

# Sistema Integral para la Clasificación y Cuantificación de Severidad en Enfermedades Foliares del Maíz usando Deep Learning y Segmentación Avanzada

Jose Emanuel Rodriguez  
*Licenciatura en Sistemas de Información*  
*Universidad Nacional de Luján*  
Luján, Buenos Aires, Argentina  
joseemaro@hotmail.com

**Resumen**—La detección de enfermedades agrícolas es crítica para la seguridad alimenticia y la gestión de pérdidas. Si bien las técnicas de Aprendizaje Profundo (Deep Learning) han demostrado alta precisión en la *clasificación* de enfermedades del maíz, la *cuantificación* de la severidad sigue siendo un desafío. Este trabajo presenta un pipeline integral que aborda ambos problemas. Primero, se implementa un preprocesamiento multi-espacio (combinando un realce HSV condicional con CLAHE y Filtro de Mediana Adaptativo) para realzar rasgos de las imágenes. Segundo, se aplica un proceso de Aumento de Datos (Data Augmentation) para mejorar la generalización del modelo, simulando ciertas características presentes en imágenes naturales que difieren de las imágenes de laboratorio del dataset. Sobre esta base, se entrena una arquitectura CNN (ResNet50V2) mediante Transfer Learning para una clasificación precisa. La contribución principal es un módulo de análisis que: (1) utiliza el Segment Anything Model (SAM) combinado con Convex Hull y operaciones morfológicas para generar una segmentación sólida y robusta de la hoja, eliminando huecos y aislándola del fondo; y (2) aplica un análisis de color en el espacio HSV, adaptado a la enfermedad clasificada, para cuantificar el porcentaje de área foliar afectada. El sistema propuesto no solo diagnostica la enfermedad, sino que también proporciona una métrica de severidad, ofreciendo una herramienta de soporte de decisiones más completa para el manejo agronómico, sin necesidad de ser un experto en la materia.

**Términos de Índice**—aprendizaje profundo, enfermedades del maíz, clasificación de imágenes, cuantificación de severidad, segmentación de imágenes, Segment Anything Model (SAM), Convex Hull, visión por computadora.

## I. INTRODUCCIÓN

### I-A. Motivación y Contexto Agrícola

La agricultura representa un pilar fundamental para la seguridad alimentaria global, enfrentando constantemente la amenaza de enfermedades que reducen significativamente la producción, generando graves pérdidas económicas. El maíz (*Zea mays*) es uno de los cultivos más importantes a nivel global, y sobre todo en Argentina. Sin embargo, está sujeto a diversas enfermedades que pueden causar pérdidas económicas significativas, como el Tizón Foliar del Norte (*Northern Leaf*

*Blight*), la Mancha Gris de la Hoja (*Gray Leaf Spot*) y la Roya Común (*Common Rust*).

En Argentina, el maíz representa el cultivo de mayor volumen de producción, superando incluso a la soja en los últimos años. Las enfermedades foliares constituyen una amenaza significativa para este sector, pudiendo causar pérdidas de rendimiento que oscilan entre el 15 % y el 40 % dependiendo de la severidad de la infección y el momento del ciclo productivo en que ocurre [13]. Estas pérdidas no solo afectan la rentabilidad de los productores, sino que también comprometen la disponibilidad de alimentos y materias primas para la industria. La detección temprana y la cuantificación precisa de la severidad son cruciales para implementar estrategias de manejo integrado que minimicen el uso de fungicidas y maximicen la eficiencia productiva.

Históricamente, la detección y diagnóstico de estas enfermedades ha dependido de la evaluación visual realizada por expertos en la materia. Este método tradicional presenta limitaciones: subjetividad, necesidad de gran experiencia y tener que estar presente en el lugar de la problemática. La necesidad de superar estas barreras ha impulsado la investigación hacia sistemas automatizados que aprovechan los avances en visión por computadora y aprendizaje profundo, ofreciendo diagnósticos rápidos y objetivos.

### I-B. Planteamiento del Problema

Este trabajo aborda el desafío de crear un sistema automatizado integral para el diagnóstico de enfermedades en hojas de maíz. El problema no se limita solo a la *clasificación* (identificación de la enfermedad), sino que se extiende a la *cuantificación* de la severidad (medir el porcentaje o área de daño).

Si bien la clasificación mediante Deep Learning ha alcanzado altas tasas de precisión en trabajos previos, la mayoría de los sistemas fallan en proporcionar una métrica de severidad y en la generalización de dichos experimentos, centrándose en casos de laboratorio que difieren en gran medida de

las situaciones cotidianas. La cuantificación es técnicamente compleja debido a dos desafíos principales:

1. **Segmentación Robusta de la Hoja:** Aislar la hoja del fondo (tierra, otras hojas, tallos) es un paso previo indispensable. Los métodos tradicionales fallan con fondos complejos, y modelos como el Segment Anything Model (SAM) [1], aunque sean potentes, generan múltiples máscaras (ruido) y a menudo discontinuidades internas que deben ser filtradas y refinadas. Este desafío no solo es difícil por el entorno de la hoja, sino también por el hecho de que debe realizarse de forma automática ante una fotografía natural.
2. **Detección Precisa de Lesiones:** Una vez aislada la hoja, es necesario un método que identifique los píxeles correspondientes a la lesión, diferenciándolos de la parte sana, lo cual varía entre diferentes patologías, por lo que debe ser tratado de diferente forma en cada caso.

