

**UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL
DE HUAMANGA**

FACULTAD DE INGENIERÍA DE MINAS, GEOLOGÍA Y CIVIL

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



TESIS:

**“Rendimiento de Redes Convolucionales y de Auto-atención
en el Reconocimiento de Enfermedades de Cacao, Omayá -
Cusco, 2025”**

Para optar el título profesional de:
INGENIERO DE SISTEMAS

PRESENTADO POR:
Bach. Geovanni Rafael LEON PEREZ

ASESOR:
Dr. José Elías YAURI VIDALÓN

AYACUCHO - PERÚ

2025

Dedicatoria

*Con profundo amor para mis padres y mis
gatos que me brindaron el apoyo necesario
para culminar mis estudios.*

*Con un profundo agradecimiento a mi
novia, quien me acompañó
incondicionalmente durante mi formación
universitaria.*

Agradecimiento

Extiendo mis más sinceras muestras de aprecio y agradecimiento a la Universidad Nacional de San Cristóbal de Huamanga.

Agradezco especialmente al Dr. Ing. José Elías Yauri Vidalón, quien no solo ha sido mi asesor en este trabajo de investigación, sino que también ha compartido generosamente su tiempo, conocimientos y experiencia para guiarme en mi desarrollo profesional y personal. Su dedicación ha sido invaluable y ha contribuido significativamente a mi crecimiento académico y personal.

Resumen

El cacao (*Theobroma cacao* L.), es originario de las selvas amazónicas de América del Sur, es una planta de gran valor económico y cultural, cuyas semillas se utilizan para la producción de chocolate. Actualmente, Perú ocupa el octavo lugar en la producción mundial de cacao, siendo este cultivo una fuente importante de ingresos para más de 90,000 familias en regiones amazónicas como Cusco, Ayacucho y Junín. El Valle de los Ríos Apurímac, Ene y Mantaro (VRAEM) desempeña un papel importante en esta producción, especialmente en zonas como Omayta, Cusco.

A pesar de su relevancia económica, la productividad del cacao en la región del VRAEM es baja (594 kg/hectáreas), debido a enfermedades y plagas como la pudrición parda (*Phytophthora palmivora*), la moniliasis (*Moniliophthora roreri*), la escoba de bruja (*Moniliophthora perniciosa*) y el mazorquero (*Carmenta foraseminis*). Estas enfermedades pueden ocasionar pérdidas de hasta el 100% de la producción, afectando directamente la economía local.

Ante esta problemática, la inteligencia artificial, en especial los modelos de aprendizaje profundo, se presenta como una herramienta eficaz para la detección temprana y clasificación de enfermedades a partir de imágenes. Esto permite una intervención rápida, reduciendo el impacto de las enfermedades y optimizando el monitoreo de cultivos.

El presente estudio tiene como objetivo evaluar y comparar el rendimiento de Redes Neuronales Convolucionales (CNN) y de Auto-Atención (Self-Attention) para el reconocimiento de enfermedades en frutos de cacao en imágenes provenientes de Omayta, Cusco, 2025. El propósito es identificar el modelo que presente la mayor tasa de detección y clasificación de enfermedades para ser utilizado en el desarrollo de sistemas de detección automática de enfermedades.

Así, esta investigación contribuye al desarrollo de sistemas inteligentes que mejoren la capacidad de los agricultores para detectar enfermedades de forma oportuna, permitiendo una producción más sostenible y de mejor calidad. Los resultados de este estudio podrán servir como base para futuras investigaciones y aplicaciones tecnológicas en otras regiones cacaoteras del país y del mundo, apoyando la sostenibilidad agrícola y la seguridad alimentaria global.

Palabras clave: Aprendizaje profundo, Visión artificial, Agricultura inteligente, Detección de enfermedades del cacao, Enfermedades del cacao peruano.

Abstract

Cacao (*Theobroma cacao* L.), native to the Amazon rainforests of South America, is a plant of great economic and cultural value, whose seeds are used to produce chocolate. Currently, Peru ranks eighth in global cacao production, with the crop being a significant source of income for more than 90,000 families in Amazonian regions such as Cusco, Ayacucho, and Junín. The Apurímac, Ene, and Mantaro River Valley (VRAEM) plays a key role in this production, especially in areas like Omayta, Cusco.

Despite its economic importance, cacao productivity in the VRAEM region is low (594 kg/hectare) due to diseases and pests such as brown rot (*Phytophthora palmivora*), moniliasis (*Moniliophthora roreri*), witches' broom (*Moniliophthora perniciosa*), and the cacao pod borer (*Carmenta foraseminis*). These diseases can cause losses of up to 100% of production, directly impacting the local economy.

Faced with this problem, artificial intelligence, especially Deep Learning models, emerges as an effective tool for the early detection and classification of diseases from images. This allows for rapid intervention, reducing the impact of diseases and optimizing crop monitoring.

This study aims to evaluate and compare the performance of Convolutional Neural Networks and self-care networks for disease recognition in cacao pods using images taken in Omayta, Cusco, in 2025. The goal is to identify the model with the highest disease detection and classification rate.

This research will contribute to the development of intelligent systems that improve farmers' ability to detect diseases in a timely manner, enabling more sustainable and higher-quality production. The results of this study can serve as a basis for future research and technological applications in other cacao-growing regions of the country and the world, supporting agricultural sustainability and global food security.

Keywords: Deep Learning, Computer Vision, Smart agriculture, Cocoa Disease Detection, Peruvian Cocoa Diseases.

ÍNDICE

Dedicatoria	i
Agradecimiento	ii
Resumen	iii
Abstract	iv

CAPÍTULO I INTRODUCCIÓN

1.1	PLANTEAMIENTO DEL PROBLEMA	1
1.2	FORMULACIÓN DEL PROBLEMA	2
1.2.1	Problema general	2
1.2.2	Problemas específicos	2
1.3	JUSTIFICACIÓN DE LA INVESTIGACIÓN	2
1.4	LIMITACIONES DE LA INVESTIGACIÓN	3
1.5	OBJETIVOS	4
1.5.1	Objetivo general	4
1.5.2	Objetivo específico	4

CAPÍTULO II MARCO TEÓRICO

2.1	ANTECEDENTES DE LA INVESTIGACIÓN	5
2.2	MARCO CONCEPTUAL	8
2.2.1	Rendimiento de la red neuronal	8
2.2.2	Paradigma de computación de la red neuronal	8
2.2.3	Capas base de aprendizaje de la red	11
2.2.4	Eficacia de la red	15
2.2.5	Eficiencia de la red	16
2.2.6	Uso de GPU/CPU/RAM	16
2.3	MARCO REFERENCIAL	17
2.3.1	Reconocimiento de enfermedades de cacao	17
2.3.2	Preprocesamiento de imágenes	24
2.3.3	Red neuronal convolucional (CCN)	24
2.3.4	Red neuronal de auto-atención	24
2.3.5	Prueba de wilcoxon	25

CAPÍTULO III MATERIAL Y MÉTODOS

3.1	TIPO Y NIVEL DE INVESTIGACIÓN	26
3.1.1	Tipo de investigación	26
3.1.2	Nivel de investigación	26
3.2	DISEÑO DE LA INVESTIGACIÓN	27

3.3	VARIABLES	27
3.3.1	Definición conceptual de variables	27
3.3.2	Definición operacional de variables	27
3.4	POBLACIÓN Y MUESTRA	28
3.4.1	Población	28
3.4.2	Muestreo	28
3.4.3	Muestra	28
3.5	OPERACIONALIZACIÓN DE LAS VARIABLES.....	30
3.6	TÉCNICAS E INSTRUMENTOS DE LA INVESTIGACIÓN	31
3.6.1	Técnicas	31
3.6.2	Instrumentos.....	31
3.6.3	Validez del instrumento	31
3.6.4	Confiabilidad de instrumento	31
3.7	PROCEDIMIENTOS	31
3.7.1	Estrategia de prueba de hipótesis.....	31
3.7.2	Técnicas de procesamiento de datos	32
3.7.3	Diseño estadístico	35
3.7.4	Análisis e interpretación de resultados	35

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1	RESULTADOS	36
4.1.1	Construcción del dataset.....	38
4.1.2	Pre-procesamiento de datos.....	42
4.1.3	Proceso de entrenamiento del modelo	43
4.1.4	Proceso de evaluación y validación del modelo	51
4.2	DISCUSIÓN.....	55

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1	CONCLUSIONES	57
5.2	RECOMENDACIONES.....	58
	BIBLIOGRAFÍA	59
	Anexos	62
	Anexo 1. Matriz de consistencia.....	62
	Anexo 2.....	64
	Publicaciones.....	64
	Artículos de conferencia	64
	Dataset libre	64

Índice de Tablas

Tabla 1 Operacionalización de variable.	30
Tabla 2 Herramientas para el desarrollo de modelos Deep Learning.	34
Tabla 3 Dataset de enfermedades en frutos de cacao.	41
Tabla 4 Selección de hiperparámetros.	46
Tabla 5 Resultados del rendimiento generalizado de los modelos (eficacia).	53
Tabla 6 Resultados de la eficiencia.	54

Índice de Figuras

Figura 1 Arquitectura YOLOv5.	10
Figura 2 Arquitectura DINOv2.	11
Figura 3 Bloque convolucional clásico.	12
Figura 4 Bloque convolucional del YOLOv5.	13
Figura 5 Bloque de auto-atención del DINOv2.	15
Figura 6 Frutos de cacao.	18
Figura 7 La <i>Phytophthora palmivora</i> en el fruto de cacao.	20
Figura 8 La moniliasis en el fruto de cacao.	21
Figura 9 La escoba de bruja en el fruto de cacao.	22
Figura 10 El mazorquero en el fruto de cacao.	23
Figura 11 Flujo de trabajo para la clasificación de enfermedades en el cacao.	37
Figura 12 Lugar de la recolección de datos.	39
Figura 13 Ilustración de frutos sanos y frutos enfermos del dataset.	42
Figura 14 Train loss y sensibilidad según las épocas de entrenamiento del modelo YOLOv5.	47
Figura 15 Train loss y sensibilidad según las épocas de entrenamiento del modelo DINOv2.	48
Figura 16 Instancias validadas de cacao en el modelo YOLOv5.	49
Figura 17 Instancias validadas de frutos de cacao en el modelo DINOv2.	50
Figura 18 Uso de recursos durante el entrenamiento de YOLOv5.	50
Figura 19 Uso de recursos durante el entrenamiento de DINOv2.	51
Figura 20 Matriz de confusión YOLOv5.	52
Figura 21 Matriz de confusión DINOv2.	53

CAPÍTULO I

INTRODUCCIÓN

1.1 PLANTEAMIENTO DEL PROBLEMA

La producción de cacao ejerce una influencia considerable en la economía del VRAEM, particularmente en Omayá. Pese al notable crecimiento reciente que lo consolida como un producto de exportación clave, el sector enfrenta un desafío significativo: diversas enfermedades de los frutos (como la mazorca negra, la escoba de bruja, la moniliasis y el mazorquero) afectan negativamente el estado de salud y la calidad de la cosecha. Por ello, el control fitosanitario es esencial para la rentabilidad y sostenibilidad de la industria (Devida, 2021).

La detección y clasificación de enfermedades en los frutos de cacao es un tema que ha sido abordado por varias investigaciones recientes. Estas investigaciones se enfocan en el desarrollo de modelos de clasificación utilizando técnicas avanzadas de inteligencia artificial y aprendizaje profundo (Godmalin et al., 2022, 2023; Yanac Montesino et al., 2021). Sin embargo, la mayoría de estos estudios no han investigado a fondo el impacto en esta región específica. Es crucial mencionar que la base de datos utilizada en muchos estudios no se discute en detalle, lo cual representa una omisión significativa. Además, la falta de especificación sobre la variedad de frutos de cacao y las condiciones controladas de captura de imágenes (como la orientación y la iluminación) pueden sesgar los resultados y limitar la aplicabilidad del modelo en condiciones de campo reales (Yanac Montesino et al., 2021).

Con lo anterior mencionado, se propone evaluar un modelo de las redes neuronales convolucionales y de Auto-Atención que tome en cuenta aspectos fundamentales de las imágenes de los frutos de cacao, el cual clasifica automáticamente los frutos de cacao en dos categorías: enfermos y sanos. Este enfoque no solo establece una base sólida para investigaciones futuras, sino que también contribuye al desarrollo de sistemas de detección temprana, mejorando la capacidad de los agricultores para proteger sus cultivos y garantizar una producción más sostenible y de una calidad alta.

1.2 FORMULACIÓN DEL PROBLEMA

1.2.1 Problema general

¿Qué tipo de red neuronal presenta el mejor rendimiento para el reconocimiento de enfermedades de frutos de cacao en imágenes, Omayá-Cusco, 2025?

1.2.2 Problemas específicos

- a. ¿Cuál es el paradigma de computación de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes?
- b. ¿Cómo es el funcionamiento de las capas base de aprendizaje de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes?
- c. ¿Cuál es la eficacia de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes?
- d. ¿Cuál es la eficiencia de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes?

1.3 JUSTIFICACIÓN DE LA INVESTIGACIÓN

El estudio impulsa el conocimiento en el área de la agricultura inteligente y la fitopatología aplicada mediante la implementación de Deep Learning para realizar la clasificación de patologías en frutos de cacao. Desde el enfoque teórico, el trabajo es un soporte para la implementación de herramientas con inteligencia artificial en el campo, al probar la eficacia de las redes convolucionales y auto-atención en la identificación de enfermedades, lo que se considera un valioso aporte a la documentación sobre tecnologías avanzadas para la agricultura de precisión.

Desde una perspectiva metodológica, este estudio ofrece un enfoque novedoso al aplicar modelos de aprendizaje profundo para la clasificación de enfermedades en tiempo real, utilizando imágenes de alta resolución capturadas bajo diversas condiciones de luz y ángulos. La combinación de diversas etapas como la construcción del dataset, pre-procesamiento y entrenamiento de los modelos Deep Learning no solo eleva la precisión, sino que también proporciona una metodología adaptable y replicable para otras enfermedades y cultivos, ampliando su utilidad en la investigación agrícola.

A nivel práctico, la investigación ofrece a los agricultores de Cusco y otras zonas un instrumento potente para la detección rápida de enfermedades del cacao, lo que facilita la intervención adelantada. Esto se traduce en una reducción de pérdidas y un incremento en la calidad y productividad, lo cual mejora la rentabilidad. A escala macro, el beneficio para la industria cacaotera puede contribuir a la seguridad alimentaria y la sostenibilidad agrícola, fortaleciendo la resiliencia de los sistemas frente a amenazas fitosanitarias.

1.4 LIMITACIONES DE LA INVESTIGACIÓN

La muestra de estudio se limitó al uso de 3104 imágenes de frutos de cacao lo que puede limitar la representatividad de los resultados y su generalización a otras poblaciones de cultivos de cacao debido al tamaño reducido de la muestra.

El enfoque limitado al centro poblado de Omayá (Cusco) no captura la diversidad de otras zonas cacaoteras, lo que restringe la generalización de los hallazgos a otros contextos productivos.

El diseño de la investigación puede no controlar completamente todas las variables externas que podrían influir en los resultados. Factores como la luz, la temperatura o la calidad del equipo utilizado al capturar imágenes pueden variar y afectar la apariencia de los frutos, influyendo en la precisión del modelo de aprendizaje profundo.

Las condiciones climáticas y el estado de salud de los cultivos de cacao pueden introducir "ruido" en las imágenes, afectando la precisión del análisis y la validez de los resultados. Variaciones en la iluminación, como días soleados o nublados, pueden alterar la apariencia de los frutos, mientras que otros problemas de salud no relacionados con la enfermedad específica pueden confundir al modelo de aprendizaje profundo, reduciendo su capacidad para identificar correctamente las enfermedades.

1.5 OBJETIVOS

1.5.1 Objetivo general

Evaluar y comparar el rendimiento de las redes neuronales convolucionales y de Auto-Atención para el reconocimiento de enfermedades de frutos de cacao en imágenes, Omayacusco, 2025.

1.5.2 Objetivo específico

- a. Determinar el paradigma de computación de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes.
- b. Describir el funcionamiento de las capas base de aprendizaje que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes.
- c. Determinar la eficacia de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes.
- d. Determinar la eficiencia de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes.

