

# Machine learning en a detección de enfermedades en plantas

## Machine learning in plant disease detection

Andrés Calderón Romero,<sup>1</sup>, Harold Hurtado Cortes <sup>2</sup>

Para citar: Calderón, A., Hurtado, H. (2019), Arquitectura basada en micro-servicios para aplicaciones web. TIA, 7 (2), pp. 55-61.

### Artículo de investigación

Fecha de recepción:  
**2019-12-09**  
Fecha de recepción:  
**2019-12-26**

ISSN: 2344-8288  
Vol. 7 No. 2  
Julio- diciembre 2019  
Bogotá-Colombia

### Resumen:

El rápido crecimiento de la población ha generado una alta demanda de alimentos, haciendo difícil cumplir con la demanda, debido a diversos factores y a la proliferación de enfermedades en los cultivos. Puesto que la identificación de enfermedades se realiza de manera manual y las técnicas utilizadas son primarias y el personal especializado requiere de mucho tiempo para realizar el diagnóstico respectivo.

Este artículo, busca mostrar la relevancia a nivel tecnológico del uso del machine learning en la identificación de enfermedades en plantas. Mediante la revisión literaria de varios trabajos y las diversas técnicas de reconocimiento, que permitirán identificar y diagnosticar las enfermedades en las plantas y cultivos, lo que permitirá tomar decisiones y acciones para evitar la proliferación de la enfermedad.

**Palabras Clave:** Plantas, Aprendizaje Automático, Algoritmo, Redes Neuronales, Enfermedades.

### Abstract:

The Rapid population growth has generated a high demand for food, making it difficult to meet the demand, due to various factors and the proliferation of diseases in crops. Since the identification of diseases is done manually and the techniques used are primary and specialized personnel require a lot of time to make the respective diagnosis. This article seeks to show the relevance at a technological level of the use of machine learning in the identification of plant diseases. Through the literature review of several works and the various recognition techniques, which will allow to identify and diagnose diseases in plants and crops, which will allow to take decisions and actions to prevent the proliferation of the disease

**Key Word:** Plants, Machine Learning, Algorithm, Neuronal Networks, diseases

---

<sup>1</sup> Ingeniero Electrónico, calderonacr@gmail.com, Bogotá, Colombia

<sup>2</sup> Ingeniero de Sistemas, harolddavidhurtado@gmail.com, Bogotá, Colombia,

## I. Introducción

La agricultura a través de la historia ha cumplido un rol muy importante en el desarrollo económico del mundo. Por ello, con el transcurrir del tiempo se han desarrollado infinidad de técnicas para mejorar el desarrollo agrícola permitiendo aumentar la productividad de los cultivos y así cumplir con la alta demanda de alimentos a nivel mundial, generada por el rápido crecimiento de la población y las zonas urbanas.

Sin embargo, sin importar los grandes avances tecnológicos de hoy en día, no es posible conseguir un rendimiento completo en todos los cultivos debido a las diferentes enfermedades ocasionadas por bacterias, virus y hongos [1].

Entre las diferentes estrategias para el tratamiento de enfermedades se encuentra la erradicación manual y el uso de pesticidas en las plantas afectadas. Para la erradicación manual, es necesario contar con la ayuda de expertos en el campo de la agricultura, ya que gracias a la experiencia reconocen los patrones de las enfermedades más fácil que una persona común. Sin embargo, este método requiere de mucho tiempo para realizar un muestreo en plantaciones de gran tamaño y de personal especializado [2].

Sin lugar a dudas, las pestes y enfermedades son la razón principal por la que una plantación no alcance su madurez en los tiempos normales de crecimiento, la reducción de la productividad por las enfermedades entre un 10 – 95% [9] o en el peor de los casos la pérdida total de las plantas.

Es por ello, que la detección de enfermedades puede ser más fácil con técnicas como el machine learning en comparación con los métodos manuales, a través del procesamiento de imágenes que permitan detectar en tallos, ramas, hojas y

flores características propias de las diferentes enfermedades que pueden afectar a una planta y con ello generar una red de entrenamiento con una gran variedad de muestras para hacer más precisa la detección de enfermedades.

Tanto algoritmos de machine learning como de deep learning han sido empleados para aumentar la tasa de reconocimiento, fiabilidad y precisión en el proceso de detección de enfermedades en plantas [3].