El conjunto de datos utilizado corresponde a un subconjunto especializado extraído del repositorio New PlantVillage [2], enfocado exclusivamente en imágenes de maíz, dejando de lado el resto de imágenes que corresponden a otras especies. La composición del dataset se presenta en la Tabla I. Este dataset, si bien es efectivo para la clasificación, presenta fondos simples que no son representativos de condiciones reales (son más bien, imágenes de laboratorio), haciendo necesario el uso de imágenes de prueba adicionales para mejorar la robustez de la segmentación.

Una limitación inherente a PlantVillage es que sus imágenes fueron capturadas en condiciones controladas, con fondos uniformes y bajo iluminación artificial, lo cual facilita la tarea de clasificación pero no refleja la variabilidad presente en el campo [3]. Esta diferencia entre las condiciones de entrenamiento y las condiciones de producción real es un desafío conocido en la aplicación práctica de modelos de visión por computadora en agricultura [14]. Para confrontar esto, se implementó una estrategia de Aumento de Datos agresiva que simula variaciones de iluminación, rotaciones, y cambios de perspectiva, con el objetivo de reducir la diferencia entre el entorno de laboratorio y el entorno natural, permitiendo que el modelo generalice mejor ante imágenes capturadas en el campo con fondos complejos y condiciones variables.

Cuadro I  
COMPOSICIÓN DEL SUBCONJUNTO DE MAÍZ DEL DATASET  
PLANTVILLAGE

Clase	Número de Muestras
Tizón Foliar del Norte	985
Mancha Gris de la Hoja	513
Roya Común	1.192
Hojas Sanas	1.162
<b>Total</b>	<b>3.852</b>

### I-C. Objetivos de Investigación

El objetivo general es desarrollar y evaluar un pipeline completo para la clasificación y cuantificación de severidad

de enfermedades en hojas de maíz.

#### I-C1. Objetivos Específicos:

1. **Implementar un pipeline de realce multi-espacio:** Diseñar y aplicar un proceso de preprocesamiento que combina un realce condicional en el espacio HSV (solo para la clase *Gray Leaf Spot*) con un realce de contraste general en el espacio  $L^*a^*b^*$  (mediante CLAHE) y un filtrado de ruido (AMF) para mejorar la representación de las características de cada enfermedad.
2. **Generar datos sintéticos para generalización:** Aplicar un robusto proceso de **Aumento de Datos** (Data Augmentation) sobre las imágenes procesadas, simulando condiciones de campo (rotaciones, zoom, cambios de brillo) para permitir que el modelo, entrenado con imágenes de laboratorio, pueda clasificar imágenes en entornos naturales.
3. **Entrenar un modelo de clasificación de alta precisión:** Entrenar y evaluar la arquitectura ResNet50V2 usando Transfer Learning para la tarea de clasificación de 4 clases (3 enfermedades + 1 sana), ajustando los parámetros de entrenamiento hasta encontrar un modelo capaz de clasificar imágenes en entornos naturales.
4. **Desarrollar un módulo de segmentación de hoja:** Implementar un método que combine el Segment Anything Model (SAM), Convex Hull y morfología matemática para aislar la hoja del fondo y generar máscaras sólidas de manera precisa y automática.
5. **Diseñar un módulo de cuantificación de área:** Crear un sistema que, basándose en la enfermedad clasificada, aplique umbrales de color específicos (en el espacio HSV) para detectar píxeles de lesión y calcular el área de la hoja que fue afectada.
6. **Integrar y validar el pipeline completo:** Ensamblar todos los módulos en un sistema funcional y probarlo en un conjunto de imágenes de validación para evaluar su rendimiento (clasificación y cuantificación).

#### I-D. Contribuciones

Las contribuciones clave de este trabajo son:

1. **Un pipeline integral (Clasificación + Cuantificación):** A diferencia de la mayoría de los trabajos que se centran solo en la clasificación, se presenta un sistema completo que ofrece un diagnóstico y una métrica de severidad.
2. **Estrategia de Generalización (Lab-a-Campo):** La aplicación de un pipeline de preprocesamiento (realce HSV/ $L^*a^*b^*$ ) combinado con data Augmentation, permitiendo que un modelo entrenado con imágenes de laboratorio (PlantVillage) pueda clasificar imágenes de entornos naturales y fondos complejos (removidos con un pipeline de segmentación).
3. **Metodología novedosa de segmentación foliar:** La combinación específica de SAM (para generación de máscaras) con Convex Hull refinado y operaciones morfológicas para aislar hojas de maíz, garantizando máscaras sólidas y continuas sin necesidad de entrenamiento de un modelo de segmentación customizado.

