Texto

Descripción generada automáticamente

Universidad Internacional de La Rioja

Escuela Superior de Ingeniería y

Tecnología

Máster Universitario en Inteligencia artificial

Categorizar imágenes de plagas en cultivos de cacao y piña utilizando métodos de aprendizaje profundo supervisado

|  |  |
| --- | --- |
| Trabajo fin de estudio presentado por: | Bryan Ricardo Armas Loyaga |
| Tipo de trabajo: | Comparativa de soluciones |
| Director/a: | Alvaro Alexander Martinez Navarro |
| Fecha: | 21-04-2025 |

Resumen

En este apartado se introducirá un breve resumen en español del trabajo realizado (extensión entre 150 y 300 palabras). Este resumen debe incluir el objetivo o propósito de la investigación, la metodología, los resultados y las conclusiones.

El resumen debe contener lo qué se ha pretendido realizar (objetivo o propósito de la investigación), cómo se ha realizado (método o proceso desarrollado) y para qué se ha realizado (resultados y conclusiones).

**Importante:** La extensión mínima en un TFE individual es de 50 páginas, sin contar portada, resumen, *abstract*, índices y anexos.

**Palabras clave:** (De 3 a 5 palabras) Descriptores del trabajo que lo enmarcan en unas temáticas determinadas. Serán los utilizados para localizar tu trabajo si llega a ser publicado.

Abstract

En este apartado se introducirá un breve resumen en **inglés** del trabajo realizado (extensión entre 150 y 300 palabras). Este resumen debe incluir el objetivo o propósito de la investigación, la metodología, los resultados y las conclusiones.

**Keywords**: (De 3 a 5 palabras en inglés)

Índice de contenidos

[1. Introducción 1](#_Toc202826334)

[1.1. Motivación 2](#_Toc202826335)

[1.2. Planteamiento del trabajo 3](#_Toc202826336)

[1.3. Estructura del trabajo 4](#_Toc202826337)

[2. Contexto y estado del arte 6](#_Toc202826338)

[2.1. Contexto del problema 6](#_Toc202826339)

[2.2. Estado del arte 7](#_Toc202826340)

[2.3. Conclusiones 11](#_Toc202826341)

[3. Objetivos concretos y metodología de trabajo 12](#_Toc202826342)

[3.1. Objetivo general 12](#_Toc202826343)

[3.2. Objetivos específicos 12](#_Toc202826344)

[3.3. Metodología del trabajo 12](#_Toc202826345)

[4. Planteamiento de la comparativa 14](#_Toc202826346)

[4.1. Identificación y Contextualización del Problema 14](#_Toc202826347)

[4.2. Soluciones Alternativas a Evaluar: Modelos de Deep Learning Supervisado 14](#_Toc202826348)

[4.3. Criterios de Éxito para la Comparativa 15](#_Toc202826349)

[4.4. Medidas y Métricas de Evaluación 16](#_Toc202826350)

[5. Desarrollo de la comparativa 17](#_Toc202826351)

[6. Discusión y análisis de resultados 17](#_Toc202826352)

[7. Conclusiones y trabajos futuros 17](#_Toc202826353)

Índice de figuras

[Figura 1. *Ejemplo de figura realizada para nuestro trabajo.* 2](#_Toc155946760)

Índice de tablas

[Tabla 1. *Ejemplo de tabla con sus principales elementos.* 2](#_Toc158282432)

# Introducción

El incremento acelerado en la generación de datos visuales incrustados en imágenes, impulsado por la propagación de dispositivos de captura de imágenes y la expansión de plataformas digitales, ha permitido posicionar la categorización de las imágenes como una tarea de importancia crítica en el panorama tecnológico actual. Esta capacidad de analizar y comprender automáticamente el contenido visual dentro las imágenes se han convertido en un componente esencial para la eficiencia operativa, la toma de decisiones estratégicas y la innovación en diversas industrias, que van desde la medicina hasta la agricultura, pasando por la seguridad, banca, finanzas y el comercio electrónico.

Con el aprendizaje supervisado profundo, al ser una rama avanzada del aprendizaje automático, se ha considerado como la metodología dominante para abordar los desafíos inherentes a la categorización de imágenes. Esto ha permitido aprovechar las bondades de los grandes conjuntos de datos etiquetados, los modelos de aprendizaje profundo supervisado, especialmente las Redes Neuronales Convolucionales (CNNs), han demostrado una capacidad sin precedentes para extraer características complejas y establecer patrones complejos dentro de los datos visuales. Esta habilidad permite, en muchos casos, superar la precisión del ojo humano en tareas específicas de clasificación, abriendo nuevas fronteras en la automatización y la inteligencia artificial.

