

## Estimación de *Podosphaera xanthii* en pepino: técnicas de aprendizaje automático con imágenes digitales

Berenice Alonso-Barrera<sup>1</sup>

Francisco Marcelo Lara-Viveros<sup>1,§</sup>

Audberto Reyes-Rosas<sup>1</sup>

1 Centro de Investigación en Química Aplicada. Blvd. Enrique Reyna Herosillo núm. 140, Saltillo, Coahuila, México. CP. 25294. Tel. 844 4389830.(berenice.alonso.ps@ciqa.edu.mx; audberto.reyes@ciqa.edu.mx).

Autor para correspondencia: francisco.lara@ciqa.edu.mx.

### Resumen

Los hongos fitopatógenos representan una amenaza considerable para el cultivo de cucurbitáceas, por lo que la detección temprana y cuantificación precisa de enfermedades resulta esencial para reducir pérdidas en la producción. En este estudio se desarrolló una metodología para estimar cuantitativamente el daño provocado por *Podosphaera xanthii* en hojas de pepino, mediante el uso de imágenes digitales y técnicas de aprendizaje automático. Se emplearon redes neuronales convolucionales para clasificar visualmente el grado de severidad en seis categorías predefinidas, utilizando secciones de hojas con síntomas evidentes del hongo. Adicionalmente, se entrenaron y compararon cuatro algoritmos de clasificación supervisada: K-NN, árboles de decisión, bosques aleatorios y redes neuronales. El modelo que obtuvo el mejor desempeño fue el de bosques aleatorios, con una precisión del 90%, mientras que K-NN alcanzó el valor más bajo (79%). Estos resultados posicionan al modelo como una herramienta útil para el monitoreo automatizado de enfermedades en campo, facilitando la toma de decisiones fitosanitarias. Además, la metodología ofrece una base sólida para investigadores interesados en diseñar e implementar sistemas automáticos de clasificación de enfermedades en plantas, brindando información clara sobre el rendimiento de distintas arquitecturas de clasificación. El algoritmo desarrollado en R permite adaptar y escalar esta solución a diferentes condiciones de cultivo y tipos de enfermedades foliares.

### Palabras clave:

algoritmo, cenicilla, enfermedad.



## Introducción

El mildiu polvoriento, causado por hongos como *Golovinomyces cichoracearum*, *Erysiphe cichoracearum*, *Sphaerotheca fuliginea*, *Podosphaera xanthii* y *Podosphaera fusca* (Mohamed *et al.*, 1995; Morejón *et al.*, 2010), es una enfermedad de amplia distribución que compromete seriamente la calidad y el rendimiento del cultivo de pepino, generando pérdidas económicas significativas (Sun *et al.*, 2022). Esta fitopatología se manifiesta inicialmente como una coloración blanquecina que progresa a manchas amarillo-cremosas, afectando principalmente las hojas (Rocha *et al.*, 2023).

Dado su impacto, la detección temprana y precisa es fundamental para una gestión oportuna; lo cual, contribuye directamente a la seguridad alimentaria (Kaushik *et al.*, 2023). En este contexto la inteligencia artificial (IA) ofrece herramientas valiosas para mejorar los sistemas agrícolas y la economía de los agricultores. Las técnicas comunes de clasificación supervisada incluyen regresión logística, análisis discriminante, K-NN, redes neuronales, árboles de decisión y bosques aleatorios (Zapata *et al.*, 2014; Paymode y Malode, 2022).

La estimación automatizada de enfermedades en plantas agiliza el monitoreo en grandes cultivos y permite la detección temprana de síntomas. Los diferentes algoritmos de aprendizaje automático generan resultados variables, por lo que es crucial identificar el más adecuado para cada problema específico. La técnica K-NN se fundamenta en la proximidad de objetos similares en el espacio de características (Zhao y Yang, 2023).

Otro método utilizado en la actualidad es el de árboles de decisión, basado en observaciones y lógica, que representan y categorizan condiciones sucesivas para resolver problemas (Ramos *et al.*, 2023). Asimismo, Paccirett *et al.* (2020) refieren que el método de clasificación bosques aleatorios (RF) es un modelo de regresión que utiliza muestreo para construir múltiples árboles de regresión y los ensambla para lograr un modelo predictivo.

Cabe mencionar, que entre los modelos de clasificación más utilizados se encontró la clasificación de red neuronal (CNN) son modelos de clasificación computacional que ofrece soluciones y validación de secuencias en reconocimiento de patrones como una extensión de métodos estadísticos clásicos (Hassoun y Threshold, 1995). Este método se adaptó a las exigencias del entorno, dado que puede combinar técnicas que procesan la información de manera paralela (Figueredo y Ballesteros, 2016).

El objetivo fue desarrollar una herramienta para la estimación automatizada del daño foliar causado por *P. xanthii* en hojas de pepino, utilizando imágenes digitales y técnicas de aprendizaje automático. Se propuso una metodología basada en redes neuronales convolucionales (CNN), árboles de decisión, bosques aleatorios (RF) y K-NN, con el fin de comparar su desempeño en la clasificación de la severidad de la enfermedad.

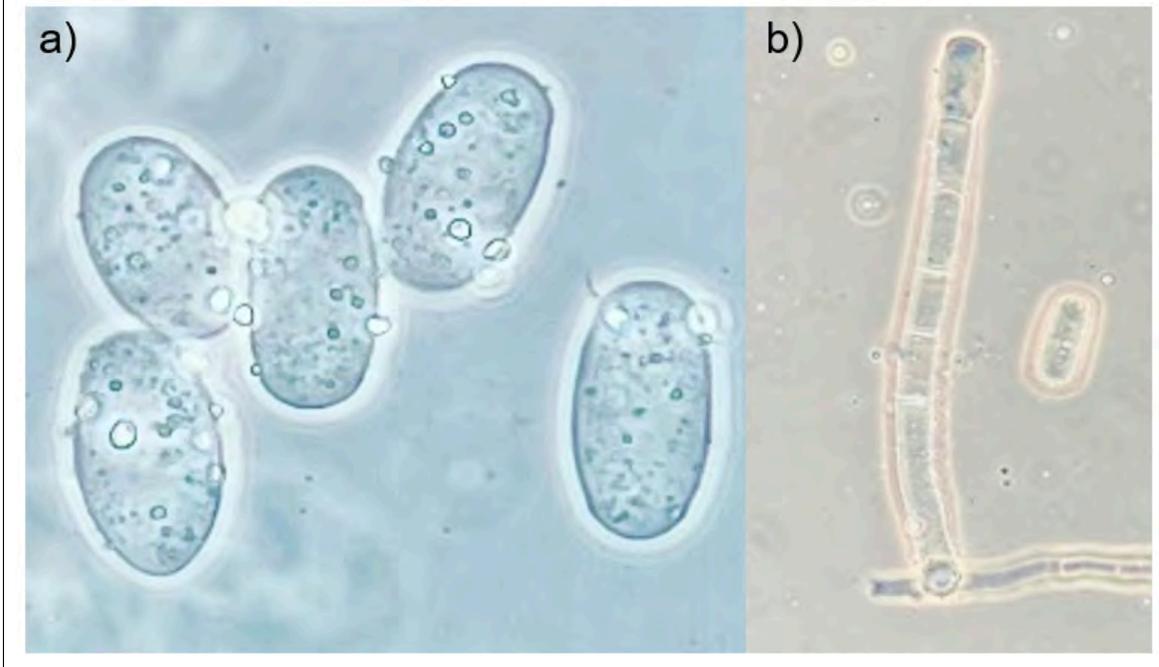
## Materiales y métodos

### Material vegetal

El material vegetal se obtuvo de una parcela comercial de pepino francés con una antigüedad de cinco años. En esta parcela se ha cultivado de manera exclusiva una variedad partenocárpica y se ha observado la manifestación de síntomas asociados a *Podosphaera xanthii* en todos los ciclos productivos.