Entre las técnicas que han tomado fuerza en la detección de enfermedades implementando machine learning se pueden encontrar bosque aleatorio (*Random Forest*), redes neuronales artificiales, máquinas de vectores de soportes (SVM), lógica difusa, redes neuronales artificiales y convolucionales, etc. [3].

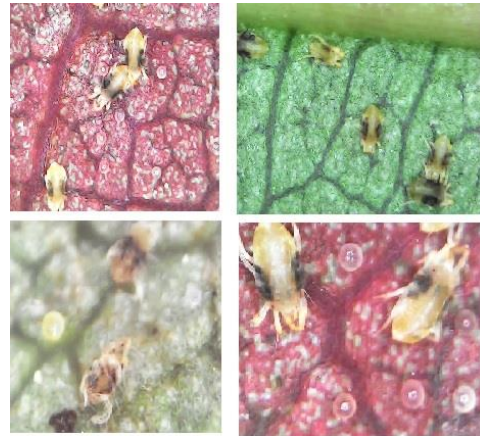


Figura 1. Enfermedades en hojas de plantas  
(Fuente: autoría Propia)

De la figura 1, se pueden observar distintas enfermedades que se presentan en las hojas de las plantas y que gracias al machine learning y sus distintas técnicas se podrían detectar de una forma más rápida y precisa que lo realizado con técnicas manuales.

A partir de esto, se plantea como objetivo mostrar la relevancia que está tomando el machine learning en la identificación de enfermedades en plantas a través de una revisión literaria de varios trabajos con diversas técnicas de machine learning para el reconocimiento de imágenes que permiten una detección temprana de enfermedades.

Teniendo claro el objetivo del trabajo, cabe aclarar que el alcance de este trabajo está dado como un artículo de revisión de tema, en donde se realizará una revisión literaria de la temática del machine learning en la detección de enfermedades en plantas y no se plantea en este trabajo hacer el desarrollo de algunas de las técnicas o algoritmos que se pueden ver reflejadas en los trabajos a los cuales se les está haciendo la revisión literaria.

## II. Marco teórico

### A. Machine Learning

Es un subcampo de la inteligencia artificial que permite crear sistemas que aprenden automáticamente. Sin embargo, hay que tener en cuenta el significado de aprender en este contexto, lo cual quiere decir que ese aprendizaje se da gracias a la identificación de patrones gracias a un análisis de datos que permitirán crear un modelo o algoritmo y así generar predicciones de muestras de datos que nunca se han visto como lo manifiestan [4], [5], [6], [7].

Estos modelos o algoritmos pueden mejorar su aprendizaje a través de la experiencia, es decir, utilizando nuevos datos de entrada para alimentar el modelo y de esa forma se podrá redefinir el modelo para que pueda hacer las predicciones [5].

En machine learning existe una variedad de algoritmos que se pueden emplear y a partir del resultado deseado se escoge que algoritmo a utilizar [6]. Sin embargo, estos algoritmos se pueden clasificar en 2 grupo: aprendizaje supervisado y aprendizaje no supervisado.

- **Aprendizaje supervisado**

Es el trabajo con un grupo de datos que ya se encuentran etiquetados. En donde para ejemplo se tiene una entrada llamada característica y una salida denominada etiqueta. A partir de muchos ejemplos de este estilo el algoritmo aprende a asociar las entradas con las salidas de manera que pueda hacer combinaciones para realizar predicciones [4], [5].

Además, existen 2 tipos de aprendizaje supervisado: La clasificación en donde el algoritmo encuentra diferentes patrones y a partir de esto clasifica la información en varios grupos. El otro tipo de aprendizaje supervisado es la regresión que permite a partir de un número de entradas realizar la predicción de la salida [8].

- **Aprendizaje no supervisado**

En este tipo de aprendizaje los datos de entrenamiento no tienen etiquetas, además tampoco tienen una señal de supervisión o profesor que les indique lo que se debe aprender. Por ello, para lograr un resultado más preciso es necesario realizar una mayor cantidad de intentos para extraer información de un grupo de datos que no requiera la intervención del ser humano para realizarlo [8].

A continuación, se realizará una breve descripción de algunos de los algoritmos utilizados en machine learning que se observaran en la revisión literaria de trabajos en la siguiente sección.

#### B. Bosque aleatorio (RF)

Es un algoritmo de aprendizaje supervisado que en su función crea un bosque y de manera aleatoria, es decir, toma varios árboles de decisión que combina para obtener una predicción más precisa y estable [8]. Dada su simplicidad y que se puede usar tanto para tareas de clasificación y regresión se ha convertido en uno de los algoritmos más poderosos disponibles hoy en día.

