



**Facultad de Ingeniería**  
**Doctorado en Ingeniería Ambiental**  
**Fundamentos de Deep Learning**  
Tatiana Camila Puentes Escobar  
Diego Alejandro Grajales González

## 1. Contexto de aplicación

La producción agrícola, eje central de la seguridad alimentaria, depende de las condiciones del medio ambiente, por lo que, se ve afectada por el cambio climático, la contaminación del recurso hídrico y la calidad del suelo (Lira-Saldivar & Lara-Viveros, 2023). Por otro lado, las plagas y enfermedades tienen un impacto negativo en el rendimiento y la calidad de las cosechas (Lira-Saldivar & Lara-Viveros, 2023), agravado por el escaso conocimiento técnico de los productores sobre el manejo de estas, y la inapropiada asistencia técnica que hacen que las aplicaciones de plaguicidas in situ sean inadecuadas (Parra et al., 2012).

En contrapartida, que el rendimiento de los cultivos dependa de las condiciones del ambiente, ha generado soluciones basadas en tecnologías de inteligencia artificial (IA) en las áreas de monitoreo y establecimiento de los cultivos, la detección variables físicas y nutricionales del suelo, presencia de malezas y plagas, entre otras (Saxena et al., 2023; Lira-Saldivar & Lara-Viveros, 2023). Sudar et al. (2022) aplicó sistemas de IA para el análisis y categorización de plagas y enfermedades del cultivo de tomate en etapa temprana, con base en conocimientos agrónomos, determinando que pueden ayudar a limitar la transmisión de patógenos en el área sembrada agrícola y al mismo tiempo, ayudar a aumentar el rendimiento de los cultivos (Lira-Saldivar & Lara-Viveros, 2023). Por tal motivo, es factible la aplicación de modelos de redes neuronales para brindar soluciones y estrategias para que los cultivos mejoren los rendimientos y reduzcan los costos de la implementación del paquete tecnológico de producción agronómica.

## 2. Objetivo de machine learning

Teniendo en cuenta el contexto anteriormente descrito y la base de datos disponible en [Conjunto de datos de PlantVillage \(kaggle.com\)](https://www.kaggle.com/datasets/plantvillage/plantvillage) se plantea el siguiente objetivo:

“Desarrollar un modelo de machine learning basado en redes neuronales convolucionales (CNN) para la clasificación de enfermedades en plantas a partir de imágenes del área foliar de hojas de diferentes especies”

## 3. Dataset

**Tipo de datos:** Imágenes anotadas y etiquetadas con información que incluye la enfermedad o el problema específico que afecta la planta

**Tamaño:** El dataset de PlantVillage consta de 54,303 imágenes entre hojas sanas y enfermas divididas en 38 categorías por especie y enfermedad. El tamaño en disco es de 4GB. La distribución de las clases (38) por especie de planta y enfermedad se presenta en la siguiente tabla:

Plant	Disease	Files
Apple	healthy	1645
Apple	Cedar_apple_rust	275
Apple	Black_rot	621
Apple	Apple_scab	630
Blueberry	healthy	1502
Cherry_(including_sour)	Powdery_mildew	1052
Cherry_(including_sour)	healthy	854



**Facultad de Ingeniería**  
**Doctorado en Ingeniería Ambiental**  
**Fundamentos de Deep Learning**  
Tatiana Camila Puentes Escobar  
Diego Alejandro Grajales González

Corn_(maize)	Common_rust_	1192
Corn_(maize)	Cercospora_leaf_spot Gray_leaf_spot	513
Corn_(maize)	healthy	1162
Corn_(maize)	Northern_Leaf_Blight	985
Grape	Leaf_blight_(Isariopsis_Leaf_Spot)	1076
Grape	Black_rot	1180
Grape	Esca_(Black_Measles)	1383
Grape	healthy	423
Orange	Haunglongbing_(Citrus_greening)	5507
Peach	Bacterial_spot	2297
Peach	healthy	360
Pepper_bell	Bacterial_spot	997
Pepper_bell	healthy	1478
Potato	healthy	152
Potato	Early_blight	1000
Potato	Late_blight	1000
Raspberry	healthy	371
Soybean	healthy	5090
Squash	Powdery_mildew	1835
Strawberry	healthy	456
Strawberry	Leaf_scorch	1109
Tomato	Leaf_Mold	952
Tomato	healthy	1591
Tomato	Spider_mites Two-spotted_spider_mite	1676
Tomato	Early_blight	1000
Tomato	Septoria_leaf_spot	1771
Tomato	Target_Spot	1404
Tomato	Bacterial_spot	2127
Tomato	Tomato_Yellow_Leaf_Curl_Virus	5357
Tomato	Tomato_mosaic_virus	373
Tomato	Late_blight	1909