CAPÍTULO II

MARCO TEÓRICO

2.1 ANTECEDENTES DE LA INVESTIGACIÓN

El trabajo realizado por (Yanac Montesino et al., 2021) se enfocó en el uso de avanzadas de Deep Learning para detectar *phytophthora palmivora* en frutos de cacao, aprovechando una revisión exhaustiva de la literatura y métodos innovadores. El propósito fundamental es proporcionar una herramienta efectiva para los cacaocultores, facilitando la identificación de la enfermedad y fomentando el cultivo de cacao saludable, lo que repercute en el bienestar económico de los agricultores. Para lograr este objetivo, el estudio se basó en el uso del modelo ResNet18, entrenando el sistema con conjuntos de datos extensos y alcanzando una exactitud del 83% en la detección de *phytophthora palmivora*, con una destacada precisión del 96% en la diferenciación de imágenes de cacao de otros frutos similares. La efectividad del método queda validada por estos resultados, lo que demuestra su relevancia para el sector cacaotero. El estudio resalta la necesidad de innovación en la agricultura y genera nuevas posibilidades para mejorar la producción y la calidad.

El estudio presentado por (Basri et al., 2023) se dedica a revisar diversas técnicas para el procesamiento de imágenes que son utilizadas para la detección temprana de plagas y enfermedades en las plantaciones de cacao actuales. En un contexto donde la Industria 4.0 está transformando la aplicación de tecnología en la agricultura del cacao, en este trabajo se resalta la significancia de abordar los problemas de plagas y enfermedades que pueden afectar la productividad. El análisis de la literatura revela que el ataque de enfermedades en el cacao es un área ampliamente estudiada, con un enfoque particular en el análisis de patrones de imágenes. Los resultados indican que los enfoques existentes tienen una exactitud media del 82.85%, lo que señala áreas para mejorar la exactitud mientras se reduce la complejidad del procesamiento. Además, se recomienda un modelo que integre procesamiento de imágenes con tecnología de aprendizaje automático para la industria del cultivo del cacao, destacando su potencial para apoyar la Industria 4.0 en este sector. Esta investigación establece una base esencial para el desarrollo de futuros trabajos en el mundo real que utilicen tecnologías de visión por computadora aplicadas al cultivo del cacao.

El trabajo de (Godmalin et al., 2023), se enfoca en abordar la amenaza que representan las enfermedades, como la pudrición de la vaina negra, para el cultivo de cacao, mediante el uso de aprendizaje profundo. El objetivo principal es entrenar una red neuronal convolucional para clasificar tres niveles de infección por vainas de cacao: baja, moderada y grave. El estudio utilizó un enfoque experimental y alcanzó una exactitud del 91% en la

clasificación correcta de la condición de la vaina de cacao en un entorno controlado. Como siguiente paso, se considera realizar a cabo más investigación para implementar el modelo en dispositivos de supervisión y monitoreo, lo que permitiría la clasificación en tiempo real en el campo. Esto facilitaría que los agricultores respondan de manera rápida y efectiva para mitigar la pérdida de producción causada por las enfermedades del cacao.

También en el trabajo de (Godmalin et al., 2022), se enfoca en desarrollar un algoritmo de aprendizaje profundo para clasificar la condición de la vaina de cacao, diferenciando entre saludable, ataque de la enfermedad de la vaina negra y ataque de plagas. Este enfoque surge como respuesta a la constante amenaza que representan las enfermedades y plagas para los productores de cacao, buscando proporcionar una herramienta automatizada para identificar y controlar estos eventos. Utilizando un diseño de investigación experimental, se emplea una red neuronal convolucional para entrenar el modelo, logrando una exactitud del 94% en la clasificación de la condición de la vaina de cacao en condiciones controladas. Se recomienda integrar el modelo con dispositivos de monitoreo en campo para permitir la clasificación en tiempo real, facilitando respuestas rápidas y eficaces para mitigar la pérdida de producción. Este estudio destaca el potencial de la visión por computadora en la industria del cacao, ofreciendo una solución innovadora para mejorar la gestión y producción de cultivos de cacao.

(Gyamfi et al., 2020), se enfoca en desarrollar un Sistema Inteligente de Salud del Cacao utilizando técnicas de aprendizaje automático y arquitectura de ingeniería de software. El objetivo de este sistema es detectar, diagnosticar y predecir enfermedades que han reducido la productividad del cacao en Ghana, como los márridos, la vaina negra y el virus de los brotes hinchados (CSSVD). El software, dividido en gestión de datos, análisis y control, utiliza técnicas de aprendizaje automático para la predicción y la anticipación de posibles enfermedades. Los usuarios interactúan con el sistema a través de la aplicación de control, basándose en la información proporcionada por la aplicación de análisis. Este estudio destaca el potencial del aprendizaje automático y la ingeniería de software para mejorar la salud y la productividad del cultivo de cacao, ofreciendo una solución innovadora para la industria.

(Eric et al., 2023) se concentra en mejorar la clasificación de granos de frutos cacao utilizando avanzadas técnicas de extracción de características de la imagen y una red neuronal artificial regularizada. Abordan la clasificación tradicional por prueba de corte, la cual es limitada por su subjetividad y las variaciones en la percepción visual, lo que genera imprecisiones y desacuerdos en los resultados. En este trabajo, se propone una técnica de extracción de color y textura para mejorar la representación de las imágenes, en conjunto con un modelo menos complejo, pero más generalizado. Los experimentos realizados en un conjunto de datos de prueba de frutos de cacao muestran que esta metodología logra una tasa de clasificación homogénea del 85,36% en exactitud, 85% en precisión, 83% en

recuperación y 83% en la medida F1. Estas técnicas representan una herramienta precisa y confiable para la clasificación de granos de cacao, ofreciendo un potencial significativo para el diagnóstico asistido por computadora en la industria cacaotera.

(Kumi et al., 2022), desarrollaron Cocoa Companion, una aplicación móvil para smartphones que se centra en utilizar Redes Neuronales Convolucionales (CNN) para detectar enfermedades en los frutos de cacao. Esta aplicación está diseñada para facilitar la detección temprana y el diagnóstico de las enfermedades del fruto de cacao, como el brote hinchado y la vaina negra. Permite a los agricultores tomar y cargar una foto de la mazorca para su diagnóstico, el cual es procesado en un servicio backend en la nube. El documento también revela que de los cuatro modelos CNN entrenados, El SSD MobileNet V2 alcanzó una precisión de detección de confianza (confidence score) por encima del 80%, lo que confirma su eficacia para la detección automática. Este enfoque innovador aprovecha la tecnología móvil y el aprendizaje profundo para ofrecer herramientas prácticas y efectivas para los agricultores y productores de cacao, contribuyendo así a mejorar la salud y la producción de cacao.

(Madupuri et al., 2023), se centra en desarrollar una aplicación automatizada para la clasificación de la enfermedad a partir de imágenes con aprendizaje profundo de las hojas del tomate mediante redes neuronales convolucionales profundas preentrenadas. Considerando la importancia del sector agrícola en la India y la creciente incidencia de enfermedades que afectan a las plantas, especialmente al tomate, se propone un modelo para identificar estas enfermedades utilizando imágenes de las hojas de tomate. El objetivo principal es desarrollar un modelo preciso y una aplicación móvil accesible para los agricultores, que funcione tanto en línea como fuera de línea. Utilizando un enfoque basado en redes neuronales convolucionales, el modelo propuesto alcanza una exactitud del 98%, lo que lo convierte en un modelo eficaz para su implementación en la aplicación. Esta aplicación puede predecir con precisión la enfermedad de la hoja del tomate y sugerir el tratamiento adecuado, lo que representa una herramienta valiosa para mejorar la salud de los cultivos de tomate y aumentar la productividad agrícola en la India.

(Yatoo & Sharma, 2021), se enfoca en desarrollar un nuevo modelo de detección automática de enfermedades de los cultivos. Considerando la relevancia de la agricultura para la seguridad alimentaria y los desafíos del crecimiento poblacional y la escasez de suelo cultivable, la detección precisa y precoz de enfermedades en los cultivos es de vital importancia. Esta detección es vital para mitigar las pérdidas de calidad y rendimiento en la producción agrícola. Bajo estas circunstancias, la agricultura de precisión ha cobrado una gran relevancia por el desarrollo de enfoques automatizados para la identificación temprana de patologías vegetales. Estas enfermedades, causadas por agentes como virus, bacterias u hongos, tienen un impacto negativo directo en la calidad del cultivo y causan pérdidas

económicas considerables a los agricultores. En este proyecto, se presenta un modelo afinado y se compara su rendimiento con VGG16 utilizando el conjunto de datos PlantPathology. La evidencia experimental indica que el modelo propuesto supera a VGG16. Este hecho confirma su utilidad para la detección automatizada de patologías de cultivos y su pertinencia en el ámbito de la agricultura.

2.2 MARCO CONCEPTUAL

2.2.1 Rendimiento de la red neuronal

Es una medida integral de su capacidad para ejecutar una tarea específica de manera óptima. A diferencia de conceptos simplistas como la precisión, el rendimiento se evalúa considerando dos criterios fundamentales: la eficacia, que se refiere a la calidad de las clasificaciones, y la eficiencia, que mide el costo computacional para obtener dichas clasificaciones. Esta dualidad es crucial para determinar la viabilidad y el valor de un modelo en aplicaciones reales.

La eficacia centrada en el contexto de la detección de enfermedades del cacao tiene las siguientes métricas que son la sensibilidad y la especificidad. Por un lado, la sensibilidad se centra en medir la capacidad del modelo para identificar correctamente los frutos enfermos (evitando falsos negativos), mientras que la especificidad mide su capacidad para clasificar correctamente los frutos sanos (evitando falsos positivos). Un modelo eficaz es aquel que minimiza ambos tipos de errores, proporcionando diagnósticos confiables (Tharwat, 2021).

La eficiencia se relaciona con los recursos y el tiempo que la red consume. Un modelo puede ser muy eficaz, pero si requiere una infraestructura computacional costosa o tiempos de procesamiento prolongados, su utilidad práctica es limitada. Las métricas clave incluyen el tiempo de entrenamiento, que indica cuánto tiempo tarda el modelo en aprender de los datos, y el tiempo de inferencia o tiempo de prueba, que es el tiempo que se necesita para obtener una clasificación en una nueva imagen. Además, se considera el uso de recursos de hardware, como CPU, GPU y RAM, que son determinantes para la implementación del modelo en dispositivos móviles o en tiempo real (Kingma & Ba, 2014).

2.2.2 Paradigma de computación de la red neuronal

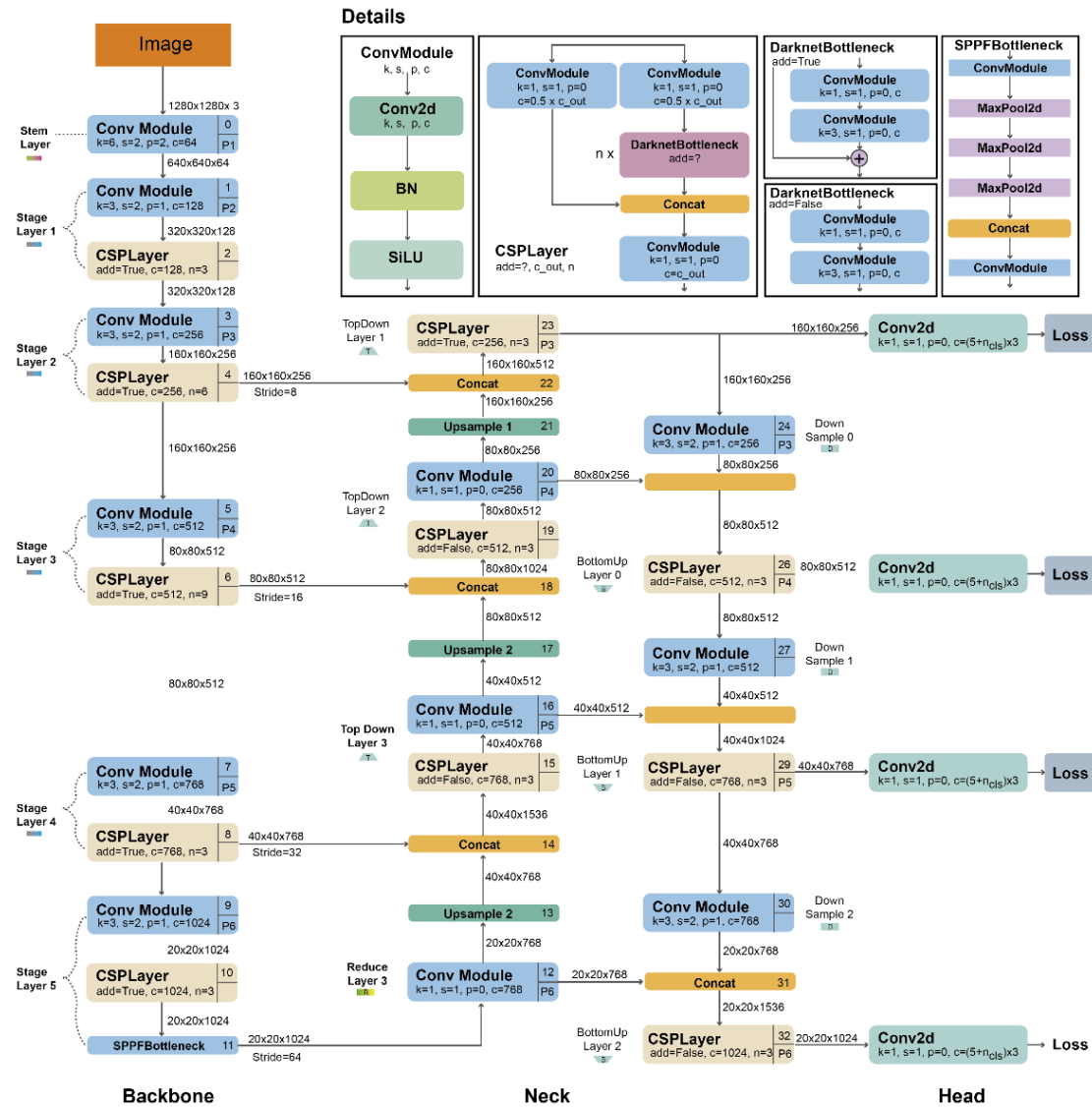
El Paradigma de Computación de la Red Neuronal se refiere al modelo elegido para la experimentación, sirviendo para el procesamiento de datos. Este paradigma se sustenta en el Deep Learning y se distingue radicalmente de los enfoques tradicionales, tal como lo señalan (LeCun et al., 2015). El modelo se basa en la capacidad intrínseca de las redes neuronales para analizar la información a través de múltiples capas interconectadas. Esta arquitectura profunda es vital, ya que permite al sistema identificar y aprender patrones complejos de forma jerárquica y autónoma.

A. Yolov5

Es una CNN (Red Neuronal Convolucional) muy bien optimizada para realizar tareas de detección de objetos en tiempo real. La diferencia este modelo frente a otros modelos es que combina las tareas de localización y clasificación en una única etapa, ofreciendo una velocidad de inferencia ultrarrápida. Esto lo hace ideal para sistemas con requisitos de tiempo real (Ultralytics, 2020).

Figura 1

Arquitectura YOLOv5.



Nota. Tomado de Terven et al. (2023).

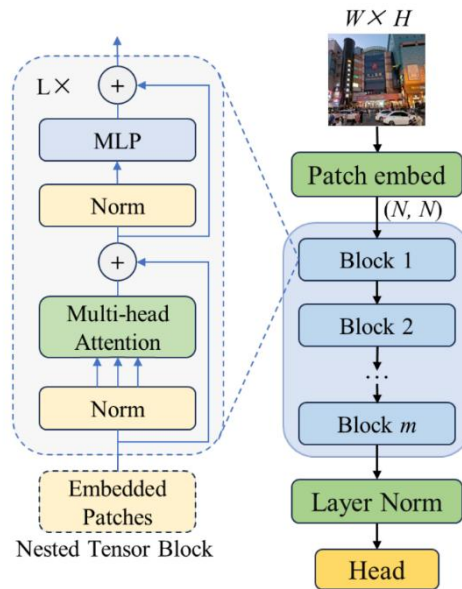
B. DINOv2

DINOv2 (Distillation of knowledge with NO labels v2) es un modelo de visión fundamental, no supervisado, desarrollado por Meta AI. La singularidad de DINOv2 reside en que aprende características visuales de propósito general de modo auto-supervisado a partir de 142 millones de imágenes sin etiquetar. Debido a su arquitectura ViT y el framework de auto-atención, el modelo produce embeddings visuales superiores. Lo más notable es su capacidad para ofrecer un rendimiento de vanguardia 'listo para usar' (out-of-the-box) en muchas de tareas de visión por computadora como clasificación, segmentación, y estimación de profundidad monocular, sin requerir el ajuste fino (fine-tuning) específico para cada tarea. Este enfoque demuestra que el escalamiento de los métodos de auto-supervisión

discriminativa puede producir características visuales transferibles comparables o superiores a las de los modelos entrenados con supervisión débil (Oquab et al., 2023).