En esta investigación se ahonda en el núcleo de la categorización de imágenes mediante el uso del aprendizaje profundo supervisado. A través de una exploración exhaustiva, se desglosa los conceptos fundamentales que sustentan este campo, se analiza las arquitecturas de CNNs más influyentes que han marcado su evolución, se examina los conjuntos de datos de referencia que sirven como pilares dentro de la investigación y la evaluación, permite la discusión de las métricas esenciales para cuantificar el rendimiento de los modelos. Además, se incluye el tema de los desafíos persistentes, las limitaciones inherentes a esta disciplina, destacando las tendencias emergentes y los avances recientes los cuales prometen dar forma al futuro en relación con la categorización de imágenes. Finalmente, se establecerá los pasos esenciales utilizados en el desarrollo de un sistema de categorización de imágenes los cuales se basan en aprendizaje supervisado profundo.

## Motivación

En el territorio ecuatoriano varias empresas dedicadas a la producción de cacao y piña se han visto afectados por pérdidas económicas y del producto, dando como resultado que el causante de esta perdida son las plagas dentro de los cultivos.

Una de las problemáticas detectadas según (Instituto Nacional Autónomo de Investigaciones Agropecuarias, n.d.) se debe a diversos problemas ocasionados por insectos, ácaros, malezas, hongos, bacterias, virus, nematodos entre otros y al no disponer de información exacta sobre sus causas no pueden tomar medidas apropiadas, con perjuicio al rendimiento, incrementado los costos y ocasionando deterioro de la calidad de sus cosechas.

A su vez (Pérez-Martínez et al., 2017) en su documento menciona que los síntomas de marchitamiento o clorosis que generalmente aparecen en los viveros de cacao se deben principalmente al ataque de hongos, los cuales llegan por vías de transmisión horizontal.

Por su parte (Sanchez Alvarado et al., 2023) indica que Diversas plagas se hacen resistentes a plaguicidas, fungicidas y herbicidas; se emplean agroquímicos porque se cree que no es viable producir arroz sin un alto nivel de sustancias químicas para controlar las plagas. Los plaguicidas causan y suscitan problemas ecológicos y de salud para humanos y animales.

Dentro de la investigación previa realizada por mi parte para encontrar trabajos con características similares dentro de Ecuador, si bien se encuentra trabajos que utilizan detección con procesamiento de imágenes multiespectrales, manejo de plagas con redes de sensores inalámbricos o análisis de plagas en rosas con redes neuronales convolucionales, sin embargo, no existen proyectos que se enfocan en la clasificación de imágenes para detección de plagas con aprendizaje profundo supervisado.

La utilización de imágenes dentro de los cultivos de cacao y piña en el día a día, la creciente difusión y la importancia estratégica de la información visual ha permitido centrar el estudio y optar por este tema de titulación. La proliferación de cámaras en dispositivos móviles, sistemas de vigilancia dentro de las plantaciones, genera un flujo constante de imágenes que contienen información valiosa, susceptible de ser analizada para obtener conocimientos profundos y facilitar la automatización de tareas.

En el ámbito de la agricultura, la categorización de imágenes permite la detección temprana de anomalías en cultivos, la clasificación en relación de la calidad de las piñas y el cacao, de los procesos de cosecha, lo que se traduce en una mayor eficiencia y una reducción de costos.

El aprendizaje profundo supervisado ha demostrado ser una herramienta excepcionalmente poderosa para abordar estos desafíos en diversos dominios. Su capacidad para aprender representaciones jerárquicas de características directamente de los datos de la imagen, sin la necesidad de una ingeniería de características manual exhaustiva, lo convierte en un enfoque atractivo y efectivo (LeCun et al., 2015). Las CNNs, en particular, han logrado avances significativos en la precisión de la categorización de imágenes a gran escala, como lo demuestran sus resultados en competiciones de referencia como ImageNet (Khan et al., 2022).

Sin embargo, a pesar de los éxitos alcanzados, la categorización de imágenes con aprendizaje profundo supervisado aún presenta desafíos importantes. La necesidad de grandes cantidades de datos etiquetados para entrenar modelos robustos, el problema del sobreajuste, la interpretabilidad limitada de las redes neuronales profundas y la demanda de recursos computacionales significativos son solo algunos de los obstáculos que aún requieren investigación y desarrollo continuo.

En este trabajo de investigación la motivación es el deseo de ahondar en la comprensión de los fundamentos teóricos y las aplicaciones prácticas del aprendizaje profundo supervisado para la categorización de imágenes en detección de plagas en los cultivos de cacao y piña. Al explorar las arquitecturas de CNNs más relevantes, los conjuntos de datos de referencia, las métricas de evaluación y los desafíos actuales, se busca proporcionar una visión integral del estado del arte en este campo y sentar las bases para futuras investigaciones y avances.