#### 4. Métricas de desempeño (de machine learning y negocio)

Las métricas de desempeño para la evaluación del modelo son:

- Accuracy: Se espera medir el modelo considerando qué tan bien clasifica las imágenes de entrada en las categorías deseadas. El indicador se basa en el número de predicciones, así:

$$Accuracy = \frac{\text{Número de predicciones correctas}}{\text{Número total de predicciones}}$$

- Análisis de error: Como complemento a la métrica de accuracy, se empleará la identificación de falsos positivos y falsos negativos, es decir, aquellos casos en que se predice una clase que no es correcta y casos en que no se predice una clase que se debería haber predicho.

Por otro lado, la métrica de negocio seleccionada corresponde a la reducción de costos, empleando un análisis de los costos que se podrían reducir dada la implementación del modelo para una hectárea de cultivo, asociados a la reducción de la pérdida en cosecha y la reducción de la inversión en plaguicidas.



## 5. Referencias y resultados previos

La tecnología de Deep Learning ha tomado protagonismo en todas las ramas del conocimiento. Su uso, en especial en agronomía y el área de trabajo abordado para el contexto anteriormente planteado ha llevado al desarrollo de aplicaciones que permiten la detección de enfermedades en tiempos cortos. En un estudio publicado por Lu et al. (2021) discutieron los principales problemas y las correspondientes soluciones de las CNN utilizadas para la clasificación de las enfermedades de las plantas. Los autores presentaron el proceso de los métodos de la CNN aplicados a la clasificación de las enfermedades de las plantas y resumieron los principios de Deep Learning relacionados en este tipo de clasificación.

Así mismo, Geetharamani & Arun Pandian (2019) proponen un nuevo modelo de identificación de las enfermedades foliares de las plantas basado en una red neuronal convolucional profunda (Deep CNN) entrenada con un conjunto de datos abierto con 39 clases diferentes de hojas de plantas e imágenes de fondo. Para aumentar el rendimiento del modelo, utilizaron técnicas de aumento de datos, como el cambio de imagen, la corrección gamma, la inyección de ruido, el aumento de color con PCA, la rotación y el escalado logrando una precisión de clasificación del 96,46%, superior a la de los enfoques tradicionales de aprendizaje automático.

Finalmente, en el artículo publicado a Falaschetti et al. (2022) se propone el uso de un detector de imágenes que utiliza una red neuronal convolucional (CNN) con recursos limitados e implementada en una plataforma de bajo coste y bajo consumo llamada OpenMV Cam H7 Plus. La red CNN se basa en dos conjuntos de datos específicos para la detección de enfermedades de las plantas: el conjunto de datos de la ESCA y el conjunto de datos aumentado por PlantVillage. Los resultados obtenidos demuestran que la implementación en el sistema de recursos restringidos elegido logra una alta precisión con un bajo costo de memoria y tiempo de inferencia para los conjuntos de datos aumentados por ESCA y PlantVillage, respectivamente.

## 6. Bibliografía

- Falaschetti, L., Manoni, L., Leo, D. Di, Pau, D., Tomaselli, V., & Turchetti, C. (2022). A CNN-based image detector for plant leaf diseases classification. *HardwareX*, 12.  
<https://doi.org/10.17605/OSF.IO/UCM8D>
- Geetharamani, G., & Arun Pandian, J. (2019). Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Computers and Electrical Engineering*, 76, 323–338.  
<https://doi.org/10.1016/j.compeleceng.2019.04.011>
- Lira-Saldivar, R. H., & Lara-Viveros, F. M. (2023). *Inteligencia artificial en agricultura - Uso de algoritmos, drones y biosensores*. Atena.
- Lu, J., Tan, L., & Jiang, H. (2021). Review on convolutional neural network (CNN) applied to plant leaf disease classification. In *Agriculture (Switzerland)* (Vol. 11, Issue 8). MDPI AG.  
<https://doi.org/10.3390/agriculture11080707>
- Parra, M., Rivera, P. A., Rodríguez, A., & Aguilar, O. E. (2012). *Acuerdo de competitividad para la cadena productiva de Pasifloras en Colombia*.