Figura 2

Arquitectura DINOv2.



Nota. Tomado de Huang et al. (2023).

2.2.3 Capas base de aprendizaje de la red

Las capas base de aprendizaje de la red se refieren a los componentes fundamentales o bloques de construcción algorítmicos que constituyen la arquitectura de una red neuronal. Estas capas son responsables de realizar transformaciones específicas sobre los datos de entrada, permitiendo a la red aprender y extraer representaciones jerárquicas y cada vez más abstractas de los mismos a medida que los datos fluye a través de ellas (Goodfellow et al., 2016).

Cada tipo de capa está diseñada para capturar diferentes tipos de patrones o características. Por ejemplo, en el ámbito de la visión por computadoras, las capas convolucionales son expertas en identificar patrones locales y jerárquicos como bordes, texturas y formas en imágenes (LeCun et al., 2015).

Mientras que los bloques de Auto-Atención (como los encontrados en los Transformers) son particularmente hábiles para modelar dependencias a largo plazo y la relativa relevancia de diferentes partes de la entrada, independientemente de su posición (Vaswani et al., 2017).

La combinación y la secuencia de estas capas base son lo que define la arquitectura

general de una red y su capacidad para ejecutar tareas específicas de aprendizaje en el mundo real, como clasificación de imágenes, detección de objetos o procesamiento de lenguaje natural (Goodfellow et al., 2016).

Estas capas cumplen la función de ser los instrumentos computacionales mediante los cuales las redes neuronales logran extraer conocimiento de los datos para luego ejecutar las clasificaciones.

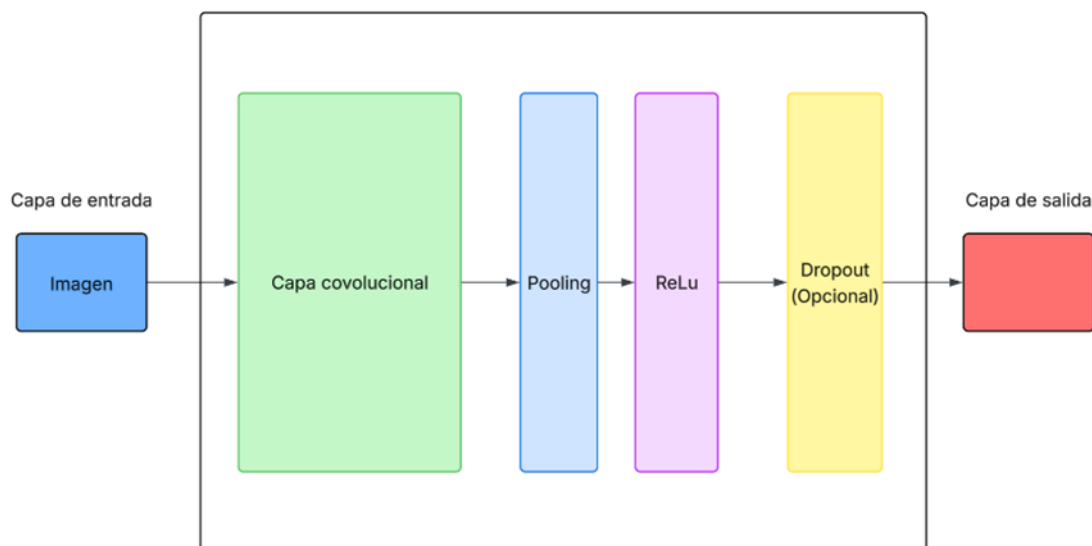
A. Bloques convolucionales

Componentes clave en las Redes Neuronales Convolucionales (CNNs) que realizan operaciones de convolución, aplicando filtros para detectar patrones locales en los datos de entrada, como bordes o texturas en imágenes (LeCun et al., 2015). Estas operaciones permiten extraer características jerárquicas de distintos niveles de abstracción, facilitando el reconocimiento y la clasificación de objetos visuales.

De acuerdo con los diseños clásicos, un bloque convolucional integra una capa de convolución, una capa de pooling (para la reducción dimensional), una función ReLU (para la no linealidad) y, ocasionalmente, una capa de dropout (para la prevención del sobreajuste). La combinación de estas operaciones es lo que dota a las CNNs de la capacidad de aprender representaciones espaciales de forma robusta y eficiente.

Figura 3

Bloque convolucional clásico.



En arquitecturas contemporáneas de detección de objetos, como es el caso de YOLOv5, la estructura del bloque convolucional ha evolucionado para maximizar la eficiencia computacional y la precisión. A diferencia de los esquemas tradicionales, el bloque fundamental (denominado ConvModule) prescinde de capas de pooling externas y opta por una integración técnica superior compuesta por tres fases críticas:

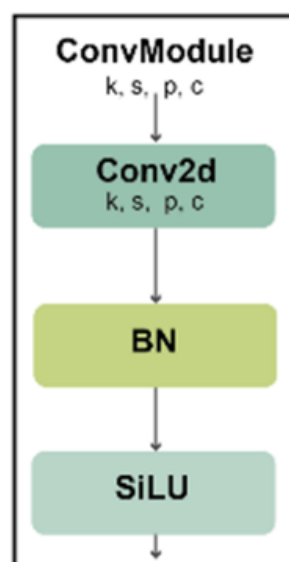
Capa Convolutiva (Conv2d): Es el componente encargado de la extracción de características mediante el uso de kernels (filtros). En este diseño, el parámetro de zancada (stride) adquiere un rol protagonista; cuando se requiere reducir la resolución espacial de la imagen, se ajusta el stride dentro de esta misma capa, eliminando la necesidad de una capa de pooling adicional. Esto permite una reducción dimensional más "aprendible" y menos agresiva con la información espacial.

Normalización por Lotes (Batch Normalization - BN): Ubicada inmediatamente después de la convolución, esta capa normaliza las activaciones de salida para que mantengan una media de cero y una varianza unitaria. Su implementación es vital en redes profundas, ya que actúa como un regularizador que reduce la dependencia del modelo hacia la inicialización de los parámetros, permitiendo un entrenamiento más estable y rápido al mitigar el desplazamiento de la covarianza interna.

Función de Activación SiLU (Sigmoid Linear Unit): Representa el componente de no linealidad del bloque. La función SiLU, definida matemáticamente como $f(x) = x \cdot \sigma(x)$ es una versión suavizada de la función ReLU. Su principal ventaja radica en que permite el flujo de pequeños gradientes negativos, lo cual evita el problema de las "neuronas muertas" y mejora la capacidad de generalización del modelo en tareas complejas de visión artificial.

Figura 4

Bloque convolutiva del YOLOv5.



Nota. Tomado de Terven et al. (2023).

B. Bloques de auto-atención

Un mecanismo que permite a una red neuronal ponderar la importancia de diferentes partes de su entrada al procesarla. Es fundamental en las arquitecturas Transformer, permitiendo al modelo enfocarse en las relaciones entre elementos distantes en una secuencia o imagen (Vaswani et al., 2017).

Tal como se establece en 'Attention Is All You Need', la red Transformer se articula a partir de múltiples bloques de atención, agrupados en sus dos módulos principales: el codificador y el decodificador. La estructura del codificador incluye un total de seis bloques de auto-atención, cada uno integrado por una capa de Multi-Head Attention y una Feed Forward Network, ambas acompañadas de mecanismos de suma residual y normalización (Add & Norm).

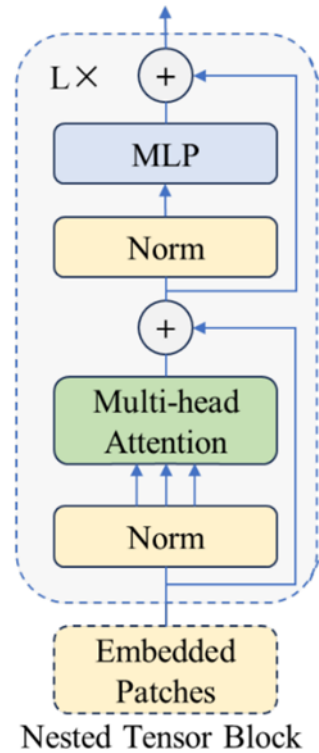
El decodificador, por su parte, también se compone de seis bloques, con una estructura que incorpora una Masked Multi-Head Attention, una Multi-Head Attention conectada con la salida del codificador, y una Feed Forward Network, todas seguidas de sus respectivas operaciones de normalización y conexiones residuales.

Cada capa de Multi-Head Attention se divide en ocho cabezas de atención (heads), lo que permite al modelo aprender múltiples relaciones de dependencia de manera paralela. Esta organización estructural convierte a los bloques de auto-atención en el núcleo funcional del Transformer, al posibilitar un procesamiento eficiente y contextualizado de la información de entrada.

Si bien el ConvModule constituye la unidad fundamental de procesamiento en arquitecturas como YOLOv5 mediante el aprendizaje de filtros locales, en modelos basados en visión por computadora de última generación como DINOv2, dicho rol es asumido por el bloque de auto-atención. El mecanismo de Multi-Head Self-Attention actúa como el núcleo de extracción de características, sustituyendo la operación de convolución por un cálculo de afinidad global que permite al modelo capturar dependencias espaciales de largo alcance desde las primeras capas de la red.

Figura 5

Bloque de auto-atención del Dinov2.



Nota. Tomado de Huang et al. (2023)

2.2.4 Eficacia de la red

La eficacia de la red es la aptitud del sistema (red neuronal) para cumplir sus metas y funcionar con efectividad, logrando así los objetivos establecidos durante su diseño y entrenamiento (Goodfellow et al., 2016). En el contexto de la visión por computadora, la eficacia se mide comúnmente por qué tan bien el modelo puede realizar su tarea de clasificación sobre datos nuevos e invisibles.

Para medir la eficacia de la red, se emplean métricas cuyo propósito es establecer, de forma numérica, siendo métrica de acierto que relaciona las clasificaciones del modelo con los resultados reales. Para evaluar el rendimiento en problemas de clasificación (tales como el reconocimiento de enfermedades o la identificación de objetos), la sensibilidad y la especificidad se consideran dos indicadores cruciales (Powers, 2020).

A. Sensibilidad

Son los verdaderos positivos que fueron identificados correctamente por la red. Es crucial en contextos donde es vital no perder ningún caso positivo (Powers, 2020).

B. Especificidad

Son los verdaderos negativos que fueron identificados correctamente por la red entre todos los casos negativos reales (Powers, 2020).

2.2.5 Eficiencia de la red

Se refiere a qué tan bien una red neuronal utiliza sus recursos para lograr sus objetivos (Goodfellow et al., 2016). A diferencia de la eficacia, que se centra en la precisión de los resultados, la eficiencia se enfoca en el costo computacional y temporal asociado con la operación de la red. En el contexto del aprendizaje profundo, la eficacia se traduce directamente en la cantidad de recursos de hardware consumidos y el tiempo requerido para las etapas clave de su ciclo de vida, como el entrenamiento y la inferencia.

Una red eficiente es aquella que puede ofrecer un rendimiento deseado minimizando el uso de recursos como la CPU o GPU y la memoria RAM, así como el tiempo requerido para el entrenamiento y la clasificación (Sze et al., 2017). Esto es crucial para la implementación práctica de modelos, especialmente en escenarios con recursos limitados (como dispositivos móviles) o donde la latencia es crítica (como sistemas en tiempo real).

A. Tiempo de entrenamiento

El tiempo que una red convolucional requiere para aprender a partir de un conjunto de datos de entrenamiento (Goodfellow et al., 2016).

B. Tiempo de inferencia

El tiempo que tarda una red entrenada en procesar una nueva entrada (por ejemplo, una imagen) y producir una clasificación. Es crucial para aplicaciones en tiempo real (Goodfellow et al., 2016).

2.2.6 Uso de GPU/CPU/RAM

El uso de GPU, CPU y RAM se refiere a la utilización de los principales recursos de hardware computacional requeridos para el funcionamiento eficiente de las redes neuronales, tanto en sus fases de entrenamiento como de inferencia. Estos componentes son fundamentales para procesar la vasta cantidad de operaciones y datos inherentes al aprendizaje profundo (Goodfellow et al., 2016).

Las GPUs (Unidades de Procesamiento Gráfico) son procesadores altamente especializados que están equipados con miles de núcleos. Esta arquitectura les confiere la capacidad de manejar múltiples cálculos de manera simultánea (procesamiento paralelo), lo cual es ideal para el entrenamiento y la inferencia de redes neuronales (INTEL Corporation, 2025).

La CPU es el "cerebro" principal de una computadora, responsable de ejecutar las instrucciones generales. Aunque menos eficientes que las GPU para las operaciones de redes neuronales a gran escala, siguen siendo cruciales para tareas de preprocesamiento de datos, control de flujo y operaciones secuenciales (Arm Ltd., 2025).

La RAM es una memoria volátil y rápida que cumple la función de almacenar de forma temporal los datos y las aplicaciones que el sistema está utilizando activamente. En el contexto de las redes neuronales, la RAM es esencialmente utilizada para cargar el modelo en sí, conteniendo el conjunto de datos de entrenamiento y maneja las activaciones transitorias que se generan en el proceso (Hennessy & Patterson, 2019).

2.3 MARCO REFERENCIAL

2.3.1 Reconocimiento de enfermedades de cacao

Identificar y reconocer adecuadamente las enfermedades en el cacao es fundamental para garantizar una producción sostenible y de alta calidad. Para evaluar el impacto real de las enfermedades en los cultivos de cacao, es fundamental considerar las distintas variedades presentes. Esto se debe a que cada variedad puede exhibir un grado de vulnerabilidad particular ante diferentes agentes patógenos. En este contexto, se da especial importancia a las variedades que poseen rasgos morfológicos y genéticos distintivos.

A. Variedad de cacao

Se refiere a las diferentes agrupaciones genéticas identificadas dentro de la especie *Theobroma cacao* L. Estas variedades se clasifican en base a su diversidad genética y se distinguen por sus características morfológicas y genéticas únicas. La diversidad genética del cacao es crucial para la selección de cultivos resistentes a enfermedades, adaptados a diferentes condiciones climáticas y con perfiles de sabor específicos para la producción de chocolate. (Motamayor et al., 2008) subrayan que las variedades de cacao, siendo grupos genéticos con características únicas (rendimiento, resistencia a enfermedades y calidad), tienen un impacto directo en la producción y el sabor del chocolate. Por consiguiente, la clasificación precisa y la comprensión de su diversidad genética y distribución son cruciales para implementar adecuadas estrategias de mejora y conservación. El VRAEM es ampliamente reconocido por la producción de fino cacao con aroma, el cual es muy solicitado por los mercados internacionales por su muy buena calidad. En esta región, es importante señalar que las variedades VRAE-99 y CCN-51 son las que presentan la mayor prevalencia en los cultivos (Devida, 2024).

Figura 6

Frutos de cacao y variedades.



Fruto de cacao CCN-51



Fruto de cacao VRAE-99



Fruto de cacao morado criollo



Fruto de cacao amarillo criollo

B. Tipo de enfermedad

El término "tipo de enfermedad" en las plantas se refiere a respuestas anormales causadas por la interacción con patógenos o factores ambientales, tales como sequías, temperaturas extremas o exceso de agua. Estas respuestas alteran la forma, función e

integridad de las plantas, y en casos severos, pueden provocar su muerte parcial o total (Agrios, 2005). En el contexto del cacao, este término engloba enfermedades que afectan específicamente los frutos del árbol *Theobroma cacao* L., agrupándolas en categorías según su impacto, síntomas y causas.

Entre las principales enfermedades del cacao destaca la pudrición parda o negra, causada por el oomiceto *Phytophthora palmivora*, clasificado dentro del reino Protista (CropLife Latin America, 2024a).

Figura 7

La Phytophthora palmivora en el fruto de cacao.



Otra enfermedad importante es la moniliasis, la cual es provocada por el hongo hemibiotrófico *Moniliophthora roreri*. Este patógeno se desarrolla favorablemente en ambientes con elevada humedad y temperatura. Se manifiesta mediante la aparición de manchas oscuras en el fruto y un distintivo polvo blanco que contribuye a su fácil dispersión (CropLife Latin America, 2024b).

Figura 8

La moniliasis en el fruto de cacao.