### Identificación del microorganismo

La identificación del microorganismo se llevó a cabo mediante la preparación de muestras en portaobjetos a partir de las lesiones foliares. La morfología observada se contrastó con las descripciones publicadas en el manual taxonómico de *Erysiphaceae* (Braun y Cook, 2012) y con la caracterización proporcionada por Cipriano y González (2022) (Figura 1).

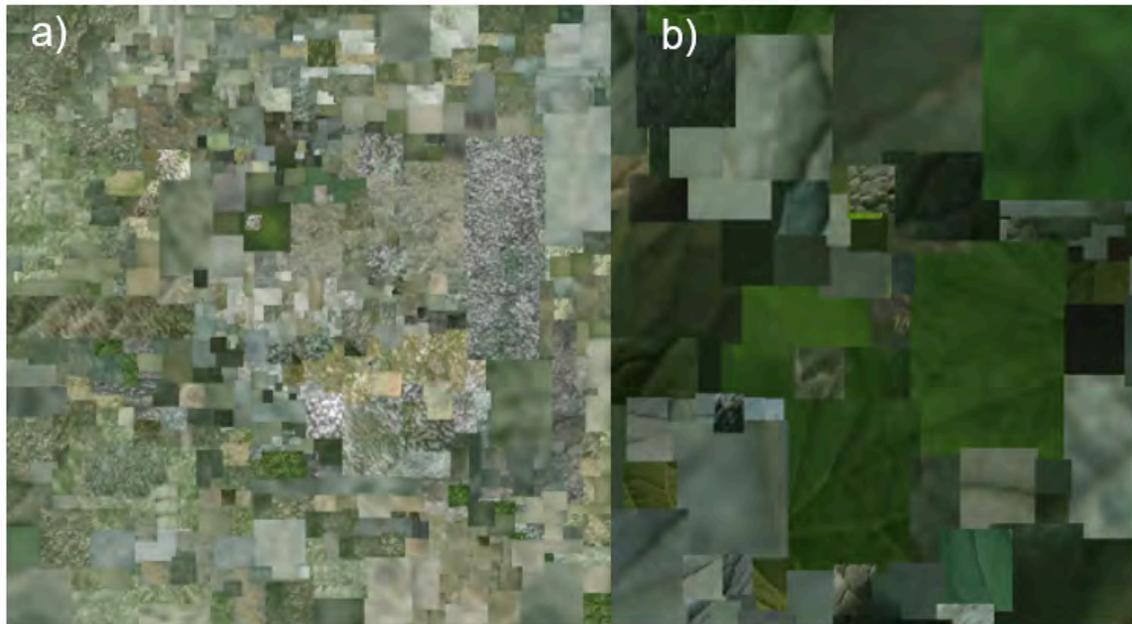
Figura 1. a) conidios ovoides de *P. xanthii* vista a microscopio y b) conidios en cadena de *P. xanthii*.

### Datos de entrenamiento

Para el análisis, se tomaron imágenes al azar y se extrajeron secciones de hojas que exhibían síntomas evidentes del hongo. Estos segmentos se unificaron para crear una imagen de 5 580 000 píxeles, que capturaba las características distintivas de las hojas afectadas. De manera análoga, se realizó el mismo proceso para obtener una imagen de control, recortando áreas de hojas sin indicios visibles de la enfermedad (Figura 2).



Figura 2. Conjunto de imágenes destinadas al entrenamiento. a) imagen con síntomas del hongo y b) imagen sin síntomas del hongo.



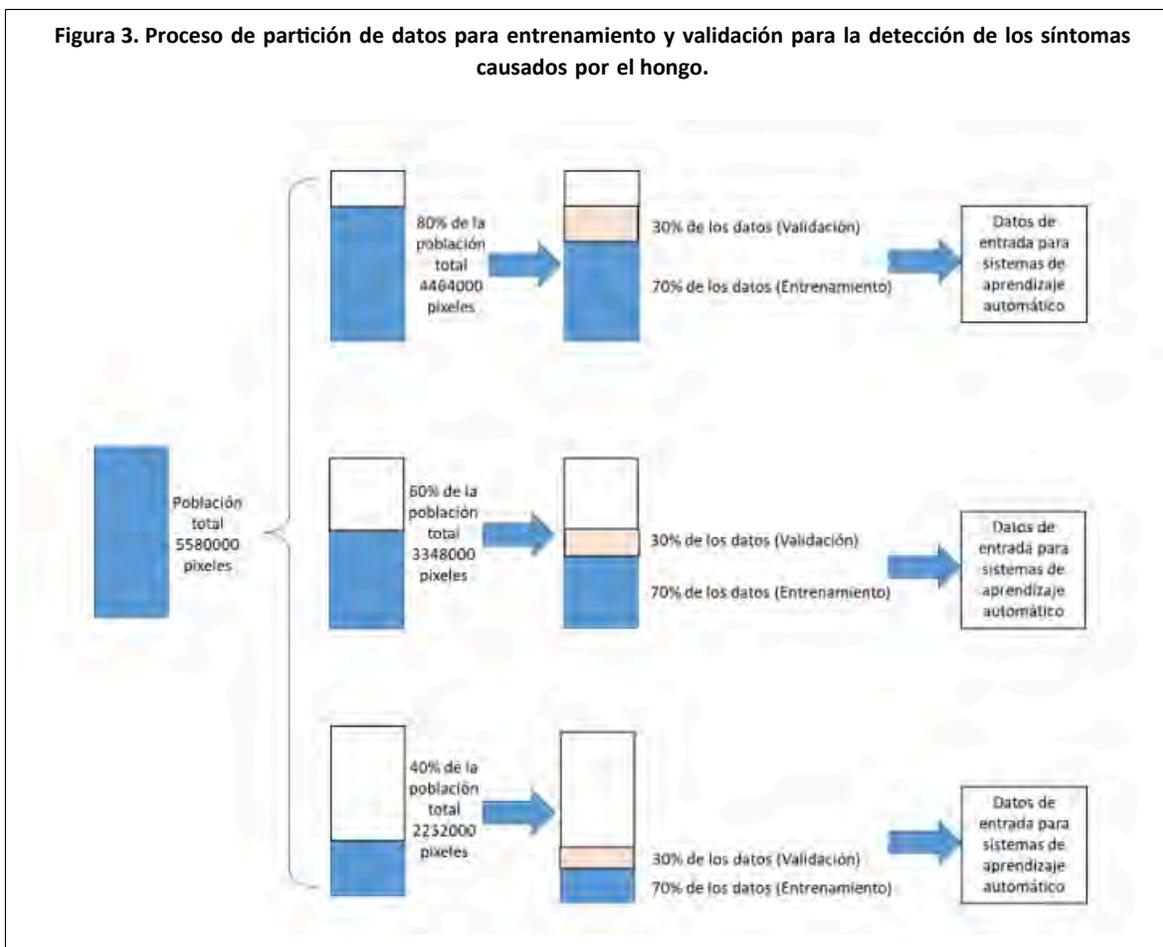
Con el fin de evitar un sesgo en el entrenamiento de los modelos de clasificación, se equilibró el conjunto de datos seleccionando un número equivalente de píxeles de ambas clases (sanos y enfermos). Este balance fue realizado mediante muestreo aleatorio uniforme, asegurando que cada clase estuviera representada por la misma cantidad de píxeles en el conjunto de entrenamiento.