## Planteamiento del trabajo

El presente trabajo se centra en la exploración de la categorización de imágenes para la detección de plagas en cultivos de cacao y piña utilizando métodos de aprendizaje profundo supervisado. La premisa de que el aprendizaje profundo, y en particular las CNNs, representan el estado del arte en la resolución de problemas complejos de clasificación visual. Lo cual permitirá evaluar y determinar el mejor método de aprendizaje profundo supervisado.

**Conceptos Fundamentales**: Definir y clarificar los conceptos esenciales relacionados con la categorización de imágenes y el aprendizaje profundo supervisado, estableciendo un marco teórico sólido para la investigación.

**Arquitecturas de CNNs**: Investigar y describir las arquitecturas de CNNs más influyentes y ampliamente utilizadas en la categorización de imágenes, analizando sus innovaciones, ventajas y limitaciones.

**Conjuntos de Datos de Referencia**: Identificar y caracterizar los conjuntos de datos de imágenes de referencia más relevantes y utilizados para el entrenamiento y la evaluación de modelos de categorización, destacando sus características y su importancia en la investigación.

**Métricas de Evaluación**: Explicar y analizar las métricas de evaluación estándar utilizadas para cuantificar el rendimiento de los modelos de categorización de imágenes, debatiendo su significado y aplicabilidad en diferentes contextos.

**Desafíos y Limitaciones**: Explorar y discutir los desafíos y limitaciones propias a la categorización de imágenes con aprendizaje profundo supervisado, incluyendo el sobreajuste, la necesidad de datos etiquetados y la interpretabilidad de los modelos.

**Tendencias y Avances Recientes**: Investigar, destacar las tendencias y los avances más recientes en la investigación de la categorización de imágenes, como el uso de mecanismos de atención y redes transformadoras.

**Pasos de Desarrollo**: Describir los pasos involucrados en el desarrollo de un sistema de categorización de imágenes utilizando aprendizaje profundo supervisado, desde la recopilación de datos hasta la evaluación del modelo.

En resumen, este trabajo proporciona una comprensión profunda y actualizada del campo de la categorización de imágenes utilizando aprendizaje profundo supervisado, sirviendo como un recurso valioso para aquellos interesados en investigar y aplicar estas técnicas en diversos dominios.

## Estructura del trabajo

• **Capítulo 1**: Introducción: Este capítulo proporciona una visión general del tema de la categorización de imágenes de plagas en los cultivos de cacao y piña utilizando métodos de aprendizaje profundo supervisado, incluyendo la motivación, el planteamiento del trabajo y la estructura del TFM.

• **Capítulo 2**: Contexto y estado del arte: En este capítulo se describe el contexto del problema de la categorización de imágenes de plagas en los cultivos de cacao y piña, se presenta un análisis del estado del arte en el campo del aprendizaje profundo supervisado aplicado a esta tarea. Se revisarán los antecedentes, los estudios actuales y las comparativas de herramientas existentes.

• **Capítulo 3**: Objetivos concretos y metodología de trabajo: Este capítulo detalla el objetivo general y los objetivos específicos de la investigación, formulados bajo los criterios SMART. Además, se describe la metodología de trabajo que se seguirá para alcanzar estos objetivos.

• **Capítulo 4**: Planteamiento de la comparativa: Este capítulo detalla el trabajo previo realizado para identificar el problema y las posibles soluciones a evaluar.

• **Capítulo 5**: Desarrollo de la comparativa: Este capítulo detalla la comparativa realizada, incluyendo resultados y mediciones obtenidos.

• **Capítulo 6**: Discusión y análisis de resultados: Este capítulo se centrará en la discusión, análisis de las ventajas y desventajas de las soluciones evaluadas.

• **Capítulo 7**: Conclusiones y trabajos futuros.

# Contexto y estado del arte

## Contexto del problema

La categorización o clasificación de imágenes, es una tarea fundamental dentro del campo de la visión por computador que se centra en la comprensión automática del contenido visual de una imagen. Implica la asignación de una categoría o etiqueta a una imagen completa basándose en su contenido visual. Esta tarea es crucial en una amplia gama de aplicaciones en diversos dominios, incluyendo la agricultura, la medicina, la seguridad, la teledetección, banca, finanzas y el comercio electrónico.

En la agricultura, las técnicas de aprendizaje supervisado se aplican para la detección y clasificación de enfermedades y defectos en imágenes de frutas, tal como se revisa en el trabajo de (Ynfante Martínez et al., 2022), quienes destacan la relevancia de estas técnicas para asegurar la calidad de productos como la fruta bomba y la guayaba en el mercado internacional. Asimismo, (Sanchez Pardo & More Villegas, 2023) han desarrollado métodos basados en procesamiento de imágenes digitales y aprendizaje automático para la clasificación por madurez de la fresa, buscando optimizar la gestión de cultivos y la comercialización.