#### C. Máquina de vectores de soporte (SVM)

Son un conjunto de algoritmos de supervisión que son utilizados para tareas de clasificación y regresión [2]. Estos son muy utilizados en aplicaciones como el procesamiento de lenguaje natural, reconocimiento de imágenes y la visión artificial [17].

Entre las ventajas que se pueden encontrar en este algoritmo se encuentran: su bajo error de generalización, costo computacional bajo y sus resultados son fáciles de interpretar. Por otro lado, sus desventajas son: sensible a los parámetros de ajuste y elección del núcleo y la clasificación binaria la maneja únicamente de manera nativa. Una de las implementaciones más comunes del SVM es el algoritmo de secuencia mínima de optimización (SMP) [7].

#### D. Redes Neuronales artificiales

Son modelos simples basados en el la estructura funcional del cerebro humano [15], además también se pueden pensar en ellas como en un modelo computacional que quiere asemejarse al cerebro [6]. Estas redes funcionan haciendo el uso simultaneo de una especie de unidades de procesamiento similares a las neuronas abstractas que se encuentran entrelazadas y de esta manera pueden transmitir la información [16].

La red se forma con la conexión de las neuronas entre sí, es decir, la salida de una neurona está conectada con la entrada de otra y así se replicarán estas conexiones con varias neuronas.

Entre las aplicaciones que se pueden encontrar hoy en día con la implementación de redes neuronales se tienen: Robótica, Monitoreo médico, Administración de centros de datos, reconocimiento de imágenes y voz entre otras [4].

#### E. Redes neuronales Convolucionales

Son un tipo de redes neuronales especializadas para el procesamiento de datos con una topología de malla [5]. Estas redes son útiles para localizar patrones en imágenes, objetos y escenas, en donde, aprenden de los datos de la imagen para la clasificación de las imágenes y con ello se elimina la necesidad de la extracción manual de cada una de las características que se están buscando [18].

Una de las características más importante de este tipo de redes, es que se pueden entrenar a para nuevas tareas de reconocimiento a partir de redes ya existentes.

Ya que las redes neuronales convolucionales están inspiradas en la visión humana, las capas de estas redes tienen neuronas dispuestas en tres dimensiones un ancho una altura y una profundidad [15].

### III. Revisión literaria

En esta sección se dará la revisión literaria de diversos trabajos enfocados hacia la temática del machine learning en la identificación de enfermedades en plantas, en donde, utilizan diversos enfoques para la identificación de enfermedades con diversas técnicas.

En [1], se muestra el rol que juega la agricultura en el crecimiento de la población dado que deben cumplir con la demanda de comida para esta. Sin embargo, hacen énfasis en las enfermedades que pueden sufrir las plantas a partir de bacterias, virus y hongos. Por ello, es importante para ellos plantear estrategias para la prevención de enfermedades a través del uso del enfoque del machine learning para la detección de enfermedades.

A partir de ello, muestras las etapas generales para la detección de enfermedades y un estudio comparativo de las técnicas utilizadas con enfoque de machine learning para la clasificación de enfermedades. Las técnicas empleadas en el estudio son:

- El SVM (Maquina de vectores de soporte)
- ANN (Redes neuronales artificiales)
- el KNN (Clasificador del vecino más cercano)
- FUZZY Classifier (Clasificador Difuso)
- CNN (Redes Neuronales Convolucionales)

Finalmente, el trabajo muestra que las Redes Neuronales Convolucionales son la técnica que muestran la mayor precisión para la detección del mayor número de enfermedades de varios cultivos como ser evidencia [2], donde se realiza el planteamiento del problema hacia las enfermedades en las plantas y las consecuencias de dichas enfermedades que pueden reducir la productividad de un cultivo en un 10 % - 95%. Po consiguiente a partir del estudio en la literatura se evidencia QUE las técnicas más empleadas son

- El árbol de decisión (Decision Tree Learning)
- Aprendizaje de reglas de asociación (Association Rule Learning)
- Redes neuronales artificiales (Artificial Neuronal Networks)
- Máquinas de vectores de soporte (Support Vector Machine)

A partir de este análisis identifican las ventajas y desventajas de cada una de las técnicas con un enfoque al machine learning, en donde, concluyen que la técnica de máquinas de vectores de soporte es la predilecta, puesto que se basa en un algoritmo que proporciona una mejor clasificación y predicción de resultados, siendo muy útil en el momento de identificar las enfermedades en las plantas.