4. **Cuantificación adaptativa de lesiones:** El diseño de un módulo de cuantificación que utiliza la salida del clasificador (ya sea una enfermedad o una hoja sana) para seleccionar el perfil de detección de lesiones (rangos HSV) más adecuado para cada patología.

El resto de este documento se organiza de la siguiente manera: la Sección II revisa los trabajos previos en el área. La Sección III detalla el pipeline propuesto, desde el pre-procesamiento de datos hasta la cuantificación. La Sección IV describe la configuración experimental y las métricas utilizadas. La Sección V presenta y discute los resultados obtenidos. Finalmente, la Sección VI resume las conclusiones del trabajo y propone líneas futuras de investigación.

## II. TRABAJOS RELACIONADOS

La aplicación de la inteligencia artificial (IA) en la agricultura de precisión ha generado un amplio campo de investigación. Este trabajo toca tres dominios clave: la clasificación de enfermedades, la segmentación de imágenes agrícolas y la cuantificación de la severidad de dichas enfermedades.

### II-A. Clasificación de Enfermedades mediante Deep Learning

El uso de Redes Neuronales Convolucionales (CNNs) revolucionó la detección de enfermedades en plantas. Un trabajo pionero en esta área fue el de Mohanty et al. [3], quienes demostraron la alta efectividad de arquitecturas como AlexNet y GoogLeNet en el dataset PlantVillage [2], logrando precisiones superiores al 99 %.

Desde entonces, la investigación se ha centrado en arquitecturas más avanzadas y eficientes. Estudios recientes han comparado sistemáticamente modelos como ResNet [4], DenseNet [5] y EfficientNet [6], destacando la robustez del aprendizaje por transferencia (*Transfer Learning*) [7]. Específicamente para el maíz, diversos autores han aplicado estas técnicas con éxito [8], [9], confirmando la viabilidad de la clasificación automática. Sin embargo, la mayoría de estos trabajos se detienen en la clasificación, sin abordar la cuantificación del daño, y sin mostrar la eficacia de dichos modelos de clasificación en imágenes de entornos reales.

### II-B. Enfoques Clásicos y Híbridos

Además de los enfoques basados puramente en Deep Learning, una línea de investigación significativa ha sido el uso de métodos híbridos. Estos sistemas combinan técnicas clásicas de procesamiento de imágenes para la extracción manual de características (descriptores de color en HSV, características de textura como GLCM, o descriptores de forma) con clasificadores de Machine Learning, incluyendo Redes Neuronales Convolucionales menos profundas [12]. Si bien estos métodos han demostrado ser efectivos [12], requieren una ingeniería de características considerable y dependen en gran medida de la calidad de la segmentación y de los descriptores seleccionados.

A diferencia de enfoques tradicionales o híbridos (como los vistos en [12]), este trabajo no implementó una etapa de extracción manual de características (como descriptores

de color, textura GLCM o forma). Se optó por un enfoque *end-to-end* utilizando ResNet50V2, donde la red aprende automáticamente representaciones jerárquicas de características directamente desde las imágenes de entrada hasta la clasificación final. La hipótesis, validada por los resultados, fue que un modelo pre-entrenado y ajustado mediante *fine-tuning* sería capaz de aprender automáticamente representaciones más ricas y discriminativas que aquellas diseñadas manualmente, eliminando la necesidad de una ingeniería de características explícita y mejorando la generalización del modelo.

### II-C. Segmentación de Hojas y Lesiones

La cuantificación de la severidad depende de una segmentación precisa. Tradicionalmente, esto se lograba con métodos de procesamiento de imágenes clásicos, como la umbralización por color (ej. en espacios HSV o  $L^*a^*b^*$ ) y algoritmos de contorno [10]. Si bien son rápidos, estos métodos son poco robustos ante cambios de iluminación, sombras o fondos complejos.

Para superar esto, se han adoptado modelos de segmentación semántica basados en DL, siendo U-Net [11] la arquitectura dominante. U-Net y sus variantes pueden aprender a segmentar con precisión tanto la hoja como las lesiones, pero requieren una gran cantidad de datos de entrenamiento meticulosamente etiquetados a nivel de píxel (máscaras), lo cual es un cuello de botella significativo.

### II-D. Modelos de Fundación para Segmentación y el “Gap” de Cuantificación

La reciente aparición de Modelos de Fundación (*Foundation Models*) para visión, como el Segment Anything Model (SAM) [1], ofrece una alternativa prometedora. SAM puede segmentar cualquier objeto en una imagen sin necesidad de re-entrenamiento específico, operando en modo de “cero-disparo” (*zero-shot*). No obstante, SAM genera múltiples máscaras candidatas para una misma imagen, creando un nuevo desafío: cómo filtrar y refinar estas máscaras para aislar un objeto de interés específico, como una hoja.