En el ámbito de la medicina, el aprendizaje profundo supervisado se utiliza para la clasificación de tractos nerviosos mediante técnicas supervisadas, como se visualiza en la tesis de (Escalante Belmonte, 2024). En esta investigación se observa la implementación de técnicas basadas en deep learning para distinguir tractos específicos en imágenes de resonancia magnética por difusión, con el propósito de su análisis posterior y potencial uso en aplicaciones clínicas. Además, (Hinojosa Merlo, 2022) investiga la aplicación de técnicas de Deep Learning para la clasificación de radiografías de COVID-19, demostrando la utilidad de estas soluciones de Inteligencia Artificial para complementar la detección y el diagnóstico de esta enfermedad. Los documentos de (Ruíz & Domínguez, n.d.) también exploran el Deep Learning aplicado en imágenes fotoacústicas para la identificación del cáncer de seno, resaltando el potencial de estas técnicas en el diagnóstico médico.

En el campo de la seguridad, se emplean técnicas de aprendizaje profundo supervisado para la detección facial e identificación de personas en imágenes, como se aborda en el trabajo de (Salvadó i Gómez, 2024). Este proyecto se enfoca en entrenar un modelo de IA que detecte e identifique personas utilizando el algoritmo YOLO y técnicas de Deep Learning.

En la teledetección, el aprendizaje profundo supervisado se aplica para la clasificación automática de cobertura de suelo en imágenes satelitales, como se presenta en el artículo de (Suárez Londoño et al., 2017). En este trabajo se puede comprender el uso de Redes Neuronales Convolucionales para la clasificación y el mapeo automático de coberturas del suelo en imágenes satelitales. Así mismo, (Jiménez Cleves et al., 2024) realiza una evaluación comparativa del efecto del tamaño de entrenamiento de algoritmos de machine learning en la clasificación de imágenes Landsat-9, destacando la relevancia de la ingeniería geomática en la aplicación efectiva de modelos de clasificación de imágenes satelitales. El trabajo de (Baquerizo & Ventocilla, 2022) también evalúan algoritmos de machine learning en la clasificación de imágenes satelitales multiespectrales, en el caso de la Amazonía peruana, comparando el rendimiento de diferentes técnicas de aprendizaje supervisado.

En el comercio electrónico, la categorización de imágenes mediante aprendizaje profundo supervisado se utiliza para la clasificación de productos, la búsqueda visual y la recomendación de artículos, mejorando la experiencia del usuario y optimizando las ventas. Estos ejemplos ilustran la vasta aplicabilidad de la categorización de imágenes utilizando métodos de aprendizaje profundo supervisado en diversos campos, lo que destaca la importancia de esta área de investigación.

## Estado del arte

El aprendizaje profundo supervisado, y en particular las Redes Neuronales Convolucionales (CNNs), han revolucionado el campo de la categorización de imágenes (Cireşan et al., 2011). A lo largo de los años, se han desarrollado diversas arquitecturas de CNNs que han permitido obtener un rendimiento notable en tareas de categorización de imágenes.

Teoría sobre el aprendizaje profundo

De acuerdo con (Díaz-Ramírez, 2021) en la actualidad, la Inteligencia Artificial, Aprendizaje Automático y el Aprendizaje Profundo están en una etapa de constante crecimiento, serán seguramente los ejes principales para el desarrollo de la ciencia de la computación y la humanidad donde estarán fundidas, no sólo hardware y software, sino también varias tecnologías, tales como nanotecnología, computación cuántica, automatización, entre otras.

Por otra parte (Díaz Rodríguez et al., 2022) en su investigación manifiesta que gracias al aprendizaje automático las máquinas son capaces de aprender la relación oculta existente entre los datos de entrada.

También (Ramos & Santiago, 2022) comenta que el aprendizaje profundo es una de las poderosas técnicas de aprendizaje automático impulsadas por la IA. Las técnicas de aprendizaje profundo pueden procesar una gran cantidad de información presente en los conjuntos de datos de manera eficiente.

(Türkoğlu & Hanbay, 2019) afirma que las CNN utilizadas como herramientas básicas de aprendizaje profundo han obtenido un éxito significativo en varios estudios de detección de enfermedades de las plantas.

