la Campaña de Plagas Reglamentadas de los Cítricos, por su invaluable apoyo y por todas las facilidades brindadas durante las actividades de campo realizadas.

A la Dra. María Elena Méndez López y la Mtra. Guadalupe Borja César, por su apoyo y orientación constante en este proceso. Valoro profundamente su guía, acompañamiento y disposición. A las profesoras y los profesores que influyeron positivamente durante esta y otras etapas, por compartir su conocimiento, motivarme a cuestionar, y sembrar en mí la pasión por el aprendizaje.

Por último, pero no menos importante, a la familia Rodríguez Benavidez, por apoyarme durante mi estancia en la Ciudad de México, por abrirme las puertas de su hogar y hacerme un espacio en sus vidas. Su generosidad, hospitalidad y cariño fueron un refugio en los momentos de vulnerabilidad, brindándome tranquilidad cuando todo era caos. Les estaré eternamente agradecida.

Índice General

1	Introducción	1
1.1	Planteamiento del problema	2
1.2	Hipótesis	3
1.3	Objetivos	4
1.3.1	General	4
1.3.2	Específicos	4
1.4	Justificación de la investigación	4
1.5	Alcances y limitaciones	5
2	Estado del Arte	7
2.1	Series de Tiempo en Agricultura	7
2.1.1	Detección, Clasificación y Pronóstico del HLB	8
3	Marco Teórico	19
3.1	Huanglongbing en cultivos de naranjo (<i>Citrus sinensis</i>)	19
3.1.1	Detección histórica y distribución geográfica	20
3.1.2	Agente causal y síntomas	21
3.1.3	Impacto socioeconómico de las pérdidas en la industria citrícola	23
3.2	Zona de estudio	24
3.3	Tecnologías de Monitoreo Agrícola	25
3.3.1	Series de tiempo de Sentinel-2	25
3.3.2	Detección de afectaciones en cultivos mediante NDVI	27
3.3.3	Vehículo Aéreo No Tripulado (VANT)	29
3.3.4	Fotogrametría digital	29
3.4	Modelo Estadístico	31

ÍNDICE GENERAL

3.4.1	Modelo Autorregresivo Integrado de Media Móvil Estacional (SARIMA)	32
3.4.2	Métricas para evaluar el modelo	35
3.5	Aprendizaje Automático	37
3.5.1	K-Vecinos más cercanos (KNN)	39
3.5.2	Validación Cruzada	40
3.5.3	Matriz de Confusión	42
4	Metodología	45
4.1	Flujo de proceso con imágenes de VANT	47
4.1.1	Obtención de datos: Levantamiento fotogramétrico	48
4.1.2	Fotogrametría y generación de ortomosaico	49
4.2	Flujo de proceso con datos Sentinel-2	51
4.2.1	Obtención de datos: Series de Tiempo NDVI	51
4.2.2	Limpieza de ruido y normalización	52
4.2.3	Identificación de clases	55
4.2.4	Implementación y Evaluación	60
4.2.4.1	Algoritmo KNN	61
4.2.4.2	Modelo SARIMA	62
5	Resultados y Discusión	66
5.1	Resultados de la clasificación con KNN	66
5.2	Resultados del pronóstico de valores futuros con SARIMA	71
5.3	Integración y Superposición	75
5.4	Discusión de los hallazgos	79
6	Conclusiones	81
6.1	Contribución de la investigación	81
6.2	Trabajo a futuro	82
	Bibliografía	84

Índice de Figuras

Figura 1	Interacción entre hospedero, patógeno, vector y ambiente para la incidencia de HLB Fuente: Elaboración propia con base en Santivañez <i>et al.</i> (2023).	21
Figura 2	Síntomas de HLB observados en hojas de árbol de naranjo (<i>Citrus sinensis</i>) Fuente: Lomellí <i>et al.</i> (2010).	22
Figura 3	Ubicación geográfica de la huerta Fuente: Elaboración propia.	24
Figura 4	Porcentajes NIR y R de vegetación sana y enferma Fuente: Elaboración propia con base en Cherlinka (2024).	27
Figura 5	Ortorrectificación de una fotografía aérea Fuente: GeoBC (2024).	31
Figura 6	Notación SARIMA Fuente: Elaboración propia.	33
Figura 7	Tipos de aprendizaje automático Fuente: Elaboración propia con base en Chauhan y Henrietta (2023) y L. Chen <i>et al.</i> (2019).	38
Figura 8	Ejemplo para nuevos datos antes y después de KNN Fuente: Elaboración propia con base en Kufel <i>et al.</i> (2023).	39
Figura 9	Ejemplo de validación cruzada Fuente: Elaboración propia.	40
Figura 10	Ejemplo de matriz de confusión Fuente: Elaboración propia.	42
Figura 11	Diagrama del proceso metodológico Fuente: Elaboración propia.	46
Figura 12	Árboles de naranjo con afecciones por HLB captados a 30 metros de altura Fuente: Elaboración propia.	48
Figura 13	Ortomosaico generado en ODM con imágenes capturadas a 30 metros con VANT Fuente: Elaboración propia.	50
Figura 14	Serie de tiempo original de un solo píxel antes de la limpieza Fuente: Elaboración propia.	53
Figura 15	Filtrado de picos y umbrales mínimos en la serie de tiempo de un solo píxel Fuente: Elaboración propia.	54

ÍNDICE DE FIGURAS

Figura 16	Serie de tiempo final de un solo píxel posterior a la limpieza Fuente: Elaboración propia.	55
Figura 17	Visualización del ortomosaico y los puntos reproyectados Fuente: Elaboración propia.	56
Figura 18	Identificación de clases en base a Distribución NDVI Fuente: Elaboración propia.	58
Figura 19	Árbol gravemente afectado por HLB capturado a 30 metros Fuen- te: Elaboración propia.	59
Figura 20	Árbol moderadamente afectado por HLB capturado a 30 metros Fuente: Elaboración propia.	59
Figura 21	Árbol levemente afectado por HLB capturado a 30 metros Fuente: Elaboración propia.	60
Figura 22	Precisión Media y Desviación Estándar de KNN para Diferentes Valores de k Fuente: Elaboración propia.	67
Figura 23	Matriz de Confusión por clase Fuente: Elaboración propia. . . .	69
Figura 24	Clasificación con KNN: Reales vs Predichos Fuente: Elaboración propia.	70
Figura 25	Gráficos de Diagnóstico del Modelo SARIMA para la Serie de Tiempo NDVI Fuente: Elaboración propia.	73
Figura 26	Datos Reales, Pronósticos SARIMA para Validación y Proyección Futura Fuente: Elaboración propia.	74
Figura 27	Grado de afectación por HLB en cultivos de naranjo, febrero 2023 y 2024 Fuente: Elaboración propia.	78

Índice de tablas

1	Trabajos relacionados	14
2	Interpretación de los valores NDVI	28
3	Comparación de Modelos SARIMA	63
4	Significancia Estadística en el Modelo SARIMA	64
5	Métricas de evaluación por clase	68
6	Incremento y decremento del NDVI en 2023, 2024 y pronóstico para 2025	76

Capítulo 1

Introducción

La citricultura es fundamental tanto para la alimentación como para el desarrollo económico global, destacando a México como uno de los principales productores (SENASICA, 2022). Sin embargo, esta industria enfrenta una amenaza crítica: el Huanglongbing (HLB), una enfermedad bacteriana que afecta severamente los cultivos de cítricos. Su rápida propagación, frecuentemente antes de ser detectada, representa un desafío complejo para su manejo.

A nivel mundial, el HLB ha impactado a 39 países, incluyendo 23 estados de México, que concentran una proporción significativa de la producción citrícola nacional (Sandoval & Avila, 2019). La dificultad para identificar la enfermedad en etapas tempranas y su alta capacidad de diseminación ponen en riesgo la sostenibilidad de esta industria.

El monitoreo efectivo de los cultivos afectados por HLB es crucial para controlar y mitigar sus impactos. Las tecnologías de teledetección, como los satélites Sentinel-2 y los Vehículos Aéreos No Tripulados (VANTs), ofrecen una excelente capacidad para recolectar datos espaciotemporales, permitiendo un análisis detallado de la evolución de la enfermedad. Para interpretar estos datos y facilitar la identificación de patrones de propagación, es necesario emplear métodos avanzados de análisis, como el aprendizaje automático y los modelos estadísticos.

Este proyecto pretende integrar estas tecnologías y métodos para optimizar el monitoreo del HLB en cultivos de naranjo, permitiendo un análisis preciso de su dinámica temporal y espacial.

1.1. Planteamiento del problema

En México, el HLB ha afectado aproximadamente 307,805 hectáreas, lo que representa el 50% de la superficie cultivada con cítricos, según el director general de Sanidad Vegetal, Francisco Ramírez y Ramírez. Esta situación resalta la urgente necesidad de desarrollar métodos más eficientes y accesibles para la detección de la enfermedad, ya que los métodos convencionales, como la reacción en cadena de la polimerasa en tiempo real (qRT-PCR), aunque precisos, presentan desventajas importantes en términos de costo y tiempo, especialmente en grandes extensiones de terreno (Garza-Saldaña *et al.*, 2017). Esto hace necesario explorar nuevas tecnologías que permitan mitigar el impacto de esta enfermedad de manera más eficaz y accesible.

Como señala la Organización de las Naciones Unidas para la Agricultura y la Alimentación (*Food and Agriculture Organization*, FAO), es crucial adoptar un “Enfoque Regional para la gestión del HLB”, que implique una serie de acciones integrales para fortalecer la investigación y las estrategias de lucha contra esta enfermedad (Könnöla *et al.*, 2023). En el Estado de San Luis Potosí, se han registrado 3,833 hectáreas afectadas por HLB, especialmente en municipios como Ciudad Fernández, un área clave para la producción de naranja que representa el 15.4% de la cosecha estatal (CESAVESLP, 2022).

Dado que no existe un tratamiento efectivo para combatir esta enfermedad, es crucial identificarla en sus etapas iniciales, posibilitando la eliminación de los cultivos infectados, ayudando a reducir y controlar la propagación del HLB tanto en los huertos como en la región en su conjunto (Pourreza *et al.*, 2015).

En este contexto, las tecnologías basadas en teledetección y sensores remotos, como los satélites y los VANTs, han demostrado ser herramientas eficaces para el monitoreo a gran escala. Estas tecnologías permiten recopilar datos espaciotemporales y observar características específicas del cultivo y su estado fenológico, facilitando la identificación del grado de afección por el HLB. Sin embargo, persisten desafíos como el acceso limitado a datos de calidad, las barreras tecnológicas y la falta de capacitación en el uso de estas tecnologías.

Por lo tanto, esta investigación busca abordar la problemática del monitoreo del HLB a gran escala utilizando datos de Sentinel-2, análisis mediante el uso de datos provenientes del satélite Sentinel-2, combinados con análisis basados en modelos estadísticos y algoritmos de aprendizaje automático. Lo anterior, con el propósito de clasificar, monitorear y pronosticar la evolución del grado de afectación de los cultivos de naranjo en una huerta ubicada en el municipio de Ciudad Fernández, San Luis Potosí.

Se espera que este enfoque contribuya a la creación de estrategias de intervención más eficientes. Además, se utilizarán imágenes de alta resolución capturadas mediante VANTs para evaluar el monitoreo, permitiendo un análisis más detallado del estado de los cultivos afectados, y proporcionando una mejor comprensión de los patrones espaciales de la enfermedad.

Las preguntas que guían esta investigación son:

1. ¿Qué información se obtiene al clasificar y pronosticar el grado de afectación por HLB en cultivos de naranjo y qué tan satisfactorios son los resultados, en términos de precisión y utilidad práctica, al integrar datos geoespaciales, algoritmos de aprendizaje automático y modelos estadísticos?
2. ¿Cómo se puede mejorar la precisión y eficiencia en la clasificación y pronóstico del grado de afectación por HLB en los cultivos de naranjo en la zona de estudio?

1.2. Hipótesis

El uso de datos geoespaciales obtenidos a partir de sensores remotos, combinado con aprendizaje automático y modelos estadísticos, mejora la clasificación y pronóstico del grado de afectación por HLB en cultivos de naranjo a gran escala. Esta integración aporta eficientemente al monitoreo espaciotemporal para la gestión de la enfermedad.

1.3. Objetivos

1.3.1. General

Desarrollar un sistema de monitoreo y pronóstico espaciotemporal del grado de afectación causado por Huanglongbing en cultivos de naranjo, que apoye a la toma de decisiones informadas. Utilizando series de tiempo de Sentinel-2, imágenes capturadas con VANTs, e integrando técnicas de aprendizaje automático y modelos estadísticos.

1.3.2. Específicos

De lo anterior, se proponen los siguientes objetivos específicos:

- **Capturar** datos en campo y **generar** un ortomosaico de alta resolución como base para el análisis espacial y la identificación de áreas afectadas por HLB.
- **Implementar** un algoritmo de aprendizaje automático para clasificar los niveles de afectación por HLB, así como un modelo estadístico basado en series de tiempo que utilice los valores del índice de vegetación de diferencia normalizada para pronosticar la evolución futura de la enfermedad.
- **Evaluar** la precisión de la clasificación mediante la superposición visual de los datos sobre el ortomosaico generado.

1.4. Justificación de la investigación

La identificación de características asociadas con el HLB en cultivos de naranjo, utilizando aprendizaje automático para clasificar el grado de afectación y modelos estadísticos para pronosticar su evolución, es fundamental para detectar patrones de propagación a gran escala. La integración de datos geospaciales de plataformas como los satélites Sentinel-2 y los VANTs mejora significativamente el monitoreo espaciotemporal, permitiendo una evaluación detallada de la extensión y severidad de la enfermedad a lo largo del tiempo y el espacio.

En comparación con métodos tradicionales como la inspección manual o el uso de sensores convencionales, que son costosos y demandan mucho tiempo y recursos, el empleo de tecnologías avanzadas optimiza la precisión y eficiencia en la detección de afectaciones causadas por HLB.

Además, la integración de diversas fuentes de datos y técnicas facilita una toma de decisiones más informada, lo que es crucial para implementar estrategias de intervención más efectivas.

1.5. Alcances y limitaciones

Este trabajo tiene como objetivo la integración de diversas herramientas y métodos para mejorar el monitoreo, la clasificación y el pronóstico de la salud de los cultivos, con un enfoque particular en la identificación temprana de áreas afectadas por HLB.

Es fundamental reflexionar sobre los alcances y limitaciones de este enfoque, que no solo busca generar resultados significativos, sino también resaltar la importancia de una integración y calibración adecuadas de cada componente para garantizar una implementación efectiva. Aunque se destacan los beneficios de estas tecnologías, también se enfatiza la necesidad de un proceso continuo de validación y ajuste para asegurar su eficacia.

Alcances

- **Detección a gran escala:** La combinación de tecnologías de percepción remota y aprendizaje automático facilita la identificación y clasificación precisa del grado de afectación por HLB.
- **Uso de software libre:** Se emplean herramientas de código abierto para promover la accesibilidad y la colaboración.
- **Monitoreo continuo:** Integración de datos satelitales con modelos estadísticos para el pronóstico temporal, junto con imágenes capturadas por VANTs para el monitoreo espacial.

- Optimización del tiempo: Mejora en los procesos de recolección, clasificación y pronóstico, superando las limitaciones de los métodos tradicionales.

Limitaciones

- Disponibilidad y calidad de datos: La calidad y accesibilidad de los datos pueden verse afectadas por factores externos, como la disponibilidad de imágenes o restricciones de acceso a ciertas áreas.
- Desafíos técnicos: Se requiere equipo especializado para realizar vuelos con VANTs. Además, los modelos sugeridos demandan grandes volúmenes de datos y capacidad de procesamiento computacional.
- Factores ambientales: Las condiciones climáticas pueden influir en la calidad de los datos recolectados.

Este enfoque, aunque prometedor, exige una implementación cuidadosa y adaptable, teniendo en cuenta tanto los avances tecnológicos como las limitaciones externas inherentes y los recursos disponibles. Es fundamental que el proceso sea flexible y se ajuste de manera continua a las necesidades del monitoreo y a las condiciones cambiantes de los cultivos.

Finalmente, este capítulo ha establecido el contexto y la relevancia del estudio, presentando el planteamiento del problema y los objetivos que guiarán la investigación. También se ha expuesto la justificación del estudio, así como los alcances y limitaciones del trabajo. Estos elementos proporcionan un marco claro para entender la problemática y los enfoques que se adoptarán para abordarla.

A partir de esta base, en el siguiente capítulo se revisará el contexto académico y científico previo relacionado con el tema, proporcionando una base sólida para el desarrollo de la investigación.

Capítulo 2

Estado del Arte

El HLB, reconocido como una de las enfermedades más devastadoras para los cítricos, ha generado un creciente interés en el desarrollo de metodologías avanzadas para su monitoreo, clasificación y pronóstico.

Este capítulo presenta una revisión del estado del arte, destacando las principales contribuciones científicas y avances tecnológicos aplicados en el sector agrícola. Se analiza el papel de estas herramientas y técnicas en la gestión sostenible de cultivos afectados, con un enfoque particular en su utilidad para enfrentar los desafíos asociados con la enfermedad.

2.1. Series de Tiempo en Agricultura

El estudio de series de tiempo es esencial en la investigación agrícola, proporcionando herramientas valiosas para modelar tendencias, variaciones estacionales y gestionar riesgos agrícolas (Sathya & Karthiban, 2021). Además, permite pronosticar parámetros meteorológicos que influyen significativamente en las actividades agrícolas (Sodha & Saha, 2016).

Las técnicas de aprendizaje automático se pueden integrar con datos históricos para crear modelos predictivos robustos sobre el rendimiento de cultivos. Estos modelos ayudan a los agricultores a tomar decisiones informadas sobre siembra, cosecha y asignación de recursos, y también apoyan a los responsables de formular políticas en el

diseño de estrategias agrícolas efectivas (Chaudhary *et al.*, 2023).

2.1.1. Detección, Clasificación y Pronóstico del HLB

La necesidad urgente de abordar los problemas causados por el HLB en la citricultura requiere estrategias de intervención sostenibles, asequibles y rápidas. Implementar metodologías avanzadas es crucial para mejorar el rendimiento y la gestión de recursos.

En este contexto, el aprendizaje automático juega un papel esencial en la clasificación de la enfermedad, mientras que los modelos estadísticos, son especialmente útiles para capturar patrones temporales, modelar tendencias y realizar proyecciones cuantitativas a partir de datos históricos. Diversas investigaciones sobre enfermedades en cítricos subrayan la importancia de desarrollar métodos efectivos de detección temprana y estrategias de control.

Algunos enfoques han sido propuestos para mejorar la detección y clasificación del HLB, con un enfoque inicial en el procesamiento de imágenes. Garza *et al.* (2020), utilizaron imágenes RGB capturadas por VANT y el Índice de Verdor Triangular (TGI) para evaluar la salud de los árboles de cítricos. Su enfoque correlacionó el TGI con mediciones de campo como el estado nutricional de los árboles, encontrando que el 61 % de las variaciones del TGI podían explicarse por factores como el sodio y el hierro.

Aunque este método demostró ser útil y de bajo costo, los autores sugirieron incluir factores adicionales en estudios futuros. Así, este enfoque resalta el potencial de las tecnologías de imágenes accesibles en el monitoreo de la salud de los cítricos. Este enfoque, aunque económico y eficiente, presenta limitaciones al depender de datos adicionales y sugiere la necesidad de explorar nuevos factores que puedan mejorar el monitoreo de los cítricos.

La utilización de tecnologías más avanzadas, como las imágenes multiespectrales y los sensores hiperspectrales, ha mostrado avances significativos en la detección de HLB. A. Mishra *et al.* (2011), evaluaron el uso de un sensor óptico activo multibanda y técnicas de imágenes hiperspectrales para detectar HLB en cítricos, alcanzando precisiones superiores al 95 % en la clasificación de árboles sanos e infectados.

Este estudio subraya el potencial de los sensores multibanda e hiperespectrales para una detección temprana del HLB, aunque señala limitaciones en la aplicación a diferentes cultivos y condiciones ambientales, sugiriendo la necesidad de explorar otros tipos de deficiencias y enfermedades, abriendo la puerta a la incorporación de modelos más robustos y adaptativos, como los que integran datos geoespaciales de satélites.

En un enfoque similar, Garcia-Ruiz *et al.* (2013) investigaron el uso de imágenes aéreas multiespectrales de alta resolución para detectar HLB en cítricos, utilizando un VANT con una cámara multiespectral de 6 bandas. Sus resultados indicaron que las imágenes recolectadas y su aplicación con máquinas de soporte vectorial (*Support Vector Machines*, SVM), ofrecieron una precisión del 85 % en la detección, pero también enfrentaron limitaciones como el corto tiempo de vuelo del VANT.

Este trabajo ilustra el avance hacia el uso de VANTs con sensores multiespectrales, pero también resalta la necesidad de mejorar la plataforma y los algoritmos, así como la implementación de enfoques de clasificación adaptativos para cubrir áreas más grandes y abordar la variabilidad en los datos geoespaciales, como los índices de vegetación derivados de satélites.

Una evolución en el uso de imágenes para la detección del HLB se refleja en el trabajo de Deng *et al.* (2016), quienes introdujeron un método que combina el procesamiento de imágenes en el espectro visible con SVM. Utilizando imágenes RGB capturadas bajo luz natural, lograron clasificar hojas como infectadas o sanas con una precisión del 91.93 %.

Este avance destaca la importancia de combinar datos visuales y técnicas de aprendizaje automático para la detección temprana de enfermedades. Sin embargo, se identificó la necesidad de incluir más variedades de cítricos y etapas de la enfermedad para mejorar la generalización de los modelos. Pone de manifiesto la efectividad de técnicas basadas en imágenes, aunque está limitado al análisis a nivel de hoja y no integra información temporal o espacial.

A su vez Sharif *et al.* (2018), propusieron un método híbrido para la detección y clasificación de enfermedades cítricas mediante técnicas de procesamiento de imágenes. Utilizando un enfoque basado en características de textura, color y geometría, y cla-

sificadores de máquina de vectores de soporte multiclase (*Multiclass Support Vector Machines*, M-SVM), lograron precisiones de hasta el 97% en bases de datos específicas.

Este enfoque resulta prometedor para mejorar la identificación de enfermedades como melanosis y HLB, pero no aborda explícitamente la integración de datos geoespaciales o el pronóstico de la tendencia temporal de estas enfermedades, lo cual podría complementar la metodología utilizada.

El uso de algoritmos de aprendizaje automático, como SVM y redes neuronales, ha sido clave en la mejora de la precisión de los métodos de detección. Lan *et al.* (2020), investigaron el uso de imágenes multiespectrales capturadas por VANT para detectar HLB, logrando precisiones de hasta 100% con AdaBoost y 97.28% con redes neuronales. Este avance destaca el potencial de las técnicas de aprendizaje automático en la detección de enfermedades, aunque los autores identificaron limitaciones en la calidad de las imágenes y sugirieron la optimización de los algoritmos.

Además, plantearon la necesidad de explorar imágenes hiperespectrales y datos térmicos para una detección más robusta. En el contexto actual, se podría fortalecer este enfoque mediante la incorporación de datos geoespaciales a nivel satelital, así como modelos estadísticos que permitan no sólo detección sino también el análisis temporal y el pronóstico de la evolución del HLB en los cultivos.

En paralelo, Stegmayer *et al.*, (2013), propusieron un modelo que combina la selección de características con un clasificador perceptrón multicapa (MLP), entrenado para identificar enfermedades cuarentenarias en cítricos. Utilizando 212 muestras de mandarinas Nova de la región del Río Uruguay, el modelo seleccionó solo 14 características relevantes de un total de 90, logrando una tasa de clasificación superior al 83%.

Este enfoque optimiza el diagnóstico visual y los procesos de control y exportación, demostrando cómo la combinación de selección de características con técnicas de aprendizaje automático mejora la detección de enfermedades en cítricos. A pesar de que este estudio no aborda directamente el HLB, el uso de técnicas similares para la identificación de enfermedades en cítricos resulta crucial para la metodología propuesta en el presente trabajo.

En un enfoque más amplio sobre Inteligencia Artificial (IA), Bu *et al.* (2020), realizaron una revisión exhaustiva de diferentes métodos para la identificación de HLB, observando mejoras significativas desde 2016, con técnicas avanzadas como redes neuronales profundas (*Deep Neural Network*, DNN) alcanzando precisiones cercanas al 100%.

Esta revisión resalta cómo los avances en satélites de alta resolución espectral, cámaras de mayor definición y algoritmos de IA han revolucionado el monitoreo y la alerta temprana del HLB. No obstante, no se menciona la integración de datos geoespaciales como los derivados de satélites ni el pronóstico de tendencias temporales.

Esta evolución hacia técnicas más accesibles y menos costosas se refleja en el trabajo de Awe y Dias (2022), quienes utilizaron Redes Neuronales Artificiales (RNA) para pronosticar la producción agrícola, que varía entre 18,000 y 23,000 billones de Naira (moneda de Nigeria), comparando su precisión con la del modelo autorregresivo integrado de media móvil (*AutoRegressive Integrated Moving Average*, ARIMA) tradicional. La raíz del error cuadrático medio (RMSE) de la RNA fue de 416.70, inferior al 437.97 de ARIMA. Estos errores representan aproximadamente un 2-2.5% del total, lo que indica un margen de error bajo. Aunque la diferencia en los RMSE es pequeña, la mejora en la precisión del modelo RNA puede ser significativa a largo plazo.

Este método demuestra el potencial de las técnicas de aprendizaje profundo para el pronóstico agrícola y su relevancia para el monitoreo espaciotemporal de enfermedades como el HLB, aunque sugiere explorar redes más avanzadas y variables adicionales, lo que podría incluir datos geoespaciales obtenidos mediante sensores remotos para un análisis más detallado de la afectación de los cultivos a lo largo del tiempo.