En el trabajo presentado por [3], se propone el uso del algoritmo de Bosque aleatorio (Random Forest) para identificar entre muestras de hojas sanas y limpias. A partir de ello, plantean varias fases de implementación del algoritmo a través de la creación de conjuntos de datos, extracción de características, el entrenamiento del

clasificador y la clasificación.

De esta forma los conjuntos de datos se entrenan con el algoritmo de bosque aleatorio para identificar las hojas buenas y malas y la extracción de características de una imagen se realiza mediante el histograma de un gradiente orientado (HOG). Con esto, en el trabajo se llega a la conclusión que esta técnica del bosque aleatorio tiene una alta efectividad con una cantidad de muestras pequeña comparada con otras técnicas de machine learning (SVM, KNN, Logic Regression, entre otros).

En [10], este trabajo describe un algoritmo con enfoque en machine learning para la detección de enfermedades en una gran variedad de plantas. Además, la precisión del algoritmo alcanza un 93% entrenado con más de 80000 imágenes con características de mucho ruido, con diferentes características y diversas coberturas en sus hojas. Lo interesante de este trabajo radica en que puede entrenarse por sí mismo el algoritmo, por lo cual a medida que se use, la precisión de este crecerá. Finalmente, el algoritmo utilizado es CNN (Redes Neuronales Artificiales) y puede ejecutarse en varias plataformas, entre ellas teléfonos móviles para que personal inexperto puedan controlar las enfermedades de manera efectiva.

En [11], se realizó una investigación de las enfermedades en plantas de maíz apoyados en el procesamiento de una gran cantidad de imágenes que gracias a unas características y patrones permitieron la identificación de las enfermedades. Características para detectar color como

- RBG
- la transformación de funciones locales invariantes en escala (SIFT)
- funciones robustas aceleradas (SURF)
- Orientado y girado BRIEF
- detector de objetos como histogramas de gradientes orientados (HOG).

Donde a través de algoritmos de machine learning como SVM (Maquinas de vectores de soporte), DT (Arboles de decisión), RF (Bosque aleatorio) y NB (Naive Bayes), se evaluó el rendimiento de cada una de ellas, en donde se encontró que dentro de los resultados obtenidos de los experimentos las características de color como RBG con SVM logran el mejor rendimiento y precisión para la mayoría de clasificaciones que se valoraron.

En [12], se propone un trabajo para el diseño de un sistema de detección de enfermedades en plantas, dado que entienden que existe una fuente problemática en la disponibilidad o falta de información relevante del conocimiento agrícola oportuno para el tratamiento de estas enfermedades por parte de los pequeños agricultores. Por ello, el trabajo realizado en este artículo propone realizar un procesamiento de imágenes a través de técnicas como las redes neuronales convolucionales con un enfoque al machine learning y el aprendizaje supervisado.

El trabajo del algoritmo se embeberá en una aplicación móvil para dispositivos Android, la cual deberá realizar el procesamiento de la imagen tomada tanto en línea (realizando el procesamiento a través de un algoritmo clasificador desarrollado en TensorFlow) y a continuación, los resultados se almacenarán en una base de datos para mantener un

repositorio de imágenes. Además, dependiendo de los resultados, se le dará al usuario algunos consejos para el manejo de la enfermedad y cómo prevenirla para un futuro. Finalmente, proponen que la aplicación pueda utilizarse también fuera de línea.

#### **IV. Conclusiones**

Sin lugar a dudas el machine learning a través de la inteligencia artificial ha tomado una gran relevancia en varios campos de la ciencia gracias al reconocimiento de imágenes que permiten dar un diagnóstico en un porcentaje muy elevado puntualmente en nuestro caso para enfermedades de plantas.

Al mismo tiempo esta herramienta a su vez proporciona métodos más rápidos de diagnóstico gracias a que las redes neuronales que se entrenan con un gran conjunto y variedad de datos permiten casi al instante hacer un análisis y clasificación de la imagen que son muy difíciles de manejar para los seres humanos inclusive para especialistas en cada uno de los sectores en donde se utilice el reconocimiento de imágenes con machine learning.

Para lograr una mayor precisión en la detección de enfermedades en plantas es necesario incrementar la cantidad de muestras con las que se entrenara la red neuronal teniendo en cuenta que estas sean sobre plantas enfermas bajo condiciones diferentes. Esto permitirá llegar a evaluar la capacidad de llegar a resultados similares bajo diferentes condiciones generando una mayor fiabilidad y precisión.