Finalmente, el “gap” en la literatura reside en la integración de estas técnicas. Existen trabajos sobre clasificación y trabajos sobre segmentación, pero pocos presentan un pipeline completo y robusto que (1) clasifique la enfermedad, (2) segmente la hoja en condiciones variables sin un modelo de segmentación entrenado a medida, y (3) utilice la información de la clasificación para (4) cuantificar la severidad de esa enfermedad específica. Este trabajo busca llenar ese vacío.

## III. METODOLOGÍA

El sistema propuesto se implementa como un pipeline secuencial de dos ramas, diseñado para procesar una imagen de entrada y entregar tanto una clasificación de enfermedad como una cuantificación de su severidad (en caso de sufrir alguna enfermedad).

Se parte de un conjunto de imágenes de hojas de Maíz, las cuales pasan por una etapa de preprocesamiento, buscando mejorar las métricas en la posterior etapa de modelado, donde

se obtiene un modelo clasificador. Por la otra rama, a partir de una imagen de una hoja de maíz en estado natural, esta pasa por una etapa de preprocesamiento y su posterior segmentación, obteniendo una muestra lista para su clasificación. Para esto último consume el modelo clasificador previamente generado, una vez clasificada se procede por último con la cuantificación del área afectada de la hoja en cuestión.

### III-A. Dataset y División de Datos

El conjunto de datos base para el entrenamiento del modelo de clasificación es el subconjunto de maíz del repositorio público **New PlantVillage** [2] (Tabla I), que consiste en 3.852 imágenes en condiciones controladas. Este fragmento de imágenes del dataset representa apenas poco más del 7 % del total de las imágenes del dataset original, esto conlleva a prestar mucha atención sobre cómo tratar un dataset mucho más pequeño en cuanto a tamaño, y sobre todo que técnicas deben ser implementadas para mitigar el efecto de esto.

El proceso de preparación de datos fue el siguiente:

1. Las 3.852 imágenes originales fueron preprocesadas.
2. Se aplicó Aumento de Datos para generar un dataset balanceado de 6.000 imágenes (1.500 por clase).
3. Este conjunto de 6.000 imágenes se dividió de forma estratificada en 80 % (4.800 imágenes) para **Train** y 20 % (1.200 imágenes) para **Test**.

### III-B. Pipeline de Preprocesamiento y Aumento de Datos

Esta etapa es fundamental para (A) realzar las características patológicas sutiles y (B) asegurar que el modelo pueda generalizar de imágenes de laboratorio a imágenes de campo. El proceso se aplica secuencialmente:

**III-B1. Paso 1: Realce Multi-Espacio:** El preprocesamiento fue una etapa crítica, diseñada para estandarizar las imágenes y, lo más importante, realzar las características patológicas sutiles antes de alimentar al modelo. El pipeline consistió en una secuencia adaptativa:

- **Realce HSV (Condicional):** Se identificó que las lesiones de *Gray Leaf Spot* (Mancha Gris) eran particularmente sutiles, con tonos marrones/grises que se confundían fácilmente con el tejido sano. Para esta clase, y solo para esta clase, se aplicó un realce en el espacio HSV. El proceso implicó: (1) un incremento en la saturación (canal S) para intensificar los colores de la enfermedad y (2) la aplicación de CLAHE (con `clipLimit=1.5`, `tileGridSize=(6,6)`) al canal de valor (V) para mejorar el contraste local de las lesiones. Esta decisión fue clave, ya que aplicar este mismo realce a otras enfermedades (como *Common Rust*) resultaba perjudicial, distorsionando sus características.
- **Realce L\*a\*b\* (General):** Tras el paso condicional, todas las imágenes (incluidas las sanas y las de *Gray Leaf Spot* ya procesadas) se convirtieron al espacio de color CIE L\*a\*b\*. Se eligió este espacio porque permite manipular la luminancia (canal L\*) independientemente de los canales de color (a\* y b\*). El objetivo era mejorar el contraste general y definir mejor los bordes de las lesiones

sin alterar el color. Se aplicó **CLAHE** (Contrast Limited Adaptive Histogram Equalization) únicamente al canal L\*, usando `clipLimit=2.0` y `tileGridSize=(8,8)`.

- **Filtrado AMF:** La aplicación de CLAHE puede amplificar el ruido de fondo existente en la imagen. Para mitigar esto sin desenfocar los bordes de las lesiones (un efecto secundario común de los filtros de desenfoco estándar), se aplicó un **Filtro de Mediana Adaptativo** (AMF). Este filtro se aplicó sobre el canal L\* ya procesado por CLAHE, utilizando una ventana que se ajusta entre 3x3 y 7x7 (`min_window=3`, `max_window=7`) para eliminar ruido preservando los detalles finos.

Finalmente, los canales L\* (procesado con CLAHE y AMF), a\* y b\* (originales) se recombinaron, y la imagen resultante se convirtió de nuevo a RGB, lista para el aumento de datos y el entrenamiento. La imagen resultante (Figura 2) se reconvierte a RGB y se redimensiona a 256x256 píxeles.

