

Análisis automatizado de imágenes para la detección de antracnosis en frutos de aguacate (*Persea americana* Mill)

Arango Ospina, Yeison¹; Gallego Thorne, Sara²; Juagibioy Satiaca, Jayder Andres³

¹Universidad Eafit / Estudiante del pregrado de Ingeniería Agronómica varangoo@eafit.edu.co

²Universidad Eafit / Estudiante del pregrado de Ingeniería Agronómica sgallegot1@eafit.edu.co

³Universidad Eafit / Estudiante del pregrado de Ingeniería Agronómica jajuagibis@eafit.edu.co

Resumen

El presente estudio aborda la detección automatizada de antracnosis en frutos de aguacate (*Persea americana* Mill.) mediante técnicas de aprendizaje automático. Utilizando un conjunto de datos de imágenes etiquetadas provenientes de la plataforma Kaggle, se entrenó un modelo de clasificación para diferenciar entre frutos sanos y afectados por antracnosis. A partir del etiquetado manual de píxeles en imágenes, se construyó un modelo basado en Random Forest, cuya validación arrojó una precisión del 99%. Los resultados obtenidos evidencian el potencial de esta metodología para apoyar el diagnóstico temprano de enfermedades en cultivos agrícolas, con aplicaciones prácticas en la agricultura de precisión.

Palabras clave: Antracnosis, Aguacate, Machine Learning, Etiquetado.

Introducción

El aguacate (*Persea americana* Mill.) es una especie arbórea originaria de Mesoamérica, cultivada por su fruto comestible [1]. En Colombia, su cultivo es de gran importancia económica y social, especialmente en regiones como Antioquia, donde se ha consolidado como uno de los principales productos agrícolas de exportación [2]. Sin embargo, la producción de aguacate enfrenta desafíos fitosanitarios, siendo la antracnosis una de las enfermedades más significativas.

La antracnosis, causada principalmente por el hongo *Colletotrichum gloeosporioides*, afecta tanto la etapa de desarrollo del fruto como la postcosecha, provocando manchas oscuras en el exocarpio (cáscara de la fruta) y pudrición, lo que conlleva a pérdidas económicas que pueden alcanzar hasta el 20% de la producción [2] [3] [4]. El control de esta enfermedad es complejo, ya que las infecciones pueden permanecer latentes y manifestarse bajo condiciones de alta humedad y temperaturas moderadas.

Ante este panorama, la implementación de tecnologías avanzadas como el aprendizaje de máquinas (Machine Learning) ofrece nuevas oportunidades para el diagnóstico y manejo de enfermedades en cultivos agrícolas. El uso de modelos de inteligencia artificial permite la detección temprana y precisa de síntomas de enfermedades como la antracnosis, facilitando intervenciones oportunas y prácticas agrícolas.

En este contexto, el presente estudio se enfocó en el desarrollo y entrenamiento de un modelo de Machine Learning para la clasificación de frutos de aguacate sanos y afectados por antracnosis, utilizando un conjunto de datos de imágenes etiquetadas. Esta investigación busca contribuir al fortalecimiento de estrategias de manejo integrado de enfermedades en el cultivo de aguacate, promoviendo prácticas agrícolas más sostenibles y eficientes.

Objetivos

Objetivo general:

Desarrollar un modelo de aprendizaje profundo para clasificar frutos de aguacate (*Persea americana* Mill.) como sanos o afectados por antracnosis, utilizando imágenes etiquetadas y validando su desempeño con un modelo basado en Random Forest.

Objetivos específicos.

1. Entrenar un modelo de Machine Learning utilizando un conjunto de datos de imágenes de aguacate, con el fin de identificar la presencia de antracnosis.
2. Validar la veracidad de las predicciones del modelo entrenado mediante la comparación de resultados con un clasificador Random Forest, evaluando métricas como precisión, sensibilidad y especificidad.

Metodología

Descripción del DataSet

Se empleó el dataset público titulado “Clasificación de enfermedades con Deep Learning”, disponible en la plataforma Kaggle, el cual está diseñado para el entrenamiento y evaluación de modelos de aprendizaje profundo en la clasificación de enfermedades en frutas. [\[Enlace\]](#). El dataset original incluye un total de 669 imágenes distribuidas en tres clases principales: Frutos sanos (S), frutos con antracnosis (A) y frutos con la enfermedad de roña (R).