El hongo *Moniliophthora perniciosa*, endémico de las regiones tropicales de Sudamérica, es el agente causal de la escoba de bruja del cacao. Se dispersa por medio de tejido vegetal como semillas, varetas, frutos, brotes y ramas, entre otros. Afecta diferentes especies de los géneros *Theobroma*, *Herrania* y de las familias Solanaceae, Bignoniaceae y Malpighiaceae, produciendo crecimientos anormales y lesiones en brotes, ramas, cojines

florales y frutos. Los ciclos sucesivos de este patógeno pueden causar la muerte de la planta y, en consecuencia, provocar pérdidas considerables en la producción de cacao (Rivera & Sneyder, 2020).

Figura 9

La escoba de bruja en el fruto de cacao.



El mazorquero del fruto de cacao, también conocido como perforador de mazorcas de cacao o *Carmenta foraseminis*, constituye una grave amenaza para la producción de cacao en las regiones tropicales de América. Esta plaga ataca diversos órganos del árbol, incluyendo ramas, brotes, cojines florales y, especialmente, los frutos (Cabezas et al., 2017).

Figura 10

El mazorquero en el fruto de cacao.



Es fundamental identificar con precisión el agente causal de cada enfermedad para un diagnóstico correcto. Específicamente, existe un desafío en la diferenciación de patologías como la escoba de bruja (*Moniliophthora perniciosa*) y la moniliasis (*Moniliophthora roreri*) y dado que ambas son causadas por hongos del género *Moniliophthora*. Adicionalmente, la pudrición negra del cacao es provocada por el patógeno fúngico *Phytophthora palmivora*,

mientras que el mazorquero es resultado de la infestación de la larva de la polilla *Carmenta foraseminis*.

2.3.2 Preprocesamiento de imágenes

Es una etapa esencial en la visión por computadora y el Deep Learning, cuya función principal es acondicionar y estandarizar los datos visuales como paso previo a su utilización en la arquitectura de la red neuronal (Krizhevsky et al., 2012). Este proceso es crucial para optimizar el rendimiento y la robustez del modelo. Las tareas habituales de preprocesamiento incluyen la redimensión de las imágenes para que cumplan con los requisitos de entrada de la red, la normalización, la eliminación de ruido para mejorar la calidad, y el ajuste de brillo o contraste con el fin de realzar las características relevantes para el análisis.

2.3.3 Red neuronal convolucional (CCN)

Es un modelo de red neuronal artificial sumamente eficaz y diseñado para el análisis de datos estructurados en cuadrícula (imágenes o videos). Su rasgo distintivo es la implementación de capas convolucionales que utilizan filtros para identificar patrones y características locales de forma jerárquica en la imagen (LeCun et al., 2015), como bordes o texturas. Tras esta extracción de características, las CNNs generalmente utilizan capas de pooling para reducir la dimensionalidad y capas completamente conectadas para llevar a cabo la tarea final de clasificación, segmentación o identificación de objetos.

2.3.4 Red neuronal de auto-atención

En una red neuronal de Auto-Atención (o, más precisamente, un mecanismo de Auto-Atención dentro de una red neuronal) Se trata de un mecanismo estructural clave que permite al modelo computacional evaluar la importancia relativa de los elementos de su entrada durante el procesamiento. En esencia, el mecanismo de Auto-Atención (Self-Attention) tiene como función calcular una representación de cada elemento de la entrada (como una palabra en una oración o un parche en una imagen) al considerar la relación y la importancia de dicho elemento con todos los demás componentes dentro de la misma secuencia o conjunto (Vaswani et al., 2017). Esto le permite a la red enfocarse dinámicamente en las características más relevantes, independientemente de su posición física, a diferencia de las redes convolucionales o recurrentes tradicionales que tienen un campo de visión local o secuencial limitado.

El mecanismo de Auto-Atención fue adaptado exitosamente al campo de la visión por computadora con la creación de modelos de vanguardia como los Vision Transformers (Dosovitskiy et al., 2021). Al calcular la relevancia mutua ('atención') entre los elementos de entrada, estas redes pueden capturar dependencias a largo alcance y relaciones complejas, lo que resulta fundamental para la comprensión del contexto global de los datos.

2.3.5 Prueba de wilcoxon

La prueba de Wilcoxon es un método estadístico no paramétrico diseñado para comparar dos muestras relacionadas o emparejadas, con el fin de determinar si existen diferencias significativas entre sus rangos. A diferencia de las pruebas paramétricas tradicionales, este procedimiento no asume que los datos sigan una distribución normal, lo que lo hace ideal para evaluar el rendimiento de algoritmos de aprendizaje profundo donde las métricas de error o confianza suelen presentar distribuciones sesgadas. Según (Siegel & Castellan Jr., 1988), esta prueba es la alternativa más robusta a la prueba t de Student para muestras dependientes, basando su análisis en la magnitud y el signo de las diferencias entre pares de observaciones. En este estudio, la prueba se utiliza para validar la consistencia del desempeño entre las arquitecturas evaluadas, considerando la Sensibilidad y los Niveles de Confianza como variables métricas principales.

CAPÍTULO III

MATERIAL Y MÉTODOS

3.1 TIPO Y NIVEL DE INVESTIGACIÓN

3.1.1. Tipo de investigación

La investigación observacional se caracteriza, según (Hernández Sampieri et al., 2014), basado en la observación directa de las ocurrencias (fenómenos) en su ambiente natural, lo que facilita una comprensión detallada de las situaciones sin intervención o manipulación por parte del investigador. En el presente estudio, se observan las variables relacionadas con las enfermedades en los frutos de cacao a través de imágenes, sin intervención directa en el cultivo, lo que clasifica a esta investigación como **observacional**.

Según (Cook & Campbell, 1979), una investigación prospectiva "se caracteriza por la recolección de datos antes de que ocurran los eventos que se desean estudiar". En el presente estudio, al recolectar imágenes de frutos de cacao en diferentes ciclos agrícolas, se inscribe dentro de este diseño investigación **prospectiva**.

Un estudio longitudinal se define como un diseño que "permite observar los cambios a lo largo del tiempo en una población específica, realizando mediciones en múltiples puntos temporales" (Menard, 2002). En coherencia con esta metodología, la presente investigación se clasifica como longitudinal, ya que la recolección de las imágenes de los frutos de cacao se realizó en diferentes temporadas de cosecha.

Un estudio descriptivo se caracteriza por tener como meta "detallar y especificar las propiedades y las características del fenómeno a analizar" (Hernández Sampieri et al., 2014). En correspondencia con esta definición, la metodología de la investigación se establece bajo un enfoque descriptivo, dado que se enfoca en la descripción en profundidad de una única variable de interés: el rendimiento de la red neuronal. Dicho rendimiento se evalúa y detalla mediante métricas clave (como la sensibilidad y la especificidad), sin que el estudio busque establecer correlaciones o explorar relaciones causales con otras variables.

3.1.2. Nivel de investigación

Según (Hernández Sampieri et al., 2014), el objetivo primordial en una investigación de nivel descriptivo es detallar los fenómenos, sucesos o situaciones de interés para el investigador. En este nivel, la meta principal es recoger información y medir las variables deseadas para conocer las propiedades y características que identifican al objeto de estudio. Con base en esto, la presente investigación se clasifica como de nivel descriptivo, pues su objetivo es monitorear el rendimiento de la red neuronal mediante la recopilación de datos de dimensiones clave tales como el paradigma de la red neuronal, las capas de aprendizaje, y la eficacia y eficiencia de la red, los cuales pueden ser evaluados.

3.2 DISEÑO DE LA INVESTIGACIÓN

Dado que el estudio sigue un diseño no experimental, implica que las variables no son intervenidas por el investigador; los datos, en cambio, son recolectados directamente de su entorno natural sin intervención directa (Hernández Sampieri et al., 2014). En concordancia con este diseño, los datos consistieron en imágenes de frutos de cacao. Las imágenes se obtuvieron tras un periodo de observación y análisis de un año, con la finalidad de diagnosticar las patologías de los frutos utilizando aprendizaje profundo.

3.3 VARIABLES

3.3.1 Definición conceptual de variables

Variable de interés

Rendimiento de la red neuronal

Capacidad del modelo de aprendizaje profundo para detectar y clasificar correctamente el estado sanitario de los frutos de cacao a partir de imágenes digitales. Refleja el grado de precisión, eficiencia y efectividad del algoritmo en el proceso de identificación de enfermedades.

Variables descriptivas (Dimensiones)

a. Paradigma de computación de la red

Tipo de arquitectura empleada en la red neuronal (por ejemplo, basada en convolución, auto-atención), la cual define el modo en que el modelo procesa las imágenes.

b. Capas base de aprendizaje de la red

Se refieren a los componentes estructurales (ya sean bloques convolucionales o de auto-atención) que tienen como función primordial permitir la extracción progresiva de las características visuales relevantes presentes en los frutos de cacao.

c. Eficacia de la red

Grado de precisión del modelo al clasificar los frutos, medido por métricas como sensibilidad, especificidad, precisión y exactitud.

d. Eficiencia de la red

Capacidad del modelo para procesar datos en un tiempo razonable y con un uso óptimo de recursos computacionales (CPU, GPU y RAM).

3.3.2 Definición operacional de variables

Variable de interés

Rendimiento de la red neuronal

Se evalúa a través del proceso de clasificación de las imágenes de los frutos de cacao,

las cuales se dividen en dos categorías binarias: saludable y enfermo. Esta evaluación se realiza utilizando un conjunto de métricas de desempeño del modelo, que incluyen la precisión, exactitud, la sensibilidad, la especificidad y el F1 (F1-score).

Variables descriptivas (Dimensiones)

a. **Paradigma de computación de la red**

Tipo de arquitectura o modelo empleado y son YOLOv5, DINOv2.

b. **Capas base de aprendizaje de la red**

Tipo de bloque utilizado (convolucional o de auto-atención) y el tipo de información que procesan (local o global).

c. **Eficacia de la red**

Sensibilidad, Especificidad, Precisión, F1-Score, se mide por porcentaje (%).

d. **Eficiencia de la red**

Tiempo de entrenamiento (minutos), tiempo de inferencia (milisegundos), uso de CPU/GPU/RAM porcentaje (%).

3.4 POBLACIÓN Y MUESTRA

3.4.1 Población

Se comprende como la colección exhaustiva de unidades que se ajustan a los criterios específicos de la investigación y constituyen todos los elementos posibles a estudiar (Hernández Sampieri et al., 2014). Para este caso, la población abarcar todos los frutos de cacao de 25 parcelas en el Centro Poblado de Omayá afectados o no por enfermedades durante el año 2024, que son de 10150 frutos. Trabajar con una población completa es a menudo inviable, por lo que se opta por la selección de una muestra representativa.

3.4.2 Muestreo

En esta investigación se utilizó un muestreo por cuotas, una técnica no probabilística donde se seleccionan muestras en función de cuotas preestablecidas según ciertas características. Según (Hernández Sampieri et al., 2014), define esta técnica como aquella que asigna una cuota a cada subgrupo identificado dentro de la población para garantizar su proporcionalidad. En este caso, las cuotas fueron establecidas en base a los tipos de frutos de cacao en diferentes parcelas.

3.4.3 Muestra

Se define como muestra representativa el subconjunto que refleja de modo preciso los atributos y rasgos de la población de la cual fue extraída (Cochran, 1977). Para este estudio, la muestra está compuesta por 3,104 imágenes de frutos de cacao. Los datos se obtuvieron de parcelas que presentaban una alta incidencia de enfermedades, siguiendo un enfoque de muestreo por conveniencia. Esta estrategia metodológica se adoptó para

asegurar la diversidad necesaria en los datos y maximizar la relevancia del análisis. En consecuencia, se considera relevante que el tamaño y la composición de la muestra son suficientes para capturar la variabilidad inherente en la población de frutos enfermos, lo cual permite realizar un análisis detallado de la relación entre las características visuales y las patologías del cacao.

3.5 OPERACIONALIZACIÓN DE LAS VARIABLES

Tabla 1

Operacionalización de variable.

VARIABLE	DIMENSIÓN	INDICADOR	PREGUNTA
Rendimiento de la red neuronal	Paradigma de la red	Modelo de red neuronal (YOLOv5, DINOv2)	¿Qué tipo de red se utilizó?
	Capas de aprendizaje	Bloques convolucionales (Local, Global)	¿Cómo procesa los datos un bloque convolucional?
		Bloques de auto-atención (Local, Global)	¿Cómo procesa los datos un bloque de auto-atención?
	Eficacia	Sensibilidad (alta, moderada y baja)	¿Qué porcentaje de sensibilidad obtuvo el modelo?
		Especificidad (alta, moderada y baja)	¿Qué porcentaje de especificidad alcanzó el modelo?
	Eficiencia	Tiempo de entrenamiento (minutos)	¿Cuánto minutos tomó entrenar el modelo?
		Tiempo de prueba o inferencia (milisegundos).	¿Cuánto segundos toma el modelo en analizar una imagen?
		Uso de CPU/GPU/RAM (bajo, medio y alto)	¿Cuánto porcentaje es el uso de recursos durante el entrenamiento?
Enfermedad en el fruto de cacao	Resultado	Estado del fruto (fruto saludable o fruto enfermo).	¿Cuál la clasificación del modelo?

3.6 TÉCNICAS E INSTRUMENTOS DE LA INVESTIGACIÓN

3.6.1 Técnicas

Según (Hernández Sampieri et al., 2014), el análisis documental es un método que sirve tanto para recopilar datos como para examinar la información previamente registrada. Este proceso permite utilizar registros y archivos que, aunque no fueron creados con propósitos investigativos primarios, son capaces de proporcionar material relevante y significativo para el estudio en curso. En el caso de tu estudio, se aplicará esta técnica para recopilar y analizar sistemáticamente las imágenes de frutos de cacao capturadas en diferentes temporadas de cosecha en el Centro Poblado de Omayá. Estas imágenes servirán como fuente primaria de datos para el análisis mediante técnicas de aprendizaje profundo.

3.6.2 Instrumentos

Para la recolección de este estudio se realizó mediante la aplicación del principio de revisión documental, almacenamiento y acumulación de registros y documentación oficial, según (Hernández Sampieri et al., 2014).

En este estudio, el instrumento específico para la recolección de datos primarios es el registro fotográfico. Este método se empleará para obtener datos visuales de los frutos de cacao afectados por enfermedades. Para garantizar la calidad de los datos, las imágenes fueron capturadas utilizando una cámara de alta resolución 24 MP, 100MP. Posteriormente, estas imágenes se someterán a un proceso de análisis mediante técnicas de Deep Learning con el objetivo de detectar y clasificar las enfermedades. Por lo tanto, la cámara del móvil actúa como un instrumento crucial para la obtención de los datos primarios requeridos.

3.6.3 Validez del instrumento

Se entiende por validez del instrumento el grado en que la herramienta mide apropiadamente el contenido específico para el cual fue diseñada (Hernández Sampieri et al., 2014). En el contexto de este estudio, la validez del registro fotográfico se asegura mediante la inclusión de todas las características visuales relevantes que sirven como indicadores de la presencia de enfermedades.

3.6.4 Confiabilidad de instrumento

No se realiza un proceso de evaluación de la confiabilidad del instrumento en esta investigación. La justificación de esta omisión es que la fuente principal de datos es el registro fotográfico. Por lo tanto, la técnica de análisis aplicada se clasifica, para los fines de esta investigación, como análisis documental.

3.7 PROCEDIMIENTOS

3.7.1 Estrategia de prueba de hipótesis

En esta investigación se omite la implementación de una prueba de hipótesis. Esto se justifica porque el alcance de la investigación es descriptivo, lo que implica que su meta

principal es caracterizar o describir los fenómenos estudiados, sin buscar establecer o verificar relaciones de causalidad o correlación entre variables.

3.7.2 Técnicas de procesamiento de datos

A. Procesamiento de datos en Deep Learning

El desarrollo de los modelos de aprendizaje profundo para el reconocimiento de fitopatologías en el cacao sigue un flujo de trabajo sistemático e iterativo. Este ciclo asegura la integridad de los datos y la robustez de las inferencias mediante los siguiente:

1. Construcción del dataset:

Construcción sistemática de datos visuales en condiciones en las parcelas. La recolección captura la variabilidad fenológica del fruto y las condiciones estocásticas de iluminación, garantizando una muestra estadísticamente representativa y diversa.