De igual manera (Wang et al., 2025) declara que los modelos de aprendizaje profundo han demostrado un rendimiento superior, ofreciendo mayor precisión, consistencia y escalabilidad en la detección de plagas y enfermedades. Estas ventajas resaltan la necesidad de explorar y sintetizar la investigación sobre aplicaciones de aprendizaje profundo en este campo, abordando las deficiencias críticas de los métodos tradicionales y sentando las bases para futuros avances. La Red Neuronal Convolucional (CNN), que aprovecha mecanismos como el aprendizaje jerárquico de características, la percepción local y la distribución de pesos, extrae eficazmente características complejas y discriminativas de imágenes sin procesar, lo que permite la identificación y segmentación de plagas y enfermedades de las plantas

Pero (Díaz-Gaxiola et al., 2019) comenta que el obstáculo principal cuando se utilizan las CNNs es la vasta cantidad de datos requeridos para su entrenamiento.

(Krizhevsky et al., 2011) marcó un hito en el campo de la visión por computador al ser una de las primeras CNNs profundas en lograr un éxito sobresaliente en el ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012. Su arquitectura, compuesta por ocho capas (cinco convolucionales y tres totalmente conectadas), introdujo varias innovaciones clave. Utilizó la función de activación ReLU, que demostró acelerar el entrenamiento y mitigar el problema del desvanecimiento del gradiente, superando a las funciones sigmoid y tanh utilizadas anteriormente; también empleó capas de max-pooling con solapamiento para reducir la dimensionalidad espacial y controlar el sobreajuste, así como la técnica de dropout en las capas totalmente conectadas para mejorar la generalización. Además, fue pionera en el uso de unidades de procesamiento gráfico (GPUs) para acelerar significativamente el entrenamiento en conjuntos de datos masivos como ImageNet. A pesar de estos avances, AlexNet podía ser susceptible al sobreajuste en conjuntos de datos más pequeños.

Desde el punto de vista de la agricultura (Nassr & Abu Naser, 2018) establece que, en muchos lugares, los especialistas en agricultura no tratan las enfermedades de la piña. De hecho, la presencia de especialistas y centros especializados para el tratamiento de enfermedades de la piña es escasa en la mayor parte del mundo.

(Ploetz, 2016) afirma que las enfermedades constituyen limitaciones significativas para la producción comercial y reducen el rendimiento en un 20 %, según estimaciones, o una producción estimada de 1,3 millones de toneladas de granos en 2012. Al ser trasladado fuera del neotrópico, el cacao se liberó de sus dos principales enemigos americanos: Moniliophthora roreri (causante de la mazorca helada) y M. perniciosa (escoba de bruja).

Podemos mencionar la Teoría de la Generalización (John Lu, 2010) ya que en el contexto del aprendizaje automático y la inteligencia artificial es un concepto fundamental que se refiere a la capacidad de un modelo para desempeñarse bien con datos nuevos y jamás antes vistos luego de haber sido entrenado con un conjunto de datos específico.

Por su parte (Goodfellow et al., 2016) en su libro menciona que el desafío principal en el aprendizaje automático es lograr que los algoritmos funcionen eficazmente con datos nuevos, no vistos previamente, más allá de los datos con los que fueron entrenados. A esta capacidad se le denomina "generalización”. La generalización está estrechamente ligada a la capacidad del modelo, así como a los fenómenos de sobreajuste (overfitting) y subajuste (underfitting).

A su vez (Vasileiou et al., 2025) explora las capacidades de generalización de las redes neuronales de paso de mensajes donde se refiere a la habilidad de las MPNNs para hacer predicciones significativas más allá de los datos de entrenamiento. Esto implica su capacidad para adaptarse eficazmente a nuevos grafos no vistos previamente que provienen de la misma distribución que el conjunto de entrenamiento.

La representación de características es el proceso de transformar los datos de entrada en un formato adecuado que sea de utilidad para un algoritmo de aprendizaje automático, en ese sentido se puede encontrar los embeddings y la comprensión de características.

En su estudio (Mikolov, 2013) determina que una limitación de los vectores de incrustación estática de word2vec es que las palabras pueden significar cosas diferentes en distintos contextos. Sin embargo, con las incrustaciones estáticas, cada palabra se representa mediante un único punto en el espacio vectorial, aunque pueda tener diversos significados.

(Li et al., 2022) en su artículo destaca la capacidad para aprender representaciones jerárquicas y de menor dimensión (embeddings) a partir de datos de imágenes, lo cual es crucial para tareas de clasificación, como la identificación de plagas.

La compresión no es solo una optimización sino una parte integral de cómo las redes profundas aprenden representaciones de alto nivel y semánticas, ignorando detalles irrelevantes de bajo nivel para enfocarse en los patrones clave que definen una plaga.

La evaluación de un modelo de aprendizaje profundo supervisado va más allá de solo ver su precisión. La robustez y la elección de métricas de evaluación adecuadas son cruciales para entender el verdadero rendimiento en un entorno real, específicamente en la detección de plagas.

