



CLASIFICACIÓN DE ESTRÉS HÍDRICO Y PHYTOPHTHORA EN CONÍFERAS MEDIANTE MODELOS CNN Y TRANSFORMERS

Josué Aldebaran González López

Escuela Superior de Cómputo del Instituto Politécnico Nacional

Jgonzalezl2006@alumno.ipn.mx

Orcid: 0009-0001-7033-479X

Benjamín Luna Benoso

Escuela Superior de Cómputo del Instituto Politécnico Nacional

blunab@ipn.mx

Orcid: 0000-0002-3519-6328

José Cruz Martínez Perales

Escuela Superior de Cómputo del Instituto Politécnico Nacional

jmartinezp@ipn.mx

Orcid: 0000-0001-9421-5923

Úrsula Samantha Morales Rodríguez

Escuela Superior de Cómputo del Instituto Politécnico Nacional

umoralesr@ipn.mx

Orcid: 0000-0001-9265-4228

Abstract

Los árboles del género *Cupressus*, pertenecientes al grupo de las coníferas, tienen gran importancia ecológica, ambiental y forestal. Por ello, es fundamental monitorear la presencia de enfermedades y condiciones de estrés. En este trabajo se clasificaron, mediante imágenes foliares, dos condiciones que afectan a *Cupressus*: el estrés hídrico y la infección por *Phytophthora*. Se compararon tres modelos de redes neuronales convolucionales (CNN): *ConvNeXt-Tiny*, *ResNet50+SE* y *DenseNet121*; y tres modelos basados en Transformers: *Vision Transformer (ViT)*, *Swin Transformer* y *DINOv2*. Los resultados mostraron que, en general, los modelos basados en Transformers presentaron mejor desempeño que las CNN evaluadas, destacando el modelo *Swin Transformer* que obtuvo los mejores resultados, alcanzando un accuracy del 99.69 %.

Palabras clave: Aprendizaje profundo, redes neuronales convolucionales, Transformers, estrés hídrico, *Phytophthora*.

Lo árboles del género *Cupressus*, pertenecientes al grupo de coníferas, desempeñan un papel fundamental en los ecosistemas, al contribuir a la captura de CO₂,

proporcionar hábitat a diversas especies y constituir una importante fuente de recursos forestales, como la madera, lo que resalta su importancia ecológica, ambiental y forestal. Por esto, es importante mantener a los

Cupressus sanos y monitorear de manera constante la posible presencia de enfermedades. Dos de las principales enfermedades indicadoras del deterioro y mortalidad de las coníferas son el estrés hídrico y la infección por *Phytophthora* (Gea-Izquierdo, Sánchez-Gómez y Aranda, 2023; Herrero, González-Gascuña, González-Díaz, Ruiz-Benito y Andivia, 2023).

Aunque el diagnóstico de estrés hídrico y la infección por *Phytophthora* puede realizarse de manera manual mediante la caracterización de sus síntomas visuales, esto requiere de un especialista del área de fitopatología, ya que, en etapas tempranas, los síntomas de ambas enfermedades son muy parecidas (Hornero et al., 2021).

El uso de técnicas de Inteligencia Artificial (IA) ha surgido como una alternativa para la detección de enfermedades en cultivos agrícolas desde sus primeras etapas visibles (Papazoglou et al., 2025; Wang y Ghatrehsamani, 2026). Algunos trabajos abordan la problemática utilizando técnicas clásicas de Aprendizaje Automático (Upasani y Goswami, 2025), mientras que otros utilizan técnicas de Aprendizaje Profundo (Thite, Suryawanshi, Patil y Chumchu, 2023).