Procesamiento de las imágenes muestra. Con el objetivo de entrenar y validar los modelos de aprendizaje automático, las imágenes segmentadas fueron agrupadas en tres conjuntos de datos, cada uno con distinta cantidad de píxeles: 4 464 000; 3 348 000 y 2 232 000. A cada conjunto se le aplicó un esquema de partición estratificada, donde el 70% de los píxeles se destinó al entrenamiento y el 30% restante a la validación y prueba.

Esta proporción aseguró una distribución equilibrada que permitió al modelo generalizar sin sobre ajustarse a los datos de entrada. Como se muestra en la Figura 3, el total de píxeles utilizados exclusivamente para el entrenamiento fue de 7 030 800, distribuidos de la siguiente manera: 3 124 800 píxeles del primer conjunto, 2 343 600 del segundo y 1 562 400 del tercero.



Figura 3. Proceso de partición de datos para entrenamiento y validación para la detección de los síntomas causados por el hongo.



Entrenamiento de sistemas de aprendizaje automático. Se utilizaron datos de entrenamiento para generar modelos de aprendizaje automático (K-NN, árboles de decisión, bosques aleatorios y redes neuronales) con el fin de clasificar píxeles como ‘sanos’ o ‘enfermos’. Para el entrenamiento, se consideraron hiperparámetros más comunes reportados en la literatura (Cuadro 1). Los datos de entrada para los modelos fueron los valores de los canales RGB de cada píxel.

Cuadro 1. Hiperparámetros reportados en la literatura para diferentes técnicas de aprendizaje automático.

Modelo de clasificación	Hiperparámetro	Referencia
Mejor vecino K-NN	Número de vecinos (K)	Zhang <i>et al.</i> (2019)
Árbol de decisión	La profundidad Número de observaciones en cada nodo	Demirović y Stuckey, (2021)
Bosques aleatorios (RF)	Numero de arboles	Benali <i>et al.</i> (2019)
Redes neuronales (CNN)	Profundidad de capas Número de capas Función de activación	Ma <i>et al.</i> (2018)

Tras ajustar los modelos de aprendizaje automático, se generó una matriz de confusión (Cuadro 2).



**Cuadro 2. Matriz de confusión utilizada para calcular las métricas de los modelos de aprendizaje automático.**

Clase observada	Clase estimada por el modelo	
	Sana	Enferma
Sana	Verdadero positivo (VP)	Falso positivo (FN)
Enferma	Falso negativo (FP)	Verdadero negativo (VN)

Con los datos anteriores se calculó la exactitud mediante la siguiente fórmula:

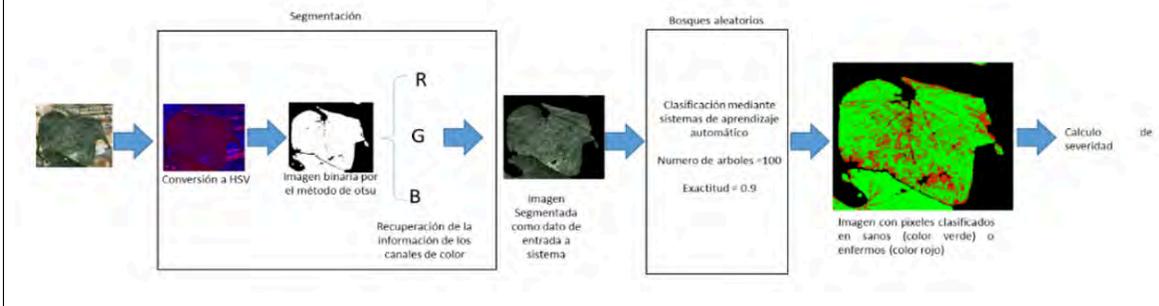
$$\text{Exactitud} = \frac{VP+VN}{VP+VN+FP+FN}$$

En todos los casos se reportó la exactitud calculada únicamente para los datos de validación.

## Procesamiento de las imágenes para el cálculo de la severidad

Para llevar a cabo la segmentación, las imágenes fueron transformadas al espacio de color HSV. Posteriormente, se aplicó el método de umbralización (Otsu, 1978), para generar una imagen binaria, la cual fue multiplicada por cada uno de los canales RGB de la imagen original. Este procedimiento permitió obtener una imagen segmentada, que posteriormente fue utilizada como entrada para el sistema de clasificación (Figura 4).

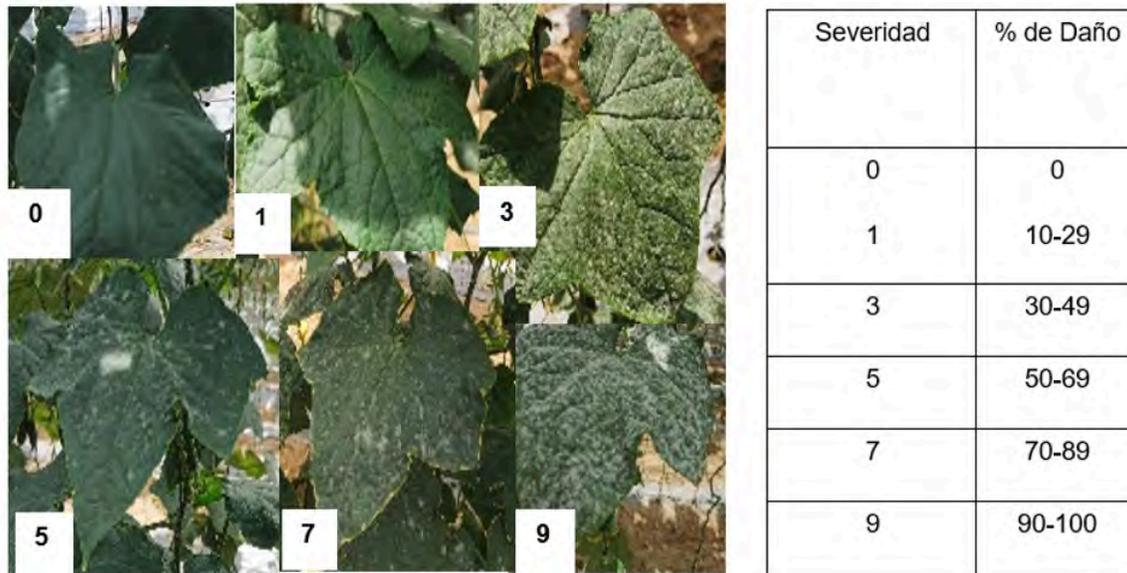
**Figura 4. Metodología de preprocesamiento y procesamiento de hojas de pepino para estimar la severidad del hongo *P. xanthii*.**



Al mismo tiempo cada imagen fue etiquetada con una de las seis clases de severidad visual previamente definidas, Mohamed *et al.* (1995) y modificada por Hernández *et al.* (2007), lo cual correspondió a distintos rangos de porcentaje de daño foliar (0%, 10-29%, 30-49%, 50-69%, 70-89%, 90-100%) (Figura 5). Se seleccionaron 50 imágenes por clase, resultó en un conjunto de datos balanceado para el entrenamiento y evaluación.