De igual manera Barman y Choudhury (2022), desarrollaron un sistema basado en aprendizaje profundo y K-Vecinos más cercanos (*K-Nearest Neighbors*, KNN) para la detección de enfermedades cítricas utilizando imágenes capturadas con smartphones en condiciones reales. Observaron que un modelo DNN alcanzó una precisión del 99.89%, siendo más efectivo que KNN (89.9%).

Este sistema permite a los agricultores identificar enfermedades tempranas como Citrus Tristeza Virus (CTV), pero en este caso, no se utiliza información temporal ni geoespacial, aspectos que podrían enriquecer la precisión del diagnóstico al incorporar la evolución espacial y temporal de la enfermedad.

Por su parte Khanramaki *et al.* (2021), presentaron un clasificador ensamblado basado en redes neuronales convolucionales (*Convolutional Neural Networks*, CNNs) para identificar tres plagas comunes de cítricos. Su modelo alcanzó una precisión del 99.04 % mediante técnicas de aumento de datos y validación cruzada.

Este sistema es clave para reducir el uso de pesticidas al mejorar la detección y manejo temprano de las plagas. Aunque su enfoque se centra en plagas y no específicamente en HLB, los resultados obtenidos destacan la eficacia de las CNNs en problemas relacionados con cultivos cítricos.

En paralelo, Khattak *et al.* (2021), implementaron un modelo CNN para clasificar hojas y frutos cítricos sanos o afectados por enfermedades como mancha negra, cancro y costra. Lograron una precisión del 94.55 % al usar conjuntos de datos como *Citrus* y *PlantVillage*, destacando el potencial de estas tecnologías para asistencia a agricultores. Este enfoque está alineado con el presente estudio en cuanto a la utilización de aprendizaje automático para identificar enfermedades en cítricos, aunque no aborda explícitamente el uso de índices de vegetación o el pronóstico temporal.

Mientras que las técnicas de imágenes y aprendizaje automático se centran en la detección y clasificación de enfermedades, los modelos estadísticos también juegan un papel crucial en el entendimiento de la propagación del HLB. Gottwald *et al.* (2010), estudiaron la epidemiología del HLB en Florida mediante un análisis espacial, revelando una relación entre árboles infectados y destacando la importancia de los modelos estocásticos para pronosticar la propagación de la enfermedad. Analizaron métodos de integración de Monte Carlo por Cadenas de Markov (MCMC).

Este análisis permitió visualizar la propagación de la enfermedad y comprender mejor las interacciones locales. Subraya la necesidad de un modelo regional para el control del HLB, y es relevante para el presente estudio que busca pronosticar la evolución temporal de la enfermedad en los cultivos de naranjo.

Aunque los estudios previos se enfocan principalmente en cítricos, algunos trabajos han explorado aplicaciones de técnicas similares en otros cultivos. Por ejemplo, Choudhary *et al.* (2022), es relevante para este trabajo debido al uso de algoritmos de aprendizaje automático en el monitoreo agrícola.

Los autores desarrollaron un modelo que utiliza datos de Sentinel-2 y Google Earth Engine para mapear y predecir el rendimiento del arroz, alcanzando precisiones superiores al 85 % en la identificación de cultivos. Este trabajo es particularmente relevante, ya que demuestra cómo la combinación de datos satelitales y algoritmos de aprendizaje automático puede mejorar significativamente las estimaciones agrícolas.

En una línea similar, Novković *et al.* (2022), analizaron la producción de trigo y maíz en Serbia entre 2005 y 2020 utilizando modelos ARIMA y extrapolación para pronosticar la producción futura. Sus resultados proyectaron incrementos del 15.54 % en la producción de trigo y del 28.41 % en maíz en comparación con los promedios históricos, resaltando la utilidad de los modelos estadísticos para evaluar tendencias agrícolas.

Aunque su enfoque no incluye datos geoespaciales ni la identificación de afectación de cultivos, demuestra la aplicabilidad de los modelos ARIMA para proyectar tendencias temporales, una metodología que se adapta al análisis de la evolución temporal del Huanglongbing (HLB) en cultivos de naranjo en este estudio.

De igual forma, Ajaykumar *et al.* (2022), aplicaron modelos ARIMA y la tendencia lineal de Holt-Winters para pronosticar la productividad de legumbres en Tamil Nadu, India, mostrando cómo integrar factores externos, como el clima y las políticas agrícolas, puede enriquecer las proyecciones.

Este trabajo no incorpora el uso de datos geoespaciales ni la clasificación del grado de afectación por enfermedades como HLB, pero ofrece un enfoque útil sobre el uso de modelos estadísticos para el pronóstico de variables agrícolas en función de series de tiempo, lo que aporta al marco metodológico de este estudio para pronosticar la tendencia temporal de la enfermedad.

Aunque los estudios presentados no emplean necesariamente la metodología propuesta en este proyecto, aportan un contexto valioso al abordar problemas similares relacionados con la detección, clasificación y pronóstico futuro de enfermedades en cultivos

agrícolas, incluyendo cítricos.

Este panorama ofrece un punto de referencia para comprender cómo diversas técnicas han sido aplicadas en el manejo agrícola, sirviendo como base para justificar y comparar la relevancia de la metodología desarrollada en este proyecto.

A continuación, se presenta la Tabla 1, que resume las investigaciones relevantes en las áreas de Percepción Remota, Aprendizaje Automático y/o Modelos Estadísticos, aplicadas a la detección, clasificación o pronóstico del HLB u otras enfermedades asociadas a los cítricos y otros cultivos.

Esta tabla incluye la categoría a la cual pertenece el estudio, autor(es), los métodos utilizados, el tipo de dato empleado y los resultados obtenidos, con el objetivo de proporcionar una visión clara de las principales contribuciones.

Tabla 1: Trabajos relacionados

Categoría	Autor(es)	Método	Datos	Resultado
Percepción remota	Garza et al. (2020)	Índice de ver- dor triangular (TGI)	Imágenes RGB VANT mediciones de campo	Correlación del TGI de y nal, explicando el 61 % de la variación.
	Mishra et al. (2009)	Técnicas hiperespec- trales	Sensores mul- tibanda, imá- genes hiperes- pectrales	Clasificación de árbo- les sanos e infectados con más del 95 % de precisión.
Percepción remota y Aprendizaje Automático	Deng et al. (2016)	Procesamiento de imágenes en el espectro visible y SVM	Imágenes RGB	Clasificación de hojas sanas o infectadas con 91.93 % de precisión.

Continúa en la siguiente página

Categoría	Autor(es)	Método	Datos	Resultado
	Choudhary et al. (2022)	Random Forest, Sentinel-2 y Google Earth Engine	Datos satelitales	Precisión superior al 85 % en el pronóstico de cultivos de arroz con Random Forest.
	García Ruiz et al. (2013)	SVM	Imágenes multiespectrales	Precisión de 85 % en detección de HLB.
	Lan et al. (2020)	AdaBoost y RNA	Imágenes multiespectrales	Precisión del 100 % con AdaBoost y 97.28 % con redes neuronales en detección de HLB.
Aprendizaje Automático	Sharif-M et al. (2018)	Multiclase SVM	Imágenes de enfermedades cítricas con características de textura, color y geometría	Precisión del 97 % en la clasificación de enfermedades cítricas.
	Barman-U et al. (2022)	DNN y KNN	Imágenes de capturas con smartphones	DNN alcanzó precisión del 99.89 %, superior a KNN (89.9 %), en la detección de enfermedades como CTV.

Continúa en la siguiente página

Categoría	Autor(es)	Método	Datos	Resultado
	Khanramaki-M et al. (2021)	CNN y aumento de datos	Imágenes de cítricos	Precisión de 99.04 % en clasificación de plagas comunes de cítricos mediante técnicas de aumento de datos.
	Khattak-A et al. (2021)	CNN para clasificación de hojas y frutos	Imágenes de cítricos	Precisión del 94.55 % en clasificación de hojas y frutos afectados por enfermedades cítricas.
Aprendizaje Automático	Stegmayer-G et al. (2013)	Selección de características con MLP	Muestras de mandarinas	Precisión superior al 83 % en la identificación de enfermedades cuarentenarias mediante técnicas de clasificación.
	Bu et al. (2020)	Revisión de métodos de identificación de HLB	Datos de diversas fuentes	Resalta el uso de técnicas avanzadas como Deep CNN, alcanzando precisiones cercanas al 100 %.
Aprendizaje Automático y Modelos Estadísticos	Awe et al. (2022)	RNA y ARI-MA	Datos históricos y series de tiempo	RNA logró RMSE de 416.70, mejor que el ARIMA con 437.97, en pronósticos agrícolas.

Continúa en la siguiente página

Categoría	Autor(es)	Método	Datos	Resultado
	Novković et al. (2022)	ARIMA	Datos de producción de cultivos	Proyección de incremento en la producción de trigo y maíz en Serbia, con un aumento del 15.54% para trigo y 28.41% para maíz.
Modelos Estadísticos	Ajaykumar et al. (2022)	ARIMA y Holt-Winters	Datos de producción agrícola	Pronóstico de la productividad de legumbres en Tamil Nadu, destacando un aumento constante con fluctuaciones mínimas influenciadas por el clima.
	Gottwald et al. (2010)	Análisis espacial y MCMC	Datos espaciales	Pronóstico de la propagación del HLB en Florida, utilizando Monte Carlo y cadenas de Markov.

La revisión de la literatura sobre la detección y monitoreo del HLB resalta avances significativos con enfoques estadísticos y de aprendizaje automático. La incorporación de VANTs y sensores multiespectrales ha demostrado mejorar notablemente la precisión y eficiencia en la identificación de esta enfermedad en cultivos.

Sin embargo, se observó que muchas investigaciones emplean métodos, técnicas y herramientas de forma aislada, lo que puede limitar su potencial para abordar la complejidad del problema de manera integral.

Aunque algunos estudios reportan precisiones extremadamente altas (cercasas al 99 % o superiores), al aplicar algoritmos de aprendizaje automático, es crucial realizar un análisis crítico de estos resultados desde una perspectiva técnica y teórica.

En particular, deben evaluarse aspectos como la metodología utilizada, la calidad y representatividad de los datos empleados, y las implicaciones del sobreajuste en los modelos. Esto es especialmente importante en un contexto práctico, como la detección de enfermedades en campo, donde los modelos deben ser capaces de generalizar y ofrecer predicciones robustas frente a datos desconocidos y condiciones adversas. El sobreajuste podría llevar a decisiones erróneas en la gestión agrícola, afectando negativamente la confiabilidad y utilidad de la tecnología en aplicaciones reales.

Por otra parte, este proyecto integra lecciones de estos trabajos y avanza en varios aspectos clave. A diferencia de estudios previos, incorpora datos geoespaciales del satélite Sentinel-2 y el índice de vegetación de diferencia normalizada, para identificar y clasificar el grado de afectación por HLB en cultivos de naranjo, utilizando técnicas de aprendizaje automático.

Asimismo, propone el pronóstico de tendencias temporales mediante un modelo ARI-MA estacional, abordando una dimensión temporal que ha sido limitada en investigaciones previas. Esto permitirá identificar patrones de afectación a lo largo del tiempo.

Para validar los resultados, en este trabajo se realizará una superposición sobre el ortomosaico generado a partir de imágenes de alta resolución obtenidas mediante VANTs. Este enfoque facilitará la integración de los métodos utilizados, mejorando la evaluación del monitoreo espaciotemporal de la enfermedad a gran escala.

Al concluir este capítulo, se ha presentado una visión actualizada de los avances y tendencias más relevantes en relación con la problemática. La información recopilada ofrece una amplia perspectiva de los métodos más empleados y sus aplicaciones prácticas en el ámbito agrícola. En el siguiente capítulo, se profundizará en el marco teórico que sustenta las metodologías utilizadas en este trabajo.

Capítulo 3

Marco Teórico

El monitoreo espaciotemporal del grado de afectación causado por HLB en cultivos de naranjo (*Citrus sinensis*) es fundamental para gestionar y mitigar esta devastadora enfermedad. Este capítulo proporciona el marco teórico que sustenta el uso de tecnologías y metodologías para detectar, clasificar y pronosticar. Se abordan desde los aspectos biológicos de la enfermedad hasta las metodologías de procesamiento y análisis de datos.

En este apartado se definen los conceptos clave relacionados con el HLB y se exploran las herramientas tecnológicas, como la percepción remota y procesos derivados. Además, se examina el uso de índices de vegetación, modelos estadísticos para el pronóstico de la evolución de la enfermedad mediante series de tiempo y algoritmos de aprendizaje automático para la clasificación de los grados de afectación.

3.1. Huanglongbing en cultivos de naranjo (*Citrus sinensis*)

El Huanglongbing, también conocido como “Enfermedad de los brotes amarillos”, “Enverdecimiento de los Cítricos”, “Dragón Amarillo” o “Citrus Greening”, es una enfermedad devastadora que afecta a los cultivos de cítricos. Causada por bacterias del género *Candidatus Liberibacter*, estas se alojan en el floema de las plantas, interrumpiendo sus funciones vitales (SENASICA, [2019](#)).

El HLB representa una amenaza significativa para la industria cítrica mundial, afectando diversas regiones de Asia, África y América, y poniendo en riesgo la producción cítrica.

Los métodos tradicionales de detección de HLB, como la inspección visual, aunque comunes, son lentos y propensos a errores humanos, lo que limita su efectividad (A. R. Mishra *et al.*, 2012). En comparación, técnicas más avanzadas, como la espectroscopía en el visible e infrarrojo cercano (Vis-NIR) y la reacción en cadena de la polimerasa (PCR), ofrecen una mayor precisión en la identificación de la enfermedad, pero suelen ser más costosas y requieren más tiempo para obtener resultados (Stover *et al.*, 2014).

3.1.1. Detección histórica y distribución geográfica

El origen de la enfermedad se remonta al continente asiático, donde fue reportada por primera vez en China a finales del siglo XIX, siendo denominada *Candidatus Liberibacter Asiaticus*; a finales de la década de 1920, se detectó otra especie en Sudáfrica conocida como *Candidatus Liberibacter Africanus* (Zhao, 1982), para 2004 en América, se reportó una nueva especie en São Paulo, Brasil asociada a *Candidatus Liberibacter Americanus*.

De acuerdo con el informe de la Organización de las Naciones Unidas para la Agricultura y la Alimentación, sólo la primera de estas especies ha reportado su expansión en diferentes años hacia otros países del continente americano, incluyendo Estados Unidos (2005); Cuba (2006); República Dominicana (2008); México, Honduras, Belice (2009); Guatemala, Nicaragua (2010); Costa Rica y Jamaica (2011) (Santivañez *et al.*, 2013).

En efecto, se ha documentado la presencia de la bacteria con sus diversas variantes en 39 países a nivel mundial (CABI/EPPO, 2017), por lo que esta enfermedad es considerada altamente perjudicial debido a los graves daños que causa en los cultivos de cítricos, además, su dificultad para el diagnóstico y la rápida propagación la convierten en un desafío para los agricultores y los sistemas de control fitosanitario.

3.1.2. Agente causal y síntomas

El agente causal del HLB es una bacteria del género *Candidatus Liberibacter* que proviene del término *Candidatus*, utilizado para organismos que no pueden ser cultivados ni clasificados con las técnicas tradicionales, *Liberibacter*, derivado del latín *Liber* (corteza) y *bacter* (bacteria) (Jagoueix *et al.*, 1994).

La propagación se debe a la interacción entre varios elementos: las plantaciones de cítricos como hospederos; las bacterias (*americanus*, *asiaticus*, y *africanus*) como patógenos; los insectos (*Diaphorina citri* para las especies asiática y americana o *Trioza erythrae* para la especie africana) como vectores de la enfermedad (Hall *et al.*, 2008), y los factores climáticos del entorno, que pueden influir en el desplazamiento de los insectos al ser arrastrados por las corrientes de aire y trasladados a distancias de 0.5 a 4 km o aún mayores (Mead & Fasulo, 2010), por lo que esta interacción espaciotemporal, es responsable de la transmisión y el incremento de los casos de HLB en los cultivos de cítricos (Figura 1).

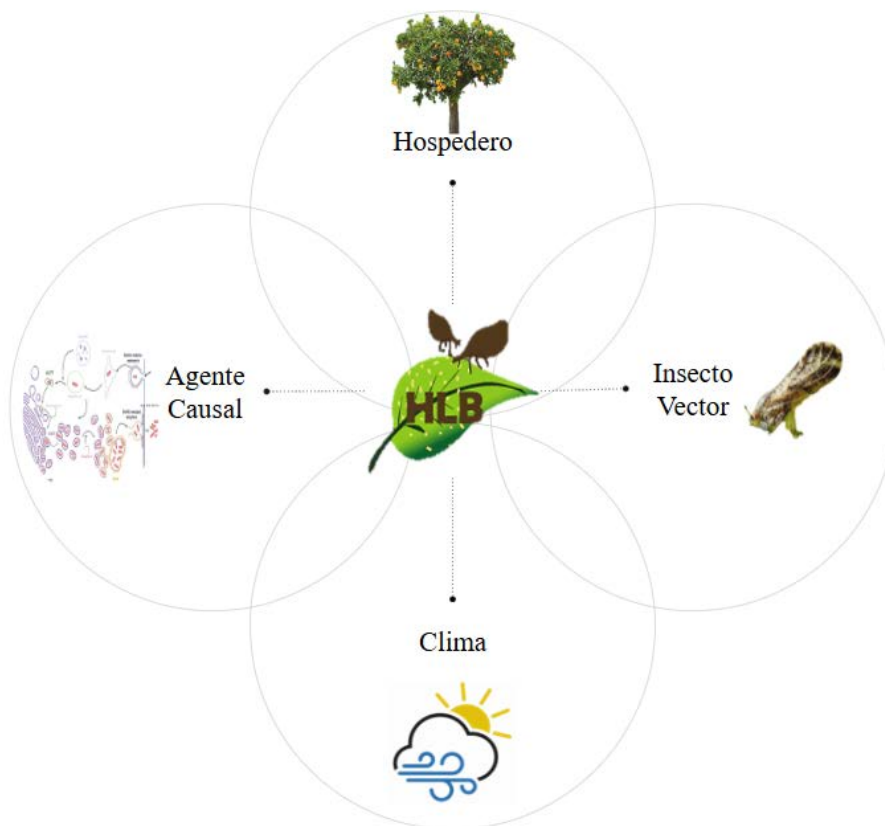


Figura 1: Interacción entre hospedero, patógeno, vector y ambiente para la incidencia de HLB
Fuente: Elaboración propia con base en Santivañez *et al.* (2023).

Por consiguiente, la presencia de estos vectores representa una amenaza directa para la propagación del patógeno tanto en la plantación como en áreas circundantes que están libres de la enfermedad, ya que se alimentan de los brotes tiernos de la planta, extrayendo una gran cantidad de savia que desarrolla fumagina, oscureciendo la superficie de las hojas y obstaculizando su función fotosintética. Con la presencia de poblaciones numerosas de insectos vectores, las ramas pueden secarse por completo (SENASICA, 2019), provocando la muerte del árbol gradualmente en un lapso de 1-2 años (Zhao, 1982).

Estudios llevados a cabo en México han observado que la gravedad de los síntomas del HLB cambia según la variedad de cítricos, siendo más pronunciados en los cítricos agrios en comparación con los cítricos dulces (Chávez, 2011). Los árboles con presencia de HLB muestran una notable disminución en el uso de agua, el potencial hídrico del tallo y el flujo de savia, en comparación con los árboles sanos, lo que afecta su crecimiento y productividad (Hamido *et al.*, 2017). Actualmente, las estrategias de manejo incluyen el control del vector que transmite la enfermedad y la eliminación de árboles infectados (CESAVESLP, 2022).

Los síntomas del HLB se visualizan principalmente en hojas y frutos. En las hojas, los síntomas pueden progresar desde manchas cloróticas y angulares hasta un acorchamiento y engrosamiento de las nervaduras, así como moteados y clorosis difusa (Figura 2). Eventualmente, las hojas pueden volverse completamente amarillas y ocurrir la defoliación del árbol (SENASICA, 2019).



Figura 2: Síntomas de HLB observados en hojas de árbol de naranjo (*Citrus sinensis*)
Fuente: Lomellí *et al.* (2010).

De acuerdo con las regulaciones establecidas por el Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria para el manejo de la enfermedad, se ha seguido principalmente el principio de erradicación en la mayoría de los árboles afectados.

Esto implica que se ha buscado eliminar la enfermedad de manera activa en los árboles positivos.

A pesar de estos esfuerzos, el HLB sigue siendo un desafío importante para los productores de cítricos, lo que subraya la necesidad de seguir investigando y desarrollando métodos de monitoreo más eficientes y sostenibles.

3.1.3. Impacto socioeconómico de las pérdidas en la industria citrícola

El HLB es reconocido como la enfermedad más devastadora para los cítricos a nivel mundial, dado su impacto en la productividad, su rápida propagación y su capacidad para afectar a todas las especies de cítricos (J. Chen *et al.*, 2009).

Este desafío ha llevado a la eliminación de más de 100 millones de árboles en todo el mundo y está generando cambios en la forma tradicional de gestionar los cultivos de cítricos. En algunas regiones del mundo, esta situación está presionando a los productores a encontrar nuevas estrategias de manejo para enfrentar la enfermedad.

De acuerdo con datos de la FAO 2021, la citricultura es una actividad agrícola de gran rentabilidad en numerosos países ya que la producción mundial supera los 130 millones de toneladas anuales. Por ejemplo, México ocupa actualmente el cuarto lugar a nivel mundial en producción de cítricos (SENASICA, 2022), con una cosecha de naranja de 4.854 millones de toneladas (SADER, 2023). Sin embargo, la diversidad de ambientes agroecológicos, culturales y económicos representa un desafío para establecer medidas de control estandarizadas contra el HLB.

La presencia del patógeno en los cultivos tiene un impacto socioeconómico considerable, traducido en pérdidas económicas significativas y una reducción estimada de entre 6 y 9 jornales en los empleos generados por la actividad citrícola (SENASICA, 2019).

Esto afecta particularmente a los pequeños productores, ya que carecen de los recursos necesarios, especialmente en términos económicos, para hacerle frente a la enfermedad y se ven obligados a buscar alternativas para subsistir.

3.2. Zona de estudio

El municipio de Ciudad Fernández destaca por su ubicación geográfica en el valle agrícola del estado de San Luis Potosí, siendo uno de los municipios con “mayor relevancia por volumen y valor de la producción agrícola” (COPLADE, 2022, p. 209), donde su economía se basa principalmente en la agricultura, con un enfoque particular en la producción de cítricos.

La huerta donde se realiza este estudio abarca 9 hectáreas (Figura 3), y se ubica en la región sureste del municipio con $21^{\circ}55'36''$ de latitud norte y $100^{\circ}03'51''$ de longitud oeste, así como una altitud de 980 metros sobre el nivel del mar.

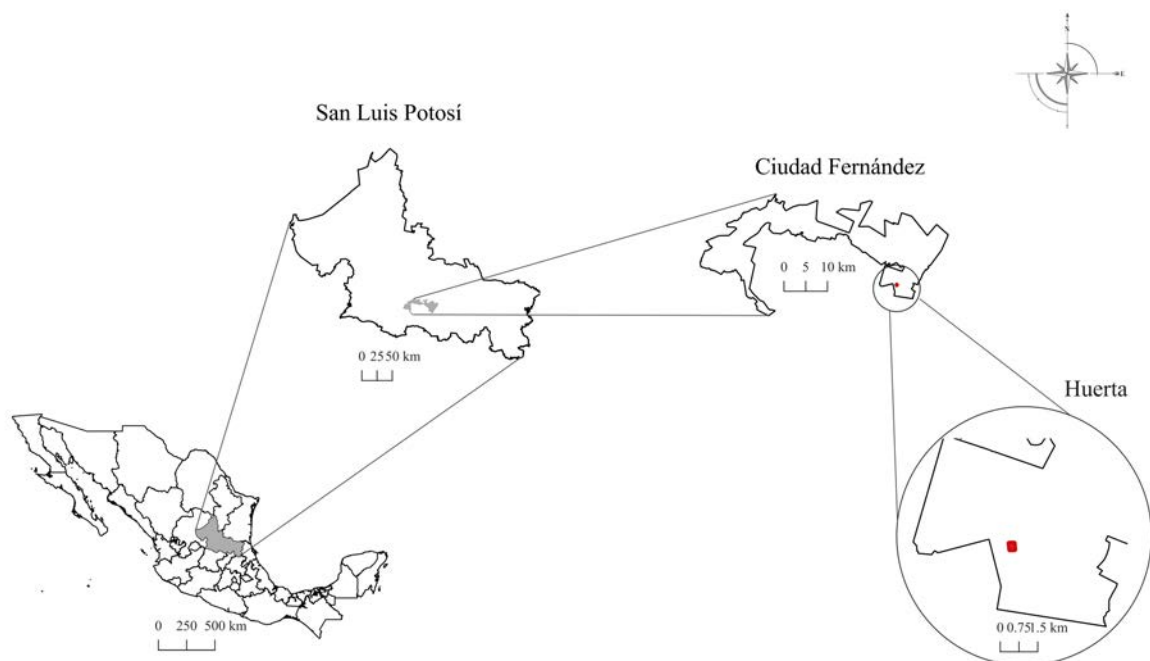


Figura 3: Ubicación geográfica de la huerta
Fuente: Elaboración propia.