Aunque las técnicas de aprendizaje profundo han sido aplicadas con éxito en la detección de enfermedades en plantas, incluyendo el estrés hídrico y las infecciones causadas por *Phytophthora*, la mayoría de los estudios se enfocan en evaluar únicamente arquitecturas basadas en CNN (Ha, Kim, Kim y Choe, 2025; Singh, Verma y Alex, 2021), modelos basados en Transformers (Li et al., 2026; Liu y Zhang, 2025) o propuestas de arquitecturas híbridas (Chakrabarty, Ahmed, Islam, Aziz y Maidin, 2024). Sin embargo, estos trabajos no realizan análisis comparativos exhaustivos entre modelos CNN y Transformers bajo una misma problemática y condiciones experimentales. Por otro lado, existen limitaciones importantes relacionadas con los tipos de datos utilizados, siendo esto más evidente en el caso de enfermedades como

la *Phytophthora*, donde el uso de imágenes RGB y sus aplicaciones han sido escasamente explorados. En este contexto, el presente trabajo propone un estudio comparativo entre tres modelos basados en CNN (ConvNeXt-Tiny, ResNet50+SE y DenseNet121) y tres arquitecturas basadas en Transformers (ViT, Swin Transformer y DINOv2) para la clasificación del estrés hídrico y la infección por *Phytophthora* utilizando imágenes RGB. Aunque el estudio se centra específicamente en estas dos enfermedades, la metodología propuesta puede ser extendida a la identificación de otras enfermedades presentes en plantas.

Metodología

Conjunto de Imágenes

Las imágenes utilizadas para este trabajo comprenden un compendio de imágenes foliares de árboles del género *Cupressus* que corresponden a un total de 9,611 imágenes, de las cuales, 3,204 corresponden a estrés hídrico, 3,203 a *Phytophthora* y 3,204 a plantas sanas. Las imágenes fueron capturadas bajo escenarios reales, y almacenadas en formato PNG, en su mayoría con una resolución de 3024 x 3024 o de 2260 x 4032. La figura 1 muestra un ejemplo representativo de *Cupressus*, en (a) con estrés hídrico, y en (b) con infección causada por *Phytophthora*.



Figura 1. Ejemplo representativo de (a) estrés hídrico y (b) *Phytophthora* en *Cupressus*.

Enfoque propuesto

La figura 2 muestra el flujo de procesamiento del enfoque propuesto. En la primera etapa, las imágenes son preparadas mediante un recorte automático para eliminar bordes irrelevantes y la redimensión de estas a 512 x 512 píxeles con la finalidad de preservar detalles relevantes.

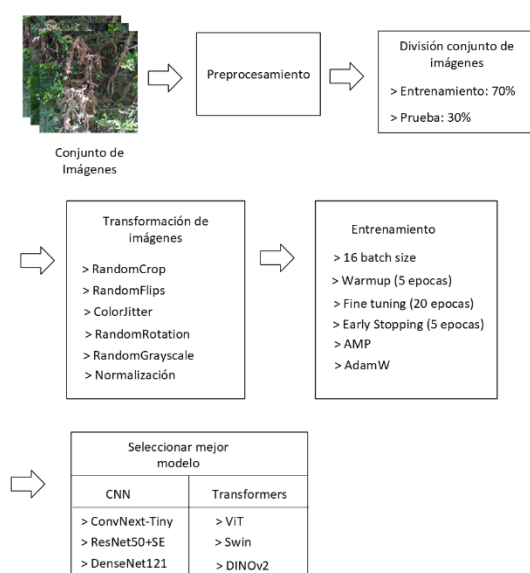


Figura 2. Flujo de procesamiento del enfoque propuesto.

En la segunda etapa, el conjunto de datos es dividido utilizando una partición estratificada, donde el 70% corresponde al conjunto de entrenamiento y el 30% al conjunto de prueba.

En la tercera etapa, se aplicaron un conjunto de transformaciones aleatorias de imágenes sobre el conjunto de entrenamiento, con el objetivo de incrementar la variabilidad de los datos mejorando la capacidad de que los modelos generalicen. Las transformaciones utilizadas fueron: RandomCrop (recorta regiones aleatorias para que los modelos no memoricen posiciones), RandomFlips (invierte imágenes horizontales y verticales),

ColorJitter (modifica color e iluminación simulando diferentes condiciones de color e iluminación), RandomRotation (rota imágenes para introducir variaciones de orientación), RandomGrayscale (convierte las imágenes a escala de grises para que los modelos aprendan texturas y formas en lugar de colores), y finalmente se normalizaron los valores de los píxeles para mejorar la convergencia de los modelos.