**III-B2. Paso 2: Aumento de Datos (Data Augmentation):** Para reducir el sobreajuste y balancear el dataset, se aplica un robusto conjunto de transformaciones aleatorias a las imágenes de entrenamiento para alcanzar el objetivo de 1.500 muestras por clase. Los parámetros específicos incluyeron: rotación (hasta 45°), zoom (hasta 0.2), desplazamiento de ancho y alto (0.2), cizallamiento (0.2), volteo horizontal y variaciones de brillo (rango [1.1, 1.5]).

### III-C. Arquitectura y Entrenamiento del Clasificador

El módulo de clasificación utiliza una CNN basada en **ResNet50V2** [4] mediante Aprendizaje por Transferencia.

**III-C1. Arquitectura del Modelo:** Se carga la arquitectura ResNet50V2 pre-entrenada en ImageNet. Se realizó un *fine-tuning* descongelando las últimas **40 capas** de la red, mientras el resto de la base permaneció congelada. Se reemplaza el cabezal de clasificación por uno personalizado que consiste en una capa de GlobalAveragePooling2D seguida de capas Dense (768 y 384 neuronas) con activación 'relu', regularización L2 (factor 0.006) y BatchNormalization, y capas de Dropout (0.4, 0.3, 0.2) para mitigar el sobreajuste. La capa de salida es una Dense de 4 neuronas con activación 'softmax'.

**III-C2. Configuración de Entrenamiento:** El modelo se entrena para 250 épocas (`batch_size=16`) usando el optimizador Adam (tasa de aprendizaje inicial de 0.0007). La métrica principal monitoreada es el **Dice Coefficient**, y la función de pérdida es una combinación de *Categorical Crossentropy* y *Dice Loss*. Se emplean callbacks de *ReduceLROnPlateau* (paciencia de 18 épocas, factor de reducción 0.5, y `min_lr=6e-5`) y *EarlyStopping* (paciencia de 55 épocas) para optimizar el entrenamiento, guardando el modelo con el mejor `val_dice_coef`.

### III-D. Segmentación Híbrida de Hoja: SAM y Convex Hull

Para aislar la hoja del fondo, un paso crítico para la cuantificación y la clasificación, se implementa un pipeline híbrido robusto. El enfoque original sufría de discontinuidades y huecos en las máscaras generadas, por lo que se implementó



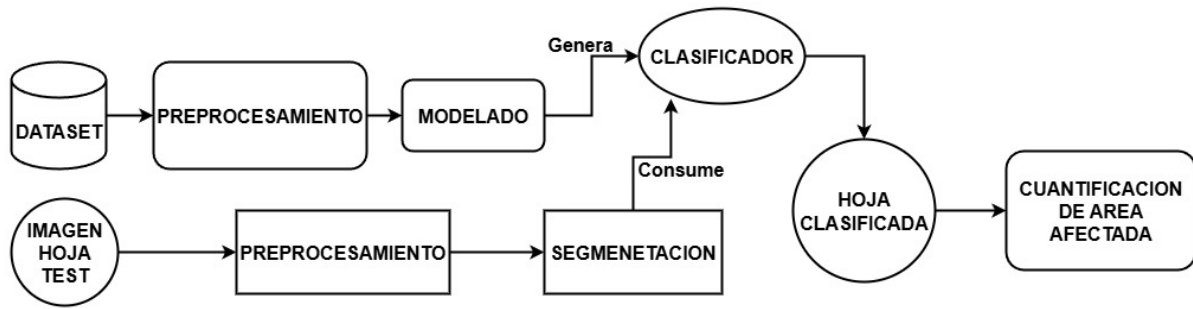


Figura 1. Diagrama de flujo del pipeline propuesto.