Figura 5. Escala de severidad utilizada para clasificar visualmente el daño de *P. xanthii* en hojas de pepino.

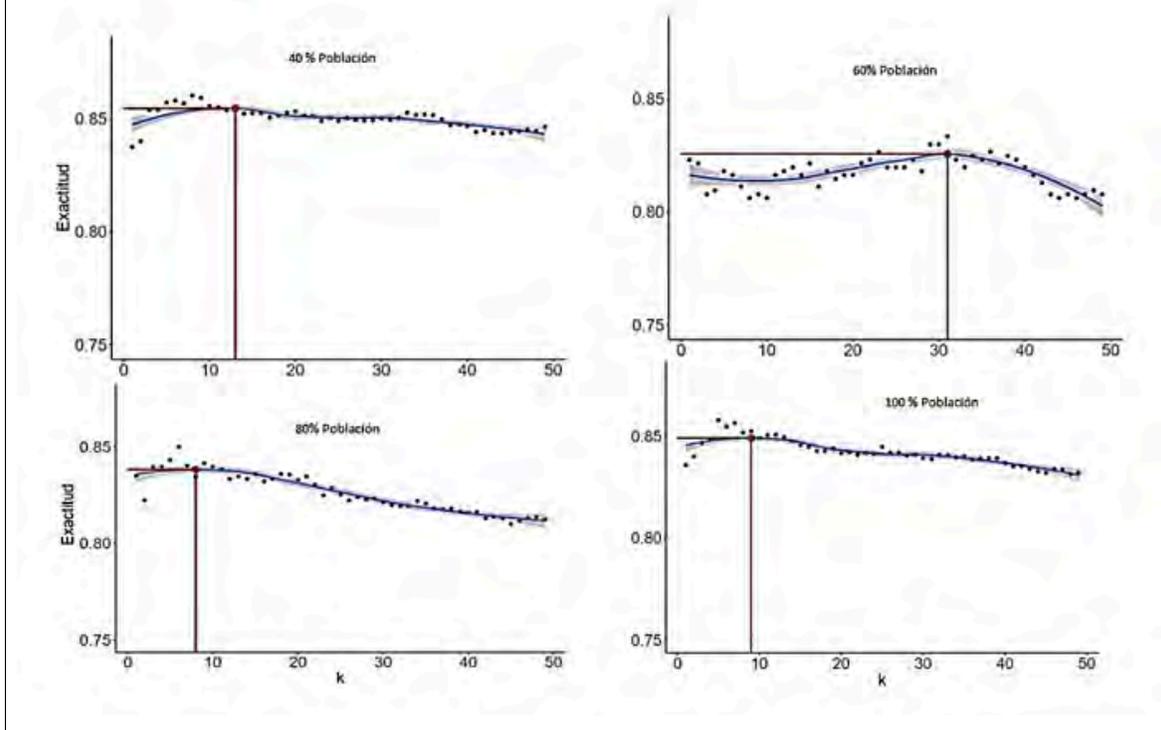


## Resultados y discusión

K-mejor vecino (K-NN). La (Figura 6) mostró que la exactitud máxima de 0.85, con una ligera tendencia a disminuir cuando se incrementó el número de datos utilizados como predictores (vecinos). A pesar de que se utilizó desde un 40% hasta un 100% de la población, dicha variable no afectó de forma importante en la exactitud del modelo.



Figura 6. Exactitud de los modelos K-NN para la estimación de *P. xanthii*.



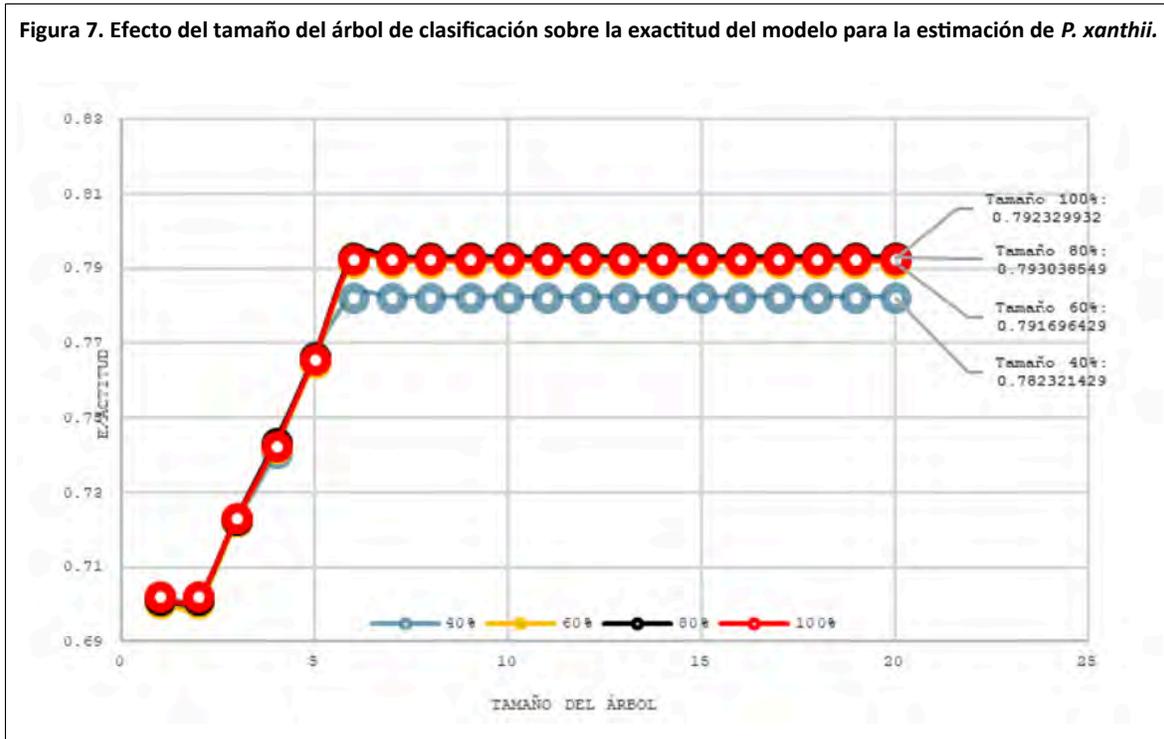
Los hallazgos de Cruz *et al.* (2020) utilizaron el método de K-NN debido a su éxito en estudios agrícolas para agilizar la detección de enfermedades. En esta investigación se realizó con imágenes de hojas de pepino seleccionadas con enfermedad y sin enfermedad para poder llegar a clasificar a los píxeles, de acuerdo con Guallazaca y Hernández (2020) utilizaron este modelo de clasificación para obtener un código que le permitiera identificar por colores y formas tres variables de identificación (producto bueno, producto regular y producto malo).