En la cuarta etapa, se llevó a cabo el entrenamiento, para ello se utilizaron las siguientes configuraciones de hiperparámetros: se utilizó un batch size de 16 para que los modelos entrenaran con imágenes de alta resolución de 512 x 512 píxeles; se empleó un Warmup de 5 épocas para que la fase de entrenamiento se estabilizara mediante el aumento de la tasa de aprendizaje; se aplicó fine tuning para adaptar el problema a la detección de estrés hídrico y *Phytophthora*; para la optimización se utilizó AdamW con la finalidad de reducir sobreajustes; para reducir tiempos de entrenamiento se utilizó Early Stopping a 5 épocas y finalmente AMP para acelerar el entrenamiento sin afectar la precisión de los modelos.

En la quinta etapa, se seleccionó el mejor modelo con base en los resultados obtenidos por las métricas de *accuracy*, *F1-score*, *precision*, *recall*, *specificity*, además de sus respectivas matrices de confusión y el resultado de los mapas de activación (Grad-CAM) aplicado a imágenes representativas por cada modelo.

Resultados

La tabla 1 muestra los resultados obtenidos por los modelos basados en CNN y Transformers, considerando las métricas *accuracy*, *F1-score*, *precision*, *recall* y *specificity*. Aunque todos los modelos mostraron un buen desempeño, se observa de manera general que los modelos basados en Transformers obtuvieron unas métricas

ligeramente superiores a los resultados obtenidos por las CNN.

Dentro de los modelos CNN, ConvNeXt-Tiny obtuvo el mejor desempeño alcanzando un *accuracy* de 0.9951, un *F1-score*, *precision* y *recall* de 0.9952, y una *specificity* de 0.9976. Por otra parte, en los modelos basados en Transformers, Swin presentó el mejor desempeño global con un *accuracy*, *F1-score*, *precision* y *recall* de 0.9969, y una *specificity* de 0.9984.

Tabla 1. Evaluación de las métricas.

Enfoque	Modelo	<i>accuracy</i>	<i>F1-score</i>	<i>precision</i>	<i>recall</i>	<i>specificity</i>
CNN	ConvNeXt-Tiny	0.9951	0.9952	0.9952	0.9952	0.9976
	ResNet50+SE	0.9896	0.9896	0.9896	0.9896	0.9948
	DenseNet121	0.9941	0.9941	0.9941	0.9941	0.9971
Transformers	Vit	0.9958	0.9958	0.9958	0.9958	0.9979
	Swin	0.9969	0.9969	0.9969	0.9969	0.9984
	DinoV2	0.9948	0.9948	0.9948	0.9948	0.9974

La Figura 3 presenta las matrices de confusión correspondientes al modelo CNN con mejor desempeño reportado en la Tabla 1 (ConvNeXt-Tiny) y al modelo basado en Transformers con mejor desempeño (Swin). Se observa que ConvNeXt-Tiny clasificó correctamente 949 de las 962 imágenes correspondientes a estrés hídrico, y 13 fueron clasificadas incorrectamente como *Phytophthora*. Por otro lado, 960 imágenes de la clase *Phytophthora* fueron clasificadas correctamente de un total de 961 imágenes, por lo que, en conjunto, ConvNeXt-Tiny presentó un total de 14 clasificaciones incorrectas. Swin Transformes por su lado, clasifico correctamente 957 imágenes de un total de 962, cuatro las clasifico como *Phytophthora* y una como saludable, mientras que para la clase *Phytophthora*, 957 de las 961 fueron clasificadas correctamente, por lo que, Swin clasifico incorrectamente nueve imágenes. Aunque ambos modelos obtuvieron errores entre las clases de estrés hídrico y *Phytophthora*, Swin redujo los errores.

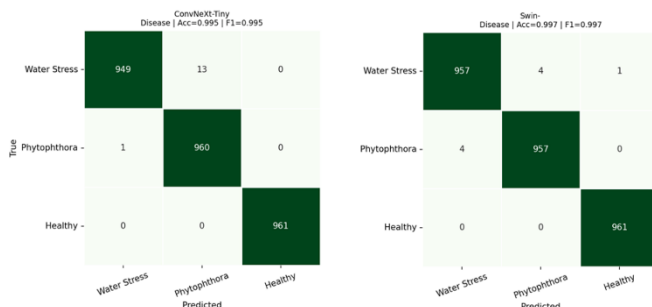


Figura 3. Matriz de confusión del modelo ConvNeXt-Tiny y del modelo Swin Transformer.