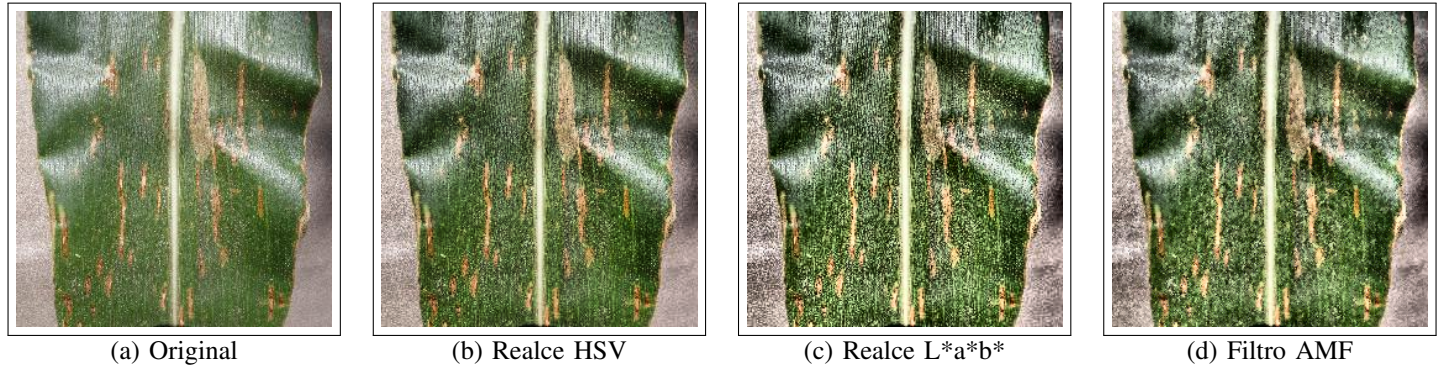


Figura 2. Visualización del pipeline de preprocesamiento. (a) Imagen original. (b) Realce condicional HSV para *Gray Leaf Spot*. (c) Realce L\*a\*b\* con CLAHE. (d) Imagen final tras Filtro de Mediana Adaptativo.

una corrección fundamental basada en el relleno del casco convexo y refinamiento morfológico. El proceso corregido es el siguiente:

1. **Generación de Máscaras (SAM):** La imagen de entrada se procesa con el `SamAutomaticMaskGenerator` del Segment Anything Model [1], el cual genera múltiples máscaras candidatas.
2. **Filtrado de Máscaras:** Se implementa un algoritmo heurístico para descartar máscaras irrelevantes (fondo, ruido). Se seleccionan máscaras basándose en el área (ni muy pequeña ni muy grande) y el *aspect ratio*, conservando la máscara de la hoja principal.
3. **Creación de Máscara Convexa Sólida:** Una vez seleccionada la máscara candidata, se procede a crear una máscara sólida mediante el siguiente proceso:
  - a) Se extrae el contorno exterior de la máscara SAM filtrada.
  - b) Se calcula el Convex Hull de dicho contorno.
  - c) Se genera una nueva máscara relleno completamente el polígono convexo mediante `cv2.fillPoly()`, eliminando así cualquier discontinuidad interna o hueco que tuviera la predicción original de SAM.
4. **Refinamiento Morfológico:** Para asegurar bordes limpios y eliminar artefactos, se aplican operaciones morfológicas sobre la máscara sólida:
  - a) **MORPH\_CLOSE** (kernel elíptico  $11 \times 11$ ): Cierra pe-

queños huecos residuales y suaviza bordes irregulares.

- b) **MORPH\_OPEN** (kernel elíptico  $3 \times 3$ ): Elimina ruido puntual externo.
- c) Finalmente, mediante análisis de componentes conexas, se retiene únicamente la región principal, descartando fragmentos aislados.
5. **Recorte y Relleno (Cropping & Padding):** Se utiliza la máscara convexa sólida resultante (no la máscara original de SAM) para recortar la hoja de la imagen original. El bounding box se calcula sobre el Convex Hull refinado, se aplica un margen adaptativo del 5 %, y la región resultante se extrae aplicando la máscara sólida para eliminar el fondo. Esta hoja aislada se redimensiona preservando la relación de aspecto y se aplica padding sobre fondo negro hasta alcanzar  $256 \times 256$  píxeles.

Este proceso asegura una máscara de segmentación continua, sólida y geométricamente coherente con la forma convexa esperada de una hoja de maíz.

### III-E. Cuantificación Adaptativa de Área Afectada

El módulo final calcula el porcentaje de área afectada. Este proceso es “adaptativo” porque utiliza la predicción del clasificador para seleccionar los parámetros de detección de lesiones.

1. **Entrada:** El módulo recibe la imagen segmentada (hoja con fondo eliminado) y la clase predicha previamente (ej. “Common Rust”).

2. **Umbral Específico por Enfermedad:** Se aplica un perfil de detección de color (umbrales) en el espacio HSV, el cual es específico para las características visuales de la enfermedad clasificada. Los perfiles de color (rangos de Tono, Saturación y Valor) se definieron empíricamente (ver Tabla II). Para *Gray Leaf Spot*, que es menos definida por el color, se utiliza un análisis de brillo en el canal L\* del espacio L\*a\*b\*.
3. **Limpieza de Máscara:** La máscara de lesión binaria resultante se refina mediante operaciones morfológicas: primero una **Apertura** (MORPH\_OPEN) para eliminar ruido pequeño (píxeles aislados), seguida de un **Cierre** (MORPH\_CLOSE) para rellenar pequeños huecos dentro de las lesiones detectadas.
4. **Cálculo de Severidad:** Se calcula el área total de la hoja (píxeles no negros) y el área total de la lesión (píxeles blancos en la máscara final). La severidad se expresa como:

$$\text{Severidad (\%)} = \frac{\text{Píxeles de Lesión}}{\text{Píxeles de Hoja (no negros)}} \times 100 \quad (1)$$

Cuadro II  
PERFILES DE COLOR HSV PARA DETECCIÓN DE LESIONES

Enfermedad	Tono (H)	Saturación (S)	Valor (V)
Common Rust	5–25	60–255	50–255
Northern LB	8–30	30–255	30–255

### III-F. Integración del Pipeline

El sistema completo (Figura 1) se orquesta como un proceso secuencial de dos ramas. Una imagen de prueba primero pasa por el pipeline de Preprocesamiento. Luego, la imagen procesada es enviada al Clasificador para obtener una predicción de clase. Paralelamente, la imagen original (o preprocesada) es enviada al módulo de Segmentación Híbrida (Sección III-D) para aislar la hoja. Finalmente, tanto la *hoja segmentada* como la *predicción de clase* se envían al módulo de Cuantificación Adaptativa (Sección III-E), que utiliza la predicción de clase para seleccionar el perfil de detección de lesiones correcto.