Dependiendo del enfoque de estudio y la aplicación que se esté realizando existirá un algoritmo de machine learning que se acomode más a las condiciones del estudio y con ello se pueda lograr un rendimiento y precisión mayor para la identificación de las enfermedades. Por ello, cabe aclarar que cada planta dada su naturaleza tiene enfermedades que en otras no se pueden presentar debido a las diversas ubicaciones geográficas y condiciones de cada una de ellas.

Además, se deben tener en cuenta los colores también en las plantas a la hora de seleccionar el algoritmo ya que es un factor fundamental en el momento de utilizar las técnicas de entrenamiento para los algoritmos.

#### **V. Recomendaciones**

Con respecto a la revisión literaria realizada se puede plantear el desarrollo de un prototipo de aplicación móvil para la detección, clasificación de enfermedades en plantas en cultivos específicos, ya que esto permitirá tener un desarrollo más eficiente y preciso del modelo a entrenar al no tener que realizar el entrenamiento para una gran cantidad de plantas, puesto que para ello sería necesario tener millones de muestras para que el modelo pueda cumplir con un reconocimiento preciso.

Conjuntamente, se puede plantear ofrecer tratamientos específicos para las enfermedades detectadas y con ello digitalizar el conocimiento de los especialistas para ofrecer tratamientos precisos partiendo de la experiencia de ellos en el sector. Esto tendrá también un valor agregado realizando esta gestión del conocimiento que no es fácil de obtener por

la naturaleza empírica en su mayoría para la detección de las enfermedades.

A partir de esto, se puede plantear si la aplicación realizara el reconocimiento con el modelo desde la aplicación móvil o si esta se conectara con un API en un servidor en la nube para hacer todo el proceso. Esto permite ofrecer la creación un repositorio de imágenes para rentrenar el modelo desarrollado y así aumentar la eficiencia y precisión de este al tener muestras en diferentes ambientes y condiciones.

Con este repositorio, se puede ofrecer una gran fuente de información para ayudar a la comunidad agrícola y científica con la identificación de enfermedades de las cuales no se tiene un amplio material. Además, a este repositorio se le puede agregar la funcionalidad para ofrecer los posibles tratamientos a las enfermedades conocidas de acuerdo a las recopilaciones de información con especialistas para cada una de las plantas que se puedan tener en el repositorio creado.

## VI. Referencias

- [1] S. U, V. Nagaveni and B. K. Raghavendra,(2019)."A Review on Machine Learning Classification Techniques for Plant Disease Detection," 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, pp. 281-284.
- [2] J. Shirahatti, R. Patil and P. Akulwar, (2018)."A Survey Paper on Plant Disease Identification Using Machine Learning Approach," 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, pp. 1171-1174.
- [3] S. R. Maniyath et al., (2018). "Plant Disease Detection Using Machine Learning," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, pp. 41-45.
- [4] Bell, Machine Learning - Hands-On for Developers and Technical Professionals. Wiley, 2014.
- [5] I. Goodfellow, y Bengio, and A. Courville, Deep Learning. (2016),(Adaptive Computation and Machine Learning series. Boston: The MIT Press,
- [6] S. Shalev-Shwartz and S. Ben-David, (2014.). Understanding Machine Learning: From Theory to Algorithms, 1st Edition. Cambridge University Press,
- [7] P. Harrington, (2012). Machine learning in action, Second Edion. New York: Manning Publications.
- [8] A. Géron, (2017). Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems, 1st Editio. Sebastopol, CA: O'Reilly Media,
- [9] F Ahmed, HA AI-Mamun, (2012). ASMH Bari, E Hossain, "Classification of crops and weeds from digital images: A SVM approach", Elsevier.
- [10] P. Sharma, Y. P. S. Berwal and W. Ghai, (2018). "KrishiMitr (Farmer's Friend): Using Machine Learning to Identify Diseases in Plants," IEEE International Conference on Internet of Things and Intelligence System (IOTAIS), Bali, 2018, pp. 29-34.
- [11] B. S. Kusumo, A. Heryana, O. Mahendra and H. F. Pardede, (2018)."Machine Learning-based for Automatic Detection of Corn-Plant Diseases Using Image Processing," International Conference on Computer, Control, Informatics and its Applications (IC3INA), Tangerang, Indonesia, pp. 93-97.
- [12] Tlhobogang and M. Wannous, (2018) "Design of plant disease detection Systems: A transfer learning approach work in progress," IEEE