La importancia de los cultivos de naranjo está asociada con factores climáticos, ya que crecen en un clima semiseco templado y semicálido con inviernos frescos. Su ubicación dentro de la cuenca del río Verde, que forma parte de la red hidrológica del sistema Pánuco y abarca subcuencas como el río Santa María Alto y el río Santa María Bajo, resalta su valor agrícola al contar con fuentes hídricas clave para la producción citrícola.

Además, se caracteriza por poseer suelos fértiles y adecuados para el cultivo (*Phaeozem* y *Chernozem*), favoreciendo una alta productividad, especialmente de naranjas (Mendoza *et al.*, 2021).

3.3. Tecnologías de Monitoreo Agrícola

En el contexto del monitoreo agrícola, las tecnologías de observación remota juegan un papel fundamental para la gestión y monitoreo de cultivos. La integración de series de tiempo de satélites, como Sentinel-2, permite obtener datos temporales que facilitan el seguimiento del estado de los cultivos a lo largo del tiempo.

Estas series de tiempo mediante índices de vegetación, son herramientas importantes para la detección de afectaciones en la vegetación, como las causadas por enfermedades, permitiendo una respuesta oportuna y eficiente.

Además de los satélites, los VANTs han emergido como una herramienta complementaria en el monitoreo agrícola. Su capacidad para capturar imágenes de alta resolución en áreas específicas de interés, con un control preciso sobre el área a observar, los hace ideales para estudios detallados de los cultivos a gran escala.

La fotogrametría digital, especialmente a través de software libre, permite procesar las imágenes capturadas por los VANTs, generando modelos tridimensionales y ortomosaicos que proporcionan información más precisa y detallada sobre el terreno y las condiciones del cultivo.

A continuación, se detallarán las tecnologías que permiten integrar estos enfoques, proporcionando una visión general sobre cómo se complementan para ofrecer una respuesta eficaz en el monitoreo de cultivos.

3.3.1. Series de tiempo de Sentinel-2

El programa Sentinel-2, parte esencial de la iniciativa Copernicus liderada por la Agencia Espacial Europea (*European Space Agency*, ESA) y la Comisión Europea, se centra en el monitoreo de la superficie terrestre mediante imágenes ópticas multiespectrales

de alta resolución. La misión incluye dos satélites idénticos, Sentinel-2A y Sentinel-2B, lanzados en junio de 2015 y marzo de 2017, respectivamente.

Operando en la misma órbita y separados por 180 grados, esta configuración asegura una cobertura global eficiente con un tiempo de revisita nominal de 5 días en el ecuador, que puede ser menor en latitudes más altas debido a la superposición orbital (ESA, 2015). Equipados con sensores multiespectrales que abarcan 13 bandas, desde el visible hasta el infrarrojo de onda corta (SWIR), estos satélites permiten capturar detalles de la superficie terrestre esenciales para aplicaciones agrícolas y medioambientales.

La misión cuenta con un diseño innovador que incluye un mecanismo de calibración y obturador, que realiza pruebas periódicas para garantizar la precisión radiométrica, mientras que la calibración geométrica inicial corrige distorsiones relacionadas con la posición orbital y las características del sensor (Andión & Olaskoaga, 2011).

Además, el mapeo del plano focal y la generación de imágenes de referencia globales contribuyen a mantener una calidad consistente de los datos (Dechoz *et al.*, 2015). Estas características, combinadas con una resolución espacial de hasta 10 metros y un campo de visión de 290 kilómetros, posicionan a Sentinel-2 como una herramienta clave para el monitoreo agrícola, ambiental y de gestión de recursos (Gascon, 2018).

En el contexto del monitoreo agrícola, las series de tiempo derivadas de imágenes de Sentinel-2 se han convertido en un recurso valioso. Una serie de tiempo, definida como un conjunto de observaciones x_t , registradas en momentos específicos t , permite analizar datos secuenciales para identificar patrones y tendencias a lo largo del tiempo (Brockwell & Davis, 2016).

Este enfoque es particularmente útil en la agricultura, donde las series de tiempo, combinadas con índices de vegetación, permiten evaluar el crecimiento de los cultivos, detectar enfermedades y optimizar la gestión de recursos como el agua y los fertilizantes.

El uso de series de tiempo basadas en Sentinel-2 permite monitorear dinámicamente los cambios en la salud de los cultivos, proporcionando una visión integral de las condiciones agrícolas a lo largo del tiempo (Bontemps *et al.*, 2015). La combinación de alta resolución espacial, cobertura frecuente y capacidades multiespectrales habilita análisis avanzados que apoyan la toma de decisiones oportunas para mejorar la productividad

agrícola.

3.3.2. Detección de afectaciones en cultivos mediante NDVI

El índice de vegetación de diferencia normalizada (*Normalized Difference Vegetation Index*, NDVI), es ampliamente utilizado para determinar la respuesta de la vegetación frente al espectro electromagnético que se destaca como una herramienta crucial para evaluar la cobertura, la clasificación y dinámica vegetal, así como sus aspectos fenológicos (Durán *et al.*, 2017), reflejando la relación entre la banda roja (R), donde la vegetación enferma refleja más luz, y la banda del infrarrojo cercano (*Near-Infrared*, NIR), donde la vegetación sana muestra mayor reflectancia (Huang *et al.*, 2020).

En la Figura 4 se muestra una representación de lo mencionado anteriormente.

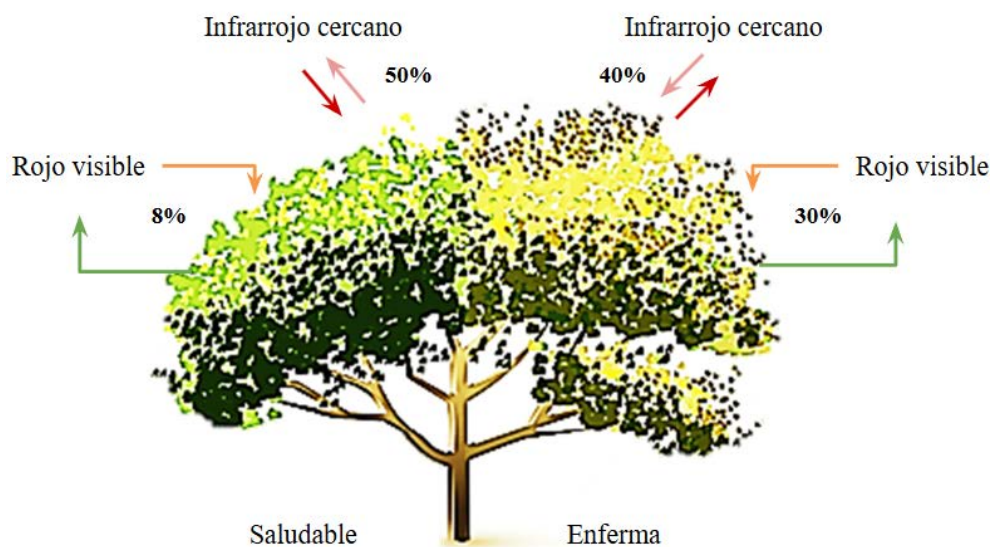


Figura 4: Porcentajes NIR y R de vegetación sana y enferma
Fuente: Elaboración propia con base en Cherlinka (2024).

La ecuación propuesta por Kriegler *et al.* (1969), se calcula mediante la razón (cociente) entre la diferencia de los valores de las bandas NIR y R y la suma de estos:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (3.1)$$

Donde:

- *NIR*: Banda del infrarrojo cercano.

- *R*: Banda roja.

La densidad y el verdor son fundamentales para evaluar la actividad fotosintética y la salud de la vegetación (Xu *et al.*, 2011), considerando que los valores de este índice oscilan entre -1 y 1, generalmente un valor más alto de NDVI indica una vegetación más densa y saludable, mientras que valores bajos pueden indicar la presencia de suelos desnudos o menor actividad vegetativa.

Tabla 2: Interpretación de los valores NDVI

NDVI	Interpretación
<0.1	Suelo desnudo
0.1–0.2	Cobertura de dosel casi inexistente
0.2–0.3	Cobertura de dosel muy baja
0.3–0.4	Cobertura de dosel baja y bajo vigor, o muy baja y alto vigor
0.4–0.5	Cobertura de dosel media-baja y bajo vigor, o baja y alto vigor
0.5–0.6	Cobertura de dosel media y bajo vigor, o media-baja y alto vigor
0.6–0.7	Cobertura de dosel media-alta y bajo vigor, o media y alto vigor
0.7–0.8	Cobertura de dosel alta y alto vigor
0.8–0.9	Cobertura de dosel muy alta y vigor muy alto
0.9–1.0	Cobertura de dosel total y vigor muy alto

Fuente: Elaboración propia con base en Antognelli (2018).

En el contexto agrícola, los valores de NDVI se interpretan según el nivel de cobertura de dosel y el vigor vegetativo (Tabla 2). Por ejemplo, valores menores a 0.1 indican suelo desnudo, mientras que rangos entre 0.7 y 0.8 reflejan cultivos con una cobertura de dosel alta y un vigor significativo. Estas categorías permiten evaluar el estado de los cultivos y detectar posibles anomalías.

El promedio del NDVI en un campo agrícola varía según factores como la variedad del cultivo y su etapa fenológica (Farzaliyev, 2021). En cultivos uniformes, donde las plantas están en la misma etapa de desarrollo, valores significativamente inferiores al promedio del campo pueden indicar problemas específicos, como estrés hídrico o nutricional, infestaciones de plagas o daños por fenómenos climáticos.

Por tanto, el análisis del NDVI no solo permite detectar áreas afectadas, sino que también proporciona información clave para la toma de decisiones correctivas.

3.3.3. Vehículo Aéreo No Tripulado (VANT)

Un VANT es una aeronave que opera sin la necesidad de un piloto a bordo y puede ser controlada de manera remota o autónoma (González *et al.*, 2017). Representan un avance tecnológico significativo en diversas áreas, con variaciones en tamaño, tipos de propulsión y capacidades (Tsach *et al.*, 2010).

Equipados con cámaras multispectrales y diversos sensores, los VANTs capturan y transmiten datos mediante fotografías y videos de alta resolución, lo que permite observar características no visibles a simple vista y facilita la toma de decisiones informadas (Ríos, 2021).

En el ámbito de la fotogrametría y los sensores remotos, los VANTs ofrecen ventajas sobre las plataformas tradicionales como satélites y aviones tripulados. Entre estas ventajas se incluyen la obtención de imágenes con alta resolución espacial y temporal, menores costos de adquisición y operación, la capacidad de operar en entornos peligrosos sin poner en riesgo a tripulantes, respuesta rápida ante situaciones emergentes y menos restricciones de vuelo en la mayoría de los países (Torrado *et al.*, 2016).

Un ejemplo destacado es su uso en la agricultura, donde ayudan a mejorar el rendimiento de los cultivos y reducen los costos de producción al ser utilizados en el monitoreo y la gestión (Sreekantha & Prasad, 2018). Recientes avances en el uso de VANTs y aprendizaje profundo han permitido estimar automáticamente la salud de los cultivos para detectar patrones en la vegetación, optimizando su evaluación y gestión (Sosa-Herrera *et al.*, 2022).

A pesar de los desafíos, como la evasión de colisiones y la duración limitada de la batería, los VANTs ofrecen soluciones prometedoras para una variedad de aplicaciones presentes y futuras.

3.3.4. Fotogrametría digital

La fotogrametría es una técnica científica que permite obtener información precisa sobre objetos y el entorno mediante el análisis de imágenes fotográficas y patrones de radiación electromagnética (Linder, 2016).

Esta técnica, que tiene sus orígenes en la fotografía aérea del siglo XIX, se ha utilizado principalmente para la elaboración de mapas topográficos (Dvivedi, 2020); (Wu, 2017).

A lo largo del tiempo, los sistemas de teledetección mediante VANT han demostrado claras ventajas sobre los métodos tradicionales, especialmente en áreas pequeñas. Estos sistemas permiten una rápida adquisición de datos a costos significativamente menores en comparación con las aeronaves tripuladas y los métodos convencionales de recolección en campo (Boon *et al.*, 2016).

Además, el desarrollo de software especializado como *OpenDroneMap* (ODM), ha optimizado los procesos fotogramétricos al generar productos de alta calidad. Este software gratuito y de código abierto, diseñado para procesar imágenes capturadas por VANT, se presenta como una alternativa económica que combina precisión y rentabilidad, posicionándose como una herramienta eficaz en diversas aplicaciones (Mokrane *et al.*, 2019).

Estudios recientes destacan el potencial de la fotogrametría con VANT y el uso de software de código abierto que implementa algoritmos de estructura a partir del movimiento (SfM), los cuales han sido cada vez más automatizados (Rinaldi & Larrabide, 2021). Estos avances permiten la generación de nubes de puntos de alta resolución, ortomosaicos, mallas 3D de alta calidad, así como modelos de elevación, tales como Modelos Digitales de Superficie (MDS) y Modelos Digitales de Terreno (MDT).

La etapa final del proceso fotogramétrico con imágenes capturadas por VANT es la ortorrectificación, que consiste en proyectar cada píxel de la imagen sobre un modelo de elevación para determinar su altura (Figura 5). Este proceso se utiliza posteriormente para recalcular la posición del píxel en el espacio tridimensional (Torrado *et al.*, 2016).

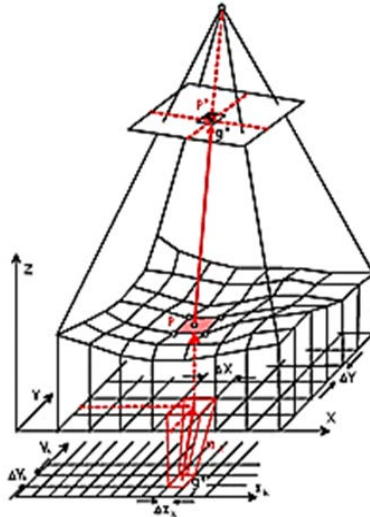


Figura 5: Ororrectificación de una fotografía aérea
Fuente: GeoBC (2024).

Las imágenes se combinan para crear un mosaico unificado que ofrece un campo de visión más amplio de la escena observada (Cheng *et al.*, 2007). Una vez corregidas las distorsiones causadas por el relieve del terreno y los objetos presentes en la escena, el mosaico resultante se denomina ortomosaico, ortoimagen, ortofoto u ortofotografía.

La calidad de los ortomosaicos puede mejorarse mediante el uso de puntos de control terrestre, aunque también es posible el procesamiento automatizado sin ellos (Liba & Berg-Jürgens, 2015). Además, técnicas avanzadas como el ensamblado en el dominio del gradiente y la optimización del movimiento utilizando homografías pueden incrementar la continuidad y la eficiencia en la producción de ortomosaicos a gran escala (Tanathong *et al.*, 2022).

3.4. Modelo Estadístico

El análisis de series de tiempo es una disciplina estadística que permite estudiar y modelar datos que presentan dependencias a lo largo del tiempo. Este enfoque es particularmente relevante para identificar patrones, tendencias y estacionalidades en conjuntos de datos secuenciales, proporcionando una base sólida para realizar pronósticos y tomar decisiones fundamentadas.

Para abordar estas características, se han desarrollado diversos modelos estadísticos que permiten analizar y pronosticar series temporales adaptándose a las particularidades de los datos.

Entre los principales enfoques se encuentran los Modelos de Suavizado Exponencial, que priorizan los datos recientes asignando menor peso a los valores más antiguos; los Modelos ARIMA, que combinan componentes autoregresivos, de diferenciación y medias móviles para capturar patrones complejos, y los Modelos de Componentes No Observados, que separan la serie en elementos como tendencia, estacionalidad y ruido, facilitando un análisis más detallado (Rahardja, 2020).

En conjunto, estos enfoques representan herramientas fundamentales para analizar y pronosticar series de tiempo en contextos diversos. Su capacidad de adaptarse a las características particulares de los datos y las necesidades del análisis los convierte en piezas clave para la modelización estadística.

3.4.1. Modelo Autorregresivo Integrado de Media Móvil Estacional (SARIMA)

Los modelos SARIMA (*Seasonal Autoregressive Integrated Moving Average*) son una extensión de los Modelos ARIMA. Mientras que los modelos ARIMA son efectivos para capturar tendencias y dependencias en datos no estacionales, los SARIMA extienden este enfoque al incorporar las características estacionales de los datos, lo que permite un pronóstico más preciso cuando los patrones regulares se repiten a intervalos fijos (Yadav *et al.*, 2024).

La estacionalidad se refiere a patrones recurrentes que aparecen en intervalos específicos y regulares, como ciclos mensuales, trimestrales o anuales. SARIMA tiene en cuenta esta estacionalidad mediante la introducción de componentes estacionales en el modelo (P. Chen *et al.*, 2018).

Esta estructura estacional se integra luego en el marco de ARIMA, que modela el comportamiento no estacional (tendencia y fluctuaciones aleatorias) de la serie. La formulación de SARIMA utiliza la notación (Figura 6):

$$\text{SARIMA } \underbrace{(p, d, q)}_{\text{No estacional}} \underbrace{(P, D, Q)}_{\text{Estacional}} s$$

Figura 6: Notación SARIMA
Fuente: Elaboración propia.

Esta notación se desglosa en partes (Mutwiri, 2019), donde:

(p, d, q) son los parámetros no estacionales de ARIMA:

- p : Representa el orden del componente autorregresivo (AR), que indica cuántos valores pasados de la serie de tiempo se utilizan para pronosticar el valor actual.
- d : Se refiere a la integración (I) y representa el número de diferencias necesarias para transformar la serie de tiempo en estacionaria. Esto implica eliminar tendencias o patrones no estacionarios, estabilizando la media y la varianza de la serie, lo que facilita su modelado y pronóstico.
- q : Es el orden del componente de media móvil (MA). Este parámetro modela la dependencia entre el valor actual de la serie y los errores residuales de períodos anteriores.

(P, D, Q) son los componentes estacionales:

- P : Es el equivalente estacional del componente autorregresivo. Indica cuántos valores previos del mismo ciclo estacional (por ejemplo, el mismo mes en el año anterior) afectan al valor actual.
- D : Representa el número de diferencias estacionales necesarias para alcanzar la estacionariedad estacional, es decir, para que los patrones estacionales sean constantes en el tiempo.
- Q : Es el equivalente estacional del componente de media móvil. Este parámetro captura la relación entre el valor actual y los errores residuales de los ciclos estacionales pasados.
- s : Es el parámetro que indica la periodicidad estacional, o el número de períodos dentro de un ciclo estacional.

A diferencia de los modelos de regresión, que relacionan una variable dependiente con otras variables independientes (Veney & Luckey, 1983), SARIMA se enfoca exclusivamente en la dependencia temporal dentro de una serie de tiempo. Utiliza únicamente valores pasados de la misma variable y términos de error anteriores para realizar pronósticos, sin incorporar factores externos al análisis (Korstanje, 2021).

De acuerdo con Hyndman y Athanasopoulos (2021), la estimación de modelos ARIMA y sus variantes, como SARIMA, implica un proceso más complejo en comparación con otros métodos debido a la necesidad de ajustar múltiples parámetros. Estos incluyen los términos autorregresivos (AR), de diferenciación (I) e integrados de media móvil (MA), así como los componentes estacionales. Este ajuste se realiza mediante métodos de optimización iterativa.

Este enfoque iterativo y la necesidad de modelar las características específicas de la serie, como tendencias, estacionalidades y patrones de ruido refuerzan el carácter especializado de SARIMA en el análisis y pronóstico de series de tiempo. Además, este modelo es particularmente útil para capturar patrones estacionales y no estacionarios, lo que lo convierte en una herramienta robusta para analizar datos temporales complejos.

El proceso de selección del modelo ARIMA más adecuado se basa en criterios de información como el Criterio de Información de Akaike (*Akaike Information Criterion*, AIC) o el Criterio de Información Bayesiano (*Bayesian Information Criterion*, BIC), los cuales permiten evaluar el desempeño del modelo al balancear su precisión con la complejidad. Estas métricas son especialmente útiles para evitar el sobreajuste, un problema que ocurre cuando un modelo se ajusta demasiado bien a los datos de entrenamiento, pero falla al generalizar para datos nuevos (Farhath *et al.*, 2016).

El AIC, una de las métricas más comunes, se calcula mediante la fórmula:

$$\text{AIC} = 2k - 2 \ln(L) \quad (3.2)$$

Donde:

- k : Es el número de parámetros estimados en el modelo (incluyendo los términos AR , MA y constantes).

- L : Es el valor de la función de verosimilitud del modelo (es decir, qué tan bien se ajusta el modelo a los datos).

El objetivo del AIC es encontrar el modelo que minimice este valor. Un AIC más bajo indica un mejor equilibrio entre la calidad del ajuste del modelo y su parsimonia, es decir, la capacidad de explicar los datos con el menor número de parámetros posibles. El modelo con el menor valor de AIC es generalmente considerado como el óptimo, ya que equilibra la calidad del ajuste y la parsimonia (Mattheakis & Protopapas, 2018).

Similar al AIC, el BIC busca balancear la precisión del modelo con su complejidad, pero con un enfoque más estricto en penalizar la inclusión de parámetros adicionales (Watanabe, 2013). Esta penalización más alta hace que el BIC tiende a favorecer modelos más simples en comparación con el AIC.

El BIC se calcula mediante la siguiente fórmula:

$$\text{BIC} = \ln(n)k - 2 \ln(L) \quad (3.3)$$

Donde:

- n : Es el número de observaciones o el tamaño de la muestra.
- k : Es el número de parámetros estimados en el modelo (igual que en el AIC, incluye los parámetros de autorregresión, media móvil y constantes).
- L : Es el valor de la función de verosimilitud del modelo, que mide cuán bien el modelo se ajusta a los datos observados.

Ambos criterios suelen emplearse de manera complementaria para identificar el modelo óptimo (Buntara *et al.*, 2023). No obstante, en la mayoría de los casos, el AIC es más utilizado debido a su menor penalización en modelos con mayor número de parámetros.

3.4.2. Métricas para evaluar el modelo

Evaluar el rendimiento de un modelo es crucial para determinar qué tan bien pronostica tanto en el conjunto de datos utilizado para el entrenamiento como en los datos no

vistos (Verma & Chakraborty, 2020). La precisión de un modelo de pronóstico puede evaluarse mediante diversas medidas estadísticas, que ayudan a cuantificar los errores entre las predicciones del modelo y las observaciones reales.

Algunas de las medidas comunes para evaluar la precisión de los pronósticos incluyen:

- **Raíz del Error Cuadrático Medio (RMSE):** Es la raíz cuadrada del promedio de los errores cuadrados entre los valores predichos y los reales.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.4)$$

Donde:

- y_i es el valor real u observado en el punto i ,
- \hat{y}_i es el valor pronosticado para el punto i ,
- n es el número total de observaciones.

El RMSE es útil cuando los grandes errores son particularmente indeseables. Esto se debe a que el error se eleva al cuadrado antes de promediarlo, lo que penaliza fuertemente las desviaciones grandes (Chai & Draxler, 2014). Un valor más bajo de RMSE indica un mejor rendimiento del modelo.

- **Error Absoluto Medio (MAE):** Es el promedio de las diferencias absolutas entre los valores predichos y los reales.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.5)$$

Donde:

- y_i es el valor real u observado en el punto i ,
- \hat{y}_i es el valor pronosticado para el punto i ,
- n es el número total de observaciones.

Esta medida calcula el error promedio en términos absolutos, es decir, mide la diferencia promedio entre los valores reales y los predichos sin importar la dirección del error (si es positivo o negativo) (Tibshirani, 2023). A diferencia del RMSE, el MAE no penaliza los errores grandes de manera tan fuerte, ya que toma las diferencias absolutas, lo que significa que cada error tiene el mismo peso. Un valor bajo de MAE indica un mejor rendimiento del modelo.

La aplicación de estas métricas permite una evaluación integral y precisa del rendimiento de los modelos ARIMA, facilitando la selección del modelo que mejor equilibra el ajuste a los datos y la complejidad del modelo estacional (Schaffer *et al.*, 2021). El AIC y BIC optimiza la selección del modelo, mientras que el RMSE y MAE proporcionan una evaluación clara del error en términos absolutos y relativos.

3.5. Aprendizaje Automático

El aprendizaje automático es un campo en expansión dentro de la ciencia de la computación, permitiendo a las computadoras aprender y mejorar por sí mismas sin necesidad de programación explícita (Raj, 2019). Su objetivo es desarrollar algoritmos capaces de manejar datos de manera eficiente para extraer conocimiento (Angra & Ahuja, 2017).

Desde sus inicios como parte de la Inteligencia Artificial en el siglo XX, el desarrollo de algoritmos ha evolucionado. Dado que el rendimiento de estos algoritmos depende en gran medida de la calidad de los datos, su desarrollo combina principios de informática con conceptos de estadística, probabilidad y optimización para mejorar su eficacia (Zaidi *et al.*, 2018).

Existen varios escenarios en el aprendizaje automático, cada uno con características específicas en cuanto a los datos utilizados para entrenamiento y evaluación (Figura 7).

De acuerdo con L. Chen *et al.* 2019, estos escenarios son: Aprendizaje Supervisado, Aprendizaje No Supervisado, Aprendizaje Semisupervisado, Aprendizaje por Refuerzo, Inferencia Transductiva y Aprendizaje Activo. De estos, los primeros cuatro son los más comúnmente aplicados.



Figura 7: Tipos de aprendizaje automático

Fuente: Elaboración propia con base en Chauhan y Henrietta (2023) y L. Chen *et al.* (2019).

Los modelos de aprendizaje automático requieren una validación rigurosa para asegurar su precisión predictiva y capacidad de generalización. Este proceso usualmente involucra la división de los datos en conjuntos de entrenamiento, validación y prueba (Burzykowski *et al.*, 2023).