La técnica del mejor vecino se usa ampliamente para clasificar enfermedades foliares con alta exactitud (Sarkar *et al.*, 2023). Existen trabajos que usan este algoritmo y han logrado una exactitud 0.9 utilizando parámetros de color y textura (Zhang y Wallace 2015), en este análisis se obtuvo una exactitud de 0.85. Uno de los principales problemas en el entrenamiento de modelos de aprendizaje automático es la optimización de los hiperparámetros (Ghawi y Pfeffer, 2019), debido a que en los modelos K-NN, la estimación de la clase se basa en la distancia euclidiana entre las observaciones más cercanas, el número de observaciones tiene un efecto en la exactitud final del modelo (Torgo, 2014).

En esta investigación se utilizaron aproximadamente 10 datos para estimar la clase de cada uno de los datos muestra. Suganya *et al.* (2020) utilizaron esta misma técnica, con una exactitud superior al 0.9; sin embargo, las imágenes que utilizaron en el entrenamiento del modelo fueron tomadas en condiciones de iluminación controladas. En contraste, en este trabajo las imágenes de entrenamiento fueron obtenidas directamente en campo, lo cual generó una disminución en la exactitud del modelo debido a las condiciones altamente variables de la iluminación.

Árbol de decisión. La exactitud en las pruebas de validación fue de 0.79 cuando se usó el 80% de los datos, adicionalmente se pudo observar que conforme se aumentó el tamaño de la población, la exactitud se incrementó. Uno de los hiperparámetros que influyó significativamente en la exactitud del modelo es el número de ramificaciones (Fernández, 2023). En la presente investigación, se observó que la precisión no mejoró al incrementar este parámetro más allá de siete ramas (Figura 7), este comportamiento es consistente con lo reportado por Ramos (2020), quien empleó cinco ramificaciones y obtuvo una exactitud de 0.84.

Figura 7. Efecto del tamaño del árbol de clasificación sobre la exactitud del modelo para la estimación de *P. xanthii*.

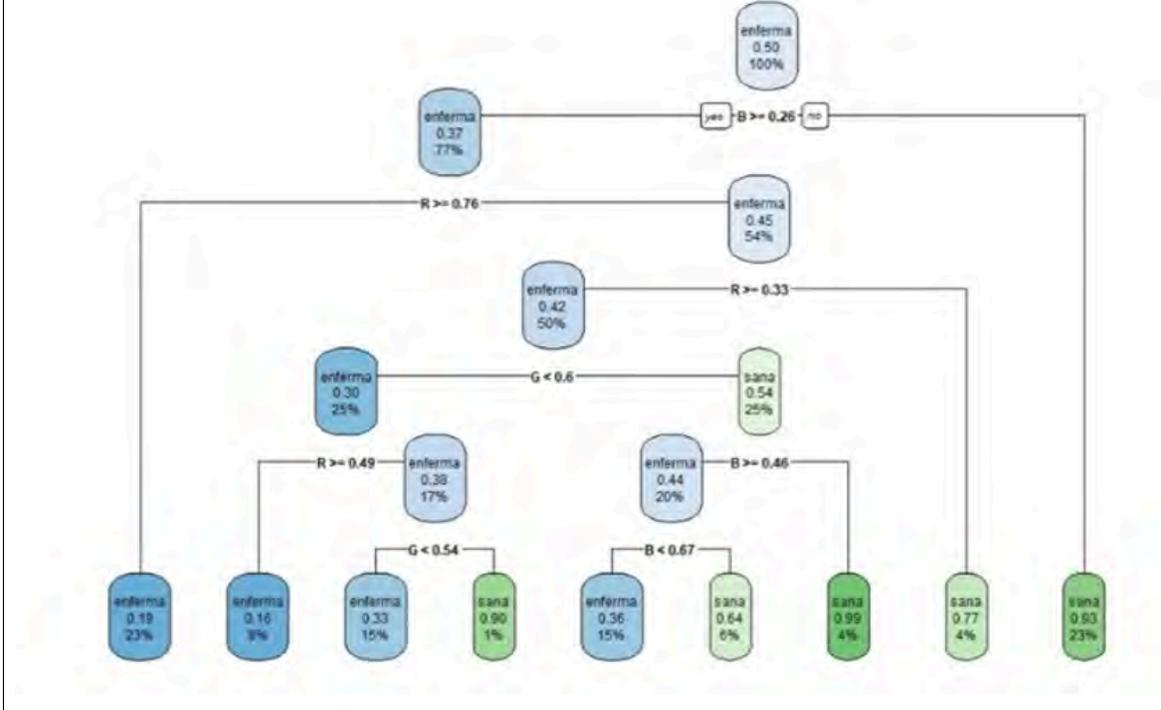


Al respecto Olivares *et al.* (2021) empleó bosques aleatorios para determinar el desarrollo de la incidencia de la marchitez del banano, obteniendo una precisión de 0.74 por lo cual no lo consideraron un modelo eficiente. La principal desventaja de este método es el elevado costo computacional que se requiere para su implementación (Alaminos, 2023).

En esta investigación, los síntomas visibles en las hojas de pepino ocasionaron cambios en el valor de los canales RGB, en comparación con los valores presentes en hojas normales, que fueron utilizadas por el modelo para clasificar los pixeles (Figura 8), estos resultados coinciden con Velázquez *et al.* (2011) que reportan para detectar cenicilla en rosa es capaz de detectarla a través del espacio de color con imágenes tomadas a una distancia cercana para tener mejor precisión.



Figura 8. Representación de las reglas de decisión generadas por el algoritmo de bosque aleatorio, para la clasificación de hojas de pepino sanas y con presencia de enfermedad.