La figura 4 muestra los mapas de activación obtenidos mediante Grad-CAM para los modelos ConvNeXt-Tiny y Swin en imágenes que corresponden a estrés hídrico y *Phytophthora*. En ConvNeXt-Tiny, los mapas Grad-CAM muestran regiones de mayor activación localizadas principalmente sobre distintas zonas de la vegetación, incluyendo áreas con cambios visibles en la coloración y apariencia del follaje. Por otro lado, los resultados obtenidos con Swin presentan un patrón predominantemente vertical y distribuido a lo largo de la imagen, sin una delimitación espacial clara de regiones específicas asociadas con los síntomas visibles. Este comportamiento se observa en ambas enfermedades, tanto en estrés hídrico como en afectaciones por *Phytophthora*. Visualmente, ConvNeXt-Tiny presenta mapas de activación más localizados y coherentes con regiones particulares de la planta, mientras que Swin genera patrones de activación menos localizados espacialmente.

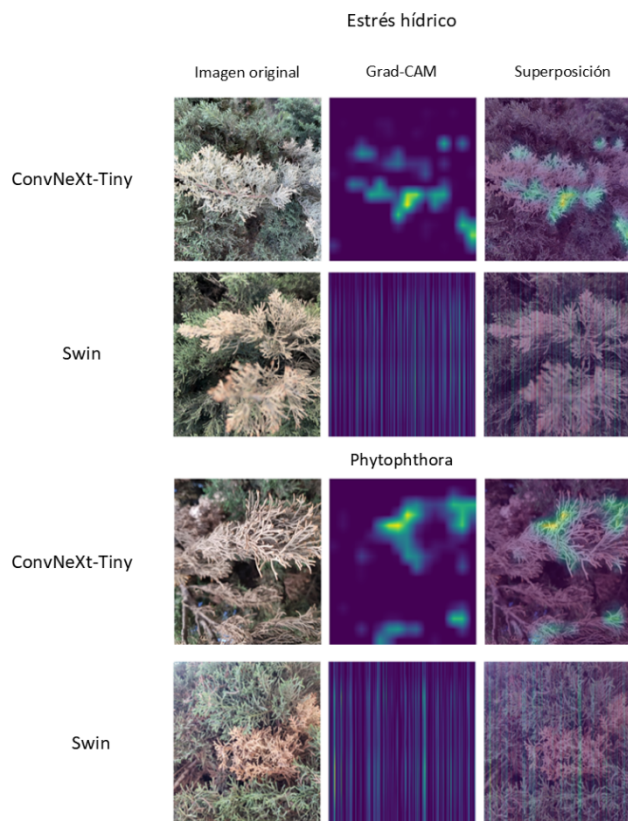


Figura 4. Visualización Grad-CAM que resaltan las regiones afectadas por estrés hídrico y *Phytophthora* para los modelos ConvNeXt-Tiny y Swin.

Discusión

Los resultados muestran que los modelos evaluados presentan un buen desempeño en la clasificación de estrés hídrico y *Phytophthora* en coníferas del tipo Cupressus. De acuerdo con los resultados mostrados en la tabla 1, Swin Transformer fue el modelo con mejor desempeño global, registrando un *accuracy* de 0.9969, mientras que ConvNeXt-Tiny fue el mejor modelo entre las arquitecturas CNN con un *accuracy* de 0.9951. Las matrices de confusión de ambos modelos muestran una elevada concentración de predicciones correctas en la diagonal principal y un reducido número de errores, confirmando la capacidad de los modelos para discriminar las clases analizadas. Por su parte, el análisis de los mapas de activación mediante Grad-CAM evidencia diferencias en la representación espacial

de ambos modelos. ConvNeXt-Tiny presenta activaciones localizadas en regiones específicas del follaje que incluyen zonas con cambios visible de coloración, mientras que Swin muestra patrones de activación más distribuidos y menos localizados espacialmente. El desempeño de Swin puede atribuirse a su mecanismo de autoatención, el cual, permite capturar relaciones locales y globales en diferentes regiones de las imágenes. En general, los resultados presentados en la tabla 1, muestran que los modelos basados en Transformers presentan un ligero mejor desempeño que las CNN evaluadas.