En su artículo (Hancox-Li, 2020) alude que este desacuerdo subyace a debates sobre otras desideratas, como la robustez de las explicaciones ante pequeñas perturbaciones en los datos de entrada. Sostengo que la robustez es deseable en la medida en que nos preocupa encontrar patrones reales en el mundo.

A su vez (Oyen et al., 2022) en su estudio enfatiza que se ha demostrado, tanto empírica como teóricamente, que los clasificadores de aprendizaje automático son robustos al ruido de etiqueta bajo ciertas condiciones; en particular, se asume que el ruido de etiqueta es independiente de las características dada la etiqueta de clase.

En el tema agrícola (Ferentinos, 2018) resalta las variaciones en las condiciones de luz, el ángulo de la cámara, la presencia de otras plantas o suciedad, y las diferencias en el desarrollo de la enfermedad pueden afectar la precisión de los modelos. Esto subraya directamente la necesidad de modelos robustos en entornos agrícolas.

La cognición visual humana según (Pinker, 1984) se refiere a los procesos complejos por los cuales nuestro cerebro interpreta la información de la luz que llega a nuestros ojos, transformándola en una comprensión significativa del mundo que nos rodea. Este proceso no es pasivo, sino que involucra atención, memoria, reconocimiento de patrones y razonamiento.

A su vez (LeCun et al., 2015) describe que las CNNs fueron concebidas con una arquitectura que imita el procesamiento jerárquico de la corteza visual. Las capas convolucionales extraen características de bajo nivel en las primeras etapas y estas a su vez se combinan en capas posteriores para formar representaciones de alto nivel, lo cual es fundamental para el reconocimiento de objetos complejos.

De la misma manera (LeCun et al., 2002) en su investigación explican cómo estas capas reducen la resolución espacial de los mapas de características, lo que les permite a las redes ser menos sensibles a pequeñas traslaciones y distorsiones en la entrada, contribuyendo así a la invariancia frente a la traslación.

Por el contrario (Smeulders et al., 2000) en su artículo formalizó el concepto de la brecha semántica. Lo define explícitamente como la diferencia entre la información de bajo nivel que se puede extraer de los datos visuales y la interpretación de alto nivel de esos datos por parte de un usuario.

A si mismo (Datta et al., 2005) en su trabajo reitera que la brecha semántica es el principal desafío en este campo, la describe como la dificultad de traducir características visuales primitivas estas pueden ser color, textura y forma, en conceptos semánticos que los humanos utilizan para describir y buscar imágenes.

Otro tema interesante es el Fairness en Machine Learning, se refiere a la aspiración de diseñar, desarrollar y desplegar sistemas de IA de manera que no discriminen, no reproduzcan ni amplifiquen sesgos existentes en los datos o en la sociedad, y traten a todos los individuos o grupos de manera justa. Según (Mehrabi et al., 2021) fairness busca garantizar que los sistemas de IA no discriminen contra individuos o grupos específicos y que traten a todos de manera justa.

El tema ético es crucial, por ello (Floridi, 2023) en su artículo aborda la ética de la IA de manera más amplia, incluye la equidad como un pilar fundamental. Resalta la responsabilidad social de los modelos de IA, la necesidad de evitar la amplificación de sesgos sistémicos cuando se despliegan en el día a día de las personas.

## Conclusiones

El estado del arte para la categorización de imágenes utilizando aprendizaje profundo supervisado ha tenido un avance significativamente en los últimos años. Las arquitecturas de CNNs han demostrado ser altamente efectivas, y la investigación continúa explorando nuevas formas de mejorar la precisión, la eficiencia y la interpretabilidad de estos modelos. La disponibilidad de grandes conjuntos de datos etiquetados y el desarrollo de potentes recursos computacionales han sido fundamentales para estos avances. Sin embargo, todavía existen desafíos relacionados con el sobreajuste, la necesidad de grandes cantidades de datos y la interpretabilidad de los modelos, que continúan impulsando la investigación en este campo. El conocimiento adquirido en este capítulo proporciona el contexto necesario para comprender los objetivos y la metodología de trabajo que se abordarán en los siguientes capítulos de esta investigación.

# Objetivos concretos y metodología de trabajo

## Objetivo general

Identificada la escases de categorización de imágenes de plagas dentro del país y al considerar que esto ocasiona perdidas significativas a los productores de cacao y piña, el presente trabajo realiza una comparativa para determinar la efectividad de diferentes arquitecturas de aprendizaje automático profundo supervisado para la categorización automática de imágenes de plagas en cultivos de cacao y piña, con el fin de mejorar la detección temprana y el manejo de estas amenazas agrícolas.