Las métricas de rendimiento clave varían según la tarea y deben considerar el desequilibrio de clases y las compensaciones entre costos y beneficios. El sobreajuste es una preocupación significativa, ya que los modelos pueden capturar patrones aleatorios de los datos de entrenamiento, resultando en un rendimiento deficiente en datos nuevos.

Aunque generalmente se evita mezclar datos de entrenamiento y validación, algunos investigadores sugieren métodos controlados para incorporar datos de validación en el entrenamiento, especialmente con algoritmos estables, para mejorar el rendimiento (Yun & Royston, 2018).

Independientemente del método, una validación adecuada es esencial para la selección de modelos, la evaluación del rendimiento y la obtención de estimaciones precisas de la precisión predictiva (Varoquaux & Colliot, 2023).

Los recientes avances en aprendizaje automático han sido impulsados por nuevos algoritmos, la mayor disponibilidad de datos y mejoras en la capacidad computacional (Jordan & Mitchell, 2015). El campo incluye diversas técnicas y algoritmos, cada uno adecuado para diferentes tipos de tareas y datos.

A medida que el aprendizaje automático continúa evolucionando, se está convirtiendo en una herramienta fundamental en la toma de decisiones basada en evidencia en numerosos dominios, transformando la forma en que abordamos problemas complejos y analizamos grandes volúmenes de datos.

3.5.1. K-Vecinos más cercanos (KNN)

El algoritmo de vecinos más cercanos (KNN, por sus siglas en inglés) es un método de aprendizaje automático supervisado, sencillo y eficaz, empleado tanto en tareas de clasificación como de regresión (Torfi, 2020). Su principio fundamental se basa en la premisa de que instancias similares en un conjunto de datos tienden a tener resultados parecidos.

En clasificación, KNN asigna una clase a un nuevo punto de datos identificando la clase predominante entre sus k vecinos más cercanos, un valor determinado por el usuario (Figura 8).

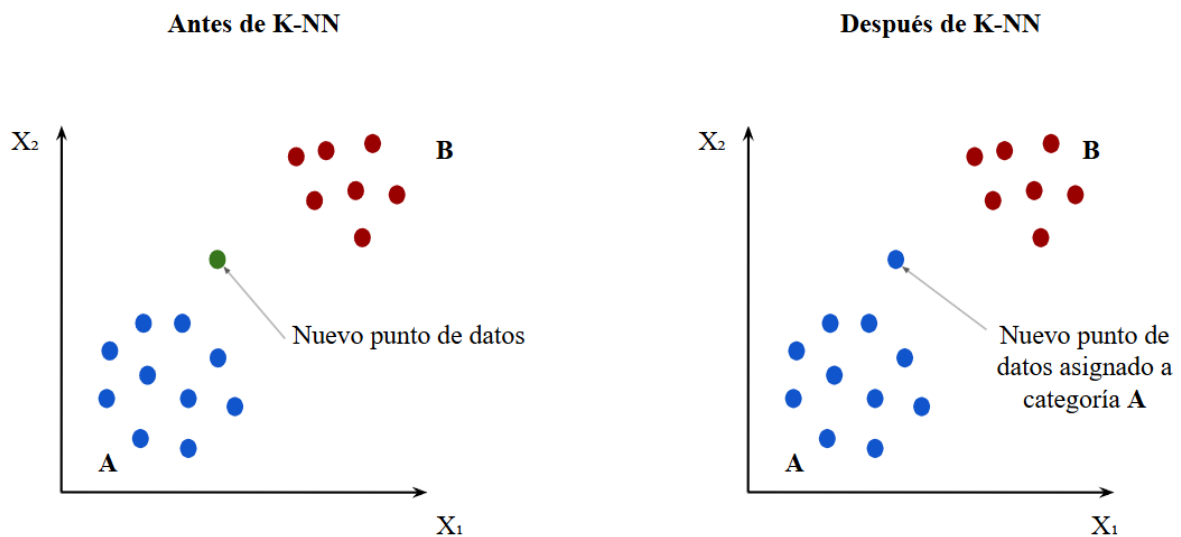


Figura 8: Ejemplo para nuevos datos antes y después de KNN
Fuente: Elaboración propia con base en Kufel *et al.* (2023).

Para medir la similitud, el algoritmo calcula distancias, generalmente empleando métricas como la distancia euclidiana, agrupando datos en clusters coherentes y clasificando nuevas entradas según su similitud con los datos entrenados previamente (Taunk *et al.*, 2019).

Una de las principales ventajas de es su simplicidad y flexibilidad, lo que facilita su uso,

requiere poco tiempo de entrenamiento y muestra robustez frente a datos ruidosos. Se adapta bien a distintos tipos de datos y puede manejar fronteras de decisión complejas (Rani & Vashishtha, 2017).

Sin embargo, su desempeño puede ser sensible a la elección del valor de k y puede enfrentar dificultades con datos de alta dimensionalidad. En tareas de regresión, KNN predice la variable objetivo promediando o tomando la mediana de los valores de sus k vecinos más cercanos, lo que lo hace adecuado para situaciones donde la relación entre características y la variable objetivo es no lineal (Zaidi *et al.*, 2018).

A pesar de su sencillez, tiene aplicaciones en diversos campos, como el reconocimiento de imágenes, sistemas de recomendación y diagnóstico médico. Aunque puede ser más intensivo computacionalmente que otros algoritmos, su enfoque intuitivo y adaptabilidad lo convierten en una herramienta valiosa en el aprendizaje automático.

3.5.2. Validación Cruzada

La validación cruzada es un método de remuestreo ampliamente utilizado para estimar errores de predicción, ajustar parámetros de modelos y evaluar la calidad de las estimaciones en diversas técnicas estadísticas (Tolosana-Delgado & Mueller, 2021).

La forma más común es la validación cruzada $k - fold$, donde los datos se dividen en k segmentos o particiones (Figura 9), con $k-1$ segmentos utilizados para el entrenamiento y uno para la validación en cada iteración.

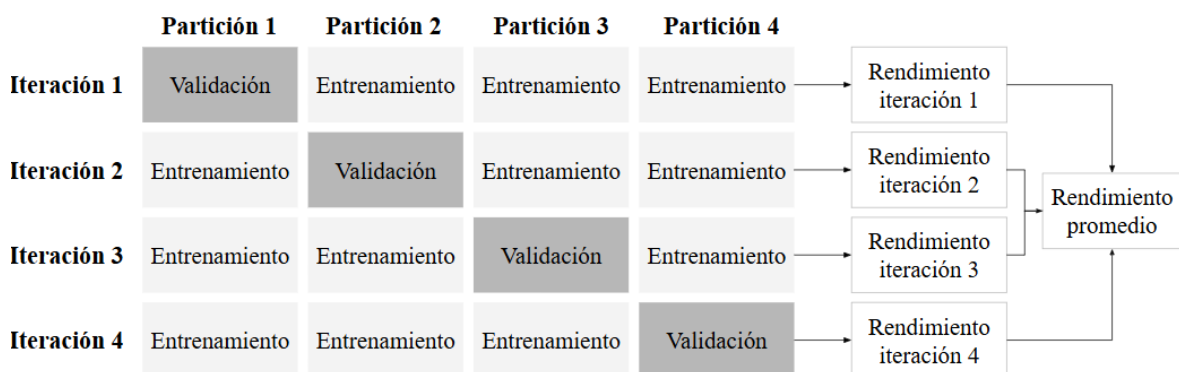


Figura 9: Ejemplo de validación cruzada
Fuente: Elaboración propia.

Este método permite evaluar de manera más robusta el rendimiento de un modelo, proporcionando una estimación confiable de su capacidad para generalizar datos nuevos (Berrar, 2019). Sin embargo, no está destinada a realizar la selección de características. Estas tareas deben abordarse de forma separada como parte del proceso de construcción del modelo.

Además, la validación cruzada tiene limitaciones, especialmente cuando se trabaja con tamaños de muestra pequeños, por lo que se han desarrollado criterios alternativos, como los de muestra única (Camstra & Boomsma, 1992), para enfrentar este desafío.

De acuerdo con Arlot y Celisse (2010), es importante elegir el método de validación cruzada que mejor se adapte a las características específicas del problema, asegurando así una evaluación óptima del modelo.

En el contexto de modelos como KNN, la validación cruzada se utiliza para evaluar la precisión media y la desviación estándar de los resultados obtenidos. Este tipo de análisis ayuda a que los resultados sean representativos de su desempeño en situaciones del mundo real (Wang *et al.*, 2023).

La precisión media es una medida del rendimiento promedio del modelo en todas las iteraciones de validación, y se considera un indicador importante de la capacidad de un modelo para hacer predicciones correctas sobre datos no vistos (Wazirali, 2020).

Además, la desviación estándar mide la dispersión de los resultados de las diferentes particiones, lo que refleja la consistencia del modelo (Hyvönen *et al.*, 2024). Un modelo con una baja desviación estándar es más confiable, ya que muestra un rendimiento consistente independientemente del subconjunto de datos en el que se pruebe.

Estas dos métricas, precisión media y desviación estándar, son esenciales para evaluar el rendimiento de modelos de clasificación, particularmente en escenarios donde la variabilidad y la generalización son factores clave.

La validación cruzada no solo ayuda a determinar el modelo con la mayor precisión media, sino también a comprender cómo este modelo puede comportarse ante diferentes muestras de datos. El uso adecuado de estas métricas permite optimizar el modelo y garantizar que las conclusiones obtenidas sean aplicables y confiables.

3.5.3. Matriz de Confusión

La matriz de confusión, introducida por Karl Pearson (1904), es una herramienta esencial en el aprendizaje automático para evaluar el rendimiento de los modelos de clasificación, especialmente en el aprendizaje supervisado.

Se aplica tanto a problemas de clasificación binaria como multiclase.

El término “confusión” se refiere a la clasificación errónea de clases por parte del modelo (Amin, 2022), mostrando la relación entre las clasificaciones reales y las predichas, lo que permite visualizar el rendimiento del algoritmo (Haghighi *et al.*, 2018).

La matriz presenta las instancias de las clases reales en filas y las clases predichas en columnas, o viceversa (Piegorsch, 2020). En la clasificación binaria, la matriz contiene cuatro elementos: Verdaderos Negativos (VN), Falsos Positivos (FP), Falsos Negativos (FN) y Verdaderos Positivos (VP) (Figura 10).

		Etiquetas reales	
		Positivo	Negativo
Etiquetas predichas	Positivo	Verdaderos Positivos VP	Falsos Positivos FP
	Negativo	Falsos Negativos FN	Verdaderos Negativos VN

Figura 10: Ejemplo de matriz de confusión
Fuente: Elaboración propia.

Las matrices de confusión son herramientas fundamentales en la evaluación de modelos de clasificación, ya que permiten analizar el desempeño del modelo al comparar las predicciones realizadas con los valores reales. A partir de esta matriz, es posible calcular métricas de rendimiento clave, como las siguientes (Dalianis, 2018):

- **Exactitud (*Accuracy*)** es la proporción de muestras clasificadas correctamente (verdaderos positivos y verdaderos negativos) con respecto al número total de muestras evaluadas:

$$\text{Accuracy} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FP} + \text{VN} + \text{FN}}$$

- **Precisión (*Precision*)** es la proporción de muestras clasificadas correctamente como positivas frente al número total de muestras clasificadas como positivas:

$$\text{Precision} = \frac{\text{VP}}{\text{VP} + \text{FP}}$$

- **Exhaustividad (*Recall*)** es la proporción de muestras clasificadas correctamente como positivas frente al número total de muestras realmente positivas:

$$\text{Recall} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

- **F1-score** se calcula como la media armónica de la *Precision* y el *Recall*, con un valor que varía de 0 a 1, siendo 1 la mejor puntuación:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Soporte (*Support*)** indica el número de instancias reales reales (observaciones) en cada clase del conjunto de datos.

Cada una de estas métricas proporciona información valiosa sobre el desempeño de un clasificador en función del tipo de error que se desea minimizar, ya sea reducir los falsos positivos, los falsos negativos o encontrar un balance adecuado entre ambas (Geng, 2024).

Por último, en este capítulo se ha revisado el marco teórico que sustenta el estudio, destacando el impacto socioeconómico del HLB y los avances en las tecnologías de percepción remota que apoyan la toma de decisiones en el análisis agrícola.

También se han analizado los modelos estadísticos y de aprendizaje automático, como el modelo SARIMA y el algoritmo KNN, que facilitan una comprensión más profunda de los datos.

A partir de este marco teórico, se establece la base para el desarrollo de la metodología utilizada. En el siguiente apartado, se describen los pasos específicos para la implementación de estas herramientas y su aplicación en el contexto de la investigación, encaminando así el logro de los objetivos planteados.

Capítulo 4

Metodología

La metodología adoptada se centra en un enfoque cuantitativo para procesar, analizar y extraer información de grandes volúmenes de datos geoespaciales, con el fin de clasificar el grado de afectación causado por el HLB y pronosticar la evolución futura de la enfermedad en cultivos de naranjo. Este enfoque es particularmente relevante debido a la necesidad de respuestas rápidas y efectivas en el manejo de esta enfermedad a gran escala.

Para clarificar el proceso, la Figura 11 presenta el esquema metodológico que organiza las etapas principales, desde la recolección de datos hasta la implementación de modelos y evaluación de resultados, los cuales se detallan en los apartados siguientes.

El proceso metodológico comienza con la recopilación de datos provenientes de dos fuentes principales: las series de tiempo NDVI obtenidas de Sentinel-2 y las imágenes capturadas mediante VANT en la zona de estudio. Estas fuentes proporcionan información con cobertura tanto temporal como espacial, permitiendo un análisis integral.

Cada conjunto de datos pasa por un preprocesamiento adaptado a sus características y al análisis requerido. En el caso de las series de tiempo NDVI, se aplican técnicas de eliminación de ruido y normalización para depurar la información, eliminando variaciones espurias y asegurando la calidad del análisis. Por otro lado, las imágenes de VANT se someten a un proceso de fotogrametría digital y ortorrectificación, lo que permite generar ortomosaicos de alta precisión.

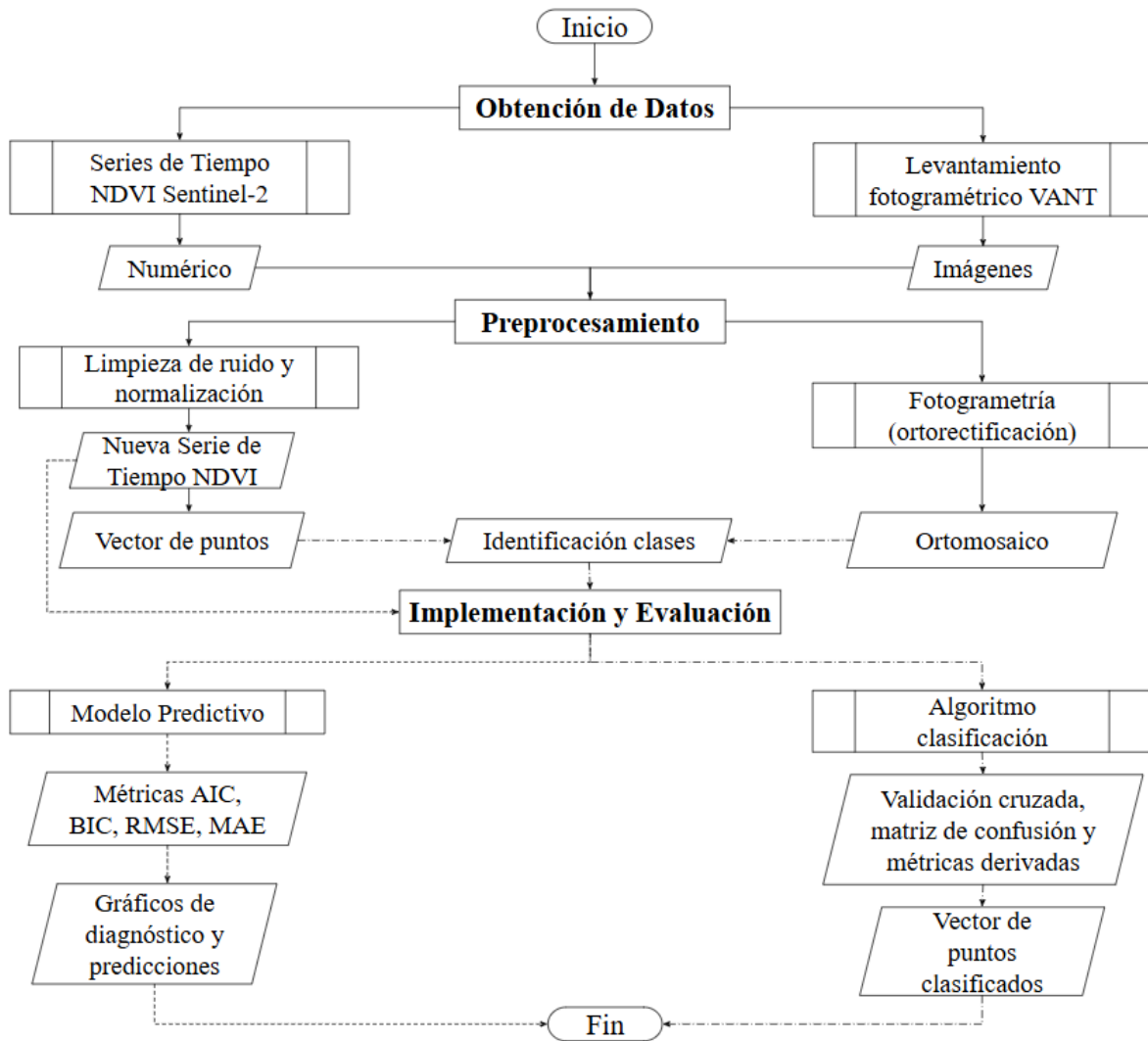


Figura 11: Diagrama del proceso metodológico
Fuente: Elaboración propia.

Posteriormente, los datos de NDVI se transforman en vectores de puntos espaciales, facilitando la identificación de clases de afectación. Esta identificación se realiza visualmente en el ortomosaico y numéricamente mediante los valores de NDVI, integrando ambos enfoques para una interpretación más robusta y precisa.

A continuación, se implementan y evalúan por separado dos modelos principales: uno estadístico para pronósticos y otro basado en un algoritmo de clasificación. El modelo estadístico, que utiliza la nueva serie de tiempo depurada, se evalúa mediante métricas que reflejan su capacidad predictiva y mediante gráficos de diagnóstico para analizar su ajuste y precisión.

Por otro lado, el algoritmo de clasificación, evaluado a través de validación cruzada, mide la precisión en la clasificación de los puntos NDVI según las clases de afectación previamente identificadas por HLB. Esta clasificación genera un conjunto de datos vectoriales que representan espacialmente los niveles de estrés en el cultivo, facilitando el análisis visual y geoespacial de la enfermedad.

4.1. Flujo de proceso con imágenes de VANT

El uso de imágenes obtenidas mediante Vehículos Aéreos No Tripulados (VANT) ha revolucionado los métodos de análisis y monitoreo en la agricultura de precisión, permitiendo capturar datos de alta resolución espacial. Este enfoque, esencial para estudios como el monitoreo de enfermedades, integra tecnologías avanzadas como la fotogrametría para procesar las imágenes capturadas y generar productos geoespaciales clave.

La obtención de datos mediante VANT se realiza en dos etapas principales dentro del flujo de proceso. La primera consiste en un levantamiento fotogramétrico, diseñado para garantizar una cobertura precisa y completa de la zona de estudio mediante la captura de imágenes aéreas. En la segunda etapa, los datos recopilados se procesan utilizando técnicas de fotogrametría para generar un ortomosaico detallado, que sirve como base para el análisis espacial y la integración con otras fuentes de información relevantes.

A continuación, se describen en detalle estas etapas clave.

4.1.1. Obtención de datos: Levantamiento fotogramétrico

Se recolectaron imágenes utilizando dos VANT diferentes: el DJI Mini 2 con cámara RGB modelo FC7303 y el DJI Air 2S con cámara RGB modelo FC3411; ambos volados a 30, 50 y 100 metros de altura sobre el polígono de la zona de estudio, obteniendo un total de 999, 735 y 176 fotos en cada caso. Las capturas se realizaron el 22 y 23 de febrero de 2024, a las 15:00 y 10:00 horas, respectivamente (Figura 12).

La diferencia de horarios y la posición del sol son factores importantes debido a cómo afecta la iluminación, las sombras y condiciones ambientales que influyen directamente en la calidad y utilidad de las imágenes, particularmente en las 9 hectáreas que están completamente infectadas por la enfermedad.



Figura 12: Árboles de naranjo con afecciones por HLB captados a 30 metros de altura
Fuente: Elaboración propia.

El uso de dos equipos resultó útil, sobre todo el primer día de vuelo, debido a que la alta temperatura afectó el dispositivo móvil designado para ejecutar el plan de vuelo programado, sobrecalentándolo y afectando su funcionamiento, problema que no se repitió el segundo día. Las imágenes de alta resolución capturaron detalles esenciales de los diferentes grados de afectación por HLB en los cultivos de naranjo, como la pérdida de hojas y ramas muertas que son difíciles de detectar con métodos tradicionales debido a su ubicación en el ápice del árbol y la necesidad de una altitud considerable para ser visibles.

4.1.2. Fotogrametría y generación de ortomosaico

Se realizó el proceso de fotogrametría para mejorar la precisión espacial de las imágenes capturadas a diferentes altitudes con VANT, utilizando OpenDroneMap (ODM), un conjunto de herramientas de línea de comandos de código abierto para procesar imágenes aéreas (OpenDroneMap, 2020).

En este apartado, es crucial configurar los parámetros de calidad de la reconstrucción, definidos como altos, y seleccionar el formato de salida como .tif (por sus siglas en inglés, *Tagged Image File Format*), que es ampliamente utilizado en imágenes geospaciales, debido a su capacidad para mantener una alta calidad. Posteriormente, se procede con la corrección en ODM, donde se alinean las imágenes utilizando algoritmos de fotogrametría para detectar y emparejar puntos en común. Esto crea una nube de puntos densa que representa la superficie del área fotografiada.

A partir de esta nube de puntos, se genera un Modelo Digital de Superficie (DSM) que incluye todas las características topográficas y objetos presentes. El siguiente paso es la ortorectificación, proceso principal para corregir distorsiones geométricas y ajustar las imágenes. Este procedimiento crea coherencia y mantiene los detalles, asegurando que cada píxel esté correctamente georreferenciado (DJI, 2021).

Los pasos anteriores permitieron crear un ortomosaico combinando imágenes ortorectificadas para formar un mosaico continuo. ODM ajusta automáticamente el color y la exposición para asegurar una transición suave entre las imágenes. El ortomosaico final se exporta en el formato previamente especificado (Figura 13).

Como último punto, se verificó la calidad de los ortomosaicos generados con imágenes a 30, 50 y 100 metros, asegurando que no hubiera errores visibles, como artefactos o desalineaciones, lo cual permitirá su correcta implementación en los siguientes pasos.



Figura 13: Ortomosaico generado en ODM con imágenes capturadas a 30 metros con VANT
Fuente: Elaboración propia.

4.2. Flujo de proceso con datos Sentinel-2

El análisis de datos Sentinel-2 para el monitoreo de vegetación se basa en la utilización de series de tiempo del índice NDVI, las cuales permiten evaluar la salud de la vegetación de manera continua. Para garantizar la calidad de los datos, es necesario aplicar procesos de limpieza de ruido y normalización, los cuales eliminan anomalías y ajustan los valores a una escala consistente. Esto sienta las bases para la identificación de clases que reflejan diferentes niveles de estrés en la vegetación, clave para entender patrones espaciales y temporales.

Posteriormente, se implementan y evalúan modelos de clasificación y pronóstico que permiten profundizar en los patrones detectados. Entre ellos, el algoritmo KNN es utilizado para clasificar los datos según sus características, mientras que el modelo SARIMA se emplea para analizar y pronosticar tendencias temporales en las series de tiempo.

A continuación, se describen en detalle los pasos del flujo de proceso, incluyendo las metodologías aplicadas en cada etapa y su importancia dentro del análisis integral.

4.2.1. Obtención de datos: Series de Tiempo NDVI

Para alcanzar el objetivo de esta investigación, se utilizó la misión multispectral de alta resolución Sentinel-2, específicamente el nivel 2B. Esta misión proporciona ventajas significativas en términos de frecuencia de imágenes, capturando datos cada 5 días, lo cual es crucial para detectar cambios rápidos en los cultivos de naranjo.

La alta frecuencia reduce las interferencias meteorológicas y asegura una cobertura consistente y de alta calidad, esencial para obtener series temporales precisas del NDVI. Con una resolución espacial de 10 metros, se facilita un monitoreo detallado de las afectaciones causadas por el HLB en los cultivos.

Para procesar las imágenes satelitales de Sentinel-2 y exportar los datos de las series de tiempo en formato *.csv* (Valores Separados por Comas), se utilizó la plataforma Google Earth Engine.

Los datos se filtraron por fecha (01/12/2017 - 01/03/2024), área de interés y un porcentaje máximo de nubes del 75%. Este umbral se estableció con el fin de evitar la eliminación arbitraria de una gran cantidad de datos. Posteriormente, en el preprocesamiento se detalla cómo se llevó a cabo la limpieza para la eliminación del ruido intrínseco.

De la misma manera, se aplicó una máscara basada en el producto QA60 de Sentinel-2 para descartar información afectada por la presencia de cirrostratos y nimbo cúmulos, lo que permitió equilibrar la disponibilidad y calidad de los datos. Esto aseguró un monitoreo continuo y detallado de los cultivos, al extraer información espacial precisa de los píxeles, a pesar de las condiciones meteorológicas variables.

