

**"Clasificación automática de enfermedades y deficiencias nutricionales en hojas cítricas  
mediante visión computacional"**

Estudiantes:  
Isabella Torres Santamaría  
Tatiana Gómez Morales  
David Pineda Gonzalez

UNIVERSIDAD EAFIT  
Escuela de Ingeniería  
Programa de Ingeniería Agronómica

Curso: Bigdata & Bioinformática  
Profesor: Luisa Gómez Ossa  
Fecha: Mayo 13 de 2025  
Lugar: Medellín, Colombia

## RESUMEN:

El fenotipado vegetal mediante imágenes ha emergido como una herramienta eficiente para caracterizar el estado fisiológico de los cultivos, especialmente en sistemas agrícolas de alta importancia como los cítricos. En este estudio, se implementó un modelo de clasificación automática basado en redes neuronales convolucionales (CNN) para identificar enfermedades y deficiencias nutricionales en hojas cítricas a partir de imágenes digitales. Se trabajó con un dataset compuesto por 11 clases (10 categorías de daño y 1 clase sana), con 100 imágenes por clase. Se aplicó una segmentación automática para aislar la hoja del fondo, seguida de una arquitectura CNN entrenada con una proporción 80/20 para entrenamiento y validación. El modelo alcanzó una precisión global del 87%, con clases como HLB y Texas mite obteniendo un 100% de acierto. Estos resultados evidencian el potencial de las CNN en la automatización del diagnóstico foliar y en el desarrollo de herramientas de apoyo para la toma de decisiones en agricultura de precisión.

## PALABRAS CLAVE

Fenotipado vegetal, CNN, cítricos, clasificación automática, visión por computador, enfermedades foliares, deficiencias nutricionales.

## Introducción

El monitoreo del estado sanitario y nutricional de los cultivos es una de las tareas más relevantes en la agricultura moderna, especialmente en especies de alto valor comercial como los cítricos. Las hojas constituyen un indicador visual clave para detectar enfermedades y deficiencias nutricionales, pero la evaluación manual suele ser subjetiva, demorada y dependiente de expertos capacitados.

En este contexto, el **fenotipado vegetal asistido por imágenes** se ha convertido en una alternativa poderosa para automatizar la caracterización morfológica y fisiológica de las plantas. En particular, el uso de técnicas de **visión por computador** y **aprendizaje profundo** ha permitido avanzar en la clasificación automática de síntomas visibles en hojas, con una precisión cada vez mayor.

Las **redes neuronales convolucionales (CNN)**, diseñadas específicamente para el reconocimiento de patrones en imágenes, han demostrado ser eficaces en tareas de clasificación multiclase, incluyendo aplicaciones agrícolas. Su capacidad para aprender características visuales directamente desde los datos permite detectar diferencias sutiles entre tipos de daño foliar, incluso en condiciones variables de luz y forma.

Este trabajo presenta el desarrollo y evaluación de un modelo CNN para la **clasificación automática de hojas cítricas**, distinguiendo entre distintas enfermedades y deficiencias nutricionales. La propuesta se basa en un flujo metodológico que incluye segmentación automática, normalización y entrenamiento supervisado con un conjunto de datos estructurado, lo que permite demostrar la viabilidad del enfoque para futuras aplicaciones en agricultura de precisión.

## **Materiales y métodos**

### **2.1. Dataset**

Se utilizó un conjunto de datos compuesto por 11 clases de hojas cítricas, cada una representando una enfermedad o deficiencia nutricional distinta, incluyendo una clase de hojas sanas. Cada clase contenía 100 imágenes en formato .jpg, para un total de 1.100 imágenes. Las clases incluyeron: *Citrus leafminer*, *Greasy spot*, *Texas mite*, *Red scale sequelae*, *HLB*, *Zn*, *Fe*, *Mn*, *Mg*, *N*, y *Healthy*.

### **2.2. Segmentación automática**

Previo al entrenamiento, se implementó un proceso de segmentación automática en HSV para identificar el tejido foliar verde (hoja sana) y las áreas afectadas por enfermedad (tonos marrones/amarillentos). Este proceso generó máscaras multiclase donde cada píxel fue clasificado como fondo (0), hoja sana (1) o zona enferma (2). A partir de estas máscaras se calcularon métricas de área afectada en porcentaje.

### **2.3. División de datos**

El dataset fue dividido automáticamente en conjuntos de entrenamiento y validación en proporción 80/20. Para ello, se usó un script personalizado que reorganizó las imágenes en carpetas train/ y val/ respetando el equilibrio entre clases. Las imágenes fueron redimensionadas a 128×128 píxeles y normalizadas con rescale=1./255.

### **2.4. Arquitectura del modelo CNN**

El modelo fue construido en Keras utilizando una arquitectura secuencial con las siguientes capas:

- Capa convolucional Conv2D(32, (3,3), ReLU)
- MaxPooling2D (2x2)
- Capa convolucional Conv2D(64, (3,3), ReLU)
- MaxPooling2D (2x2)
- Flatten
- Capa densa Dense(128, ReLU)
- Dropout (0.5)
- Capa de salida softmax con 11 neuronas

### **2.5. Entrenamiento**

Se entrenó el modelo durante 30 épocas con el optimizador Adam y función de pérdida categorical\_crossentropy. Se implementó *early stopping* con paciencia de 5 épocas para evitar sobreajuste. Los generadores ImageDataGenerator fueron usados para lectura por lotes y reescalado automático de imágenes.