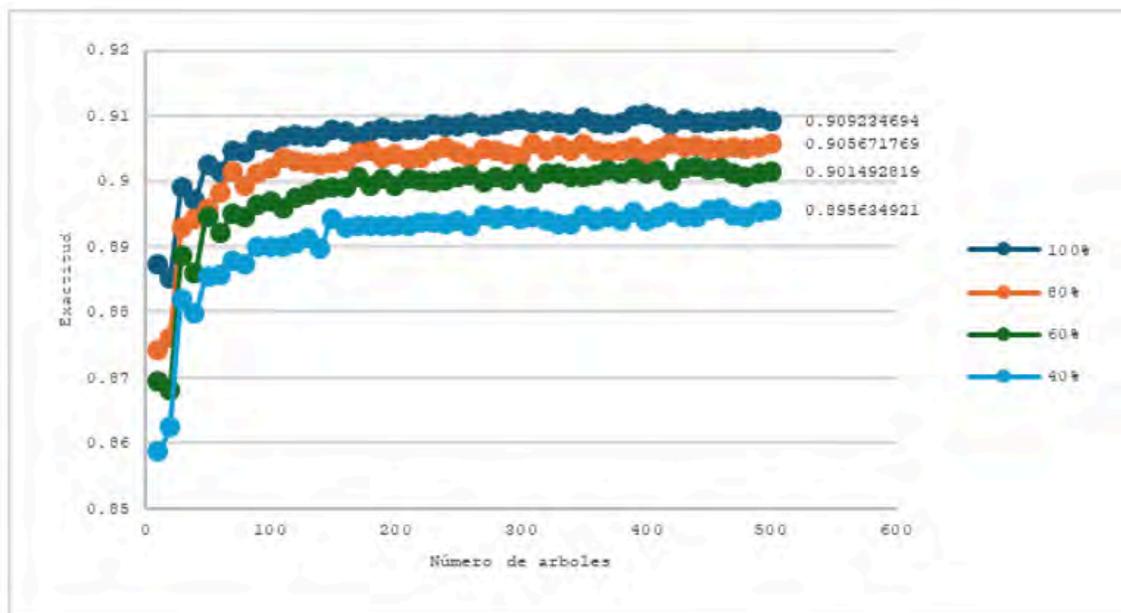


Los resultados obtenidos muestran que el modelo de bosque aleatorio permitió alcanzar una clasificación satisfactoria de las clases analizadas. Este enfoque ha sido previamente utilizado por diversos investigadores en el análisis de datos (Flores *et al.*, 2016), debido a su capacidad de adaptarse a diferentes tipos y escalas de bases de datos. Su principal ventaja radica en que no requiere asumir una distribución de normalidad y ofrece una notable flexibilidad para modelar relaciones no lineales entre las variables predictoras y las clases objetivo.

Bosques aleatorios. El número de árboles y el porcentaje de la población utilizada en los datos de entrenamiento afectaron la exactitud del modelo (Figura 9), los valores más elevados se obtuvieron cuando se utilizó un 80% de la población para realizar el entrenamiento del modelo. En general la exactitud no se incrementó después de 100 árboles, con los que se alcanzó una exactitud máxima de 0.9.



Figura 9. Efecto del número de árboles sobre la exactitud de un modelo de bosques aleatorios con diferentes proporciones de población.



Este método de aprendizaje automático se basa en la utilización de grupos de árboles aleatorios para estimar la clase a la que los datos pertenecen, por lo tanto, el número de árboles se considera como un parámetro de control importante que afecta significativamente a la exactitud final con la que cada clase es estimada (Sujatha *et al.*, 2021), esto coincide con lo reportado en este análisis, en el cual la exactitud se incrementó hasta mantenerse estable a pesar de aumentar el número de árboles empleados.

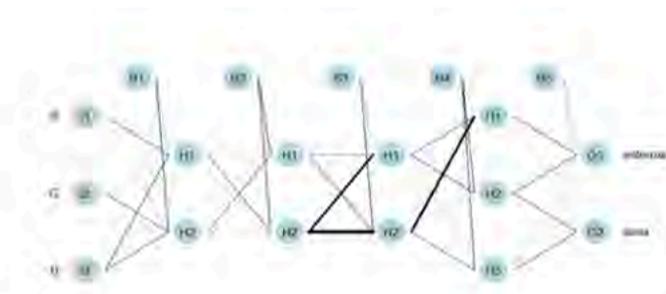
Los bosques aleatorios se han usado por otros autores para encontrar relaciones no lineales entre variables con fines de detección de enfermedades en plantas por medio de imágenes; por ejemplo, Wójtowicz *et al.* (2021) utilizaron datos de entrenamiento para generar un modelo basado en bosques aleatorios obteniendo tasas de acierto superiores al 90%, lo cual es similar a lo reportado en este trabajo.

### Redes neuronales

La estructura de la red neuronal que mostró valores más altos de exactitud fue cuando se utilizaron tres capas ocultas con dos neuronas cada una utilizando un 80% de la población total disponible (Figura 10).



Figura 10. Estructura de la red neuronal utilizada para la estimación.



Arquitectura de Red Neuronal (CNN)	Población %	Exactitud
3 capas con 2 neuronas	80	0.80 – 0.81
3 capas con 2 neuronas	100	0.79 - 0.80
1 capa con 3 neuronas	80	0.79 – 0.81
1 capa con 3 neuronas	100	0.80 – 0.82

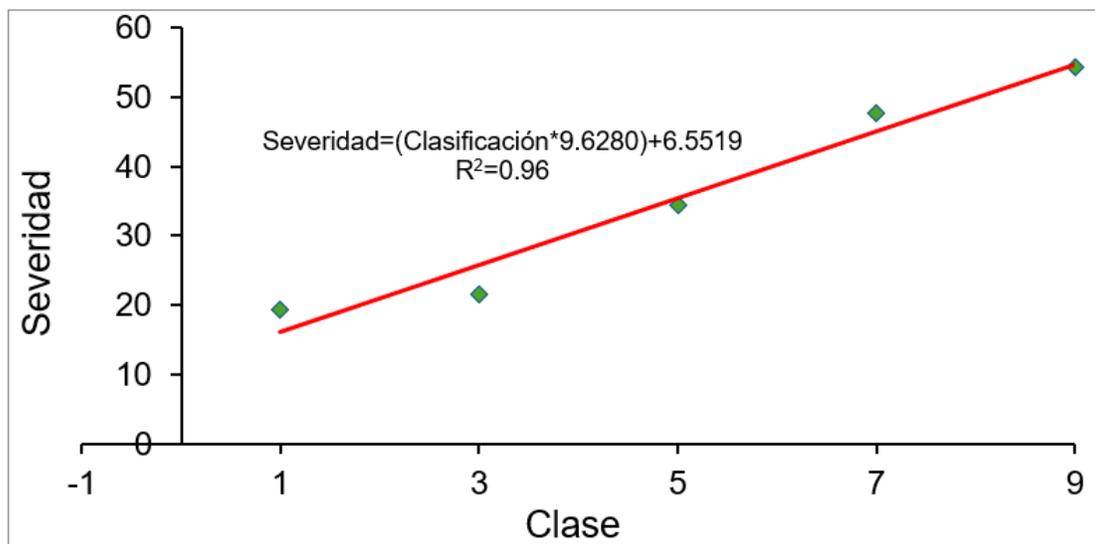
Las redes neuronales han sido ampliamente utilizadas en la detección de enfermedades; por ejemplo, Ma *et al.* (2018); Larijani *et al.* (2019) utilizó esta técnica para detectar enfermedades en hojas de arroz utilizando como variables de entrada los valores de color Lab, en el presente estudio se utilizó el espacio de color RGB, por lo que no fue necesaria una conversión. Otros autores Sujatha *et al.* (2021), utilizaron diferentes arquitecturas de redes neuronales para detectar virus en plantas con una exactitud superior al 0.95, que es muy superior a la exactitud que se reporta en este trabajo utilizando la misma técnica (redes neuronales); sin embargo, los síntomas que producen los virus es una clorosis muy acentuada y de relativa fácil detección mediante modelos basados en los cambios de color.