## IV. CONFIGURACIÓN EXPERIMENTAL

### IV-A. Entorno de Desarrollo

El pipeline completo, desde el preprocesamiento hasta la cuantificación, se desarrolló en lenguaje Python (v3.10). El preprocesamiento de imágenes y la detección de lesiones se implementaron utilizando **OpenCV**, mientras que la generación de datos sintéticos y el entrenamiento del modelo de Deep Learning se realizaron con el framework **TensorFlow** (v2.10) y la librería **Keras**. La segmentación de hojas se basó en el modelo *vit\_b* de **Segment Anything** [1].

### IV-B. Métricas de Evaluación

Dada la naturaleza dual del sistema (clasificación y cuantificación), se utilizaron diferentes métricas:

**IV-B1. Métricas de Clasificación:** Para evaluar el modelo de clasificación (ResNet50V2), la métrica principal monitoreada durante el entrenamiento fue el **Coefficiente de Dice (Dice Coefficient)**, optimizando para `val_dice_coef`. El Coeficiente de Dice es ideal para problemas de clasificación multiclase, ya que maneja bien el desbalance de clases y mide la superposición entre las predicciones y la realidad. Se calcula como:

$$\text{Dice} = \frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (2)$$

donde  $X$  es el conjunto de predicciones correctas y  $Y$  es el conjunto de etiquetas reales. Adicionalmente, se reportan las métricas estándar de *Accuracy*, *Precision*, *Recall* y *F1-Score* en el conjunto de Test.

Aquí se presenta una diferencia clave con respecto a otros trabajos de la materia que se basaban únicamente en la métrica *Accuracy*, la cual puede ser un poco engañosa a la hora de evaluar el modelo (ej: si el 95 % de píxeles son fondo, un modelo que siempre predice "fondo" tendría 95 % de accuracy pero sería inútil). Además el *Accuracy* puede ser alto solo prediciendo la clase mayoritaria, en cambio *Dice* es más robusto al ignorar los verdaderos negativos. De esta forma, termina siendo útil cuando se quiere medir qué tan bien el modelo captura la extensión real de la enfermedad en la hoja.

**IV-B2. Métricas de Cuantificación:** La evaluación del módulo de cuantificación se centra en el **Porcentaje de Área Afectada (%)** y el **Nivel de Severidad** (Leve, Moderada, Severa), derivado de dicho porcentaje.

## V. RESULTADOS Y DISCUSIÓN

### V-A. Rendimiento del Modelo de Clasificación

El modelo ResNet50V2, entrenado durante 127 épocas, demostró un rendimiento moderado. El entrenamiento se detuvo por *EarlyStopping* en la época 72 al no mejorar la métrica `val_dice_coef`.

Evaluado sobre el conjunto de Test (1.200 imágenes), el modelo alcanzó un **Dice Coeficiente global de 0.8044** y una **Accuracy de 0.8183**. La Tabla III muestra el rendimiento de Dice para cada clase, mostrando una alta efectividad en la detección de *Common Rust* y *Healthy*.

Cuadro III  
RESULTADOS DEL COEFICIENTE DE DICE POR CLASE (TEST SET)

Clase	Dice Coefficient
Tizón Foliar del Norte	0.7628
Mancha Gris de la Hoja	0.7422
Roya Común	0.8462
Hojas Sanas	0.8796
<b>Promedio</b>	<b>0.8077</b>

La matriz de confusión (Tabla IV) y el reporte de clasificación (Tabla V) muestran detalles importantes. El modelo es muy efectivo identificando hojas sanas (Recall 0.83, Precision 1.00). La principal fuente de confusión ocurre entre *Northern Leaf Blight* y *Gray Leaf Spot*, lo cual es esperado dada su