Se utilizó el Índice de Vegetación de Diferencia Normalizada (NDVI), calculado a partir de las bandas B8 (infrarrojo cercano) y B4 (rojo) de Sentinel-2, las cuales son sensibles a la reflectancia de la vegetación. Esto facilita la detección de cambios en la salud de los cultivos.

Se obtuvieron 315 imágenes que contienen un total de 286,650 registros de fechas, latitud, longitud y valores de NDVI (columnas) correspondientes a 910 píxeles (filas). Cada píxel tiene una resolución de 10x10 metros, lo que equivale a una superficie de 100 metros cuadrados en el terreno, cubriendo un área total de 9 hectáreas.

4.2.2. Limpieza de ruido y normalización

Se llevó a cabo una identificación y eliminación de *outliers* en la serie de tiempo NDVI de Sentinel-2, con el objetivo de detectar y remover valores atípicos causados por errores en la captura de datos o condiciones atmosféricas anómalas, como nubes o sombras. Este proceso forma parte de la limpieza de ruido y normalización de los datos. Para ilustrar este procedimiento y facilitar su visualización, se presentará un solo píxel del conjunto de datos mencionado (Figura 14).

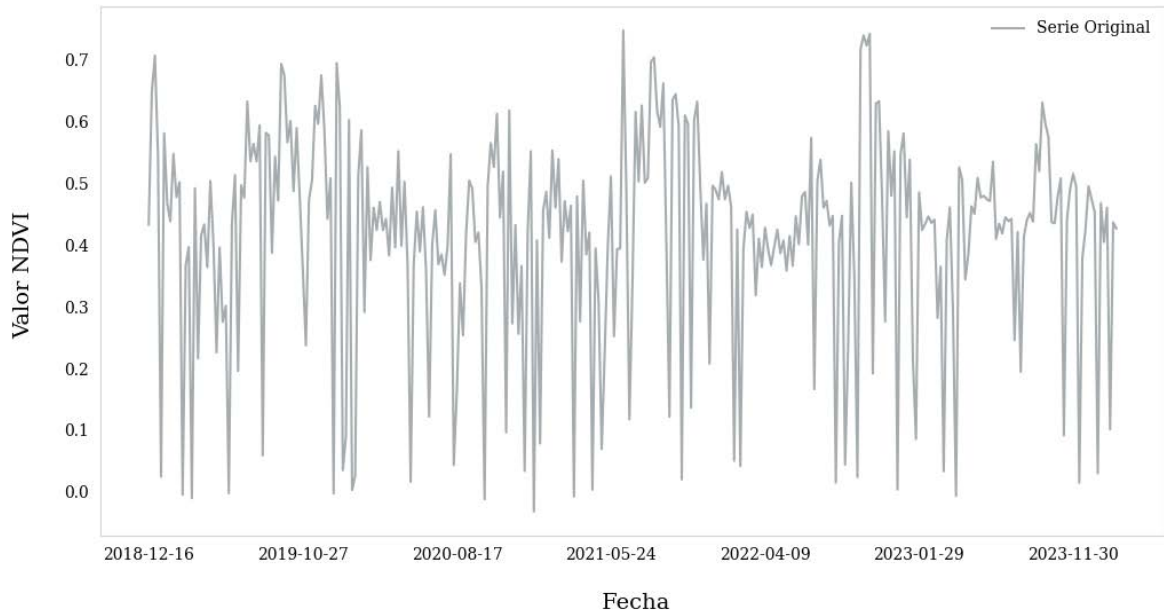


Figura 14: Serie de tiempo original de un solo píxel antes de la limpieza
Fuente: Elaboración propia.

Para detectar estos valores atípicos, se estableció un umbral relativo de 0.1, seleccionado para identificar picos anómalos significativos sin capturar fluctuaciones menores que podrían ser ruido o errores de medición.

Este umbral del 10% resalta cambios abruptos en los datos que no corresponden a variaciones naturales esperadas en los valores de NDVI, asegurando que solo los picos más notables sean considerados anómalos y facilitando la identificación de errores sin sobrecargar el análisis.

Para mejorar la detección, se invirtió la serie de tiempo multiplicando los valores por -1, convirtiendo los picos anómalos en mínimos locales. Así, la función *find_peaks* pudo identificar automáticamente estos mínimos, que corresponden a los picos anómalos en la serie original.

Una vez detectados los picos, se eliminaron los valores correspondientes a estos mínimos, limpiando los datos de valores anómalos. Esto ayuda a reducir el ruido en la serie de tiempo, asegurando que los datos reflejen más fielmente los cambios reales en la vegetación, sin las distorsiones introducidas por condiciones atmosféricas adversas o errores en la captura de los datos.

Después de la limpieza de picos anómalos, se aplicó un umbral mínimo de 0.2 para

filtrar valores de NDVI inferiores a este nivel, los cuales suelen estar asociados con reflectancias de nubes, sombras u otros factores que no representan correctamente la vegetación.

Este umbral se determinó considerando la distribución de los valores NDVI observados en la región de estudio. Específicamente, se verificó mediante un análisis visual que los valores por debajo de este límite se correlacionan consistentemente con áreas sin vegetación.

Al aplicar este filtro, se eliminan datos irrelevantes, contribuyendo a la normalización de la serie y mejorando la confiabilidad del análisis de los patrones de vegetación a lo largo del tiempo (Figura 15).

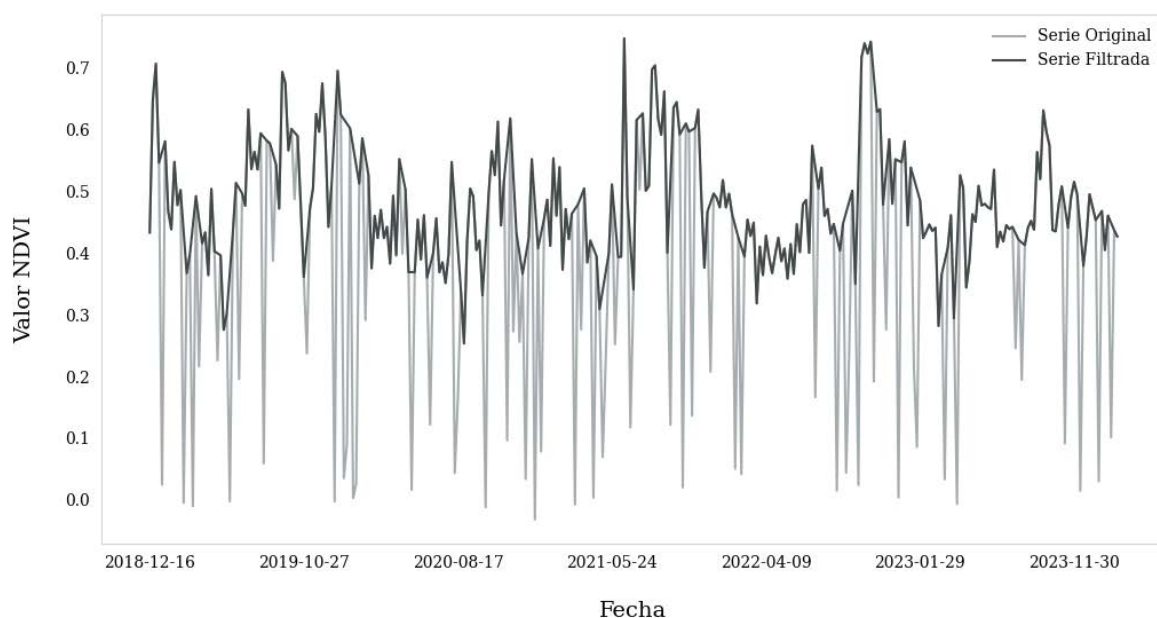


Figura 15: Filtrado de picos y umbrales mínimos en la serie de tiempo de un solo píxel
Fuente: Elaboración propia.

Como resultado del proceso, se identificaron 53 picos anómalos y 8 valores por debajo del umbral mínimo, eliminando un total de 61 valores. Finalmente, la serie de tiempo se compone de 910 filas y 254 columnas que incluyen fechas, latitud, longitud y valores de NDVI asociados (Figura 16).

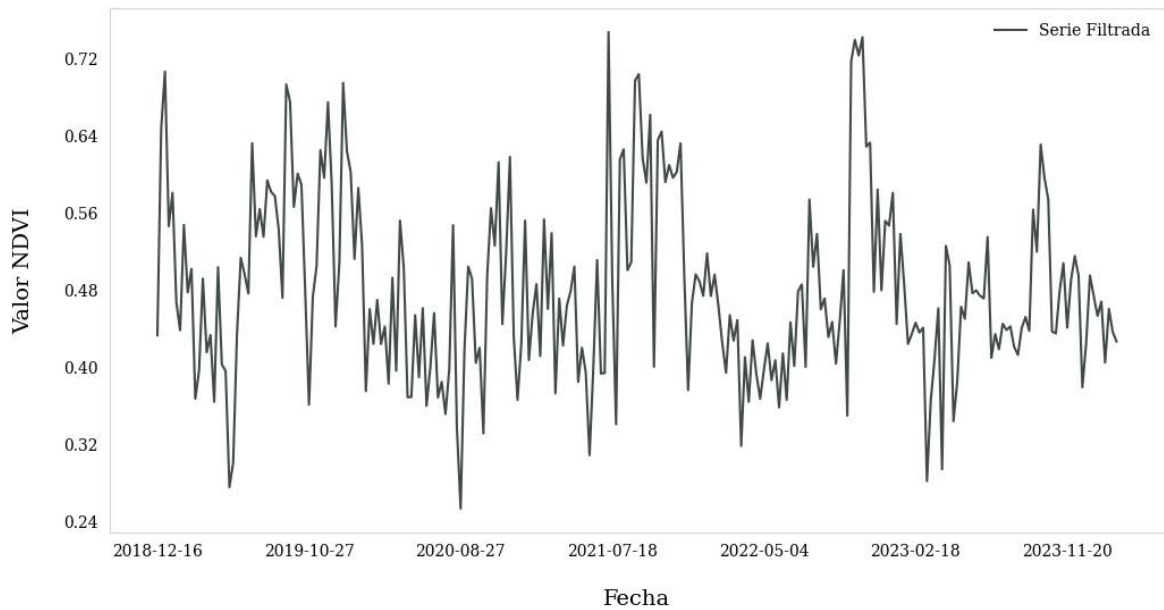


Figura 16: Serie de tiempo final de un solo píxel posterior a la limpieza
Fuente: Elaboración propia.

Este enfoque evita la eliminación arbitraria de fechas o píxeles, lo que proporciona un conjunto de datos más limpio y robusto para el análisis.

4.2.3. Identificación de clases

En este apartado, se definieron las coordenadas del conjunto de datos NDVI utilizando el Sistema de Referencia de Coordenadas (CRS) EPSG:4326, que corresponde a coordenadas geográficas (latitud/longitud). Posteriormente, se reprojectaron al EPSG:32614 que representa coordenadas proyectadas (X, Y) compatibles con el ortomosaico de la zona de estudio.

Luego, las coordenadas transformadas se plotean sobre el ortomosaico en forma de puntos y sus valores asociados al NDVI, evidenciando la ubicación espacial de cada señal en la huerta (Figura 17).

Esta integración permite corroborar la capacidad de Sentinel-2 para captar información relevante a nivel detallado en los cultivos, facilitando la correlación entre los datos temporales de las series de tiempo y el mapeo espacial de alta resolución obtenido mediante el vuelo fotogramétrico con VANT.



Figura 17: Visualización del ortomosaico y los puntos reproyectados
Fuente: Elaboración propia.

Posteriormente, se exportó en formato *shapefile*, que puede ser utilizado en software de Sistemas de Información Geográfica (SIG). Para procesar y analizar estos datos espaciales, se utilizó el software QGIS, que facilitó la visualización, edición y comparación del *shapefile* de puntos sobre el ortomosaico.

La comparación visual permitió identificar diferentes grados de afectación en los cultivos, revelando una correlación directa entre los valores del NDVI y el estado de salud de los árboles en diferentes zonas de la huerta.

La metodología de clasificación se adaptó específicamente a la zona de estudio, un área de 9 hectáreas completamente afectada por HLB. Esta premisa se fundamenta en observaciones realizadas durante el trabajo de campo, en colaboración con el Dr. Moisés Roberto Vallejo Pérez, especialista en fitopatología, y el Ing. José Rene Moreno López, coordinador de la Campaña “Plagas Reglamentadas de los Cítricos” del Comité Estatal de Sanidad Vegetal de San Luis Potosí (CESAVESLP). Estas observaciones fueron complementadas con información proveniente de informes fitosanitarios previos.

El diagnóstico inicial y las condiciones de la zona afectada permitieron confirmar que, aunque toda la superficie estaba comprometida por HLB, existían variaciones significativas en la intensidad de los síntomas visibles en hojas y ramas. Esto se reflejó en el rango de valores de NDVI observado en la serie de tiempo, que osciló entre 0.2 y 0.8, facilitando la identificación de clases con distintos niveles de estrés vegetativo.

Para clasificar los grados de afectación en base al NDVI, se empleó la técnica de terciles, dividiendo los valores en tres clases de igual tamaño que representan aproximadamente el 33.33% de las observaciones en cada grupo. Esta técnica no depende de un límite fijo, sino que se basa en la distribución de los datos, lo que permite una segmentación equilibrada. Al evitar la influencia de valores extremos, se asegura que las clases representen fielmente los diferentes grados de afectación sin desbalancear el análisis.

Este proceso se alinea con las necesidades prácticas del proyecto, lo que permite una distribución equilibrada de los datos y facilita el monitoreo continuo de la salud de los cultivos a lo largo del tiempo.

Se generó un histograma (Figura 18) para visualizar la distribución de los valores de NDVI.

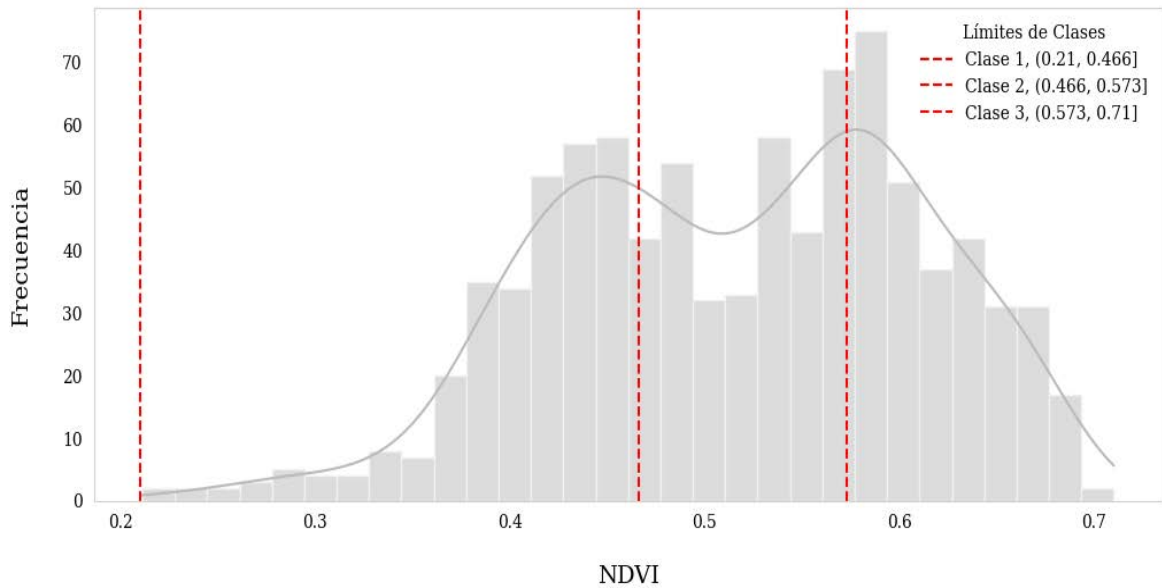


Figura 18: Identificación de clases en base a Distribución NDVI
Fuente: Elaboración propia.

En este gráfico, se señalaron con líneas verticales los límites de cada clase, permitiendo una representación clara de los rangos de clasificación: Clase 1 (tercil inferior), Clase 2 (tercil medio) y Clase 3 (tercil superior) de NDVI.

Al aplicar la técnica de terciles, se garantiza que cada clase contenga un número similar de observaciones, evitando sesgos en rangos de NDVI con más o menos datos. Esta clasificación es fundamental en el contexto de los árboles infectados, ya que las condiciones pueden variar considerablemente entre individuos.

La clasificación por terciles facilita un análisis representativo de la salud de los cultivos, permitiendo evaluar la superposición de los límites de clase con la distribución de los datos. Los rangos establecidos reflejan de manera adecuada el estado de salud de los árboles, logrando que la clasificación empírica sea consistente con las características específicas de la zona de estudio.

Las características específicas de cada clase son las siguientes:

- **Clase 1 (Afectación severa):** $0.2 < \text{NDVI} \leq 0.46$
 - Los síntomas afectan todo el árbol, con hojas que presentan amarillamiento severo, manchas y deformaciones.
 - El árbol muestra un estado de salud muy deteriorado, con crecimiento re-

ducido y presencia de ramas secas o en proceso de morir (Figura 19).



Figura 19: Árbol gravemente afectado por HLB capturado a 30 metros
Fuente: Elaboración propia.

■ **Clase 2 (Afectación moderada):** $0.46 < \text{NDVI} \leq 0.57$

- Los síntomas afectan una parte considerable del árbol, pero no todo. Se observa un amarillamiento ligero y moteado en las hojas.
- El árbol presenta un crecimiento reducido, con algunas ramas mostrando signos de estrés, como reducción en el tamaño de las hojas y pérdida de vigor (Figura 20).



Figura 20: Árbol moderadamente afectado por HLB capturado a 30 metros
Fuente: Elaboración propia.

- **Clase 3 (Afectación leve):** $NDVI > 0.57$
 - Los síntomas son localizados y afectan áreas específicas del árbol, con hojas que muestran un leve amarillamiento.
 - El árbol mantiene una densidad foliar significativa, y solo algunas ramas muestran signos de estrés leve (Figura 21).



Figura 21: Árbol levemente afectado por HLB capturado a 30 metros
Fuente: Elaboración propia.

Se utilizó la calculadora de campos en QGIS para crear una columna en la tabla de atributos que identificara automáticamente las clases según los rangos de NDVI establecidos. Los valores bajos (> 0.2 y ≤ 0.46) indican un deterioro severo, los valores intermedios (> 0.46 y ≤ 0.57) reflejan un estrés moderado, y los valores altos (> 0.57) corresponden a zonas con árboles infectados que mantienen cierta actividad fotosintética, es decir, afectación leve.

Posteriormente, los datos fueron exportados a un archivo *.csv* para su uso en la fase de implementación.

4.2.4. Implementación y Evaluación

La implementación y evaluación de métodos para el análisis de datos se fundamentan en la combinación de técnicas de clasificación y modelos estadísticos, permitiendo identificar patrones y realizar pronósticos.

Se emplearán dos técnicas clave: el algoritmo de clasificación KNN y el modelo SARI-MA. Ambas buscan mejorar el monitoreo extensivo del HLB en los cultivos de naranjo, abarcando tanto su evolución temporal como su distribución espacial.

A continuación, se profundizará en los detalles de cada metodología y su aplicación en este estudio.

4.2.4.1. Algoritmo KNN

Para clasificar la severidad del HLB en cultivos de naranjo, se utilizó el algoritmo de clasificación supervisada KNN, adecuado para este tipo de problemas debido a su capacidad para agrupar datos en función de la proximidad en el espacio de características.

En este caso, se empleó el parámetro de distancia ponderada, lo que otorga mayor peso a los vecinos más cercanos, mejorando la precisión en las predicciones de los valores ubicados cerca de las fronteras entre clases. Este ajuste permitió que el algoritmo diera más relevancia a las observaciones cercanas, lo que resultó en una clasificación más precisa, especialmente para los valores próximos a los límites de severidad del HLB.

Se utilizó la serie de tiempo con valores NDVI correspondiente a un solo píxel, con mediciones realizadas entre el 16/12/2018 y el 28/02/2024. El conjunto de datos está compuesto por 910 observaciones, a las cuales se les asignaron clases específicas basadas en el nivel de estrés vegetativo asociado a la afectación por HLB, de acuerdo con rangos predefinidos de NDVI.

De estos, 637 instancias (70 %) se asignaron al conjunto de entrenamiento, y 273 (30 %) se destinaron al conjunto de prueba. Durante el entrenamiento, el algoritmo aprende a identificar las relaciones entre los valores de NDVI y los grados de afectación. Posteriormente, se evaluó su capacidad para predecir correctamente la clase de afectación (leve, moderada o severa) al comparar las características de un punto no clasificado con las de sus k vecinos más cercanos previamente clasificados.

Para optimizar el rendimiento del modelo, se aplicó validación cruzada con 5 particiones. Esta técnica asegura una adecuada capacidad de generalización del algoritmo, minimizando el riesgo de sobreajuste y mejorando la fiabilidad de las predicciones sobre datos no vistos. La validación se repitió para diferentes valores de k , entre 1 y 20, y

para cada uno de estos se calcularon la precisión media y la desviación estándar.

Para evaluar los resultados del modelo implementado, se calculó la precisión global junto con métricas complementarias, como la raíz del error cuadrático medio (RMSE), que permiten cuantificar de manera más precisa las diferencias entre los valores predichos y los reales. Además, se construyó la matriz de confusión, herramienta fundamental para el análisis detallado del rendimiento del clasificador, al proporcionar información sobre las predicciones correctas e incorrectas para cada clase.

A partir de esta matriz, se derivaron métricas clave como *precision*, *recall*, *F1-score* y *support*, las cuales permiten evaluar el equilibrio entre la sensibilidad y la exactitud en la detección de cada clase. Estos resultados serán presentados con mayor detalle en el siguiente capítulo, donde se analizará el desempeño del modelo a partir de los indicadores mencionados.

4.2.4.2. Modelo SARIMA

Este procedimiento, aunque independiente del proceso anterior, es fundamental para asegurar la coherencia en el siguiente capítulo, específicamente en el apartado de integración. Su correcta implementación garantiza que los pronósticos se relacionen y alineen adecuadamente con los análisis previos.

Para realizar el pronóstico con base en la serie de tiempo NDVI, se utilizó un modelo SARIMA, porque permite realizar pronósticos al identificar las relaciones intrínsecas dentro de la propia serie de tiempo, sin la necesidad de incluir otras variables externas, lo que lo distingue de los enfoques de regresión.

El método aplicado en este estudio para determinar los parámetros óptimos de pronóstico es la optimización de hiperparámetros, un proceso que selecciona los valores de los parámetros que maximizan el rendimiento del modelo.

La optimización de los hiperparámetros se llevó a cabo en dos etapas: primero, una búsqueda aleatoria y luego una evaluación exhaustiva de las mejores combinaciones de parámetros identificadas.

En la primera etapa, se realizó una búsqueda aleatoria que abarcó un amplio rango de configuraciones, permitiendo identificar combinaciones prometedoras sin evaluar todas las posibilidades. Esto ayuda a reducir el tiempo de procesamiento y a centrar la atención en aquellas combinaciones que pueden mejorar el rendimiento del modelo. Una vez finalizada la búsqueda aleatoria, se seleccionaron las mejores combinaciones de parámetros para una evaluación más detallada.

En esta segunda etapa, se probaron un conjunto predefinido de combinaciones de órdenes para el modelo SARIMA y se ajustó el modelo a la serie de tiempo, evaluando el rendimiento mediante métricas como el criterio de información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC).

Los resultados de los modelos SARIMA evaluados se muestran en la Tabla 3:

Tabla 3: Comparación de Modelos SARIMA

p, d, q	P, D, Q, m	AIC	BIC
1, 0, 1	1, 1, 0, 12	-399.601	-385.646
1, 1, 1	1, 1, 0, 12	-386.173	-372.234
1, 1, 0	1, 1, 0, 12	-351.685	-341.231
1, 0, 1	0, 1, 0, 12	-339.696	-329.229
0, 1, 0	1, 1, 0, 12	-320.701	-313.732

El modelo SARIMA (1, 0, 1) x (1, 1, 0, 12) presentó los valores de AIC (-399.601) y BIC (-385.646) más bajos, indicando su idoneidad para capturar los patrones de la serie de tiempo con una estructura simple, evitando así el sobreajuste y maximizando la capacidad de pronóstico.

Este modelo seleccionado emplea un término autorregresivo (AR) de primer orden y un término de media móvil (MA) de primer orden en los componentes regulares (1, 0, 1), sin necesidad de diferenciación adicional.

En el componente estacional (1, 1, 0, 12), utiliza un término autorregresivo estacional de primer orden y una diferenciación estacional de primer orden, con periodicidad de 12 meses para reflejar la estacionalidad anual de los datos, lo que permite capturar la dependencia y estacionalidad anual en la serie de tiempo. Para validar el modelo, se examinan los coeficientes estimados en la Tabla 4, junto con su significancia estadística:

Tabla 4: Significancia Estadística en el Modelo SARIMA

Parámetros Estimados	Coef.	std err	z	$P > z $	[0.025	0.975]
AR.L1	0.8716	0.051	17.015	0.000	0.771	0.972
MA.L1	-0.5310	0.088	-6.028	0.000	-0.704	-0.358
AR.S.L12	-0.4881	0.048	-10.265	0.000	-0.581	-0.395
sigma ²	0.0107	0.001	13.053	0.000	0.009	0.012

Los coeficientes de los términos *AR* y *MA* son significativos, con *p*-valores por debajo de 0.05, lo que indica que el modelo captura efectivamente la estructura temporal y estacional de los datos. El bajo valor de la varianza residual ($\text{sigma}^2 = 0.0107$) sugiere una buena precisión en los pronósticos del modelo.