2. Preprocesamiento de datos:

a. Estandarización de la resolución y normalización: Los recortes fueron redimensionados a una resolución uniforme de 224x224 píxeles y normalizados en sus espacios de color. Esto adapta la información visual a los requisitos matriciales de las arquitecturas YOLOv5(CNN) y DINOv2 (ViT) sin degradar las características morfológicas clave para la detección de patógenos.

b. Integridad y calidad del dataset: Protocolos para asegurar la ausencia de datos nulos, duplicados o defectos técnicos que puedan inducir sesgos.

3. Proceso de entrenamiento del modelo:

Fase de transferencia de conocimiento donde los modelos ajustan sus parámetros internos. Se compone de.

a. Selección de modelos: Contraste de paradigmas entre redes convolucionales (YOLOv5) y mecanismos de auto-atención (DINOv2) para evaluar la eficacia en la extracción de features locales y globales.

b. División de la dataset: Segmentación del dataset en bloques de entrenamiento, validación y prueba (80/10/10), asegurando una evaluación imparcial en cada etapa del aprendizaje.

c. Configuración de hiperparámetros: Optimización mediante Grid Search de variables críticas como learning rate y optimizadores (Adam/AdamW) para alcanzar la convergencia óptima.

d. Entrenamiento del modelo: Supervisión del aprendizaje en tiempo real, utilizando el conjunto de validación para mitigar el sobreajuste (overfitting).

4. Proceso de evaluación y validación:

Evaluación definitiva del rendimiento utilizando el conjunto de prueba (test ciego). Los resultados son sometidos a la Prueba de rangos con signo de Wilcoxon, otorgando

rigor estadístico a la comparación de métricas y validando la superioridad de la arquitectura seleccionada para su despliegue real.

B. Herramientas para el procesamiento de datos

El conjunto de herramientas comprende las librerías especializadas para el desarrollo y entrenamiento de modelos de Deep Learning. Estas herramientas facilitan la abstracción de complejas operaciones matemáticas, optimizando el ciclo completo de desarrollo, desde la adquisición y manipulación de datos hasta la fase de entrenamiento y optimización del modelo. A continuación, en la Tabla 2 se detallan las librerías empleadas.

Tabla 2*Herramientas para el desarrollo de modelos Deep Learning.*

Proceso	Librerías	Función específica
Preprocesamiento y manipulación	pandas (pd)	Estructuración y manejo de datos para el análisis de balance de clases.
	numpy (np)	Soporte para operaciones numéricas y manipulación de arrays de datos e imágenes.
	PIL (Image)	Manipulación y apertura de archivos de imágenes.
Arquitectura y entrenamiento	torch (PyTorch)	Marco principal para el Deep Learning, manejo de tensores y entrenamiento.
	torchvision	Utilidades para la carga de datos (Dataset, DataLoader) y transformaciones (transforms).
	tqdm	Monitoreo del progreso del entrenamiento en bucles (loops).
Validación y división de Datos	pathlib, os	Gestión de rutas de archivos y directorios de forma organizada.
	shutil	Operaciones de alto nivel para mover y organizar imágenes entre conjuntos (entrenamiento/validación).
Evaluación y métricas	sklearn.metrics	Cálculo de métricas finales: Matriz de Confusión (confusion_matrix), Sensibilidad y F1-Score (f1_score).
	matplotlib.pyplot (plt)	Generación de gráficos (curvas de pérdida/métrica) y visualización de la Matriz de Confusión.

C. Técnicas de análisis de datos en imágenes

En esta investigación, el análisis de las imágenes no se basa en técnicas de procesamiento digital tradicional (filtros, segmentación manual o umbralización), sino en la capacidad analítica de los modelos de Deep Learning. La técnica de análisis consiste en la inferencia automatizada mediante las arquitecturas YOLOv5 y DINOv2, las cuales actúan como instrumentos de extracción de características complejas y clasificación de fitopatologías.

Este procedimiento de análisis se estructuró en tres etapas clave: primero, la preparación y normalización del dataset para asegurar la calidad de la entrada; segundo, el entrenamiento de los modelos para la optimización de los pesos sinápticos; y tercero, la evaluación del rendimiento del modelo como clasificador, cuantificando su eficacia y eficiencia operativa. Este enfoque permite que el análisis del contenido de las imágenes sea objetivo, escalable y validado estadísticamente mediante la prueba de Wilcoxon, determinando así la viabilidad de la solución en entornos agrícolas reales.

3.7.3 Diseño estadístico

La presente investigación se enmarca en un nivel descriptivo, por lo cual no se incluye un diseño destinado a la verificación de hipótesis. Los resultados relativos al rendimiento de las redes neuronales serán presentados y analizados de forma clara mediante el uso de tablas, figuras y pruebas de contraste que permitan validar la consistencia de las métricas obtenidas.

3.7.4 Análisis e interpretación de resultados

La evaluación comparativa de las arquitecturas YOLOv5 y DINOv2 se centra en el análisis de métricas de rendimiento clave: Precisión, Sensibilidad, Especificidad y F1-Score. Se prioriza la maximización de la Sensibilidad (Recall), con el objetivo de reducir al mínimo los Falsos Negativos, dado que en el contexto fitosanitario omitir la detección de un fruto enfermo conlleva un riesgo epidemiológico crítico. Para garantizar la validez estadística de la comparación, se aplicará la prueba de rangos con signo de Wilcoxon, permitiendo determinar si las variaciones en el desempeño son significativas. Finalmente, se interpretarán los tiempos de inferencia, la duración del entrenamiento y la eficiencia de recursos para identificar la solución que mejor se ajuste a las exigencias operativas de un entorno real.

CAPÍTULO IV

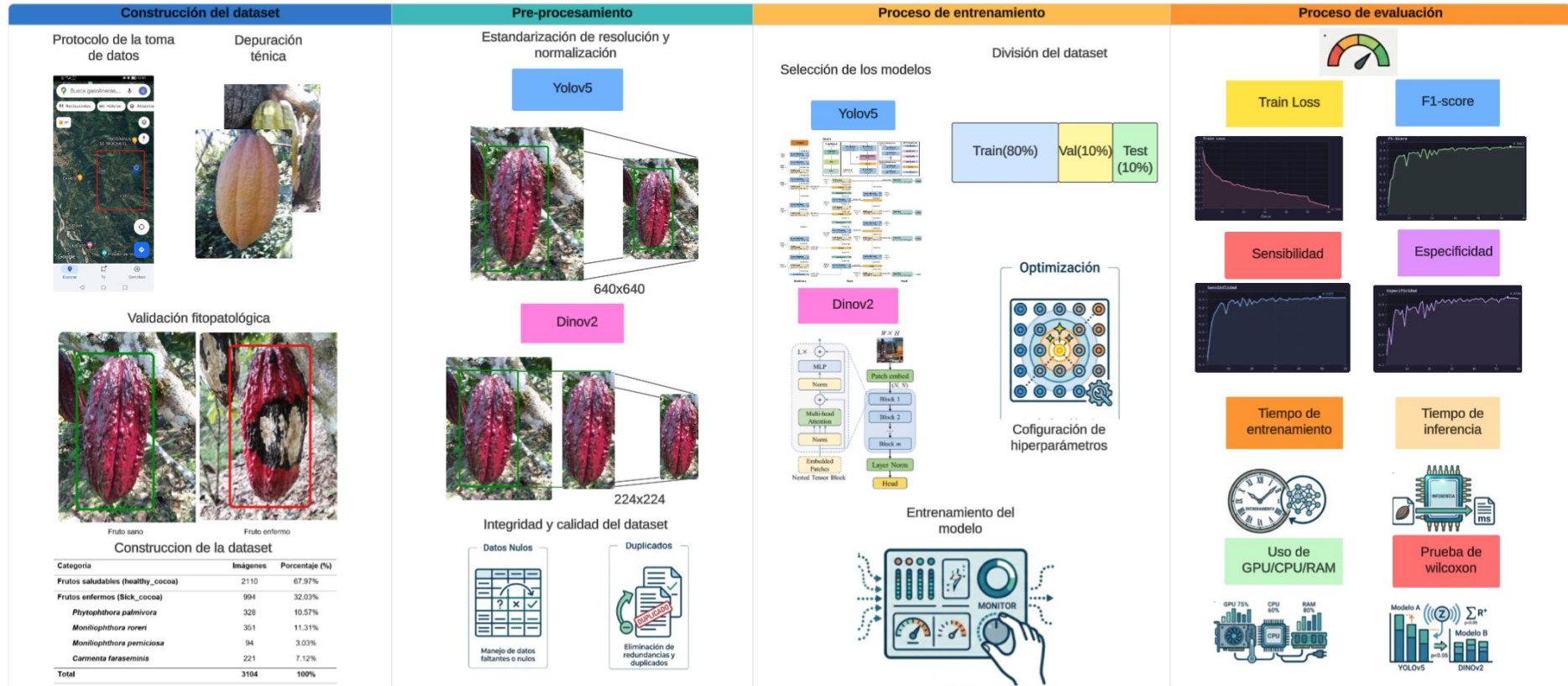
RESULTADOS Y DISCUSIÓN

4.1 RESULTADOS

Como culminación de la fase metodológica, la Figura 11 exhibe la arquitectura consolidada del flujo de trabajo experimental. Esta representación gráfica articula la progresión sistemática desde la generación del corpus de datos primarios en condiciones de campo hasta la validación estadística de los modelos, garantizando la trazabilidad e interconexión de cada fase experimental.

Figura 11

Flujo de trabajo para la clasificación de enfermedades en el cacao.



Descripción de los componentes del flujo de trabajo

A continuación, se ofrece una breve explicación de cada uno de los bloques del flujo de trabajo:

- a) **Construcción del dataset (Bloque I):** Describe el origen geográfico de los datos en el Centro Poblado de Omayá y el protocolo de depuración técnica aplicado. Se enfatiza la transición de la información en bruto hacia un conjunto de datos etiquetado bajo supervisión experta.
- b) **Pre-procesamiento (Bloque II):** Define el tratamiento diferenciado de los datos según la arquitectura. Las imágenes se normalizan a 640×640 para el modelo convolucional (YOLOv5) y se transforman mediante recortes de Regiones de Interés (RoI) a 224×224 para el modelo de auto-atención (DINOv2), integrando además protocolos de integridad de datos.
- c) **Proceso de entrenamiento (Bloque III):** Detalla la configuración de las redes y la implementación de la estrategia de Triple Partición (80/10/10). Incluye la optimización de hiperparámetros y el monitoreo de recursos de hardware (GPU/CPU) para asegurar la estabilidad del aprendizaje.
- d) **Proceso de validación (Bloque IV):** Presenta los indicadores de evaluación final, abarcando métricas de eficacia (Sensibilidad, F1-score), eficiencia (tiempos de procesamiento) y la validación de significancia mediante la prueba de Wilcoxon.

4.1.1 Construcción del dataset

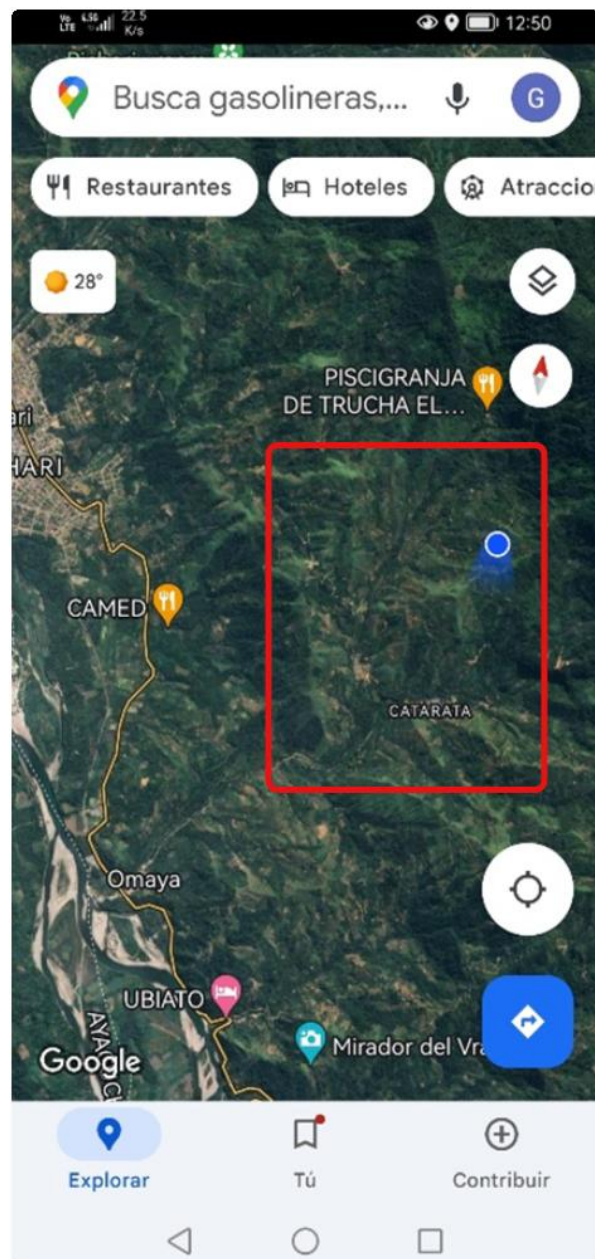
La generación del dataset se fundamentó en una metodología sistemática de adquisición y curación visual, requisito indispensable para asegurar la capacidad de generalización en arquitecturas de *Deep Learning*. Este procedimiento se alinea con los estándares del ciclo de vida de datos propuestos por la literatura académica especializada para la validación del *ground truth*. Este trabajo sigue el ciclo de vida de datos propuesto por (Everingham et al., 2009) y (Goodfellow et al., 2016).

A. Protocolo de toma de datos

La producción de cacao ejerce una influencia considerable en la economía del VRAEM, particularmente en Omayá. Ante el desafío fitosanitario que representan enfermedades como la mazorca negra y la moniliasis, el control de calidad es esencial para la industria (Devida, 2021). En este contexto, los datos fueron obtenidos mediante visitas de campo a parcelas del Centro Poblado de Omayá, Cusco. El campo de recolección está localizado en las coordenadas UTM: -12.5584191, -73.8062487 y la Figura 12 ilustra su locación.

Figura 12

Lugar de la recolección de datos.



Se empleó un muestreo por cuotas para garantizar la representatividad de los subgrupos patológicos. Para la estandarización de la captura, se establecieron los siguientes parámetros:

- **Restricción horaria:** Capturas realizadas entre las 11:00 am y las 3:00 pm para homogeneizar la iluminación natural y minimizar el ruido visual por sombras.
- **Distancia focal:** Adquisición a una distancia constante de 30 cm respecto al fruto para estandarizar la escala y el encuadre de las lesiones.

B. Especificaciones de hardware de captura de imágenes

Para la captura de imágenes de frutos de cacao, se utilizó un dispositivo móvil con sensores de alta resolución (40 MP y 100 MP), logrando una resolución promedio de 2394×3158 píxeles, óptima para identificar texturas patológicas complejas. Este proceso de recolección masiva generó un total inicial de 4,969 imágenes en bruto.

C. Depuración técnica

Posteriormente, se aplicó un protocolo de depuración técnica bajo criterios estrictos:

Criterios de exclusión: Se eliminaron imágenes con desenfoque (blur), oclusiones severas o saturación cromática que impidieran la visibilidad de las enfermedades.

Criterios de inclusión: Se conservaron de imágenes con nitidez técnica y alta fidelidad representativa de la patología.

Este proceso de curación técnica consolidó una muestra técnica final de **3,104** imágenes de frutos de cacao, cumpliendo con los criterios de suficiencia estadística para la generalización de resultados (Cochran, 1977).

D. Validación fitopatológica y establecimiento del ground truth

Una vez consolidada la muestra técnica final, se procedió a la fase de validación diagnóstica para establecer el *ground truth* (etiquetas verdaderas) de cada imagen. Este proceso consiste en la revisión rigurosa visual por un experto fitopatólogo para identificar y confirmar la presencia de sintomatología característica de las patologías prevalentes en la zona. Bajo este esquema de anotación supervisada, se asignaron etiquetas semánticas basadas en el estado fitosanitario real del fruto:

- **Cacao saludable (healthy_cocoa):** imágenes de ejemplares de frutos de cacao que presentan integridad en la corteza, ausencia de manchas necróticas y coloración óptima según su estadio de maduración.
- **Cacao enfermo (sick_cocoa):** imágenes de ejemplares de frutos de cacao con evidencia clínica de infección, incluyendo lesiones por *Phytophthora palmivora*, esporulación de *Moniliophthora roreri*, deformaciones por *Moniliophthora perniciosa* o daños mecánicos por *Carmenta faraseminis*.