## Objetivos específicos

Establecer mediante un estudio documental las técnicas de Machine Learning a comparar.

Desarrollar la comparativa de métodos de aprendizaje profundo supervisado para la detección de plagas en cultivos de cacao y piña.

Establecer ventajas y desventajas de los métodos de aprendizaje profundo supervisado comparados para la detección de plagas en cultivos de cacao y piña.

## Metodología del trabajo

De cara a alcanzar los objetivos específicos se explicará en este capítulo las fuentes de información utilizadas, la descripción de las técnicas e instrumentos utilizados para la recopilación de los datos, los enfoques para el análisis de la información recabada y los resultados previstos para completar cada propósito de investigación.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Objetivo Específico** | **Fuentes** | **Técnica de Recolección de Información** | **Instrumento** | **Técnica de Análisis de Información** | **Resultado** |
| **1. Establecer mediante un estudio documental las técnicas de Machine Learning a comparar.** | Artículos | Revisión documental | Ficha de revisión documental | Análisis documental | Listado top 4 de modelos de Deep Learning supervisado |
| **2. Desarrollar la comparativa de métodos de aprendizaje profundo supervisado para la detección de plagas en cultivos de cacao y piña.** | Datasets de imágenes de plagas de cacao y piña | Experimentación controlada | Bitácora | Análisis | Informe de resultados |
| **3. Establecer ventajas y desventajas de los métodos de aprendizaje profundo supervisado comparados para la detección de plagas en cultivos de cacao y piña.** | Resultados de la comparativa, Métricas de evaluación estándar en visión por computadora | Evaluación y Análisis | Bitácora | Análisis estadístico | Informe de ventajas y desventajas |

# Planteamiento de la comparativa

Este capítulo tiene como propósito fundamental contextualizar la investigación, delineando el problema específico a abordar y las soluciones tecnológicas alternativas que serán evaluadas. Asimismo, se establecerán los criterios explícitos de éxito para la comparativa de modelos de Deep Learning, las métricas cuantitativas que se emplearán para evaluar su rendimiento y las consideraciones generales que guiarán el estudio experimental.

## Identificación y Contextualización del Problema

La agricultura es un pilar económico fundamental en diversas regiones, incluyendo aquellas donde el cultivo de cacao y piña es predominante. Sin embargo, estos cultivos son constantemente amenazados por diversas plagas, que pueden causar pérdidas significativas en la producción, degradación de la calidad del producto, en casos extremos, la destrucción total de las cosechas. Tradicionalmente, la detección de estas plagas se ha basado en la inspección visual por parte de agrónomos o técnicos especializados. Este método es intrínsecamente subjetivo, consume mucho tiempo, es costoso y, a menudo, la detección se produce cuando el daño ya es considerable, limitando la efectividad de las intervenciones.

La identificación tardía de plagas no solo incrementa el uso de pesticidas, con sus consecuentes impactos ambientales y en la salud humana, sino que también dificulta la implementación de estrategias de manejo integrado de plagas (MIP) que buscan soluciones más sostenibles. Por lo tanto, existe una necesidad apremiante de desarrollar sistemas de detección de plagas que sean más rápidos, precisos, automatizados y capaces de operar a gran escala en el campo.

## Soluciones Alternativas a Evaluar: Modelos de Deep Learning Supervisado

En la última década, los avances en el campo del Deep Learning, particularmente en las Redes Neuronales Convolucionales, han revolucionado el área de la visión por computadora. Estos modelos han demostrado una capacidad excepcional para aprender características complejas directamente de los datos de imagen, superando a menudo los métodos de procesamiento de imágenes tradicionales y el Machine Learning clásico en tareas de clasificación y detección de objetos.

Para abordar el problema de la detección de plagas en cultivos de cacao y piña, esta investigación se centrará en la evaluación comparativa de las siguientes soluciones alternativas, todas ellas basadas en arquitecturas de Deep Learning supervisado, reconocidas por su estado del arte en tareas de visión por computadora:

YOLO (You Only Look Once): Reconocido por su alta velocidad y buena precisión en la detección de objetos en tiempo real. Ya lo dice (Redmon et al., 2015) YOLO es extremadamente rápido. Dado que enmarcamos la detección como un problema de regresión, no necesitamos una secuencia compleja.

Faster R-CNN (Region-based Convolutional Neural Network): Conocido por su alta precisión y robustez en la detección de objetos, aunque generalmente más lento que YOLO. Como lo indica (Ahmed & Das, 2025) los resultados muestran que Faster R-CNN logra la máxima precisión en la detección de peatones y ciclistas, mientras que YOLOv5 tiene un buen equilibrio entre velocidad y precisión.