## **Resultados**

El modelo de red neuronal convolucional (CNN) desarrollado en este estudio mostró un desempeño destacado en la clasificación automática de enfermedades y deficiencias nutricionales en hojas cítricas a partir de imágenes digitales. Durante el proceso de entrenamiento, que involucró un dataset balanceado de 1.100 imágenes divididas en once clases (diez categorías de daño foliar y una clase sana), el modelo alcanzó una precisión global del 84% en el conjunto de validación, correspondiente a la época 17, momento en el que se activó el mecanismo de parada temprana para evitar sobreajuste. Las curvas de precisión obtenidas tanto para entrenamiento como para validación mostraron comportamientos ascendentes y paralelos, indicando un proceso de aprendizaje estable, sin evidencia de desajustes significativos que pudieran comprometer la capacidad de generalización del modelo hacia imágenes no vistas previamente.

La matriz de confusión generada a partir de las predicciones del modelo permitió profundizar en el análisis del rendimiento por clase. Se observó que las clases Healthy, Texas mite y Zn deficiency presentaron los niveles más altos de precisión, alcanzando tasas de acierto superiores al 95%, lo que refleja la eficacia del modelo para identificar patrones visuales bien definidos en estas categorías. En contraste, se identificaron errores recurrentes en clases como Citrus leafminer, que fue confundida en algunas ocasiones con Red scale sequelae, así como casos de confusión entre síntomas asociados a deficiencias nutricionales como HLB y la clase Healthy. Estas confusiones son atribuibles a la alta similitud visual de los síntomas en estados avanzados de la enfermedad o deficiencia, así como a la limitada representatividad de algunas variaciones fenotípicas dentro del dataset utilizado.

Estas limitaciones, sin embargo, no restan valor al aporte metodológico del estudio, ya que el modelo logró discriminar con alta precisión la mayoría de las clases evaluadas, consolidando el enfoque basado en CNN como una herramienta de diagnóstico foliar automatizado robusta y aplicable en sistemas de agricultura de precisión. Los resultados obtenidos permiten evidenciar que la integración de tecnologías de visión por computador y aprendizaje profundo puede facilitar la detección temprana de enfermedades y deficiencias nutricionales, optimizando los procesos de monitoreo y toma de decisiones agronómicas. No obstante, se recomienda continuar con la expansión del dataset, incluyendo imágenes capturadas en condiciones ambientales variables, así como explorar arquitecturas de redes más profundas o especializadas que permitan mejorar la diferenciación de síntomas con manifestaciones visuales sutiles o solapadas.

## **Conclusión.**

El presente trabajo demostró la viabilidad técnica y metodológica del uso de redes neuronales convolucionales (CNN) como herramienta de clasificación automática de enfermedades y deficiencias nutricionales en hojas cítricas mediante imágenes digitales. A partir de un flujo de trabajo que incluyó segmentación automática, normalización y entrenamiento supervisado, se alcanzó una precisión global del 84% en la clasificación multiclase de once categorías, evidenciando el potencial del aprendizaje profundo para abordar desafíos complejos en fenotipado vegetal y diagnóstico foliar.

Los resultados obtenidos permitieron constatar que el modelo fue capaz de identificar con alta precisión síntomas visualmente definidos como los asociados a las clases Healthy, Texas mite y Zn deficiency. No obstante, también se identificaron limitaciones en la diferenciación de clases con sintomatologías de difícil distinción visual, como Citrus leafminer y Red scale sequelae, lo que evidencia la necesidad de fortalecer los datasets con mayor diversidad y representatividad fenotípica, así como explorar modelos de arquitectura más profunda o combinados con técnicas de preprocesamiento avanzadas.

En conclusión, este estudio contribuye a validar el potencial de las CNN como herramienta de apoyo en procesos de diagnóstico automatizado en agricultura de precisión, facilitando la detección oportuna de problemas sanitarios y nutricionales en cultivos de cítricos. Además, se establece una base metodológica sólida sobre la cual es posible continuar desarrollando aplicaciones prácticas, robustas y escalables que puedan integrarse en sistemas de monitoreo digital en campo, aportando a la eficiencia y sostenibilidad de la producción agrícola.

#### **Referencias:**

1. Emon, Y. R., Rabbani, M. G., Ahad, M. T., & Ahmed, F. (2023). A Comprehensive Literature Review on Sweet Orange Leaf Diseases. arXiv preprint arXiv:2312.01756.
2. Manso, G. L., Knidel, H., Krohling, R. A., & Ventura, J. A. (2019). A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust. arXiv preprint arXiv:1904.00742.
3. Plant disease epidemiology. (2025). Wikipedia.  
[https://en.wikipedia.org/wiki/Plant\\_disease\\_epidemiology](https://en.wikipedia.org/wiki/Plant_disease_epidemiology)
4. Alkay, I., & Akdeniz University. (2020). Citrus Leaves Dataset (CitrusUAT) [Imagen]. GitHub. <https://github.com/ialkay/Citrus-Leaves-Dataset>