Esta etapa de etiquetado por un experto es fundamental para garantizar la fiabilidad del dataset, asegurando que las características morfológicas y de textura capturadas correspondan estrictamente a la realidad biológica del cultivo. La transformación de estas etiquetas semánticas en valores binarios o categorías multiclase para el procesamiento computacional se detalla en las fases posteriores de diseño de arquitectura y clasificación. Final-

mente, la Tabla 2 presenta el dataset de frutos de cacao con su respectivo *ground truth* (etiqueta) por clase. Asimismo, la Figura 13 presenta una ilustración de frutos sanos y frutos enfermos de nuestro dataset.

El conjunto de datos de imágenes de frutos de cacao utilizado en esta investigación, recolectado en el Centro Poblado de Omayá, se encuentra disponible para su consulta y replicación en el repositorio público de la IEEE DataPort mediante el enlace <https://dx.doi.org/10.21227/ktfb-2132>. Asimismo, los resultados preliminares obtenidos durante la fase de construcción de este dataset y las pruebas iniciales con modelos YOLOv5 fueron presentados y publicados en el VIII International Congress of Systems Engineering (CIIS 2025), bajo el sello de la Association for Computing Machinery (Leon & Yauri, 2025).

Tabla 3

Dataset de enfermedades en frutos de cacao.

Categoría	Imágenes	Porcentaje (%)
Frutos saludables (healthy_cocoa)	2110	67.97%
Frutos enfermos (Sick_cocoa)	994	32.03%
<i>Phytophthora palmivora</i>	328	10.57%
<i>Moniliophthora roreri</i>	351	11.31%
<i>Moniliophthora perniciosa</i>	94	3.03%
<i>Carmenta faraseminis</i>	221	7.12%
Total	3104	100%

Figura 13

Ilustración de frutos sanos y frutos enfermos del dataset.



Fruto sano

Fruto enfermo

4.1.2 Pre-procesamiento de datos

La fase de preprocesamiento constituye el nexo crítico entre la adquisición de datos y el modelado neuronal. En este proceso, las 3,104 imágenes del dataset son sometidas a protocolos de normalización y ajuste dimensional, transformando la información espectral en tensores de entrada. Esta transición es fundamental para asegurar que las redes neuronales operen sobre representaciones numéricas que maximicen la detección de patrones fitopatológicos y minimicen el coste computacional.

A. Estandarización de la resolución y normalización

Debido a que los modelos de visión artificial requieren dimensiones fijas para realizar operaciones matriciales consistentes en las capas de entrada, se procedió a un reajuste de escala diferenciado según la naturaleza de la arquitectura:

- **Pre-procesamiento para YOLOv5 (Basado en convoluciones):** Las imágenes originales fueron redimensionadas globalmente a 640x640x3 píxeles. Este ajuste permite que la red mantenga el contexto espacial completo de la escena, facilitando la detección de los frutos en relación con su entorno.

- **Pre-procesamiento para DINOv2 (Basado en auto-atención):** A diferencia del modelo anterior, el procesamiento se centró en la extracción de Regiones de Interés (RoI). Utilizando las coordenadas de anotación, se realizó un recorte (cropping) individual de cada fruto, aislándolo del fondo. Posteriormente, estos recortes se redimensionaron a 224x224x3 píxeles. Esta estrategia permite que el mecanismo de auto-atención se concentre exclusivamente en las texturas y morfologías patológicas, eliminando el ruido visual del contexto.

B. Integridad y calidad del dataset

Debido a que el proceso de recolección de datos fue realizado de manera sistemática y controlada en las parcelas de Omayá, el dataset presenta niveles óptimos de integridad, abordados bajo los siguientes criterios:

- **Manejo de datos faltantes o nulos:** Al tratarse de un dataset generado internamente con supervisión directa, se garantiza la inexistencia de registros incompletos o valores nulos. Cada entrada en el conjunto de datos contiene la totalidad de metadatos y etiquetas necesarias, lo que asegura un flujo de procesamiento ágil y sin interrupciones por gestión de errores de datos.
- **Eliminación de redundancias y duplicados:** El control riguroso durante la fase de captura y etiquetado eliminó el riesgo de duplicidad de imágenes. Esta limpieza previene posibles sesgos de sobreajuste (overfitting) que podrían derivarse de la presencia de múltiples registros de una misma instancia en el conjunto de entrenamiento.

4.1.3 Proceso de entrenamiento del modelo

Esta fase describe la configuración experimental y la ejecución del aprendizaje para las arquitecturas seleccionadas. El objetivo es garantizar que los modelos capturen los patrones patológicos del cacao mediante un proceso controlado.

A. Selección de los modelos

Para la clasificación de los frutos de cacao, se optó por una estrategia de experimentación comparativa seleccionando dos arquitecturas de recientes y de alto desempeño:

- **YOLOv5 (You Only Look Once):** Seleccionada por su madurez tecnológica y su equilibrio probado entre precisión y velocidad de inferencia en dispositivos de borde (edge computing). A pesar de ser primariamente un detector, su capacidad para la extracción de características locales mediante bloques convolucionales (CSP-Darknet53) lo hace ideal para entornos agrícolas donde se requiere una respuesta en tiempo real. Según (Jocher et al., 2020), YOLOv5 permite una convergencia rápida y una alta robustez en la detección de objetos con texturas irregulares, características

propias de las lesiones necróticas del cacao. Su amplia adopción en la literatura reciente para el monitoreo de cultivos (Zhu et al., 2021) respalda su eficacia como estándar de la industria.

- **DINOv2 (Distilled Self-Supervised Learning with No Labels):** Esta elección se fundamenta en su capacidad para generar representaciones visuales de alta calidad mediante un pre-entrenamiento auto-supervisado. Según (Oquab et al., 2023), DINOv2 supera a los modelos convolucionales tradicionales en tareas de comprensión semántica gracias a sus mecanismos de auto-atención global propios de los Vision Transformers. Se seleccionó este modelo específicamente para evaluar si la captura de dependencias globales permite identificar patologías sutiles que podrían ser omitidas por modelos locales.

Esta dualidad metodológica busca contrastar la eficiencia computacional de las redes convolucionales frente a la riqueza semántica de los transformadores, basándose en la premisa de que modelos diversos ofrecen una mayor capacidad de generalización en entornos no controlados (Caron et al., 2021).

B. División del Dataset

El conjunto total de imágenes $n = 3,104$ se distribuyó en el conjunto de entrenamiento, validación y test, de manera estratificada. Esta división de los datos es fundamental para garantizar una evaluación confiable de los resultados, separando los datos en tres subconjuntos mutuamente excluyentes:

- **Conjunto de entrenamiento (80%):** Correspondiente a 2,514 imágenes, destinadas exclusivamente al ajuste de los pesos sinápticos y parámetros internos de las redes neuronales.
- **Conjunto de validación (10%):** Conformado por 280 imágenes, utilizadas para la sintonización de hiperparámetros y el monitoreo del sobreajuste (overfitting) durante las épocas de entrenamiento.
- **Conjunto de test (10%):** Reservado como un conjunto de prueba ciego ($n = 310$). Este bloque no intervino en ninguna etapa del desarrollo ni ajuste del modelo, actuando como una auditoría externa para evaluar la capacidad real de generalización ante patologías del cacao en condiciones de campo.

C. Configuración de hiperparámetros

La configuración de hiperparámetros fue esencial para el ajuste fino (fine-tuning) bajo el paradigma de Transfer Learning debido a los datos escasos siendo mejor empezar con una red que "sabe algo". Se aplicó una búsqueda controlada para maximizar la convergencia, cuyos valores finales se resumen en la Tabla 4.

- **Algoritmos de optimización (optimizers):** Se seleccionó Adam para YOLOv5 por su

eficiencia en gradientes adaptativos en redes convolucionales. Para DINOv2, se optó por AdamW, estándar en arquitecturas debido a su capacidad superior de regularización mediante el desacoplamiento de la caída de peso (weight decay).

- **Número de épocas (epochs):** Se utilizó 100 épocas, con un monitoreo continuo de las curvas de aprendizaje para asegurar la estabilidad de los modelos. Se observó que el punto de convergencia óptimo, definido por la estabilización de la pérdida de validación (Validation Loss) y el valor máximo alcanzado en la métrica F1-Score, se produjo en la época 60 para la arquitectura YOLOv5 y en la época 50 para DINOv2. En consecuencia, se procedió a la parada del proceso de entrenamiento para prevenir el sobreajuste y garantizar la capacidad de generalización de ambos modelos.
- **Tasa de aprendizaje (learning rate):** Se utilizaron tasas diferenciadas 1×10^{-3} y 1×10^{-4} , permitir un ajuste preciso de las capas superiores sin degradar los pesos pre-entrenados.
- **Tamaño de lote (batch size):** Se estableció un tamaño de lote de 16 para ambos modelos. Esta elección respondió a un equilibrio técnico entre la capacidad de memoria de la GPU (VRAM) y la necesidad de estimaciones de gradiente suficientemente estables para una convergencia suave durante el entrenamiento del dataset de cacao.

Tabla 4*Selección de hiperparámetros.*

Hiperparámetros	Modelo	
	YOLOv5	DINOV2
Épocas	60	50
Tasa de aprendizaje	1.00E-03	1.00E-04
Optimizador	Adam	AdamW
Tamaño de imagen	640x640x3	224x224x3
Weight Decay	0.0005	0.05
Pesos pre-entrenados	yolov5m.pt	dinov2_vitb14
Tamaño de lote	16	16

D. Entrenamiento del modelo

Durante el entrenamiento, se evaluó el comportamiento de las funciones de pérdida y las métricas de desempeño:

- **Comportamiento en YOLOv5:** Como se ilustra en la Figura 14, la función de pérdida (Training Loss) mostró una tendencia descendente y estable. La estabilización de la pérdida de validación confirmó que el modelo aprendió a generalizar sin incurrir en un sobreajuste significativo, alcanzando una sensibilidad superior a 0.90.

Figura 14

Train loss y sensibilidad según las épocas de entrenamiento del modelo YOLOv5.



- **Comportamiento en DINOv2:** La arquitectura basada en auto-atención demostró una rápida convergencia (Figura 15). La métrica F1-Score y la Especificidad (superior a 0.90) confirmaron la robustez del modelo para discriminar texturas patológicas complejas.

Figura 15

Train loss y sensibilidad según las épocas de entrenamiento del modelo DINOv2.



Para asegurar que las redes neuronales identificaran correctamente los patrones morfológicos del cacao, se inspeccionaron visualmente las instancias validadas durante el proceso de aprendizaje. En la Figura 16, se presentan las inferencias realizadas por el modelo YOLOv5, donde se observa una delimitación precisa de los frutos y una alta confianza en la clasificación de las clases healthy y sick. Por su parte, la Figura 17 muestra las instancias procesadas por DINOv2, evidenciando cómo el mecanismo de auto-atención logra segmentar las texturas patológicas incluso en condiciones de fondo complejo, validando así la efectividad

del preprocesamiento de la Región de Interés (RoI).

Figura 16

Instancias validadas de cacao en el modelo YOLOv5.



Figura 17

Instancias validadas de frutos de cacao en el modelo DINOv2.



Asimismo, el análisis del uso de recursos de hardware fue determinante para evaluar la viabilidad técnica de los modelos. En la Figura 18, se detalla el consumo de memoria GPU y la carga de procesamiento durante el entrenamiento de YOLOv5, mostrando un uso intensivo pero estable de los recursos, característico de las arquitecturas convolucionales medianas. De manera comparativa, la Figura 19 exhibe el perfil de recursos de DINOv2, donde se observa una demanda computacional significativa debido a la complejidad de las operaciones de Auto-Atención, aunque dentro de los parámetros de estabilidad térmica y operativa de la estación de trabajo. Este monitoreo integral garantiza que el rendimiento reportado no solo sea estadísticamente sólido, sino técnicamente sostenible.

Figura 18

Uso de recursos durante el entrenamiento de YOLOv5.

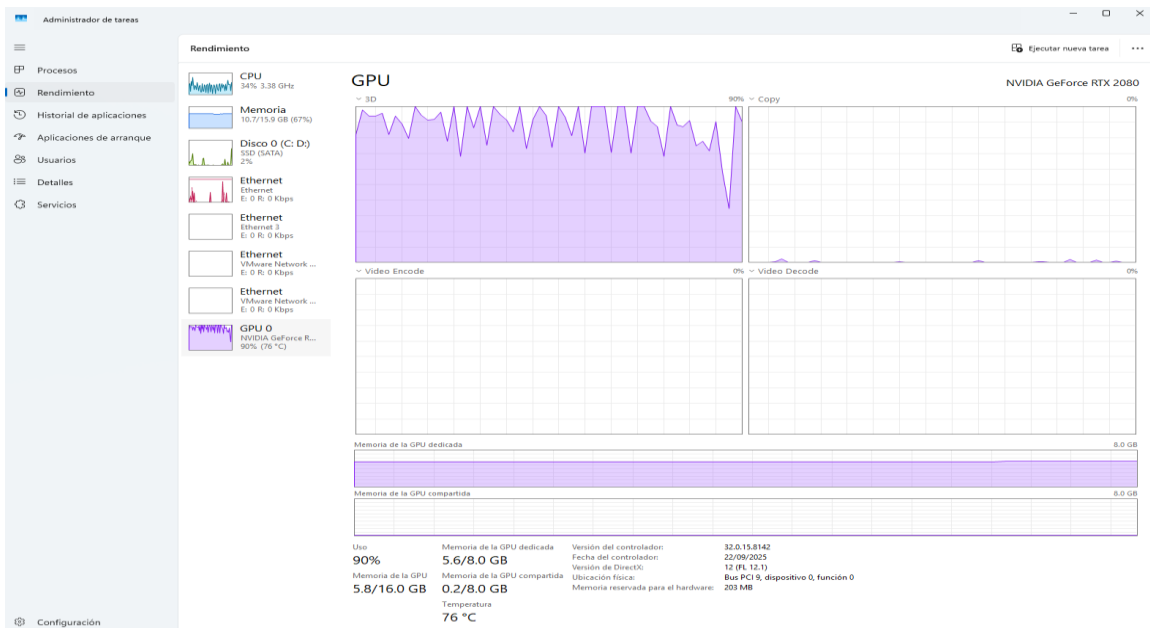
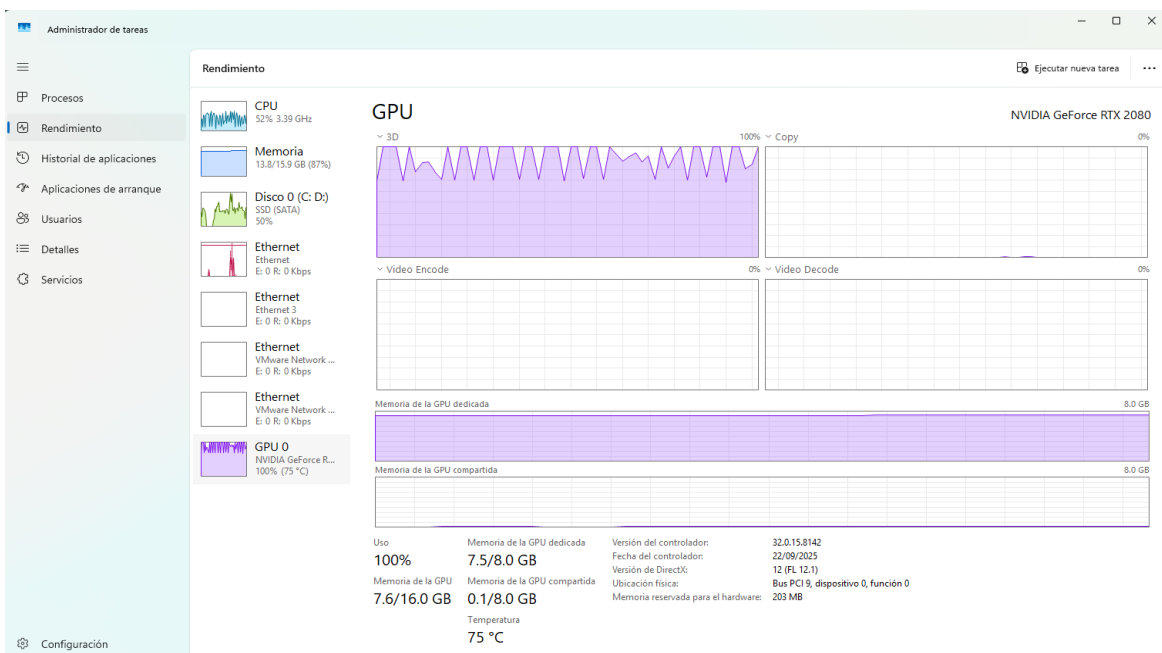


Figura 19

Uso de recursos durante el entrenamiento de DINOv2.