Para evaluar el modelo, la serie de tiempo NDVI se dividió en un conjunto de entrenamiento (80% de los datos) y uno de validación (20% restante). Esta división permite al modelo aprender de los datos históricos y evaluar su rendimiento en datos no vistos, crucial para medir la capacidad de generalización.

En la fase de pronóstico, se ajustó el modelo SARIMA con configuración $(1, 0, 1) \times (1, 1, 0, 12)$ utilizando los datos del conjunto de entrenamiento, capturando patrones temporales y estacionales presentes en la serie. Posteriormente, se generaron pronósticos para el conjunto de validación con el objetivo de evaluar el desempeño del modelo. Los resultados obtenidos, incluyendo los pronósticos, fueron exportados a un archivo *.csv* para facilitar su análisis y uso posterior.

Para evaluar la implementación del modelo SARIMA, se calcularon las métricas de Error Cuadrático Medio (RMSE) y Error Absoluto Medio (MAE), con el objetivo de medir la precisión del modelo en el pronóstico de la serie de tiempo NDVI. Estas métricas proporcionan una evaluación cuantitativa del desempeño del modelo, indicando la magnitud de los errores entre los valores predichos y los observados.

Además, se realizó un análisis exhaustivo de los residuos generados por el modelo, utilizando gráficos de diagnóstico para verificar su comportamiento y comprobar que cumplen con los supuestos fundamentales de independencia, normalidad y homocedasticidad.

La interpretación de estos resultados no solo permite validar la eficacia del modelo, sino también identificar posibles áreas de ajuste o mejora.

El análisis de los residuos desempeña un papel clave en la comprensión de las limitaciones del modelo y en la detección de patrones no capturados en la serie de tiempo, lo cual resulta fundamental al trabajar con datos complejos relacionados con la salud de los cultivos. A través de esta evaluación, se garantiza que los pronósticos obtenidos sean consistentes y puedan utilizarse como herramienta fiable en la toma de decisiones.

Los resultados y las implicaciones de estas métricas, así como los hallazgos derivados del análisis de los residuos, serán explicados con mayor detalle en el siguiente capítulo, donde se abordará su impacto directo en la gestión y pronóstico de los cultivos de naranjo afectados por HLB.

Estos fundamentos metodológicos permiten no solo obtener datos clave para la gestión agrícola, sino también sentar las bases para la interpretación de los resultados.

A continuación, se abordarán los hallazgos derivados de la aplicación de estas herramientas, evaluando su precisión y utilidad a través de la discusión de un caso práctico que ejemplifican cómo los resultados pueden apoyar la toma de decisiones en el manejo agrícola.

La conexión entre la metodología aplicada y los resultados obtenidos será crucial para comprender las implicaciones prácticas de los enfoques utilizados y su potencial impacto.

Capítulo 5

Resultados y Discusión

En este capítulo se presentan en detalle los resultados obtenidos del monitoreo espacio-temporal de los cultivos de naranjo afectados por HLB en la zona de estudio, mediante el uso de los modelos KNN y SARIMA, así como la superposición de estos datos en el ortomosaico.

Se proporciona una evaluación visual de las afecciones causadas por la enfermedad, considerando su distribución espacial. La discusión aborda cómo estos resultados contribuyen a una comprensión más profunda del impacto del HLB en el manejo de los cultivos, destacando cómo la integración de análisis espacial y temporal puede optimizar las estrategias de manejo en la industria citrícola.

5.1. Resultados de la clasificación con KNN

El análisis de los resultados del algoritmo KNN muestra un rendimiento consistente, indicado por una desviación estándar baja (menos de 0.02) en los experimentos realizados.

Esto sugiere que el modelo tiene una buena capacidad para generalizar, ya que su rendimiento no depende en gran medida de las particiones específicas del conjunto de datos. Si la desviación estándar fuera alta, significaría que el modelo es sensible a las particiones, lo que afectaría su capacidad de generalización.

En la Figura 22, se muestra cómo varía el rendimiento del modelo según los diferentes valores de k . Los gráficos revelan que, a medida que k se acerca al valor óptimo, tanto la precisión media como la desviación estándar se estabilizan.

Esto resalta la importancia de elegir un valor adecuado para k , ya que un valor bien seleccionado reduce la variabilidad del modelo, mejorando su capacidad para generalizar y clasificar de manera más precisa los grados de afectación por HLB en los cultivos de naranjo.

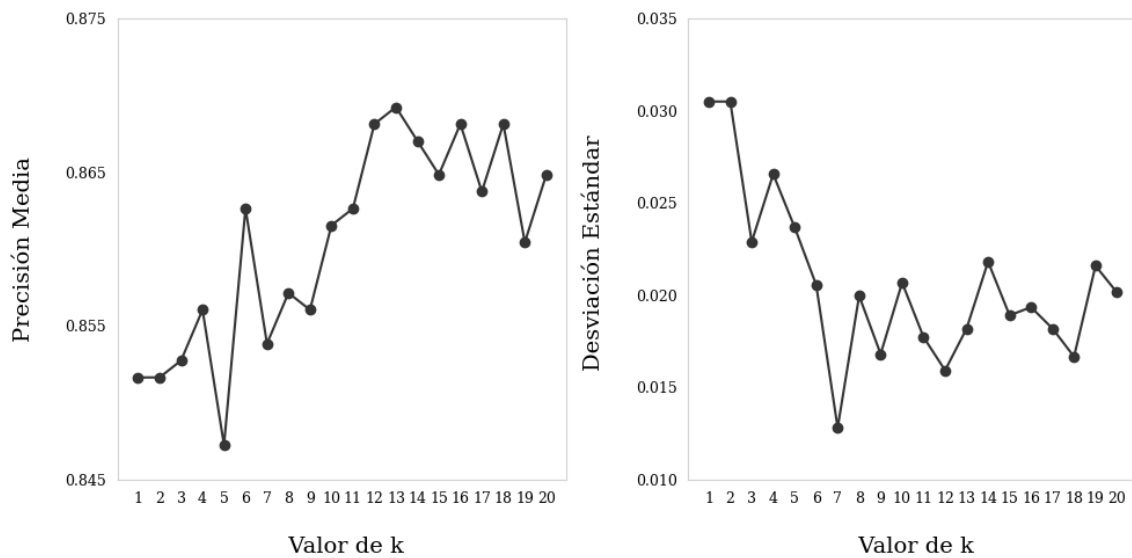


Figura 22: Precisión Media y Desviación Estándar de KNN para Diferentes Valores de k
Fuente: Elaboración propia.

Es fundamental distinguir que la k de KNN indica el número de vecinos cercanos utilizados en la clasificación para determinar la clase de un dato, mientras que la k de K -Folds se refiere al número de divisiones en que se particiona el conjunto de datos durante la validación cruzada para estimar el desempeño del modelo.

En este contexto, se determinó que el valor óptimo de k para el modelo KNN, que representa el número de vecinos más cercanos utilizados para realizar la clasificación, fue 7.

Este valor se seleccionó utilizando validación cruzada con 5 particiones (K -Folds), un procedimiento estadístico que evalúa el rendimiento del modelo al dividir el conjunto de datos en subconjuntos de entrenamiento y prueba, asegurando una evaluación más robusta y menos dependiente de una única partición.

Para evaluar la consistencia y fiabilidad del modelo, se calculó el intervalo de confianza de la precisión media obtenida durante la validación cruzada. El modelo presentó una precisión media de 0.853, con una desviación estándar de 0.0128. Con un nivel de confianza del 95 %, el intervalo de confianza obtenido indica que la precisión media se mantiene entre el 84.3 % y el 86.5 %.

Este intervalo estrecho sugiere que el modelo mantiene una alta estabilidad y consistencia en la clasificación de los grados de afectación en los árboles de naranjo a través de diferentes muestras de datos. Además, demuestra que el modelo no está sobreajustado y que generaliza adecuadamente.

El modelo KNN logró una precisión general de 86.08 %, lo que evidencia su capacidad considerable para diferenciar entre las tres clases de afectación: severa (Clase 1), moderada (Clase 2) y leve (Clase 3). Además, el modelo presentó un RMSE de 0.373, indicando que los errores de predicción son moderados y que las predicciones generadas son, en general, buenas.

Para complementar este análisis, se generó la matriz de confusión, que permitió calcular métricas clave como *precision* (proporción de predicciones correctas), *recall* (capacidad para identificar instancias positivas), *F1-score* (balance entre *precision* y *recall*) y *support* (número de instancias reales por clase); las cuales son fundamentales para evaluar el rendimiento del modelo en la tarea de clasificación.

Tabla 5: Métricas de evaluación por clase

<i>Afectación</i>	<i>precision</i>	<i>recall</i>	<i>F1-Score</i>	<i>Support</i>
Severa (Clase 1)	0.86	0.90	0.88	87
Moderada (Clase 2)	0.84	0.79	0.81	105
Leve (Clase 3)	0.89	0.91	0.90	81

Al examinar el rendimiento en detalle, el reporte de clasificación (Tabla 5) y la matriz de confusión (Figura 23) revelan algunos puntos de interés.

Etiquetas Reales	Severa	78	9	0
	Moderada	13	83	9
	Leve	0	7	74
		Severa	Moderada	Leve
		Etiquetas Predichas		

Figura 23: Matriz de Confusión por clase
Fuente: Elaboración propia.

La Clase 1, correspondiente con la afectación severa, muestra una precisión de 0.86 y un *recall* de 0.90, lo que significa que el modelo identificó correctamente el 90 % de los árboles severamente afectados. Esta capacidad de detección es crucial para el manejo de la enfermedad, ya que una identificación precisa de los casos graves permite tomar medidas preventivas más rápidas y efectivas. Sin embargo, se observó una tasa de error en 9 casos en los que la Clase 1 fue clasificada erróneamente, lo cual podría estar relacionado con las características de borde entre las clases de severidad.

La Clase 2, que representa la afectación moderada, alcanzó una precisión de 0.84 y un *recall* de 0.79, siendo esta clase la que presentó más dificultad en términos de clasificación. El modelo confundió la Clase 2 tanto con la Clase 1 (en 13 casos) como con la Clase 3 (en 9 casos), lo que podría indicar una similitud en las características de los árboles moderadamente afectados con los de otras clases.

Esto puede estar vinculado a la naturaleza intermedia de los síntomas en los árboles con afectación moderada, los cuales pueden compartir características tanto con las fases de afectación leve como severa. Estos resultados sugieren que la Clase 2 podría beneficiarse de un ajuste de parámetros o incluso de la inclusión de características adicionales para mejorar la discriminación en esta categoría.

Por último, la Clase 3, que representa la afectación leve, demostró un excelente desempeño con una precisión de 0.89 y un *recall* de 0.91, reflejando una alta capacidad del modelo para identificar los casos leves. Esta precisión en la clasificación de la afectación leve es importante para evitar intervenciones innecesarias y optimizar los recursos en el manejo de los árboles infectados. La tasa de error fue baja para esta clase, con solo 7 casos clasificados incorrectamente como Clase 2, lo cual indica que el modelo tiene una buena capacidad para diferenciar la afectación leve de las demás clases.

El vector de puntos clasificados con el modelo KNN se visualiza sobre el ortomosaico, diferenciando con colores los grados de afectación por HLB. En la Figura 24 (a), se muestran los puntos correspondientes a las clases reales utilizadas durante el entrenamiento del modelo. En contraste, la Figura 24 (b), presenta los puntos clasificados por el modelo, destacando tanto las predicciones correctas como los errores de clasificación.

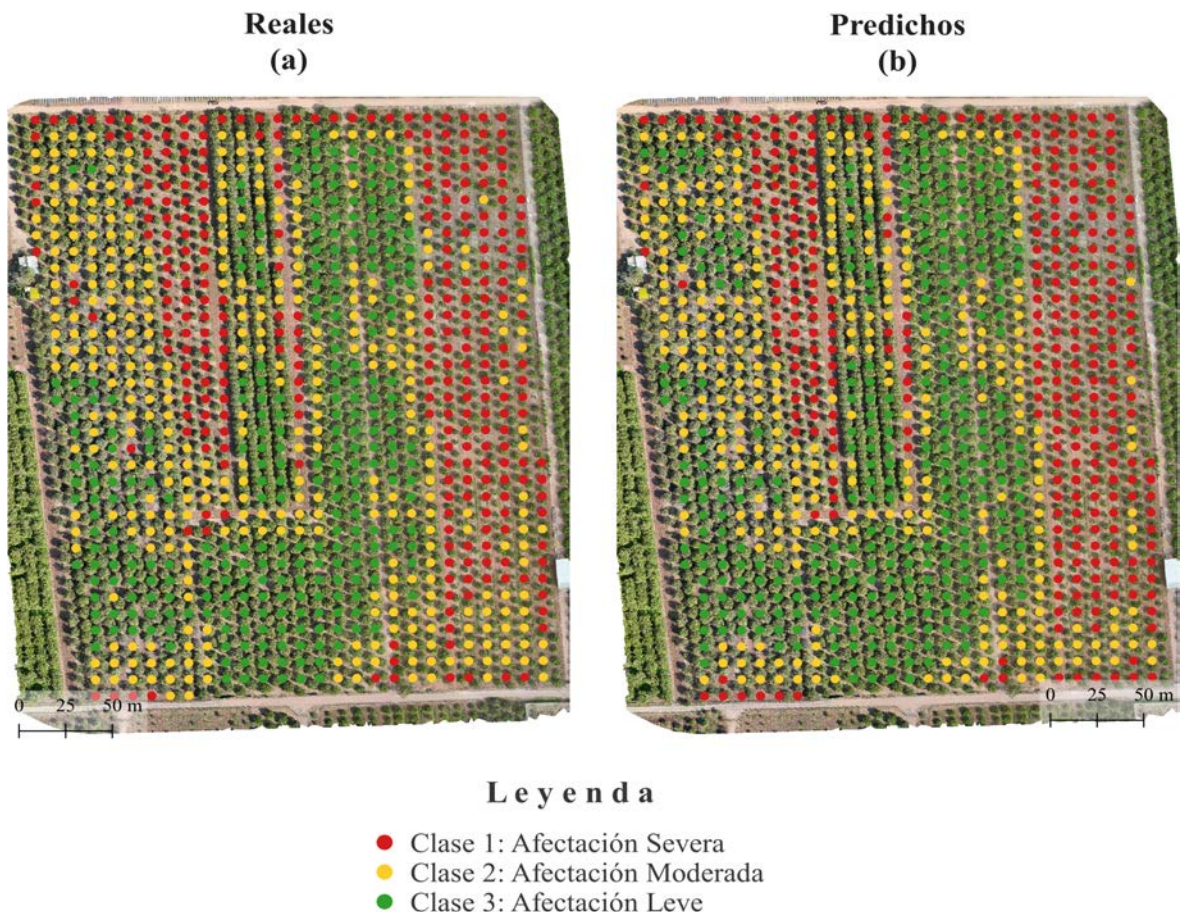


Figura 24: Clasificación con KNN: Reales vs Predichos
Fuente: Elaboración propia.

Las zonas identificadas en rojo corresponden a cultivos severamente afectados, donde los árboles presentan escasa foliación, algunos ya talados y otros en proceso de crecimiento, además de suelos desnudos. Esta clasificación refleja una baja actividad fotosintética, lo cual es coherente con los valores bajos de NDVI que indican una salud vegetal comprometida.

A su vez, las áreas clasificadas en amarillo corresponden a cultivos moderadamente afectados. En estas zonas, los árboles muestran un verdor reducido, con ramas muertas en el ápice y suelos parcialmente desnudos. Aunque presentan signos de estrés, la afectación no es tan grave como en las áreas rojas. La disminución de la actividad fotosintética es evidente, pero los árboles aún conservan una cantidad significativa de follaje y crecimiento, lo que indica una fase intermedia de afectación.

Por último, las zonas clasificadas en verde corresponden a cultivos levemente afectados, los cuales presentan una mayor frondosidad y un verdor más pronunciado. En estas áreas se puede observar que la actividad fotosintética es relativamente alta.

Si bien se identifican desafíos en la clasificación de la afectación moderada (color amarillo), debido a su naturaleza intermedia y características compartidas con otras clases, la visualización de los resultados sobre el ortomosaico confirma que el modelo logra identificar la mayoría de las zonas críticas afectadas por HLB.

5.2. Resultados del pronóstico de valores futuros con SARIMA

El modelo SARIMA $(1, 0, 1) \times (1, 1, 0, 12)$, ajustado sobre la serie de tiempo de NDVI, ha mostrado un rendimiento sólido al pronosticar la condición de los árboles de naranjo infectados por HLB.

Este modelo fue elegido por su capacidad para capturar tanto la estacionalidad anual como las dependencias temporales a corto plazo, lo que permite evaluar la salud de la vegetación con métricas de error adecuadas.

Con un RMSE de 0.0829 y un MAE de 0.0619, el modelo presenta baja dispersión en sus pronósticos frente a los valores reales del conjunto de validación, sugiriendo un buen ajuste a las características de la serie de tiempo.

La elección del modelo SARIMA responde a la necesidad de capturar los patrones estacionales anuales inherentes al ciclo de crecimiento de los cultivos, permitiendo al modelo ajustar fluctuaciones naturales en el NDVI relacionadas tanto con la fenología del cultivo como con el impacto del HLB.

Este ajuste del modelo logra una mayor precisión en la identificación de patrones de deterioro en la vegetación, los cuales reflejan distintos grados de afectación: leve, moderado y severo.

Para evaluar la adecuación del modelo SARIMA ajustado, se llevó a cabo un análisis exhaustivo de los residuos mediante gráficos de diagnóstico. Este análisis examinó elementos clave, como la distribución, independencia y estructura de los residuos, para confirmar que el modelo cumple con los supuestos estadísticos fundamentales (Figura 25).

En primer lugar, el gráfico de residuos estandarizados de la Figura 25 (a), muestra cómo estos residuos se distribuyen en torno a cero a lo largo del tiempo. La ausencia de patrones definidos y la distribución aleatoria alrededor de cero indican que no hay heterocedasticidad en los datos, es decir, la varianza de los errores es constante. Asimismo, la falta de tendencias en los residuos sugiere que estos pueden considerarse como ruido blanco, lo cual es crucial para la validez del modelo, ya que implica que el modelo ha capturado correctamente las estructuras de los datos sin dejar dependencias no modeladas.

El histograma de residuos con estimación de densidad Kernel (KDE) y la curva de distribución normal estándar $N(0,1)$ mostrado en la Figura 25 (b), permite examinar la distribución de los residuos. La KDE (curva naranja) se aproxima considerablemente a la curva normal (verde), lo que sugiere que los residuos siguen una distribución aproximadamente normal. Este cumplimiento de la normalidad en los residuos es un supuesto clave para muchos modelos estadísticos, ya que garantiza que las inferencias realizadas sobre los parámetros del modelo sean válidas y robustas.

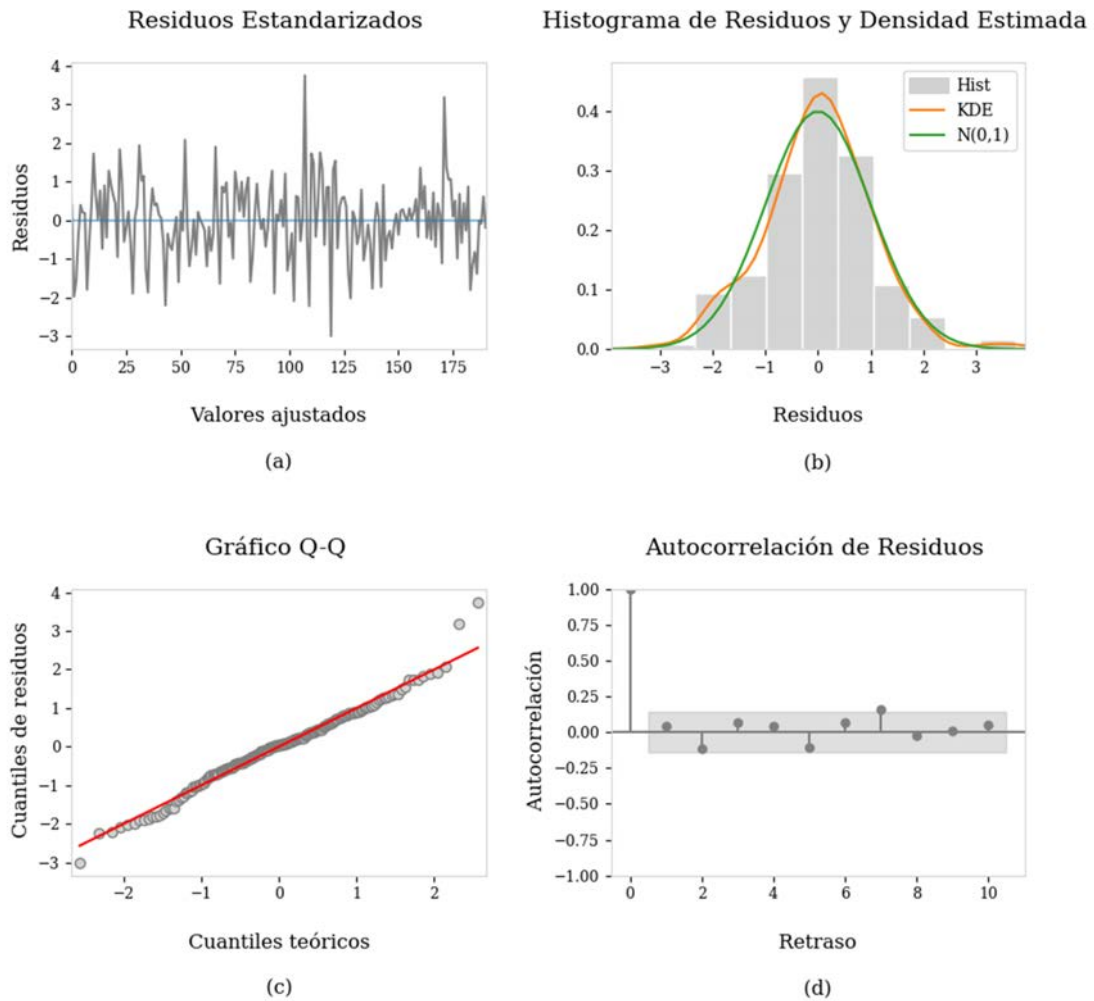


Figura 25: Gráficos de Diagnóstico del Modelo SARIMA para la Serie de Tiempo NDVI
Fuente: Elaboración propia.

El gráfico Q-Q de la Figura 25 (c), compara los cuantiles de los residuos con los cuantiles teóricos de una distribución normal. La alineación de los puntos a lo largo de la línea de referencia (que representa la distribución normal teórica) refuerza la idea de que los residuos se distribuyen normalmente. Este ajuste es importante, ya que sugiere que se cumplen los supuestos de normalidad, lo cual es esencial para la precisión de las pruebas estadísticas y la solidez de las inferencias generadas por el modelo.

El gráfico de autocorrelación de residuos que se observa en la Figura 25 (d), permite observar la independencia de los errores a través de diferentes rezagos. La mayoría de los coeficientes de autocorrelación se encuentran dentro de los intervalos de confianza, lo cual indica la ausencia de autocorrelación significativa en los residuos. Esto implica que los errores son independientes, lo que es un buen indicador de que el modelo ha

capturado la estructura temporal de los datos sin dejar correlaciones residuales.

En conjunto, estos análisis sugieren que el modelo SARIMA ajustado cumple con los supuestos de normalidad, homocedasticidad e independencia de los residuos, validando su idoneidad y su capacidad para realizar pronósticos confiables a futuro. La adecuada distribución y estructura de los residuos garantizan que el modelo puede usarse con seguridad para proyecciones futuras, respaldando así la precisión y solidez de sus resultados.

Para evaluar visualmente el desempeño del modelo SARIMA elegido, se presenta un gráfico comparativo entre los datos históricos, así como los pronósticos en el periodo de validación y futuro (Figura 26).

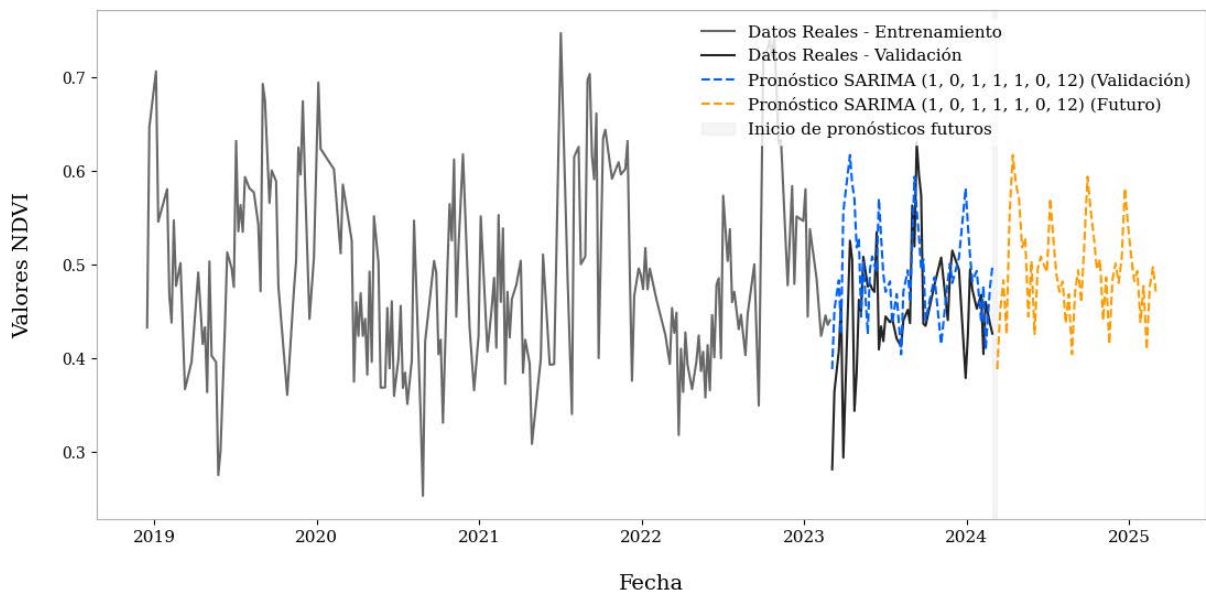


Figura 26: Datos Reales, Pronósticos SARIMA para Validación y Proyección Futura
Fuente: Elaboración propia.