En el caso específico de pepino Zhang *et al.* (2019) ensayaron distintos sistemas basados en redes neuronales para identificar diferentes enfermedades en hojas de pepino partiendo de un conjunto de imágenes en condiciones relativamente controladas con exactitudes superiores al 0.95 y con tiempos de entrenamiento de 6 y 14 h. En esta investigación se obtuvieron tiempos de entrenamiento de entre 60 y 85 min, pero el número de clases fue significativamente menor, por lo que se requirió de menos capacidad de procesamiento.

El preprocesamiento de las imágenes y la obtención a partir de los datos de entrenamiento es uno de los factores que más afecta al desempeño de los modelos matemáticos (Li *et al.*, 2022) en el caso de las imágenes utilizadas como base para generar el conjunto de datos de entrenamiento de este trabajo, fueron obtenidas en condiciones no controladas, lo cual generó una gran diversidad de condiciones en color y forma de aquellas partes de la imagen que no representaban hoja.

De todos los métodos de aprendizaje automático ensayados en este estudio fue el de bosques aleatorios el que mostró el valor de exactitud más alto, por lo cual se utilizó este modelo para comparar la relación entre la clasificación de las plantas mediante una escala hedónica realizada por una persona, y la severidad reportada por el modelo. Los resultados mostraron que existió una relación entre la clase otorgada por una persona utilizando una escala hedónica y la severidad (Figura 11).

Figura 11. Comparación de escala hedónica estimada de *P. xanthii* y severidad utilizando el modelo de clasificación bosques aleatorio.



## Conclusiones

Se evidencia que los algoritmos de aprendizaje automático constituyen una herramienta eficaz para estimar con alta precisión la severidad del daño foliar causado por *Podosphaera xanthii* en hojas de pepino. En particular, el modelo basado en bosques aleatorios alcanzó una exactitud del 90%, destacándose por su capacidad de generalización y su robustez frente a condiciones variables de iluminación y captura en campo. Este enfoque automatizado representa una alternativa viable al método tradicional de evaluación visual, aportando objetividad, reproducibilidad y eficiencia al monitoreo fitosanitario.

La validación del modelo mediante una escala hedónica revela una correspondencia significativa entre las predicciones computacionales y las evaluaciones humanas, lo que respalda su aplicación práctica en programas de manejo integrado de enfermedades. Con base en estos hallazgos, se proponen varias líneas de trabajo futuras. Se recomienda ampliar el conjunto de datos para mejorar la generalización del modelo, incorporando imágenes de distintas variedades de pepino, fases fenológicas diversas y condiciones ambientales heterogéneas.

Asimismo, la integración del modelo en plataformas móviles permitirá realizar diagnósticos en tiempo real, directamente en el entorno agrícola. Estas acciones podrían contribuir significativamente a la adopción de tecnologías de inteligencia artificial en la agricultura de precisión, mejorando la detección oportuna de enfermedades y la toma de decisiones en campo.

## Bibliografía

- 1 Alaminos, F. A. F. 2023. Árboles de decisión en R con Random Forest. Obts Ciencia Abierta Alicante: limecop. 47-50 pp.
- 2 Benali, L.; Notton, G.; Foulloy, A.; Voyant, C.; Dizene, R.; Boum, H.; Ene, E.; Alia, E.; Ezzouar, B. and Algiers, A. 2019. Solar radiation forecasting using artificial neural network and random forest methods: application to normal beam, horizontal diffuse and global components. Renewable Energy. 132:871-884. <https://doi.org/10.1016/j.renene.2018.08.044>.

- 3 Braun, U. and Cook, A. R. T. 2012. Taxonomic manual of the Erysiphales (powdery mildews). Centraalbureau voor Schimmelcultures. 11. CBS Biodiversity: 86-644 pp.
- 4 Cipriano, G. R. y González, D. 2022. Identificación molecular de los tipos de compatibilidad en poblaciones de *Podosphaera xanthii* (Erysiphaceae) infectando cucurbitáceas en Veracruz, México. Acta Botánica Mexicana. 128(129):1-11. <https://doi.org/10.21829/abm129.2022.2068>.
- 5 Cruz, S. H.; Sanchez, M. G.; Rivera, C. J. P. and Avila, G. H. 2020. Identification of phenological stages of sugarcane cultivation using Sentinel-2 images. applications in software engineering proceedings of the 9<sup>th</sup>. International Conference on Software Process Improvement, CIMPS. 110-116 pp. <https://doi.org/10.1109/CIMPS52057.2020.9390095>.
- 6 Demirović, D. and Stuckey, P. J. 2021. Optimal decision trees for nonlinear metrics. 3733-3744. <http://www.aaai.org>.
- 7 Fernández, A. 2023. Árboles de decisión en R con Random Forest. Obets Ciencia Abierta. Alicante: limencop. 134 p.
- 8 Figueredo, A. A. y Ballesteros, R. J. 2016. Identificación del estado de madurez de las frutas con redes neuronales artificiales, una revisión. Ciencia y Agricultura 13(1):117-132. <https://www.redalyc.org/journal/5600/560062814010/html/>.
- 9 Flores, P. G.; López, I. F.; Kemp, P. D.; Dörner, J. y Zhang, B. 2016. Modelo de árbol de decisión: una herramienta para el manejo de la pradera. Agro Sur. 44(2):3-10. <https://doi.org/10.4206/agrosur.2016.v44n2-02>.
- 10 Ghawi, R. and Pfeffer, J. 2019. Efficient hyperparameter tuning with grid search for text categorization using knn approach with BM25 similarity. Open Computer Science. 9(1):160-180. <https://doi.org/10.1515/COMP-2019-0011/machinereadablecitation/ris>.
- 11 Guailazaca, G. C. A. y Hernández, A. V. 2020. Clasificador de productos agrícolas para control de calidad basado en machine learning e industria 4.0. Revista Perspectivas. 2(2):21-28. <https://doi.org/10.47187/perspectivas.vol2iss2.pp21-28.2020>.
- 12 Hassoun, M. H. 1995. Fundamentals of artificial neural networks book. MIT Press. Cambridge, Massachusetts. 417-452 pp. <https://kupdf.net/download/48375906-fundamentals-of-artificial-neural-networks-book-1-598b1ef3dc0d601b67300d18-pdf>.
- 13 Hernández, F. Y.; González, Z. E.; Marrero, T. A. y Dueñas, G. M. J. 2007. Uso de escala para determinar severidad de enfermedades fungosas en híbridos de pepino bajo cultivo protegido. INIFAT. 11(31):49-51.
- 14 Kaushik, H.; Khanna, A.; Singh, D.; Kaur, M. and Lee, H. N. 2023. TomFusioNet: a tomato crop analysis framework for mobile applications using the multi-objective optimization based late fusion of deep models and background elimination. Applied Soft Computing. 133:1-24. <https://doi.org/10.1016/j.asoc.2022.109898>.
- 15 Larijani, M. R.; Asli-Ardeh, E. A.; Kozegar, E. and Loni, R. 2019. Evaluation of image processing technique in identifying rice blast disease in field conditions based on KNN algorithm improvement by K-means. Food Science and Nutrition. 7(12):3922-3930. <https://doi.org/10.1002/FSN3.1251>.
- 16 Li, S.; Li, K.; Qiao, Y. and Zhang, L. 2022. A multi-scale cucumber disease detection method in natural scenes based on YOLOv5. Computers and Electronics in Agriculture. 202:1-12. <https://doi.org/10.1016/J.COMPAG.2022.107363>.
- 17 Ma, J.; Du, K.; Zheng, F.; Zhang, L.; Gong, Z. and Sun, Z. 2018. A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. Computers and Electronics in Agriculture 154:1-7. <https://doi.org/10.1016/j.compag.2018.08.048>.