La roña es una enfermedad causada por los hongos *Sphaceloma perseae* o *Elsinoe perseae*. Esta enfermedad se caracteriza por provocar lesiones de apariencia corchosa en el epicarpio de los frutos, lo que disminuye la calidad y precio; consiste en manchas oscuras redondas o irregulares que pueden cubrir parcial o totalmente el fruto, dando un aspecto corchoso café [5].

Para los propósitos de este proyecto, se seleccionaron únicamente las imágenes correspondientes a frutos sanos (S) y frutos con síntomas de antracnosis (A), comprendiendo un total de 288 y 178 imágenes respectivamente.

Procesamiento.

La manera en la que se etiquetaron los puntos se observa en la Tabla 1.

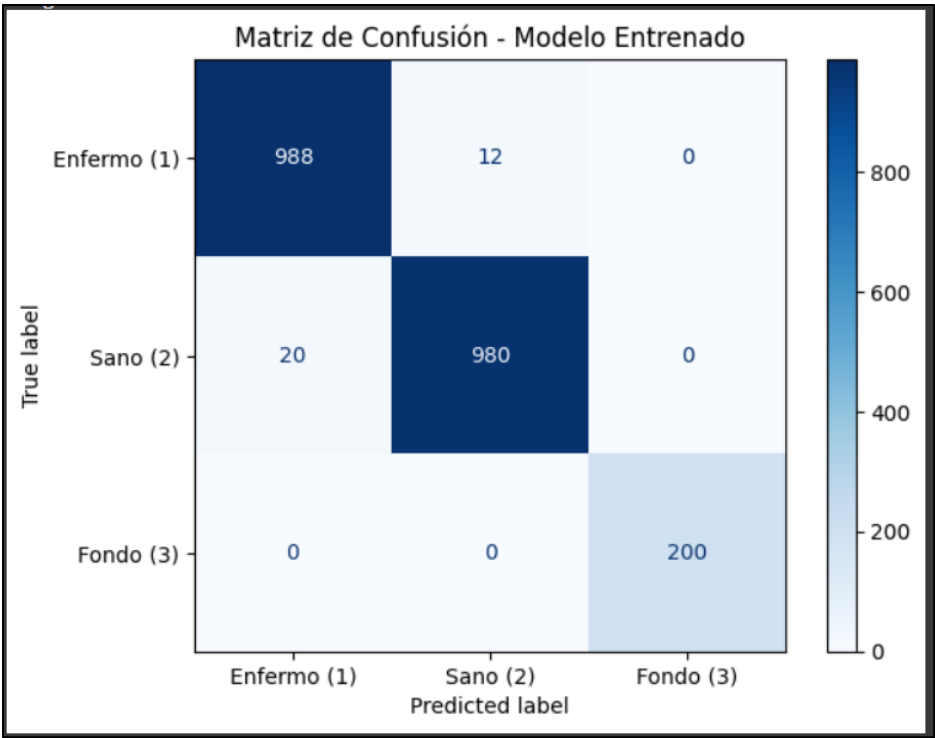
Etiqueta	Característica	Cantidad de datos
1	Enfermedad	1000
2	Sano	1000
3	Fondo	200

Tabla 1. Distribución de Píxeles Etiquetados por Clase para el entrenamiento del modelo.

Se construyó un archivo Excel con la información etiquetada de las imágenes para un total de 2201 líneas (incluyendo el encabezado) con datos RGB de cada píxel etiquetado según la Tabla 1.

El código implementado describe un flujo completo para el entrenamiento y evaluación de un modelo de clasificación basado en Random Forest, orientado a segmentar imágenes de aguacates en tres clases: enfermo, sano y fondo, a partir de características de color (valores RGB). Para ello, se cargan los datos desde un archivo Excel, se preprocesan las columnas, y se construye un pipeline de machine Learning que incluye la normalización de datos y el clasificador. A través de una búsqueda exhaustiva de hiperparámetros mediante validación cruzada estratificada (GridSearchCV), se optimiza el rendimiento del modelo. Finalmente, se evalúa el desempeño del mejor modelo obtenido usando métricas como la precisión global y un reporte detallado de clasificación, lo cual permite validar su capacidad predictiva sobre los datos de entrenamiento. Todo el código para este flujo de trabajo se desarrolló en el entorno de Google Colab, empleando el lenguaje Python.