Referencias

- Chakrabarty, A., Ahmed, S. T., Islam, M. F. U., Aziz, S. M., & Maidin, S. S. (2024). An interpretable fusion model integrating lightweight CNN and transformer architectures for rice leaf disease identification. *Ecological Informatics*, 82, 102718. doi:10.1016/j.ecoinf.2024.102718
- Gea-Izquierdo, G., Sánchez-Gómez, D., & Aranda, I. (2023). Functional adjustment under lethal drought and physiological memory to water stress of two phylogenetically close and coexisting conifers. *Environmental and Experimental Botany*, 214, 105484. doi:10.1016/j.envexpbot.2023.105484
- Ha, U., Kim, H., Kim, S., & Choe, H. (2025). Enhanced pine wilt disease outbreak prediction: Integrating deep learning-detected infected trees with species distribution modeling. *Ecological Informatics*, 91, 103421. doi:10.1016/j.ecoinf.2025.103421
- Herrero, A., González-Gascuña, R., González-Díaz, P., Ruiz-Benito, P., & Andivia, E. (2023). Reduced growth sensitivity to water availability as potential indicator of drought-induced tree mortality risk in a Mediterranean *Pinus sylvestris* L. forest. *Frontiers in Forests and Global Change*, 6, 1249246. doi:10.3389/ffgc.2023.1249246
- Hornero, A., Zarco-Tejada, P. J., Quero, J. L., North, P. R. J., Ruiz-Gómez, F. J., Sánchez-Cuesta, R., & Hernandez-Clemente, R. (2021). Modelling hyperspectral- and thermal-based plant traits for the early detection of *Phytophthora*-induced symptoms in oak decline. *Remote Sensing of*



Environment, 263, 112570.
doi:10.1016/j.rse.2021.112570

Li, Y., Fayaz, M., Danish, S., Tightiz, L., Wang, H., Nguyen, T. N., & Dang, L. M. (2026). End-to-end plant disease detection using transformers with collaborative hybrid assignment training. *Applied Soft Computing*, 186, 114137.
doi:10.1016/j.asoc.2025.114137

Liu, W., & Zhang, A. (2025). Plant disease detection algorithm based on Efficient Swin Transformer. *Computers, Materials & Continua*, 82(2), 3045–3068. doi:10.32604/cmc.2024.058640

Papazoglou, P., Navrozidis, I., Testempasis, S., Pantazi, X. E., Lagopodi, A., & Alexandridis, T. (2025). Early detection of bacterial canker in tomato plants using spectroscopy for smart agriculture applications. *Biosystems Engineering*, 251, 1–10.
doi:10.1016/j.biosystemseng.2025.01.009

Singh, P., Verma, A., & Alex, J. S. R. (2021). Disease and pest infection detection in coconut tree through deep learning techniques. *Computers and Electronics in Agriculture*, 182, 105986.
doi:10.1016/j.compag.2021.105986

Thite, S., Suryawanshi, Y., Patil, K., & Chumchu, P. (2023). Coconut (*Cocos nucifera*) tree disease dataset: A dataset for disease detection and classification for machine learning applications. *Data in Brief*, 51, 109690.
doi:10.1016/j.dib.2023.109690

Upasani, N., & Goswami, C. (2025). Plant disease detection using machine learning techniques: A comprehensive approach. *International Journal of Environmental Sciences*, 11(12s), 1468–1480.
doi:10.64252/qm5z0g23

Wang, W., & Ghatreh Samani, S. (2026). Early detection of plant pathogens in the asymptomatic

phase: A scoping review of hyperspectral imaging combined with machine learning. *Smart Agricultural Technology*, 14, 102123.
doi:10.1016/j.atech.2026.102123