- 18 Mohamed, Y. F.; Bardin, M. N. P. C. and Pitrat, I. 1995. Causal agents of powdery mildew of cucurbits in Sudan. *Plant Disease*. 79(6):635-636. <https://www.apsnet.org/publications/plantdisease/backissues/Documents/1995Articles/PlantDisease79n06-634>.
- 19 Morejón, G. N.; Coca, M. B. y Martínez, I. D. 2010. Mildiu polvoriento en las cucurbitáceas. *Revista de Protección Vegetal*. 25(1):44-50. <https://revistas.censa.edu.cu/index.php/RPV/article/view/282>.
- 20 Olivares, B. O.; Vega, A.; Angélica, M.; Calderón, R.; Rey, J. C. and Lobo, D. 2021. Classification of areas affected by banana wilt: an application with machine learning algorithms in Venezuela. *REICIT: Revista especializada de ingeniería y ciencias de la tierra* <https://revistas.up.ac.pa/index.php/REICTORCID>.
- 21 OTSU, N. A. 1978. Threshold selection method from gray level histogram. *IEEE. Transactions on Systems, Man and Cybernetics*. 9(1):62-66.
- 22 Pacciorett, P. A.; Kurina, F. G. y Balzarini, M. G. 2020. Muestreo de sitios a escala regional para mapeo digital basado en propiedades de suelo. *Ciencia del Suelo*. 38(2):310-320. <https://www.scielo.org.ar/scielo.php?script=sciarttext&pid=S185020672020000200310&Ing=es&lng=es>.
- 23 Paymode, A. S. and Malode, V. B. 2022. Transfer learning for multi-crop leaf disease image classification using convolutional neural Network VGG. *Artificial Intelligence in Agriculture*. 6:23-33. <https://doi.org/10.1016/j.aiia.2021.12.002>.
- 24 Ramos, R. T. V.; Castillo, A. P. J.; Ticona, J. B. y Velasco, B. J. G. 2023. Predicción del éxito del telemarketing bancario mediante el uso de árboles de decisión. *Innovación y Software*. 4(1):122-137. <https://doi.org/10.48168/innosoft.s11.a84>.
- 25 Rocha, J. F. L.; Reyes, D. Y.; Días, L. E.; Francisco, F. N. y Juárez, C. J. A. 2023. El mildiu polvoriento en calabaza: identificación y manejo bajo las condiciones de Tehuacán, México. *Cultivos Tropicales*. 44(2): <https://cu-id.com/2050/v44n2e09>. <https://ediciones.inca.edu.cu/index.php/ediciones/article/view/1731>.
- 26 Sarkar, C.; Gupta, D.; Gupta, U. and Hazarika, B. B. 2023. Leaf disease detection using machine learning and deep learning: review and challenges. *Applied Soft Computing*. 145:1-61. <https://doi.org/10.1016/J.ASOC.2023.110534>.
- 27 Suganya, D. K.; Srinivasan, P. and Bandhopadhyay, S. 2020. H2K-A robust and optimum approach for detection and classification of groundnut leaf diseases. *Computers and Electronics in Agriculture*. 178. <https://doi.org/10.1016/j.compag.2020.105749>.
- 28 Sujatha, R.; Chatterjee, J. M.; Jhanjhi, N. Z. and Brohi, S. N. 2021. Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocessors and Microsystems*. 80:103615. <https://doi.org/10.1016/J.MICPRO.2020.103615>.
- 29 Sun, Z.; Hu, S. Y. and Wen, Y. 2022. Biological control of the cucumber downy mildew pathogen *Pseudoperonospora cubensis*. *Horticulturae*. 8:1-15. MDPI. <https://doi.org/10.3390/horticulturae8050410>.
- 30 Torgo L. 2014. *Data mining using R: learning with case studies (CRC) Press, Ed. Second*. Minneapolis, Minnesota, USA. 87-165 pp. ISBN: 9781439810187.
- 31 Velázquez, L. N; Sasaki, Y.; Nakano, K.; Mejía, M. J y Romanchik, K. E. 2011. Detección de cenicilla en rosa usando procesamiento de imágenes por computadora. *Revista Chapingo Serie Horticultura*. 17(2):151-160. <http://sourceforge.net/projects/opencvlibrary>.
- 32 Wójtowicz, A.; Piekarczyk, J.; Czernecki, B. and Ratajkiewicz, H. 2021. A random forest model for the classification of wheat and rye leaf rust symptoms based on pure spectra at leaf scale. *Journal of Photochemistry and Photobiology B. Biology*. 223:1-11. <https://doi.org/10.1016/j.jphotobiol.2021.112278>.
- 33 Zhang, S.; Zhang, S.; Zhang, C.; Wang, X. and Shi, Y. 2019. Cucumber leaf disease identification with global pooling dilated convolutional neural network. *Computers and Electronics in Agriculture* 162:1-9. <https://doi.org/10.1016/j.compag.2019.03.012>.

- 34 Zhang, Y. and Wallace, B. 2015. A sensitivity analysis and practitioners' guide to convolutional neural networks for sentence classification. 256-263 pp. <http://arxiv.org/abs/1510.03820>.
- 35 Zhao, Y. and Yang, L. 2023. Distance metric learning based on the class center and nearest neighbor relationship. Neural Networks. 164:631-644. <https://doi.org/10.1016/j.neunet.2023.05.004>.
- 36 Zapata, T. A.; Pérez, L. S. y Mora, F. J. 2014. Método basado en clasificadores k-NN parametrizados con algoritmos genéticos y la estimación de la reactancia para localización de fallas en sistemas de distribución. Revista Facultad de Ingeniería Universidad de Antioquia. 70:220-232. <http://www.scielo.org.co/scielo.php?script=sci-arttext&pid=S012062302014000100021&lng=pt&tlng=es>.



## Estimación de *Podosphaera xanthii* en pepino: técnicas de aprendizaje automático con imágenes digitales

Journal Information
Journal ID (publisher-id): remexca
Title: Revista mexicana de ciencias agrícolas
Abbreviated Title: Rev. Mex. Cienc. Agríc
ISSN (print): 2007-0934
Publisher: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias

Article/Issue Information
Date received: 00 February 2025
Date accepted: 00 August 2025
Publication date: 15 October 2025
Publication date: Sep-Oct 2025
Volume: 16
Issue: esp30
Electronic Location Identifier: e4039
DOI: 10.29312/remexca.v16i30.4039
Article Id (other): 00002

### Categories

Subject: Artículos

### Palabras clave

#### Palabras clave

algoritmo  
cenicilla  
enfermedad

### Counts

Figures: 11  
Tables: 2  
Equations: 0  
References: 36