4.1.4 Proceso de evaluación y validación del modelo

Una vez entrenado el modelo, este está listo para ser evaluado y validado. Así, en esta parte utilizamos el conjunto de test compuesto por 310 imágenes independientes. Vale la pena recordar, que este conjunto de test no fue presentado al modelo durante la etapa de entrenamiento para garantizarse una evaluación limpia.

La evaluación y validación del modelo permite determinar la capacidad de generalización de los modelos ante escenarios reales que no intervinieron en el entrenamiento ni en la validación interna.

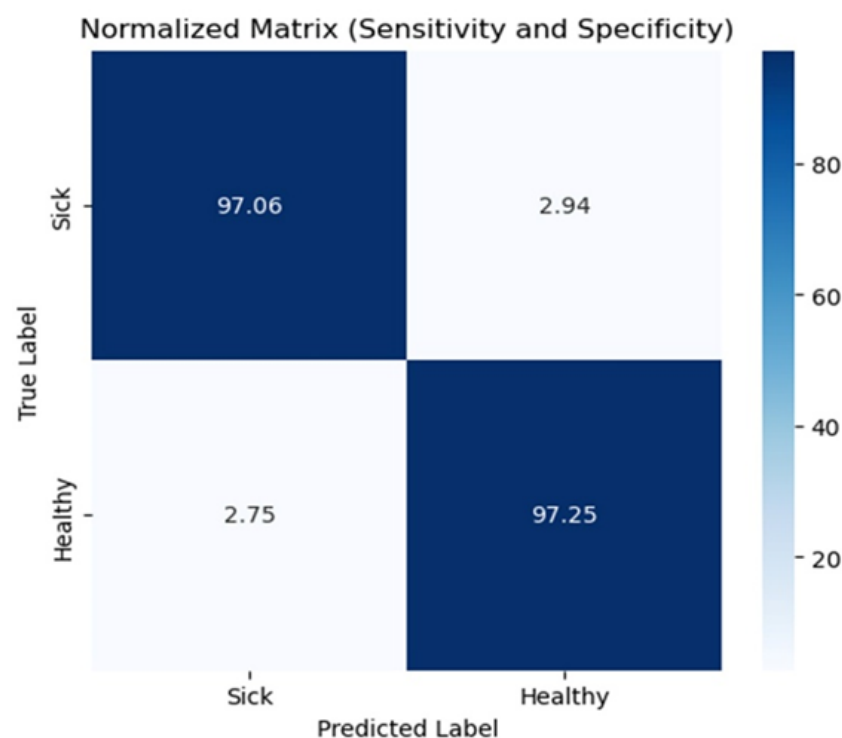
A. Análisis de rendimiento mediante matrices de confusión

El desempeño de cada modelo en la clasificación final se analizó a través de matrices de confusión normalizadas, las cuales permiten visualizar la capacidad de discriminación entre las clases 'Sick' (Enfermo) y 'Healthy' (Saludable).

- **Rendimiento de YOLOv5:** Como se observa en la Figura 20, alcanzó una Sensibilidad del 97.06%, lo que implica una tasa de falsos negativos de apenas 2.94%. Asimismo, reportó una Especificidad del 97.25%. Estos resultados confirman una sobresaliente precisión diagnóstica, fundamental para el reconocimiento temprano de patologías en campo.

Figura 20

Matriz de confusión YOLOv5.

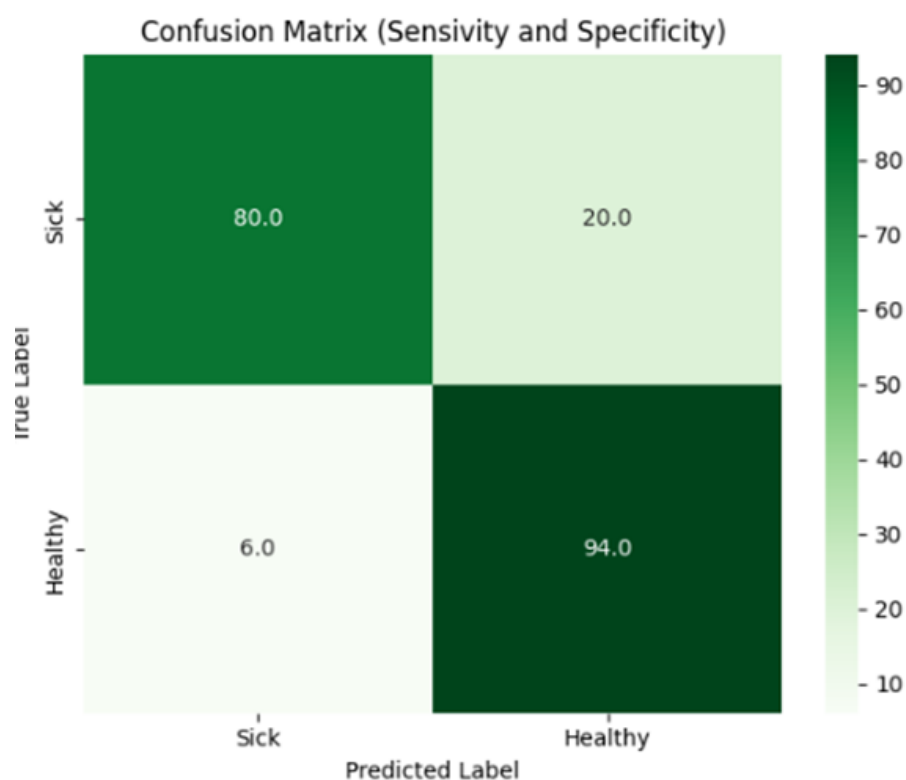


- **Rendimiento de DINOv2:** Según la Figura 21, aunque robusto, mostró una Sensibilidad del 80.00%, lo que conlleva una tasa de error del 20.0% en la detección

de frutos enfermos (falsos negativos). Su Especificidad fue del 94.00%, indicando un rendimiento inferior al paradigma convolucional en este dataset específico.

Figura 21

Matriz de confusión DINOv2.



B. Análisis de eficiencia y tiempos de inferencia de modelos

La eficiencia se evaluó en términos de consumo de recursos y velocidad de inferencia, factor crítico para la viabilidad de la solución en dispositivos móviles. Los resultados consolidados se presentan en las Tablas 5 y 6.

Tabla 5

Resultados del rendimiento generalizado de los modelos (eficacia).

Modelo	Sensibilidad	Especificidad	Precisión	Exactitud
YOLOv5	97.06	97.25	94.29	97.19
DINOv2	80.00	94.00	86.59	89.37

Tabla 6

Resultados de la eficiencia.

Modelo	Entrenamiento (minutos)	Inferencia (ms)	Uso de Hardware
YOLOv5	93 min	4.6 ms	Alto
DINOv2	52 min	14.83 ms	Alto

C. Análisis comparativo de modelos y significancia estadística

Para validar que la superioridad de YOLOv5 no es producto de fluctuaciones aleatorias, se aplicó la prueba de rangos con signo de Wilcoxon sobre el conjunto de prueba ($n = 310$). Los resultados proporcionaron evidencia cuantitativa contundente:

Valor $p < 0.001$, indicando que la diferencia de rendimiento es estadísticamente significativa.

Estadístico $W = 268.0$, confirmando que YOLOv5 mantuvo una certeza diagnóstica superior en la práctica totalidad de las muestras emparejadas.

Se interpreta que la brecha entre la confianza promedio de YOLOv5 (0.9088) y DINOv2 (0.6940) se argumenta desde la eficiencia de las capas convolucionales para extraer características locales (texturas y bordes de manchas de moniliasis), superando la capacidad de atención global en este dominio.

4.2 DISCUSIÓN

La evaluación de las arquitecturas YOLOv5 y DINOv2 en el conjunto de prueba evidenció que, aunque ambos modelos son competentes en la clasificación de frutos de cacao, YOLOv5 alcanzó un desempeño superior. La matriz de confusión muestra una alta sensibilidad del 97.06% y una especificidad del 97.25%, superando ampliamente a DINOv2 (Sensibilidad del 80.0% y Especificidad del 94%). Esta superioridad descriptiva fue ratificada mediante la prueba de los rangos con signo de Wilcoxon, la cual arrojó un $p < 0.001$ y un estadístico W de 268.0. Estos valores confirman que la diferencia entre la confianza promedio de YOLOv5 (0.9088) y DINOv2 (0.6940) es estadísticamente significativa y consistente a través de las 310 muestras emparejadas del test set. Esta diferencia indica que YOLOv5 logra un reconocimiento más robusto, con una tasa de falsos negativos reducida (2.94%), lo cual es crítico en el diagnóstico de enfermedades agrícolas, donde los errores de omisión pueden facilitar la propagación del patógeno.

El mejor desempeño de YOLOv5 puede atribuirse a varias razones técnicas. En primer lugar, YOLOv5 utiliza una arquitectura convolucional optimizada con bloques CSP (Cross Stage Partial connections) y una estructura tipo single-shot detector, permitiendo un balance efectivo entre la exactitud y la velocidad de procesamiento, lo que permite un equilibrio eficiente entre velocidad y precisión de inferencia (Jocher et al., 2020). Esta eficiencia resulta ventajosa en contextos agrícolas donde se requiere procesamiento en tiempo real o en dispositivos de pocos recursos. La alta confianza predictiva observada (superior al 90%) sugiere que el modelo ha logrado una extracción de características más discriminante para las texturas de las enfermedades en comparación con su contraparte.

Por otro lado, DINOv2, basado en mecanismos de autoatención (Transformers), está diseñado para aprender representaciones visuales generalistas mediante auto-supervisión (Oquab et al., 2023). Sin embargo, este tipo de modelos requiere grandes volúmenes de datos y alta capacidad de cómputo para alcanzar su máximo potencial. En conjuntos de datos relativamente pequeños o con variabilidad limitada, como los de frutos de cacao, los modelos convolucionales pre-entrenados tienden a adaptarse mejor, ya que su sesgo inductivo favorece la extracción de patrones locales relevantes (Liu et al., 2022). La menor confianza estadística registrada por DINOv2 refleja una mayor incertidumbre en la clasificación de bordes de lesiones necróticas, donde los mecanismos de atención global pueden ser menos efectivos que los filtros locales convolucionales ante datos de grano fino.

En comparación con otros detectores de una sola etapa, como SSD o RetinaNet, YOLOv5 presenta ventajas significativas. Su diseño modular, la utilización del mosaic data

augmentation y su mecanismo de auto learning bounding box anchors permiten mejorar la generalización y la precisión en objetos de diferentes tamaños (Jocher et al., 2020; Redmon et al., 2016). Estas características explican por qué YOLOv5 ha sido ampliamente adoptado en aplicaciones agrícolas recientes para el reconocimiento de plagas y enfermedades en cultivos (Zhao et al., 2023).

En síntesis, los resultados demuestran que YOLOv5 constituye una alternativa más efectiva que DINOv2 para el diagnóstico visual de enfermedades en frutos de cacao, especialmente en escenarios donde los datos son limitados y se requiere una detección precisa en tiempo real. La validación mediante la prueba de Wilcoxon permite concluir con rigor científico que la arquitectura basada en convoluciones sigue siendo una estrategia más competitiva y confiable frente a modelos de autoatención en dominios específicos y con restricciones de datos como el cacao peruano.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

La presente investigación evaluó y comparó satisfactoriamente el rendimiento de las arquitecturas YOLOv5 y DINOv2 para el reconocimiento de enfermedades en frutos de cacao en Omayta, Cusco. Se concluye que la implementación de modelos de Deep Learning posee una eficacia significativa para esta tarea, estableciendo a la arquitectura YOLOv5 como la solución óptima. Mediante la prueba estadística de Wilcoxon, se validó que la superioridad de este modelo no es producto del azar, obteniendo un valor p ($p < 0.001$) que confirma una diferencia de rendimiento trascendental, alcanzando una Exactitud del 97.19% en el análisis del estado fitosanitario de los frutos en su entorno natural.

Este rendimiento superior de YOLOv5 se fundamenta en la capacidad de sus capas convolucionales para la extracción jerárquica de características locales y patrones espaciales. Este mecanismo resultó ser más efectivo para capturar y discriminar las texturas y morfologías sutiles de las patologías del cacao en comparación con el enfoque de contexto global de DINOv2. La robustez de YOLOv5 quedó demostrada con una Sensibilidad del 97.06% y una Especificidad del 97.25%, logrando una tasa mínima de falsos negativos (2.94%). Estos niveles de precisión son vitales para el sector agrícola, ya que garantizan la detección de casi la totalidad de los frutos enfermos, permitiendo un control temprano de la propagación de plagas.

En términos de eficiencia operativa, YOLOv5 se consolidó como la red más viable para su implementación práctica. A pesar de un proceso de entrenamiento exhaustivo, el modelo registró un tiempo de inferencia de solo 4.6 ms por imagen. Esta velocidad de procesamiento es fundamental para asegurar el diagnóstico en tiempo real mediante dispositivos móviles en el campo, superando las restricciones de latencia de modelos más pesados.

Finalmente, este trabajo establece las bases para el desarrollo de algoritmos de clasificación multiclase que permitan diferenciar con mayor granularidad enfermedades específicas como la Monilia o la Phytophthora. Este avance representa el siguiente paso lógico para maximizar la utilidad del sistema en el sector agrícola peruano y aumentar la precisión diagnóstica en entornos no controlados.

5.2 RECOMENDACIONES

Se recomienda el desarrollo e implementación de una aplicación móvil orientada al usuario final, con el objetivo de evaluar el desempeño de los modelos en condiciones de operación real y obtener retroalimentación directa de los agricultores. Esta etapa es fundamental para validar la usabilidad de la herramienta y ajustar los parámetros de interfaz según las necesidades específicas del sector agrícola en campo.

Para garantizar la viabilidad técnica de esta solución, se propone el despliegue de una infraestructura que soporte el diagnóstico mediante capacidades de Edge Computing. Este enfoque permitiría el procesamiento inmediato de las imágenes en el dispositivo móvil, minimizando la dependencia de la conectividad a internet en zonas rurales. Asimismo, se sugiere establecer un sistema de aprendizaje continuo (Active Learning); al sincronizar nuevas imágenes capturadas con un servidor central, se facilitaría un ciclo de mejora constante que fortalecería la robustez de los modelos ante la aparición de nuevas variantes de patógenos.

Como complemento a la solución móvil y con el fin de centralizar la gestión de la información, se recomienda la integración de estos modelos en plataformas web interactivas basadas en la nube. Este ecosistema multiplataforma permitirá a los productores, técnicos e investigadores cargar de manera remota no solo imágenes, sino también secuencias de video para un diagnóstico fitopatológico masivo y de baja latencia. Además, actuará como un repositorio centralizado de telemetría y mapeo de alertas fitosanitarias, facilitando el monitoreo epidemiológico a gran escala en las parcelas del VRAEM.

Finalmente, con el propósito de transitar de una clasificación binaria hacia una clasificación multiclase, es imperativo ampliar el volumen del conjunto de datos y explorar técnicas de aumento de datos sintéticos. Esta estrategia permitiría balancear artificialmente las clases de enfermedades menos frecuentes y mejorar la capacidad de generalización de los modelos ante la variabilidad fenológica del cacao en las diversas regiones productoras del Perú, asegurando una herramienta diagnóstica de alta precisión y cobertura nacional.