Resultados



Gráfica 1. Matriz de confusión del modelo Random Forest entrenado para la clasificación.

La matriz de confusión obtenida del modelo entrenado refleja un desempeño sólido en la clasificación de las tres clases definidas: frutos enfermos (clase 1), frutos sanos (clase 2) y fondo (clase 3). El modelo identificó correctamente 988 de 1000 frutos enfermos, y 980 de 1000 frutos sanos, con errores mínimos de clasificación cruzada entre ambas clases (12 y 20 casos, respectivamente) lo que indica una alta sensibilidad y especificidad. La clase fondo fue clasificada correctamente en su totalidad (200 muestras), sin confusiones con las clases de interés. Estos resultados confirman la capacidad del modelo para discriminar eficazmente entre frutos afectados por antracnosis y sanos, incluso en presencia de ruido visual como el fondo.

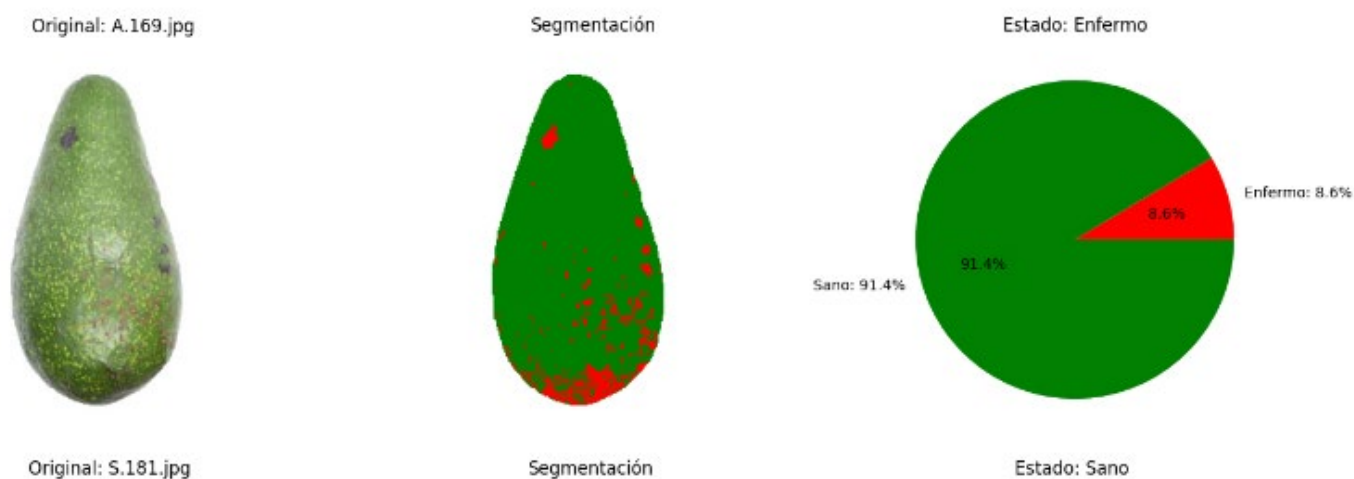
Tabla 2. Reporte de clasificación del modelo Random Forest entrenado para clasificación.

	Precision	Recall	F1-score	Support
1	0.98	0.99	0.98	1000
2	0.99	0.98	0.98	1000
3	1	1	1	200
accuracy	-	-	0.99	2200
Macro avg	0.99	0.99	0.99	2200
Weighted avg	0.99	0.99	0.99	2200

Se obtuvieron altos valores de precisión, recall y F1-score para las clases de interés. La clase 1 alcanzó una precisión de 0.98 y un recall de 0.99, mientras que la clase 2 obtuvo valores de 0.99 y 0.98, respectivamente. La clase 3 también fue identificada con alta efectividad, reflejando la solidez del modelo en la segmentación general de las imágenes.