El gráfico muestra la evolución de los valores NDVI a lo largo del tiempo, con el conjunto de entrenamiento en gris y el conjunto de validación en negro, lo que permite evaluar la capacidad del modelo para capturar patrones y realizar pronósticos precisos.

Los pronósticos para el periodo de validación se representan con una línea azul punteada. La cercanía entre estos pronósticos y los datos reales de validación indica que el modelo logra capturar adecuadamente las tendencias y fluctuaciones estacionales del NDVI en el periodo reciente.

Asimismo, se incluyen pronósticos para los próximos 12 meses, representadas por una línea naranja punteada, que proyecta los valores de NDVI más allá de los datos observados. La transición entre el periodo de validación y el periodo proyectado se resalta con un sombreado gris claro, que indica la incertidumbre inherente a los pronósticos al asumir que las condiciones históricas persistirán en el futuro.

Este proceso integral asegura no solo un modelo SARIMA validado mediante la comparación de residuos, sino también una visualización clara de su capacidad de pronóstico, esencial en el contexto de este estudio.

De esta manera, el modelo SARIMA demuestra confiabilidad para representar la condición de la vegetación en los meses posteriores al entrenamiento, ofreciendo una herramienta valiosa para la toma de decisiones basada en proyecciones del estado de los cultivos.

5.3. Integración y Superposición

En este estudio, los modelos SARIMA y KNN se implementan de manera complementaria para abordar el análisis de la severidad y distribución del HLB desde dos perspectivas clave: temporal y espacial. Esta integración busca ofrecer una visión completa que facilite tanto la anticipación de patrones temporales de estrés como la identificación de las zonas más afectadas.

SARIMA se utiliza específicamente para el monitoreo temporal. Este modelo permite capturar patrones estacionales y tendencias a largo plazo con la serie de tiempo NDVI, lo que es crucial para pronosticar los periodos en los que es más probable observar un aumento del estrés en los cultivos de naranjo.

A continuación, se ejemplifica su funcionamiento y se evalúa su desempeño centrándose en el mes de febrero de los años 2023, 2024 y en el pronóstico para 2025. La elección de estas fechas responde a su importancia en capturar los eventos extremos de aumento y disminución del NDVI, proporcionando una evaluación directa de la capacidad del modelo para replicar estas fluctuaciones críticas y anticipar tendencias.

Estas tendencias se interpretan como indicadores de estrés en los cultivos, clasificadas

en tres niveles: severo, moderado y leve.

La estacionalidad del HLB, aunque presente todo el año, tiene implicaciones fisiológicas que varían según el clima y la etapa fenológica. Febrero, como inicio de la actividad fisiológica post-invierno, proporciona un marco temporal consistente para evaluar las diferencias en los niveles de estrés detectados mediante el NDVI.

Además, usar este mes como referencia elimina la influencia de variaciones estacionales más marcadas, como las lluvias intensas del verano o las temperaturas extremas del invierno, permitiendo así un análisis más uniforme y comparativo entre los años evaluados.

Evaluar el modelo en tres años consecutivos como se observa en la Tabla 6, asegura que los pronósticos se basen en la evolución del HLB bajo condiciones similares, por lo que esta valoración se centró en identificar y modelar los patrones asociados a los incrementos y decrementos más significativos registrados dentro de dicho mes, permitiendo evaluar la capacidad del modelo para capturar variaciones críticas en los valores de NDVI y su relación con las condiciones de estrés en los cultivos.

Tabla 6: Incremento y decremento del NDVI en 2023, 2024 y pronóstico para 2025

Fecha incremento	NDVI incremento	Fecha decremento	NDVI decremento	Decremento total
18/02/2023	0.446	08/02/2023	0.424	0.022
03/02/2024	0.468	08/02/2024	0.404	0.064
23/02/2025	0.513	16/02/2025	0.434	0.079

Durante febrero de 2023, el día 18 se observó un incremento en el NDVI hasta alcanzar un valor de 0.446, mientras que el día 8 de febrero del mismo año se registró un decremento hasta 0.424, resultando en una disminución total de 0.022.

Este rango de valores ubica los cultivos dentro de la categoría de deterioro severo, según la clasificación utilizada, ya que ambos valores son menores o iguales a 0.46. El modelo refleja una estabilidad relativa dentro de una condición crítica para los cultivos, lo que es característico de escenarios con daño generalizado y una mínima recuperación durante este periodo.

En febrero de 2024, el incremento se registró el día 3, alcanzando un valor de NDVI de 0.468, mientras que el decremento se dio el día 8 de febrero, con un NDVI de 0.404, resultando en un decremento total más significativo de 0.064.

Este comportamiento muestra una dinámica más variable en comparación con el año anterior, con valores que oscilan dentro del rango de deterioro severo.

Sin embargo, las fluctuaciones en el NDVI podrían estar relacionadas con la respuesta del cultivo a factores externos como cambios en las condiciones ambientales (por ejemplo, humedad del suelo). La capacidad del modelo para capturar estas fluctuaciones es crucial para entender la progresión del daño causado por el HLB y los factores que influyen en su comportamiento.

Finalmente, para febrero de 2025, el día 23 se espera un incremento en el NDVI alcanzando un valor máximo de 0.513, lo que lo posiciona en la transición entre deterioro severo y estrés moderado. Por otro lado, el día 16 se registra un decremento hasta un valor de 0.434, permaneciendo dentro del rango de deterioro severo.

La diferencia total entre ambos valores fue de 0.079, lo que indica una mayor variabilidad en comparación con los años anteriores (0.022 en 2023 y 0.064 en 2024). Este comportamiento sugiere fluctuaciones más amplias en la salud del cultivo, lo que podría estar relacionado con condiciones ambientales más dinámicas o un impacto heterogéneo de las medidas de manejo implementadas.

Las fechas de decremento del NDVI para los años 2023, 2024 y 2025 son fundamentales para este análisis, ya que marcan los puntos de mayor afectación y enfatizan la necesidad de un monitoreo constante en dichas fechas críticas.

Por otro lado, KNN se implementa para realizar un análisis espacial detallado. Este modelo se encarga de clasificar los valores de NDVI asociados con los diferentes niveles de afectación. Basándose en la información obtenida de Sentinel-2, KNN permite identificar las áreas geográficas con mayor riesgo de afectación.

En este sentido, KNN actúa como una herramienta para detectar patrones espaciales que complementan los pronósticos temporales de SARIMA, proporcionando junto con el ortomosaico, un mapa de riesgo que puede facilitar la toma de decisiones en campo.

Por lo tanto, la superposición de los puntos clasificados mediante el algoritmo KNN en el ortomosaico para los años 2023 y 2024, utilizando una escala de colores intuitiva que muestra los tres grados de afectación por HLB, proporciona una perspectiva visual integral de los resultados (Figura 27).

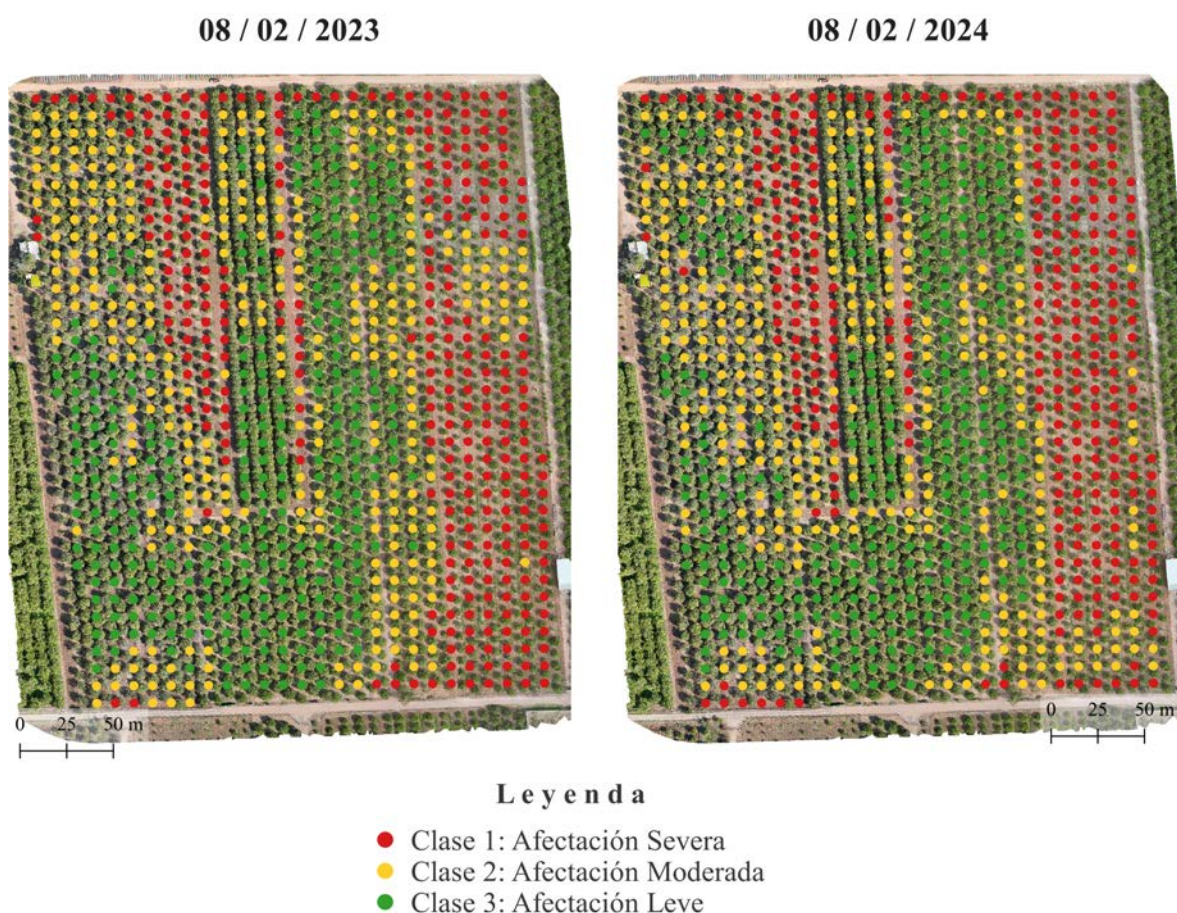


Figura 27: Grado de afectación por HLB en cultivos de naranjo, febrero 2023 y 2024
Fuente: Elaboración propia.

Al observar estos puntos sobre el ortomosaico, es posible correlacionar visualmente las áreas con mayor afectación y proyectar su evolución en el futuro. Esto también facilita la comparación de la situación futura con la situación actual, lo que permite a los gestores identificar las zonas más críticas y priorizar su atención.

La evaluación de los puntos clasificados de los años 2023 y 2024 sobre el ortomosaico ayuda a identificar patrones recurrentes de deterioro, proporcionando una base sólida para anticipar las áreas de mayor riesgo en febrero de 2025, considerando también el pronóstico generado por el modelo SARIMA.

Estos datos previstos para 2025, ofrecen una perspectiva anticipada de las posibles condiciones del cultivo en ese año, basándose en las tendencias históricas de los años anteriores. Al analizar los decrementos pasados y su relación con las áreas afectadas, se puede prever con mayor certeza qué zonas del cultivo podrían experimentar mayores niveles de afectación por HLB.

Al integrar los resultados de esta manera, se facilita la interpretación visual de los datos, ofreciendo una herramienta eficiente para corroborar la correspondencia entre los patrones de afectación detectados, la distribución espacial de la enfermedad y la evolución a futuro.

La combinación de ambos enfoques se justifica porque, mientras SARIMA ayuda a pronosticar cambios en el tiempo, KNN identifica las zonas específicas en las que esos cambios tendrán un mayor impacto. Esta integración resulta particularmente valiosa para diseñar estrategias de monitoreo y manejo del HLB, ya que permite priorizar tanto los periodos críticos como las áreas más vulnerables.

5.4. Discusión de los hallazgos

La integración del algoritmo de clasificación KNN con el modelo SARIMA para evaluar las afectaciones por HLB en cultivos de naranjo, representa un avance importante en el monitoreo espaciotemporal de la enfermedad.

El análisis detallado de los pronósticos y la clasificación, complementadas con una integración precisa de los resultados de las distintas fuentes de datos, ha mostrado la efectividad de las metodologías empleadas. La superposición de estos datos ha revelado patrones consistentes que refuerzan la validez de los enfoques utilizados.

Sin embargo, el estudio también ha identificado ciertas limitaciones, especialmente en las zonas de transición, donde la complejidad fenotípica de los cultivos dificulta la precisión de los modelos implementados. Estos desafíos resaltan la necesidad de mejorar su capacidad para capturar con mayor fidelidad las variaciones en los patrones de afectación.

A pesar de estas dificultades, el enfoque propuesto ofrece una herramienta útil y escalable para la gestión integrada de cultivos, favoreciendo la toma de decisiones informadas y basadas en datos.

Al visualizar tanto el “cuándo” como el “dónde” del grado de afectación por HLB, esta combinación de análisis permite a los tomadores de decisiones diseñar estrategias de intervención más focalizadas y oportunas.

Más allá de las métricas numéricas y los hallazgos técnicos, es fundamental reflexionar sobre las implicaciones de estos resultados para la gestión sostenible de los cultivos. Estos hallazgos abren una puerta a la optimización de estrategias de monitoreo y manejo en tiempo real, lo cual será clave en la evolución de las soluciones basadas en inteligencia artificial, modelos estadísticos para pronóstico mediante series de tiempo y sensores remotos en la agricultura.

Este capítulo no solo se limita a presentar los resultados de los modelos implementados, sino que también marca el punto de partida para las conclusiones, donde se consolidarán los aprendizajes y se discutirán las oportunidades para futuras investigaciones y mejoras.

Capítulo 6

Conclusiones

Este capítulo final presenta las principales contribuciones de esta investigación, resaltando el impacto de la integración de los modelos SARIMA y KNN en el monitoreo espacial y temporal de las afectaciones causadas por HLB en cultivos de naranjo. También se exploran las oportunidades de mejora y expansión del trabajo, con un enfoque en optimizar las metodologías empleadas e incorporar nuevas tecnologías que puedan perfeccionar el manejo y promover la sostenibilidad de los cultivos.

6.1. Contribución de la investigación

Este estudio demuestra cómo la integración SARIMA y KNN, junto con técnicas de teledetección y procesamiento de imágenes, puede optimizar la gestión del HLB en cultivos de naranjo. La generación de ortomosaicos a partir de vuelos con VANT y la clasificación de los valores de NDVI han permitido un análisis detallado de las afectaciones del cultivo en el espacio y el tiempo, proporcionando una visión integral de los patrones de estrés en los árboles.

El modelo SARIMA ha sido clave para identificar periodos críticos en que los cultivos son más vulnerables al estrés, ofreciendo una herramienta valiosa para programar el monitoreo y las intervenciones preventivas en momentos óptimos. Al mismo tiempo, el modelo KNN ha permitido una clasificación espacial con buenos resultados, identificando áreas específicas con mayor severidad de HLB.

La clasificación de NDVI en tres niveles de estrés (severo, moderado y leve) ha facilitado una comprensión granular de la distribución y progresión de la enfermedad. Esta categorización mejora la precisión en las decisiones de manejo del cultivo, permitiendo intervenciones más focalizadas y adecuadas a las condiciones reales de la zona de estudio.

Por otro lado, la visualización de resultados a través de SIG, ha sido clave para validar empíricamente la efectividad de los modelos aplicados. La superposición de ortomosaicos con las clasificaciones de NDVI en QGIS facilita la validación visual, proporcionando a los gestores del cultivo una herramienta intuitiva para identificar áreas de atención prioritaria y monitorear el estado de los árboles.

La generación de indicadores de severidad, que integran las tendencias temporales de SARIMA y la distribución espacial de KNN, constituye un recurso valioso para la planificación de actividades de manejo a lo largo del tiempo. Este índice de riesgo permite priorizar intervenciones y adoptar medidas preventivas antes de que el estrés alcance niveles críticos, contribuyendo a una gestión más sostenible y eficiente de los cultivos.

Estos hallazgos destacan el potencial de combinar métodos de análisis temporal y espacial para abordar los desafíos de la agricultura moderna. Al integrar múltiples enfoques y tecnologías, se obtiene una visión más completa de los factores que afectan la salud de los cultivos, facilitando decisiones basadas en datos y adaptando las estrategias a las necesidades específicas de cada periodo y área del cultivo.

Este trabajo sienta las bases para futuras investigaciones orientadas a optimizar el monitoreo de enfermedades en plantaciones, promoviendo prácticas agrícolas más eficientes y sostenibles.

6.2. Trabajo a futuro

A partir de esta investigación, se identifican múltiples oportunidades para trabajos futuros y áreas de mejora que pueden perfeccionar y ampliar los alcances del estudio actual. En primer lugar, sería valioso profundizar en el desarrollo de modelos híbridos

que integren métodos de análisis temporal y espacial, optimizando la precisión en la detección y clasificación de la severidad de HLB.

La combinación de SARIMA y KNN podría extenderse incorporando otros modelos de series de tiempo y algoritmos de aprendizaje automático, que permitan capturar patrones complejos no lineales y dinámicos en la salud de los cultivos. Esto incluiría, por ejemplo, técnicas de redes neuronales recurrentes, que podrían aportar mejoras en la precisión predictiva de los periodos críticos de estrés en el cultivo.

Otra línea prometedora sería la ampliación de las fuentes de datos mediante la inclusión de sensores adicionales, como cámaras hiperespectrales o multiespectrales de mayor resolución y cobertura, que complementen los datos NDVI actuales. La incorporación de datos de humedad, temperatura, y otras variables ambientales permitiría una comprensión más detallada de los factores que influyen en la propagación de HLB y en los niveles de estrés en las plantas.

A través de esta integración de datos multivariados, se podría realizar un análisis más robusto de las condiciones que favorecen la aparición de la enfermedad, lo cual contribuiría a una gestión más precisa y preventiva. Además, se identificó como área de mejora la posibilidad de optimizar la precisión en la clasificación espacial del modelo KNN, especialmente en áreas donde la cantidad de datos es limitada.

En este sentido, se podrían explorar métodos de interpolación y generación de datos sintéticos, tales como el uso de algoritmos de aumento de datos o Redes Generativas Adversariales (GANs), para superar la limitación de datos en zonas con menor cobertura o accesibilidad. También podría implementarse un análisis de sensibilidad que evalúe el impacto de diferentes valores de parámetros en la precisión del modelo, lo que ayudaría a entender mejor el comportamiento del KNN en función de distintos contextos y características espaciales del cultivo.

Una oportunidad adicional radica en mejorar la integración de resultados para que sean aplicables en plataformas de toma de decisiones en tiempo real, permitiendo a los productores acceder fácilmente a mapas de riesgo y recomendaciones de manejo basadas en datos actualizados. Esto implica trabajar en el diseño de una interfaz de usuario intuitiva y un sistema de actualización de datos en tiempo real, posiblemente

mediante aplicaciones móviles o sistemas de alerta automatizados, que puedan guiar las decisiones de monitoreo y tratamiento en campo.

Por otro lado, en el ámbito de la validación metodológica, sería ideal realizar pruebas en diferentes regiones y con distintos tipos de cultivos, explorando cómo la presencia de HLB o enfermedades similares afecta otras especies y en qué medida los modelos aquí desarrollados pueden adaptarse a nuevos contextos. Estas validaciones adicionales ayudarían a robustecer la metodología, asegurando su aplicabilidad en distintos entornos agrícolas y para diferentes tipos de cultivos susceptibles a enfermedades de estrés biótico.

Esta investigación abre múltiples vías para el desarrollo de herramientas avanzadas y soluciones innovadoras en la gestión de enfermedades agrícolas. Al explorar modelos predictivos más complejos, enriquecer las fuentes de datos, optimizar la clasificación espacial y hacer que las herramientas sean accesibles en tiempo real para los usuarios, se podría establecer una metodología de monitoreo y control más integral y robusta.

Esto no solo contribuiría a mejorar la eficiencia en la gestión de HLB en los cítricos, sino que también podría convertirse en una referencia útil para enfrentar otras problemáticas fitosanitarias en el sector agrícola, promoviendo la sostenibilidad y resiliencia de los cultivos a largo plazo.

Bibliografía

- Ajaykumar, Shankar, Navinkumar, Sekar, Sivasabari & Harishankar. (2022). Time Series Modelling And Forecasting Of Pulses Productivity In Tamil Nadu, India. *Agriculture Association of Textile Chemical and Critical Reviews Journal*, 10, 1-9. <https://doi.org/10.58321/aatccreview.2022.10.01.01>
- Amin, M. F. (2022). Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial. *Journal of Engineering Research*, 6, 12. <https://doi.org/10.21608/erjeng.2022.274526>
- Andión, J., & Olaskoaga, X. (2011). Sentinel-2 Multi-Spectral Instrument Calibration and Shutter Mechanism. *Remote Sensing of Environment*. <https://api.semanticscholar.org/CorpusID:209898258>
- Angra, S., & Ahuja, S. (2017). Machine learning and its applications: A review. *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, 57-60. <https://doi.org/10.1109/icbdaci.2017.8070809>
- Antognelli, S. (2018, mayo). NDVI and NDMI vegetation indices: Instructions for Use [Tags: Satellite imagery, Tobacco growing, Vegetation indices, Viticulture]. Consultado el 11 de mayo de 2024, desde <https://www.agricolus.com/en/vegetation-indices-ndvi-ndmi/>
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4. <https://doi.org/10.1214/09-ss054>
- Awe, O. O., & Dias, R. (2022). Comparative Analysis of ARIMA and Artificial Neural Network Techniques for Forecasting Non-Stationary Agricultural Output Time Series. *Agris on-line Papers in Economics and Informatics*, 14, 3-9. <https://doi.org/10.7160/aol.2022.140401>

- Barman, U., & Choudhury, R. D. (2022). Smartphone assist deep neural network to detect the citrus diseases in Agri-informatics. *Global Transitions Proceedings*, 3, 392-398. <https://doi.org/10.1016/j.gltip.2021.10.004>
- Berrar, D. (2019). Cross-Validation. *Encyclopedia of Bioinformatics and Computational Biology*, 542-545. <https://doi.org/10.1016/b978-0-12-809633-8.20349-x>
- Black Sabbath. (1971). Sweet Leaf [Song] on *Master of Reality*, Vertigo/Warner Bros [Songwriters: Ozzy Osbourne, Tony Iommi, Geezer Butler, Bill Ward; Producer: Rodger Bain]. <https://www.blacksabbath.com/discography.html>
- Bontemps, S., Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Morin, D., Rabaute, T., Savinaud, M., Sepulcre, G., Valero, S., Defourny, P., & Koetz, B. (2015). Sentinel-2 for agriculture: Supporting global agriculture monitoring. *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. <https://doi.org/10.1109/igarss.2015.7326748>
- Boon, M. A., Greenfield, R., & Tesfamichael, S. (2016). Wetland Assessment Using Unmanned Aerial Vehicle (UAV) Photogrammetry. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B1, 781-788. <https://doi.org/10.5194/isprsarchives-xli-b1-781-2016>
- Brockwell, P. J., & Davis, R. A. (2016). Introduction to Time Series and Forecasting. *Springer Texts in Statistics*. <https://doi.org/10.1007/978-3-319-29854-2>
- Bu, Y., Gan, H., & Gan, Y. (2020). Current Status and Perspectives of the Artificial Intelligence-Based Identification of Citrus Huanglongbing. *Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence*, Not available, 175-180. <https://doi.org/10.1145/3438872.3439077>
- Buntara, M. S., Napitupulu, H., & Gusriani, N. (2023). Pemrograman Python Untuk Peramalan Data Deret Waktu Menggunakan Metode Seasonal Autoregressive Integrated Moving Average (Sarima). *In Search*, 22, 354-362. <https://doi.org/10.37278/insearch.v22i2.774>
- Burzykowski, T., Geubbelmans, M., Rousseau, A.-J., & Valkenborg, D. (2023). Validation of machine learning algorithms. *American Journal of Orthodontics and Dentofacial Orthopedics*, 164, 295-297. <https://doi.org/10.1016/j.ajodo.2023.05.007>