Mask R-CNN: Una extensión de Faster R-CNN que añade la capacidad de segmentación de instancias a nivel de píxel. (He et al., 2016) menciona que el método genera eficientemente objetos en una imagen mientras simultáneamente produce una máscara de segmentación de alta calidad para cada instancia.

ResNet / Inception (como backbones para clasificadores): Estas arquitecturas son fundamentales como extractores de características y como base para clasificadores de imágenes. Ya lo afirma (He et al., 2015) donde explica cómo las redes residuales profundas ResNets facilitan el entrenamiento de redes mucho más profundas, lo que conduce a una mayor precisión en la clasificación de imágenes. El concepto de saltos residuales permite que la información fluya directamente a través de capas, mitigando el problema de la desaparición o explosión de gradientes.

“Estas arquitecturas representan un espectro de soluciones desde las más rápidas (YOLO) hasta las más precisas y detalladas (Mask R-CNN), permitiendo una comparativa exhaustiva de sus trade-offs en el contexto específico de la detección de plagas en los cultivos de interés.”

## Criterios de Éxito para la Comparativa

El éxito de esta comparativa se definirá mediante un conjunto de criterios cuantitativos y cualitativos, que permitirán determinar cuál o cuáles modelos de Deep Learning son los más adecuados para la detección de plagas en cultivos de cacao y piña. Los criterios principales son:

Precisión de Detección/Clasificación: Los modelos deben ser capaces de identificar correctamente las plagas con un alto grado de fiabilidad. Esto se medirá mediante métricas estándar como el mean Average Precision (mAP) para detección de objetos, y la Precisión (Accuracy), Precisión (Precision), Recall y F1-Score para clasificación. Un modelo exitoso deberá alcanzar un mAP superior al 80% o una precisión general superior al 90% en el conjunto de prueba.

Eficiencia Computacional y Velocidad de Inferencia: La capacidad de los modelos para procesar imágenes rápidamente es crucial para aplicaciones en tiempo real o en dispositivos embebidos en el campo. El tiempo de inferencia (medido en milisegundos por imagen o Frames Per Second - FPS) será un criterio clave. Se buscará un balance óptimo entre precisión y velocidad.

Robustez y Generalización: Los modelos deben demostrar un rendimiento consistente y fiable en diversas condiciones de imagen, iluminación variable, diferentes ángulos, presencia de ruido, distintas etapas de desarrollo de la plaga y del cultivo. Esto se evaluará a través del rendimiento en el conjunto de prueba y la capacidad de los modelos para generalizar a nuevas imágenes no vistas durante el entrenamiento.

Complejidad del Modelo y Requerimientos de Recursos: Se considerará el tamaño del modelo, número de parámetros, MB de almacenamiento y los recursos de hardware necesarios para el entrenamiento e inferencia. Un modelo más ligero y menos demandante en recursos podría ser preferible para un despliegue práctico.

Facilidad de Implementación y Mantenimiento: Aunque más cualitativo, se evaluará la complejidad de la arquitectura y la disponibilidad de herramientas y librerías para su implementación y futuro mantenimiento.

## Medidas y Métricas de Evaluación

Para cuantificar los criterios de éxito, se emplearán las siguientes métricas detalladas, que serán calculadas en el conjunto de datos de prueba, asegurando una evaluación imparcial:

Matriz de Confusión: Representación tabular de los resultados de clasificación que muestra los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos para cada clase.

Para Detección de Objetos y Segmentación de Instancias:

Intersection over Union (IoU): Métrica de solapamiento entre la bounding box predicha y la ground truth.

Average Precision (AP) por clase: El área bajo la curva de Precisión-Recall para cada clase.

Mean Average Precision (mAP): El promedio de AP sobre todas las clases. Se calculará mAP@0.5 (IoU=0.5) y mAP@[0.5:0.95] promedio de mAP en diferentes umbrales de IoU si es relevante.

Para Mask R-CNN, mAP\_mask: Similar a mAP pero evaluando la precisión de las máscaras de segmentación.

Métricas de Tiempo:

Tiempo de Inferencia por Imagen: El tiempo promedio que tarda cada modelo en procesar una única imagen desde la entrada hasta la salida de la predicción, medido en milisegundos (ms) o calculando los Frames Per Second (FPS).

Estas métricas permitirán una comparativa rigurosa y objetiva de los modelos de Deep Learning, proporcionando una base sólida para determinar su idoneidad para la detección de plagas en los cultivos de cacao y piña.

# Desarrollo de la comparativa

En este capítulo se muestra la información referente a los modelos evaluados así como la información referente a los datasets utilizados en el desarrollo de esta comparativa de soluciones.

# Discusión y análisis de resultados

# Conclusiones y trabajos futuros