BIBLIOGRAFÍA

- Agrios, G. N. (2005). *Plant pathology*. Elsevier.
- Arm Ltd. (2025). *What is a Central Processing Unit?* [https://www.arm.com/glossary/cpu#:~:text=Control%20Processing%20Unit%20\(CPU\),based%20on%20those%20apps'%20requirements](https://www.arm.com/glossary/cpu#:~:text=Control%20Processing%20Unit%20(CPU),based%20on%20those%20apps'%20requirements).
- Basri, B., Indrabayu, Achmad, A., & Sari, I. (2023). A Review of Image Processing Techniques for Pest and Disease Early Identification Systems on Modern Cocoa Plantation. *International Journal of Computing and Digital Systems*, 14(1), 10277–10286. <https://doi.org/10.12785/ijcds/1401100>
- Cabezas, O. E., Gil, J. L., Gómez, R., Dávila, C., Morón, S., & Ramírez, C. (2017). *Estado fitosanitario en la producción de cacao (Theobroma cacao L.) en la región de Huánuco (Perú): Incremento del impacto de Carmenta foraseminis Eichlin*.
- Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). *Emerging Properties in Self-Supervised Vision Transformers*. <https://arxiv.org/abs/2104.14294>
- Cochran, W. G. (1977). *Sampling Techniques*. Wiley.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin.
- CropLife Latin America. (2024a). *La moniliasis del cacao: características, colores, causas y fungicidas para su control - CropLife Latin America*. <https://croplifela.org/es/plagas/listado-de-plagas/moniliasis-del-cacao>
- CropLife Latin America. (2024b). *Mazorca Negra del Cacao-pudricion parda o negra: Causas y control | mancha negra phytophthora sp - CropLife Latin America*. <https://croplifela.org/es/plagas/listado-de-plagas/mazorca-negra-del-cacao-pudricion-parda-o-negra-mancha-negra-phytophthora-sp>
- Devida. (2021). *VRAEM: Cerca de tres mil productores cacaoteros son asistidos en manejo integrado de plagas en 10 distritos*. <https://www.gob.pe/institucion/devida/noticias/482852-vraem-cerca-de-tres-mil-productores-cacaoteros-son-asistidos-en-manejo-integ>
- Devida. (2024). *EcoCacao VRAEM 2024 generó proyección de venta de S/13.9 millones durante rueda de negocios*. <https://www.gob.pe/institucion/devida/noticias/965651-ecocacao-vraem-2024-genero-proyeccion-de-venta-de-s-13-9-millones-durante-ru>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, S., Heigold, G., Gelly, S., Hounsby, N., & Simon, T. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *International Conference on Learning Representations (ICLR)*.
- Eric, O., Gyening, R., Appiah, O., Takyi, K., & Appiahene, P. (2023). Cocoa beans classification using enhanced image feature extraction techniques and a regularized Artificial Neural Network model. *Engineering Applications of Artificial Intelligence*, 125, 106736. <https://doi.org/10.1016/j.engappai.2023.106736>
- Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2009). The Pascal Visual Object Classes (VOC) challenge. *International Journal of Computer Vision*, 88(2), 303–338. <https://doi.org/10.1007/s11263-009-0275-4>
- Godmalin, R. A., Aliac, C. J., & Feliscuzo, L. (2022). Classification of Cacao Pod if Healthy or Attack by Pest or Black Pod Disease Using Deep Learning Algorithm. *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 1–5. <https://doi.org/10.1109/IICAIET55139.2022.9936817>

- Godmalin, R. A., Aliac, C. J., & Feliscuzo, L. (2023). Cacao Pod Infection Level Classification Using Transfer Learning. *2023 IEEE Open Conference of Electrical, Electronic and Information Sciences (EStream)*, 1–6. <https://doi.org/10.1109/eStream59056.2023.10135062>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. <https://books.google.com.pe/books?hl=es&lr=&id=omivDQAAQBAJ>
- Gyamfi, A., Iddrisu, S. A.-J., & Adegbola, O. (2020). Machine Learning-Based Cocoa E-Health System. *2020 13th CMI Conference on Cybersecurity and Privacy (CMI) - Digital Transformation - Potentials and Challenges (51275)*, 1–7. <https://doi.org/10.1109/CMI51275.2020.9322689>
- Hennessy, J. L., & Patterson, D. A. (2019). *Computer architecture: A quantitative approach* (6th ed.). Morgan Kaufmann.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (6th ed.). McGraw-Hill.
- Huang, G., Zhou, Y., Hu, X., Zhang, C., Zhao, L., Gan, W., & Hou, M. (2023). DINO-mix: Enhancing visual place recognition with foundational vision model and feature mixing. *ArXiv Preprint ArXiv:2311.00230*. <https://arxiv.org/abs/2311.00230>
- Jocher, G., Stoken, A., Borovec, J., Chaurasia, A., & Laughlin, M. (2020). *YOLOv5: Open-source neural network for object detection*. GitHub. <https://github.com/ultralytics/yolov5>
- Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization* (Number 1412.6980). <http://arxiv.org/abs/1412.6980>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105. https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- Kumi, S., Kelly, D., Woodstuff, J., Lomotey, R. K., Orji, R., & Deters, R. (2022). *Cocoa Companion: Deep Learning-Based Smartphone Application for Cocoa Disease Detection* (Vol. 203). Elsevier. <https://doi.org/10.1016/j.procs.2022.07.013>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Leon, G., & Yauri, J. (2025). *Dataset of Peruvian Cocoa Fruits for Automatic Disease Detection*. IEEE Dataport. <https://doi.org/10.21227/ktfb-2132>
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2022). Swin Transformer V2: Scaling up capacity and resolution. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12009–12019. <https://doi.org/10.1109/CVPR52688.2022.01169>
- Madupuri, R., Vemula, D. R., Chettupally, A. C., Sangi, A. R., & Ravi, P. (2023). Deep learning image-based automated application on classification of tomato leaf disease by pre-trained deep convolutional neural networks. *Mehran University Research Journal of Engineering and Technology*, 42(3), 52. <https://doi.org/10.22581/muet1982.2303.06>
- Menard, S. (2002). *Longitudinal research*. Sage.
- Motamayor, J. C., Lachenaud, P., da Silva e Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PloS One*, 3(10), e3311. <https://doi.org/10.1371/journal.pone.0003311>
- NVIDIA Corporation. (2024). *Tecnología de computación por GPU para inteligencia artificial*. <https://www.nvidia.com/es-la/>

- Oquab, M., Darcet, T., Moutakanni, M., Massa, F., Simon, M., Helou, P., Labatut, T., Mairal, J., Malisiewicz, T., Jouret, A., Ballas, L., Lenczner, P., Rame, A., Coucke, A., Jégou, H., Mordan, Y., Tissier, R., Szafraniec, L., Rabbat, M., & Roudière, T. (2023). *DINOv2: Learning Robust Visual Features without Supervision* (Number 2304.07193). <https://arxiv.org/abs/2304.07193>
- Powers, D. M. W. (2020). Evaluation: from precision, recall and F-measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 1(2), 29–41.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Rivera, C., & Sneyder, A. (2020). *Epidemiología de Moniliophthora perniciosa (Stahel) Aime y Phillips-Mora en seis clones de cacao de Santander, Colombia*. Nombre de la Institución o Universidad.
- Siegel, S., & Castellan Jr., N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences* (2nd ed.). McGraw-Hill.
- Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12), 2295–2329.
- Terven, J. R., Córdova-Esparza, D.-M., & Romero-González, J. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5, 1680–1716. <https://doi.org/10.3390/make5040083>
- Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Ultralytics. (2020). *ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements (v3.1)*. Zenodo. <https://doi.org/10.5281/zenodo.4154370>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30.
- Yanac Montesino, R., Rosales-Huamani, J. A., & Castillo-Sequera, J. L. (2021). Detection of Phytophthora palmivora in cocoa fruit with deep learning. *Iberian Conference on Information Systems and Technologies, CISTI*. <https://doi.org/10.23919/CISTI52073.2021.9476279>
- Yatoo, A. A., & Sharma, A. (2021). A Novel Model for Automatic Crop Disease Detection. *2021 Sixth International Conference on Image Information Processing (ICIIP)*, 6, 310–313. <https://doi.org/10.1109/ICIIP53038.2021.9702553>
- Zhao, Y., Li, J., Wang, Y., & Xu, H. (2023). Application of YOLOv5 for agricultural pest and disease detection: A comprehensive review. *Computers and Electronics in Agriculture*, 210, 107956. <https://doi.org/10.1016/j.compag.2023.107956>
- Zhu, X., Lyu, S., Wang, X., & Zhao, Q. (2021). *TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios*. <https://arxiv.org/abs/2108.11539>

Anexos

Anexo 1. Matriz de consistencia

Título: “Rendimiento de Redes Convolucionales de Auto-atención en el Reconocimiento de Enfermedades de Cacao, Omayá-Cusco, 2025”

PROBLEMA	OBJETIVO	HIPÓTESIS	METODOLOGÍA
PROBLEMA PRINCIPAL ¿Qué tipo de red neuronal presenta el mejor rendimiento para el reconocimiento de enfermedades de frutos de cacao en imágenes, Omayá-Cusco, 2025? PROBLEMAS SECUNDARIOS a. ¿Cuál es el paradigma de computación de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes? b. ¿Cómo es el funcionamiento de las capas base de aprendizaje de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes? c. ¿Cuál es la eficacia de la red que presenta el mejor rendimiento?	OBJETIVO GENERAL Evaluar y comparar el rendimiento de las redes neuronales convolucionales y de auto-atención para el reconocimiento de enfermedades de frutos de cacao en imágenes, Omayá-Cusco, 2025. OBJETIVOS ESPECÍFICOS a. Determinar el paradigma de computación de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes b. Describir el funcionamiento de las capas base de aprendizaje que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes c. Determinar la eficacia de la red que presenta el mejor rendimiento	HIPÓTESIS GENERAL No se tiene ninguna hipótesis por ser un estudio de nivel de investigación descriptivo. VARIABLES E INDICADORES VARIABLE DE ESTUDIO Rendimiento de la red neuronal Variables descriptivas a) Paradigma de computación de la red Indicadores: - Basado en convolución (YOLOv5) - Basado en auto-atención (DINOv2) b) Capas base de aprendizaje de la red Indicadores - Bloques convolucionales - Bloques de auto-atención c) Eficacia de la red Indicadores: - Sensibilidad	1. TIPO DE INVESTIGACIÓN Observacional, prospectivo, longitudinal y descriptivo. 2. NIVEL DE INVESTIGACIÓN Descriptiva 3. DISEÑO No experimental, prospectivo y Longitudinal. 4. POBLACIÓN En este caso, la población de la investigación incluye todos los frutos de cacao de 25 parcelas en el Centro Poblado de Omayá afectados o no por enfermedades durante el año 2024, que son de 10150 frutos. 5. MUESTRA Se recopilieron 3104 imágenes de frutos de cacao en diferentes momentos del año 2024, seleccionadas por conveniencia de 25 parcelas específicas donde

miento para reconocer enfermedades de frutos de cacao en imágenes? d. ¿Cuál es la eficiencia de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes?	dimiento para reconocer enfermedades de frutos de cacao en imágenes d. Determinar la eficiencia de la red que presenta el mejor rendimiento para reconocer enfermedades de frutos de cacao en imágenes	• Especificidad d) Eficiencia de la red Indicadores: • Tiempo de entrenamiento • Tiempo de inferencia • Uso de CPU/GPU/RAM	la incidencia de enfermedades es alta. 6. TÉCNICAS Análisis documental 7. INSTRUMENTOS Análisis de contenido Registro fotográfico
---	---	--	---

Anexo 2.

Publicaciones

Artículos de conferencia

- Leon, G., & Yauri, J. (2025). A benchmark dataset and baseline model for disease detection in Peruvian cocoa fruits using YOLOv5. En Proceedings of the 2025 8th International Conference on Systems Engineering - Cybersecurity & AI: Building a Reliable Digital Future (pp. 23–30). Association for Computing Machinery. <https://doi.org/10.1145/3771678.3771682>.

Dataset libre

- Geovanni Leon, Jose Yauri (2025). Dataset of Peruvian Cocoa Fruits for Automatic Disease Detection. IEEE Dataport. <https://dx.doi.org/10.21227/ktfb-2132>.

**UNSCH****FACULTAD DE
INGENIERÍA**
DE MINAS, GEOLOGÍA Y CIVIL**ACTA DE SUSTENTACIÓN DE TESIS N° 37-2026-FIMGC****PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS**

En la Universidad Nacional de San Cristóbal de Huamanga, en la ciudad de Ayacucho, en cumplimiento a la **Resolución Decanal No 169-2026-FIMGC-D**, a los **diecisiete días del mes de junio de 2026**, siendo las **10:13 a.m.**, reunidos en el **Auditorio de la Escuela Profesional de Ingeniería Civil**, bajo la presidencia del **MSc. José Ernesto ESTRADA CARDENAS**, y los miembros: **Mg. Ing^a. Edith Felicitas GUEVARA MOROTE**, **Dr. José Elías YAURI VIDALON** y **Dr. Hubner JANAMPA PATILLA**, actuando como secretario docente el **Ing. Saul Walter RETAMOZO FERNANDEZ**, para proceder a la sustentación de tesis para optar el **Título Profesional de Ingeniero de Sistemas**, del Bachiller en Ingeniería de Sistemas:

GEOVANNI RAFAEL LEON PEREZ

Quien presentó la tesis denominada:

" Rendimiento de Redes Convolucionales y de Auto-Atención en el Reconocimiento de Enfermedades de Cacao, Omayá-Cusco, 2025"

Los señores miembros del jurado luego de expuesta la tesis y absueltas las preguntas, deliberaron y declararon:

Aprobado con 17 (diecisiete)

Siendo las **11:27 a.m.** del día **17 de junio del 2026**, culmina el acto de sustentación de tesis, y en conformidad de lo actuado los miembros del jurado firmamos al pie del presente.

MSc. José Ernesto ESTRADA CARDENAS
Presidente

Mg. Ing^a. Edith Felicitas GUEVARA MOROTE
Miembro

Dr. Hubner JANAMPA PATILLA
Miembro

Dr. José Elías YAURI VIDALON
Miembro - Asesor

Ing. Saul Walter RETAMOZO FERNANDEZ
Secretario docente de la FIMGC

FACULTAD DE INGENIERÍA
DE MINAS Y CIVIL
Av. Independencia S/N
Ciudad Universitaria
Central Tel. 066 312510
Anexo 151



CONSTANCIA DE ORIGINALIDAD DE TRABAJO DE INVESTIGACIÓN

CONSTANCIA N° 043-2026-KPS-FIMGC/UNSCH

El que suscribe; responsable verificador de originalidad de trabajos de tesis de pregrado con el software Turnitin, en segunda instancia para las **Escuelas Profesionales** de la **Facultad de Ingeniería de Minas, Geología y Civil**; en cumplimiento a la **Resolución de Consejo Universitario N° 039-2021-UNSCH-CU**, Reglamento de Originalidad de Trabajos de Investigación de la Universidad Nacional San Cristóbal de Huamanga y **Resolución Decanal N° 697-2024-FIMGC-D**, deja constancia de originalidad de trabajo de investigación, que el/la Sr./Srta.

Nombres y Apellidos : Geovanni Rafael Leon Perez
Escuela Profesional : INGENIERÍA DE SISTEMAS
Título de la Tesis : "Rendimiento de Redes Convolucionales y de Auto-atención en el Reconocimiento de Enfermedades de Cacao, Omayta-Cusco, 2025"
Evaluación de la Originalidad : 4% Índice de Similitud
Identificador de la entrega : 3025542707

Por tanto, según los Artículos 12, 13 y 17 del Reglamento de Originalidad de Trabajos de Investigación, es **PROCEDENTE** otorgar la **Constancia de Originalidad** para los fines que crea conveniente.

En señal de conformidad y verificación se firma la presente constancia

Ayacucho, 24 de setiembre de 2026



Firmado digitalmente por:
PERALTA SOTOMAYOR Karel
FAU 20143660754 soft
Motivo: Soy el autor del documento
Fecha: 25/09/2026 21:38:42-0500

“Rendimiento de Redes Convolucionales y de Auto-atención en el Reconocimiento de Enfermedades de Cacao, Omayá-Cusco, 2025”

por GEOVANNI RAFAEL LEON PEREZ

Fecha de entrega: 06-sept-2026 08:15p. m. (UTC-0500)

Identificador de la entrega: 3025542707

Nombre del archivo: _Rendimiento_de_Redес_Convolucionales_y_de_Auto-atención_en_el_Reconocimiento_de_Enfermedades_de_Cacao_Omayá-Cusco_2025_.pdf (3.57M)

Total de palabras: 16258

Total de caracteres: 97490

“Rendimiento de Redes Convolucionales y de Auto-atención en el Reconocimiento de Enfermedades de Cacao, Omayá-Cusco, 2025”

INFORME DE ORIGINALIDAD

4%	4%	1%	2%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	Submitted to Universidad Nacional de San Cristóbal de Huamanga	1%
	Trabajo del estudiante	
2	repositorio.unsch.edu.pe	1%
	Fuente de Internet	
3	repositorio.uncp.edu.pe	<1%
	Fuente de Internet	
4	dspace.ups.edu.ec	<1%
	Fuente de Internet	
5	repository.uniagraria.edu.co	<1%
	Fuente de Internet	
6	www.campustudionline.com	<1%
	Fuente de Internet	
7	www.gob.mx	<1%
	Fuente de Internet	
8	worldwidescience.org	<1%
	Fuente de Internet	