- CABI/EPPO. (2017). *Candidatus Liberibacter asiaticus* [Distribution map]. *Distribution Maps of Plant Diseases*, 4, 766. <https://doi.org/10.1079/dmpd/20173134805>
- Camstra, A., & Boomsma, A. (1992). Cross-Validation in Regression and Covariance Structure Analysis. *Sociological Methods Research*, 21, 89-115. <https://doi.org/10.1177/0049124192021001004>
- CESAVESLP. (2022). *Plagas Reglamentadas De Los Cítricos*. Comité Estatal de Sanidad Vegetal de San Luis Potosí (CESAVESLP). Consultado el 19 de febrero de 2024, desde <https://www.cesaveslp.org.mx/PRC.html>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(4), 1247-1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chaudhary, N., Singh, G. P., Dagula, V., & Mahajan, A. (2023). Business Analytics for farmers in Crop Yield Prediction. *2023 IEEE Technology & Engineering Management Conference - Asia Pacific (TEMSCON-ASPAC)*, Not available, 1-5. <https://doi.org/10.1109/temscon-aspac59527.2023.10531317>
- Chauhan, A. S., & Henrietta, H. M. (2023). Machine Learning Basics: A Comprehensive Guide. A Review. *Babylonian Journal of Machine Learning*, 2023, 31-34. <https://doi.org/10.58496/bjml/2023/006>
- Chávez, F. E. (2011). *Análisis histológico y epidemiológico del limón mexicano y limón persa asociados a síntomas del HLB (Huanglongbing)*. <http://hdl.handle.net/10521/532>
- Chen, J., Pu, X., Deng, X., Liu, S., Li, H., & Civerolo, E. (2009). A Phytoplasma Related to ‘*Candidatus Phytoplasma asteris*’ Detected in Citrus Showing Huanglongbing (Yellow Shoot Disease) Symptoms in Guangdong, P. R. China. *Phytopathology*, 99, 236-242. <https://doi.org/10.1094/phyto-99-3-0236>
- Chen, L., Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2019). Foundations of Machine Learning, Second Edition. *Stat Papers*, 60, 1793-1795. <https://doi.org/10.1007/s00362-019-01124-9>
- Chen, P., Niu, A., Liu, D., Jiang, W., & Ma, B. (2018). Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing. *IOP Conference Series:*

- Materials Science and Engineering*, 394, 052024. <https://doi.org/10.1088/1757-899x/394/5/052024>
- Cheng, Y., Xue, D., Cheng, Y., & Li, Y. (2007). A Fast Mosaic Approach for Remote Sensing Images. *2007 International Conference on Mechatronics and Automation*, 2009-2013. <https://doi.org/10.1109/icma.2007.4303859>
- Cherlinka, V. (2024, marzo). *NDVI FAQ: All You Need To Know About Index - EOS Data Analytics* [Blog post]. Consultado el 27 de abril de 2024, desde <https://eos.com/blog/ndvi-faq-all-you-need-to-know-about-ndvi/>
- Choudhary, K., Shi, W., Dong, Y., & Paringer, R. (2022). Random Forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine. *Advances in Space Research*, 70, 2443-2457. <https://doi.org/10.1016/j.asr.2022.06.073>
- COPLADE. (2022, marzo). *Plan Estatal de Desarrollo San Luis Potosí 2021-2027* [Página citada: 209]. Comité de Planeación para el Desarrollo del Estado - Gobierno de San Luis Potosí. <https://slp.gob.mx/finanzas/Documentos%20compartidos/220408-PED-2021-2027-Completo.pdf>
- Dalianis, H. (2018). Evaluation Metrics and Evaluation. En *Clinical Text Mining*. Springer. https://doi.org/10.1007/978-3-319-78503-5_6
- Dechoz, C., Poulain, V., Massera, S., Languille, F., Greslou, D., de Lussy, F., Gaudel, A., L'Helguen, C., Picard, C., & Trémas, T. (2015). Sentinel 2 global reference image. *SPIE Proceedings, Image and Signal Processing for Remote Sensing XXI*, 9643, 96430A. <https://doi.org/10.1117/12.2195046>
- Deng, X., Lan, Y., Hong, T., & Chen, J. (2016). Citrus greening detection using visible spectrum imaging and C-SVC. *Computers and Electronics in Agriculture*, 130, 177-183. <https://doi.org/10.1016/j.compag.2016.09.005>
- DJI. (2021). A close look at drone orthomosaics in surveying and how to get the most from your data. <https://enterprise-insights.dji.com/blog/orthomosaics>
- Durán, E. R., Hernández, M. E. O., Ramírez, R. G., & Arteaga, G. Á. (2017). Cambio de uso de suelo e índice de vegetación de diferencia normalizada (NDVI), subcuenca del Río Salado, México [ISSN 1852-8031]. *Geografía y Sistemas de Información Geográfica (GEOSIG)*, 9(9), 39-50. <http://www.gesig-proeg.com.ar>

- Dvivedi, N. (2020). Photogrammetry is used in surveying and mapping by using photography [Received: July 10, 2020; Accepted: July 20, 2020; Published: July 27, 2020]. *Geoinfor Geostat: An Overview*, 8(3), 261. [https://doi.org/10.37532/gigs.2020.8\(3\).261](https://doi.org/10.37532/gigs.2020.8(3).261)
- ESA. (2015, junio). *SENTINEL 2*. Consultado el 8 de abril de 2024, desde https://www.esa.int/Space_in_Member_States/Spain/SENTINEL_2
- FAO. (2021). *Citrus Fruit Statistical Compendium 2020* (1.^a ed.) [Issued annually by the Market and Policy Analysis of Raw Materials, Horticulture and Tropical Products Team, Markets and Trade Division, FAO, Rome. Prepared for Members and Observers of the CCP Intergovernmental Group on Citrus Fruit.]. Rome, Italy, Food; Agriculture Organization of the United Nations. Consultado el 17 de febrero de 2024, desde <https://openknowledge.fao.org/handle/20.500.14283/cb6492en>
- Farhath, Z. A., Arputhamary, B., & Arockiam, L. (2016). A Survey on ARIMA Forecasting Using Time Series Model [Impact Factor: 5.258]. *International Journal of Computer Science and Mobile Computing*, 5(8), 104-109. <http://www.ijscmc.com>
- Farzaliyev, E. (2021). Classification of Lands of Remote Sensitive Data by NDVI Method in Smart Agriculture. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-4/W5-2021, 73-77. <https://doi.org/10.5194/isprs-archives-xlvi-4-w5-2021-73-2021>
- Food and Agriculture Organization [FAO]. (2021). *Año Internacional de la Sanidad Vegetal - Informe final. Proteger las plantas, proteger la vida* (1.^a ed.) [Publicado en nombre de la Secretaría de la Convención Internacional de Protección Fitosanitaria]. FAO. <https://doi.org/10.4060/cb7056es>
- Garcia-Ruiz, F., Sankaran, S., Maja, J. M., Lee, W. S., Rasmussen, J., & Ehsani, R. (2013). Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. *Computers and Electronics in Agriculture*, 91, 106-115. <https://doi.org/10.1016/j.compag.2012.12.002>
- Garza, B. N., Ancona, V., Enciso, J., Perotto-Baldivieso, H. L., Kunta, M., & Simpson, C. (2020). Quantifying Citrus Tree Health Using True Color UAV Images. *Remote Sensing*, 12, 170. <https://doi.org/10.3390/rs12010170>

- Garza-Saldaña, J. J., Varela-Fuentes, S., & Gómez-Flores, W. (2017). Métodos para la detección presuntiva de Huanglongbing (HLB) en cítricos. *CienciaUAT*, 11, 93. <https://doi.org/10.29059/cienciauat.v11i2.783>
- Gascon, F. (2018). Sentinel-2 for Agricultural Monitoring. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 8166-8168. <https://doi.org/10.1109/igarss.2018.8517942>
- Geng, S. (2024). Analysis of the Different Statistical Metrics in Machine Learning. *Highlights in Science, Engineering and Technology*, 88, 350-356. <https://doi.org/10.54097/jhq3tv19>
- GeoBC. (2024, junio). Orthophotos of B.C. [Last updated on June 27, 2024]. Consultado el 20 de enero de 2024, desde <https://www2.gov.bc.ca/gov/content/data/geographic-data-services/digital-imagery/orthophotos>
- González, J., Sánchez, J. M., Bueno, M., & Arias, P. (2017). Jorge A Review. *Drones*, 1, 2. <https://doi.org/10.3390/drones1010002>
- Gottwald, T. R., Irey, M. S., Gast, T., Parnell, S. R., Taylor, E., & Hilf, M. (2010). Spatio-temporal analysis of an HLB epidemic in Florida and implications for spread. *International Organization of Citrus Virologists Conference Proceedings (1957-2010)*, 17, Not available. <https://doi.org/10.5070/c50jq6d375>
- Haghighi, S., Jasemi, M., Hessabi, S., & Zolanvari, A. (2018). PyCM: Multiclass confusion matrix library in Python. *Journal of Open Source Software*, 3, 729. <https://doi.org/10.21105/joss.00729>
- Hall, D. G., Hentz, M. G., & Adair, R. C. (2008). Population Ecology and Phenology of *Diaphorina citri* (Hemiptera: Psyllidae) in Two Florida Citrus Groves. *Environmental Entomology*, 37, 914-924. <https://doi.org/10.1093/ee/37.4.914>
- Hamido, S. A., Morgan, K. T., & Kadyampakeni, D. M. (2017). The Effect of Huanglongbing on Young Citrus Tree Water Use. *HortTechnology*, 27, 659-665. <https://doi.org/10.21273/horttech03830-17>
- Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2020). 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. <https://doi.org/10.1007/s11676-020-01155-1>

- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd) [Online version last updated on 17 December 2024]. OTexts. <https://otexts.com/fpp3>
- Hyvönen, V., Jaasaari, E., & Roos, T. (2024). A Multilabel Classification Framework for Approximate Nearest Neighbor Search (S. Kumar, Ed.) [Submitted 3/23; Revised 12/23; Published 2/24]. *Journal of Machine Learning Research*, 25, 1-51. <https://jmlr.org/papers/v25/23-0286.html>
- Jagoueix, S., Bové, J.-M., & Garnier, M. (1994). The Phloem-Limited Bacterium of Greening Disease of Citrus Is a Member of the Subdivision of the Proteobacteria. *International Journal of Systematic Bacteriology*, 44, 379-386. <https://doi.org/10.1099/00207713-44-3-379>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349, 255-260. <https://doi.org/10.1126/science.aaa8415>
- Khanramaki, M., Asli-Ardeh, E. A., & Kozegar, E. (2021). Citrus pests classification using an ensemble of deep learning models. *Computers and Electronics in Agriculture*, 186, 106192. <https://doi.org/10.1016/j.compag.2021.106192>
- Khattak, A., Asghar, M. U., Batool, U., Asghar, M. Z., Ullah, H., Al-Rakhami, M., & Gumaei, A. (2021). Automatic Detection of Citrus Fruit and Leaves Diseases Using Deep Neural Network Model. *IEEE Access*, 9, 112942-112954. <https://doi.org/10.1109/access.2021.3096895>
- Könnöla, T., del Río González, P., Carrillo-Hermosilla, J., & López, F. J. D. (2023). Innovación verde en América Latina y el Caribe: Estudio de casos sobre la gestión de programas de apoyo. *Not available, Not available, Not available*. <https://doi.org/10.18235/0004891>
- Korstanje, J. (2021). The SARIMA Model. En *Advanced Forecasting with Python* (pp. 115-122). Apress. https://doi.org/10.1007/978-1-4842-7150-6_7
- Kriegler, F. J., Malila, W. A., Nalepka, R. F., & Richardson, W. (1969). Preprocessing transformations and their effects on multispectral recognition. *Proceedings of the Sixth International Symposium on Remote Sensing of Environment, Volume II*, 97.
- Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., Czogalik, Ł., Dudek, P., Magiera, M., Lis, A., Paszkiewicz, I., Nawrat, Z., Cebula,

- M., & Gruszczyńska, K. (2023). What Is Machine Learning, Artificial Neural Networks and Deep Learning? - Examples of Practical Applications in Medicine. *Diagnostics (Basel)*, 13(15). <https://doi.org/10.3390/diagnostics13152582>
- Lan, Y., Huang, Z., Deng, X., Zhu, Z., Huang, H., Zheng, Z., Lian, B., Zeng, G., & Tong, Z. (2020). Comparison of machine learning methods for citrus greening detection on UAV multispectral images. *Computers and Electronics in Agriculture*, 171, 105234. <https://doi.org/10.1016/j.compag.2020.105234>
- Liba, N., & Berg-Jürgens, J. (2015). Accuracy of Orthomosaic Generated by Different Methods in Example of UAV Platform MUST Q. *IOP Conference Series: Materials Science and Engineering*, 96, 012041. <https://doi.org/10.1088/1757-899x/96/1/012041>
- Linder, W. (2016). *Digital Photogrammetry: A Practical Course* (4.^a ed.). Springer Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-50463-5>
- Lomellí, J., Rodríguez, E., Valdez, J., & Ortega, L. (2010). Géneros de Coccinellidae asociados a Diaphorina citri (Hemiptera: Psyllidae) en México. *Resúmenes del 1er Simposio Nacional sobre Investigación para el Manejo del Psílido Asiático de los Cítricos y el Huanglongbing en México*. Monterrey, México: [INIFAP] Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, 64-77.
- Mattheakis, M., & Protopapas, P. (2018). Model Selection & Information Criteria: Akaike Information Criterion [CS 109A: Advanced Topics in Data Science, Protopapas, Rader, Pages: 15, Last modified: September 30, 2018]. <https://harvard-iacs.github.io/2018-CS109A/a-sections/a-section-2/presentation/a-sec2-MLEtoAIC.pdf>
- Mead, F. W., & Fasulo, T. R. (2010). Asian Citrus Psyllid, *Diaphorina citri* Kuwayama (Insecta: Hemiptera: Psyllidae). *EDIS*, 2010. <https://doi.org/10.32473/edis-in160-2010>
- Mendoza, M. G. G., Galicia, N. Y. A., Servín, C. C., Salas, G. S., & Rocha, S. D. A. (2021). Articulación territorial de la gestión de plaguicidas en el marco de la protección fitosanitaria. El caso del valle agrícola de Rioverde y Ciudad Fernández, San Luis Potosí. *Investigaciones Geográficas*. <https://doi.org/10.14350/rig.60415>

- Mishra, A. R., Karimi, D., Ehsani, R., & Lee, W. S. (2012). Identification of Citrus Greening (HLB) Using a VIS-NIR Spectroscopy Technique. *Transactions of the ASABE*, 55, 711-720. <https://doi.org/10.13031/2013.41369>
- Mishra, A., Karimi, D., Ehsani, R., & Albrigo, L. G. (2011). Evaluation of an active optical sensor for detection of Huanglongbing (HLB) disease. *Biosystems Engineering*, 110, 302-309. <https://doi.org/10.1016/j.biosystemseng.2011.09.003>
- Mokrane, A., Choukchou-Braham, A., & Cherki, B. (2019). DEM Generation Based On UAV Photogrammetry. *2019 International Conference on Advanced Electrical Engineering (ICAEE)*, 1-5. <https://doi.org/10.1109/icaee47123.2019.9014805>
- Mutwiri, R. M. (2019). Forecasting of Tomatoes Wholesale Prices of Nairobi in Kenya: Time Series Analysis Using Sarima Model. *International Journal of Statistical Distributions and Applications*, 5, 46. <https://doi.org/10.11648/j.ijstd.20190503.11>
- Novković, N., Vukelić, N., Šarac, V., & Nikolić, S. (2022). State and tendencies of production characteristics of wheat and maize in Serbia. *Journal on Processing and Energy in Agriculture*, 26, 68-70. <https://doi.org/10.5937/jpea26-37904>
- OpenDroneMap. (2020). ODM – A command line toolkit to generate maps, point clouds, 3D models and DEMs from drone, balloon or kite images. <https://github.com/OpenDroneMap/ODM>
- Pearson, K. (1904). *Mathematical contributions to the theory of evolution: XIII. On the theory of contingency and its relation to association and normal correlation* [Price Four Shillings]. Dulau; Co.
- Piegorsch, W. W. (2020). Confusion Matrix. *Wiley StatsRef: Statistics Reference Online*, 1-4. <https://doi.org/10.1002/9781118445112.stat08244>
- Pourreza, A., Lee, W. S., Ehsani, R., Schueller, J. K., & Raveh, E. (2015). An optimum method for real-time in-field detection of Huanglongbing disease using a vision sensor. *Computers and Electronics in Agriculture*, 110, 221-232. <https://doi.org/10.1016/j.compag.2014.11.021>
- Rahardja, D. (2020). Statistical methodological review for time-series data. *Journal of Statistics and Management Systems*, 23(8), 1445-1461. <https://doi.org/10.1080/09720510.2020.1727618>

- Raj, A. (2019). A Review on Machine Learning Algorithms. *International Journal for Research in Applied Science and Engineering Technology*, 7, 792-796. <https://doi.org/10.22214/ijraset.2019.6138>
- Rani, P., & Vashishtha, J. (2017). An Appraise of KNN to the Perfection. *International Journal of Computer Applications*, 170, 13-17. <https://doi.org/10.5120/ijca2017914696>
- Rinaldi, P., & Larrabide, I. (2021). DSM Generation with UAVs and OpenDroneMap toolkit: A parametric analysis. *2021 XIX Workshop on Information Processing and Control (RPIC)*, 1-5. <https://doi.org/10.1109/rpic53795.2021.9648415>
- Ríos, R. (2021). Uso de los Drones o Vehículos Aéreos no Tripulados en la Agricultura de Precisión. *Revista Ingeniería Agrícola*, 11(4), 75-84. <https://revistas.unah.edu.cu/index.php/IAgric/article/view/1469>
- SADER. (2023, diciembre). Cuenta México con un fuerte sector citrícola para atender el mayor consumo por temporada invernal [Comunicado de prensa]. Consultado el 20 de enero de 2024, desde <https://www.gob.mx/agricultura/prensa/cuenta-mexico-con-un-fuerte-sector-citricola-para-atender-el-mayor-consumo-por-temporada-invernal>
- Sandoval, K. V., & Avila, D. D. (2019). Los cítricos en México: Análisis de eficiencia técnica [Citrus in Mexico: Technical Efficiency Analysis] [Primer envío: 28/febrero/2019; Segundo envío: 11/mayo/2019; aceptación: 20/junio/2019]. *Análisis Económico*, 34(87), 269-283. <https://doi.org/10.xxxx/gmd-7-1247-2014>
- Santivañez, T., Vernal, P., Mora-Aguilera, G., Padilla, G., & López-Arroyo, J. I. (2013, diciembre). *Citrus: Marco Estratégico para la Gestión Regional del Hanglongbing en América Latina y el Caribe*. FAO.
- Sathya, K., & Karthiban, R. (2021). Time Series Analysis on Agricultural Commodity Prices. *Asian Journal of Applied Science and Technology*, 05, 36-41. <https://doi.org/10.38177/ajast.2021.5205>
- Schaffer, A. L., Dobbins, T. A., & Pearson, S.-A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: A guide for evaluating large-scale health interventions. *BMC Medical Research Methodology*, 21. <https://doi.org/10.1186/s12874-021-01235-8>

- SENASICA. (2019). *Ficha técnica Huanglongbing ‘Candidatus Liberibacter spp.’ Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria - Dirección General de Sanidad Vegetal - Programa de Vigilancia Epidemiológica Fitosanitaria. Con la colaboración del Laboratorio Nacional de Referencia Epidemiológica Fitosanitaria (LaNREF)* [Ficha técnica 78, última actualización: abril, 2019]. Ciudad de México, Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria. https://www.gob.mx/cms/uploads/attachment/file/463426/78._Ficha_t_cnica_Huanglongbing._Mayo_2019..pdf
- SENASICA. (2022). Destina Agricultura recursos para proteger citricultura nacional de plagas y enfermedades. Consultado el 23 de mayo de 2024, desde <https://www.gob.mx/senasica/prensa/destina-agricultura-recursos-para-proteger-citricultura-nacional-de-plagas-y-enfermedades-302708?idiom=es>
- Sharif, M., Khan, M. A., Iqbal, Z., Azam, M. F., Lali, M. I. U., & Javed, M. Y. (2018). Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. *Computers and Electronics in Agriculture*, 150, 220-234. <https://doi.org/10.1016/j.compag.2018.04.023>
- Sodha, D., & Saha, G. (2016). Crop management of agricultural products using time series analysis. *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Not available, 1456-1460. <https://doi.org/10.1109/rteict.2016.7808073>
- Sosa-Herrera, J. A., Alvarez-Jarquín, N., Cid-García, N. M., López-Araujo, D. J., & Vallejo-Pérez, M. R. (2022). Automated Health Estimation of *Capsicum annum* L. Crops by Means of Deep Learning and RGB Aerial Images [This article belongs to the Topic Grand Challenges of Advanced Technologies in Sustainable Agriculture 4.0: Future Farming, Harvesting and Preservation]. *Remote Sensing*, 14(19), 4943. <https://doi.org/10.3390/rs14194943>
- Sreekantha, D. K., & Prasad, K. (2018). Applications of Unmanned Aerial Vehicles (UAV) in Agriculture: A Study. *International Journal for Research in Applied Science and Engineering Technology*, 6, 1162-1166. <https://doi.org/10.22214/ijraset.2018.5188>

- Stegmayer, G., Milone, D. H., Garran, S., & Burdyn, L. (2013). Automatic recognition of quarantine citrus diseases. *Expert Systems with Applications*, *40*, 3512-3517. <https://doi.org/10.1016/j.eswa.2012.12.059>
- Stover, E., McCollum, G., Driggers, R., Duan, Y., R. Shatters, J., Ritenour, M., Hall, D. G., & Chaparro, J. (2014). Huanglongbing Resistance and Tolerance in Citrus. *Journal of Citrus Pathology*, *1*, Not available. <https://doi.org/10.5070/c411025259>
- Tanathong, S., Smith, W. A. P., & Remde, S. (2022). SurfaceView: Seamless and Tile-Based Orthomosaics Using Millions of Street-Level Images From Vehicle-Mounted Cameras. *IEEE Transactions on Intelligent Transportation Systems*, *23*, 3482-3497. <https://doi.org/10.1109/tits.2020.3036928>
- Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 1255-1260. <https://doi.org/10.1109/iccs45141.2019.9065747>
- Tibshirani, R. (2023). *Lecture 6: Autoregressive Integrated Moving Average Models* [Introduction to Time Series: 2023]. Consultado el 20 de diciembre de 2023, desde <https://www.stat.berkeley.edu/~ryantibs/timeseries-f23/lectures/arima.pdf>
- Tolosana-Delgado, R., & Mueller, U. (2021). Cross-Validation. *Use R!, Geostatistics for Compositional Data with R*, 133-153. https://doi.org/10.1007/978-3-030-82568-3_7
- Torfi, A. (2020). Nearest Neighbor Classifier—From Theory to Practice. *Machine Learning Mindset*.
- Torrado, J. O. E., Jiménez, J. J. C., & Díaz, H. P. (2016). Ortomosaicos y modelos digitales de elevación generados a partir de imágenes tomadas con sistemas UAV [ISSN-e 2248-7638, ISSN 0123-921X]. *Tecnura: Tecnología y Cultura Afirmando el Conocimiento*, *20*(50), 119-140.
- Tsach, S., Tatievsky, A., & London, L. (2010). Unmanned Aerial Vehicles (UAVs). *Encyclopedia of Aerospace Engineering*. <https://doi.org/10.1002/9780470686652.eae385>
- Varoquaux, G., & Colliot, O. (2023). Evaluating Machine Learning Models and Their Diagnostic Value [Published on July 23, 2023]. En O. Colliot (Ed.), *Machine*

- Learning for Brain Disorders*. Humana. https://doi.org/10.1007/978-1-0716-3195-9_20
- Veney, J. E., & Luckey, J. W. (1983). A Comparison of Regression and ARIMA Models for Assessing Program Effects: An Application to the Mandated Highway Speed Limit Reduction of 1974. *Social Indicators Research*, 12(1), 83-105. <https://doi.org/10.1007/BF00428862>
- Verma, A. P., & Chakraborty, B. S. (2020). Performance Estimation of ARIMA Model for Orographic Rainfall Region. *2020 URSI Regional Conference on Radio Science (URSI-RCRS)*, 1-4. <https://doi.org/10.23919/ursircrs49211.2020.9113522>
- Wang, Y., Khodadadzadeh, M., & Zurita-Milla, R. (2023). Spatial: A new cross-validation method to evaluate geospatial machine learning models. *International Journal of Applied Earth Observation and Geoinformation*, 121, 103364. <https://doi.org/https://doi.org/10.1016/j.jag.2023.103364>
- Watanabe, S. (2013). A Widely Applicable Bayesian Information Criterion (M. Oppe, Ed.) [Submitted 8/12; Revised 2/13; Published 3/13]. *Journal of Machine Learning Research*, 14, 867-897. <https://www.jmlr.org/papers/volume14/watanabe13a/watanabe13a.pdf>
- Wazirali, R. (2020). An Improved Intrusion Detection System Based on KNN Hyperparameter Tuning and Cross-Validation [Received 12 May 2020, Accepted 19 August 2020, Published 31 August 2020, Issue Date: December 2020]. *Arab Journal of Science and Engineering*, 45(12), 10859-10873. <https://doi.org/10.1007/s13369-020-04907-7>
- Wu, B. (2017). Photogrammetry: 3-D From Imagery. *Wiley Online Library*. <https://doi.org/10.1002/9781118786352.wbieg0942>
- Xu, L., Samanta, A., Costa, M. H., Ganguly, S., Nemani, R. R., & Myneni, R. B. (2011). Widespread decline in greenness of Amazonian vegetation due to the 2010 drought. *Geophysical Research Letters*, 38. <https://doi.org/10.1029/2011gl046824>
- Yadav, D., Soumya, K., & Goswami, L. (2024). Autoregressive Integrated Moving Average Model for Time Series Analysis. *2024 International Conference on Op-*

- timization Computing and Wireless Communication (ICOCWC)*, 1-6. <https://doi.org/10.1109/icocwc60930.2024.10470488>
- Yun, X., & Royston, G. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, 2. <https://doi.org/10.1007/s41664-018-0068-2>
- Zaidi, N., Singh, B., & Yadav, S. (2018). The evolution of machine learning algorithms: A comprehensive historical review. *International Journal of Applied Research*, 4, 49-55. <https://doi.org/10.22271/allresearch.2018.v4.i9a.11451>
- Zhao, X. Y. (1982). Citrus yellow shoot disease (Huanglongbing) in China - a review. *Citrus Huanglongbing: Review, Present status and Future Strategies*. <https://api.semanticscholar.org/CorpusID:81068526>