En términos generales, se muestra la precisión del modelo con un valor de accuracy (precisión) de 0.99, es decir que el modelo tuvo un 99% de aciertos en la clasificación con promedios macro y ponderado de 0.99 y 0.99,

respectivamente, esto puede observarse en la gráfica 1. Estos resultados indican una alta confiabilidad del modelo propuesto para distinguir entre frutos sanos y aquellos afectados por antracnosis.



Gráfica 2. Ejemplo de Resultado de la Segmentación y Clasificación de imágenes de aguacates.

Incluso en imágenes etiquetadas como correspondientes a frutos sanos por el creador del dataset, el modelo identifica regiones que clasifica como tejido enfermo, lo que evidencia una alta sensibilidad del modelo.

Análisis

El entrenamiento del modelo resultó exitoso con un porcentaje de aciertos de 99%, lo que demuestra que la detección de la enfermedad de antracnosis fue identificada con éxito en un 99% de las imágenes.

El entrenamiento del modelo arrojó una precisión del 99%, lo que indica una alta efectividad en la identificación de la antracnosis en las imágenes. Si bien el volumen de datos de entrenamiento (669 imágenes) es considerablemente menor en comparación con estudios que emplean hasta 200,000 imágenes para modelos de *machine Learning* y *Deep Learning* [6], la robustez del presente modelo se fundamenta en la granularidad del etiquetado a nivel de píxel. Este proceso manual y exhaustivo permitió una captura precisa de las características relevantes, demostrando la efectividad del enfoque basado en datos RGB por píxel para discernir patrones sutiles entre las clases.

Conclusiones

El modelo desarrollado logró una clasificación efectiva de frutos sanos y enfermos con una precisión del 98%, demostrando alta sensibilidad y especificidad en la detección de antracnosis. A pesar de contar con un número relativamente limitado de imágenes de entrenamiento, la calidad del etiquetado manual y el aprovechamiento de la información a nivel de píxel contribuyeron significativamente al buen desempeño del modelo.

Esta investigación demuestra que el uso de técnicas de machine Learning, incluso con datasets sencillos pero bien estructurados, puede ser una herramienta viable para mejorar la sanidad vegetal y apoyar decisiones en el manejo fitosanitario del aguacate.

Bibliografía

[1] Romero Sánchez, M. A. (2012). Comportamiento fisiológico del aguacate (*Persea americana* mill.) Variedad Lorena en la zona de Mariquita, Tolima. Universidad Nacional de Colombia Facultad de Agronomía, Escuela de Posgrados Bogotá, Colombia.

- [2] Salazar-García, S. (2002). Nutrición del Aguacate, Principios y Aplicaciones. Queretaro, Méjico: INPOFOS. Instituto de la potasa y el fosforo A.C.
- [3] Rodríguez, A. T., Dávila, J. F. R., Siclán, M. L. S., Vildózola, Á. C., Zamora, F. I. M., & Díaz, A. V. L. (2020). Distribución espacial de antracnosis (*Colletotrichum gloeosporioides* Penz) en aguacate en el Estado de México, México. *Revista Argentina de Microbiología*, 52(1), 72-81. <https://doi.org/10.1016/j.ram.2019.07.004>
- [4] Antracnosis en el Cultivo de Aguacate | Intagri S.C. (s. f.). <https://www.intagri.com/articulos/fitosanidad/antracnosis-en-el-cultivo-de-aguacate>
- [5] Morales-García, J. L., López-Cornejo, C. I., Pedraza-Santos, M. E., Chávez-Bárcenas, A. T., Esquivel-Miguel, E., García-Morales, S., & Pineda-Guillermo, S. (2023). Morpho-molecular identification of the causal agent of avocado scab in Michoacán. *Revista Mexicana de Fitopatología*(En Línea)/*Revista Mexicana de Fitopatología*, 41(2). <https://doi.org/10.18781/r.mex.fit.2302-4>
- [6] Warman, C., & Fowler, J. E. (2021). Deep learning-based high-throughput phenotyping can drive future discoveries in plant reproductive biology. *Plant Reproduction*, 34(2), 81-89. <https://doi.org/10.1007/s00497-021-00407-2>