### Curvas de Entrenamiento y Validación - ResNet50V2

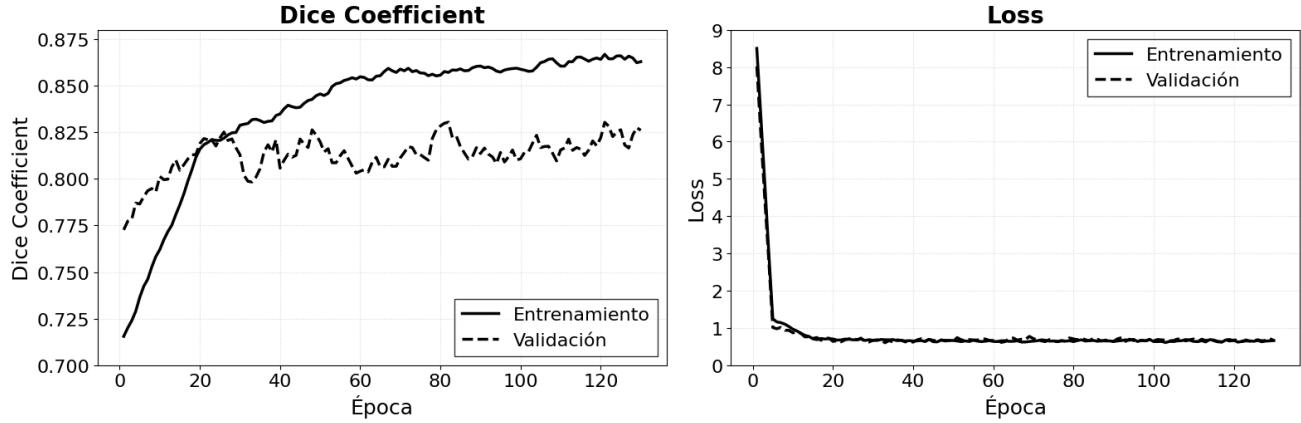


Figura 3. Curvas de entrenamiento y validación (Dice Coefficient y Loss) del modelo ResNet50V2.

similitud visual en ciertas etapas. También se observa una confusión, aunque más leve, entre *Common Rust* y *Gray Leaf Spot*. Esto era esperable debido a la similitud entre las enfermedades, que presentan ciertos rasgos que se repiten en estas 3 patologías.

Esta similitud visual, particularmente la presencia de lesiones necróticas o de coloración similar, es la causa principal de la confusión del modelo. La Figura 4 ilustra esta superposición de características entre las tres patologías, demostrando por qué la tarea de clasificación es inherentemente compleja.

Cuadro IV  
MATRIZ DE CONFUSIÓN

True Label	Predicted Label			
	Northern LB	Gray LS	Common Rust	Healthy
Northern LB	203	90	7	0
Gray LS	8	288	4	0
Common Rust	4	53	242	1
Healthy	2	48	1	249

Cuadro V  
REPORTE DE CLASIFICACIÓN (TEST SET)

Clase	Precision	Recall	F1-Score	Support
Northern LB	0.94	0.68	0.79	300
Gray LS	0.60	0.96	0.74	300
Common Rust	0.95	0.81	0.87	300
Healthy	1.00	0.83	0.91	300
<b>Weighted Avg</b>	<b>0.87</b>	<b>0.82</b>	<b>0.83</b>	<b>1200</b>

#### V-B. Resultados del Pipeline de Segmentación y Cuantificación

El pipeline híbrido de segmentación (SAM + Convex Hull + Refinamiento) demostró ser efectivo para aislar la hoja del fondo, incluso en imágenes de campo complejas. La introducción

de la máscara convexa sólida y las operaciones morfológicas eliminó exitosamente los huecos internos y suavizó los bordes irregulares que presentaban las máscaras originales de SAM. La Figura 5 ilustra este proceso de aislamiento y limpieza.

Posteriormente, el módulo de cuantificación adaptativa procesa la hoja segmentada. La Figura 6 ilustra el resultado final del pipeline: la imagen de entrada es correctamente clasificada (en este caso, *Northern Leaf Blight*) y el sistema aplica la máscara HSV correspondiente para identificar y calcular la severidad del daño.

#### V-C. Discusión

Los resultados confirman la viabilidad de un pipeline integral. La estrategia de Aumento de Datos fue clave para que el modelo pudiera clasificar imágenes naturales, cuyas condiciones son muy diferentes a las originales del dataset. El modelo ResNet50V2, pre-entrenado en ImageNet, demostró una buena capacidad de adaptación, y el *fine-tuning* con los datos aumentados fue esencial para especializarlo en la tarea.

A diferencia de enfoques tradicionales o híbridos (como los vistos en [12]), este trabajo no implementó una etapa de extracción manual de características (como descriptores de color, textura GLCM o forma). Se optó por un enfoque *end-to-end* utilizando una arquitectura CNN profunda (ResNet50V2). La hipótesis, validada por los resultados, fue que un modelo pre-entrenado y ajustado mediante *fine-tuning* sería capaz de aprender automáticamente representaciones jerárquicas de características más ricas y discriminativas que aquellas diseñadas manualmente, eliminando la necesidad de una ingeniería de características explícita y mejorando la generalización del modelo.

Otra diferencia importante con respecto a otros trabajos fue el uso de la métrica *Dice*. Dice es una métrica más sensible y representativa para problemas donde el área/región afectada importa, no solo la clasificación binaria de "correcto/incorrecto" que da accuracy. Es especialmente útil en casos como este donde las clases pueden estar desbalanceadas





Figura 4. Comparación visual de las tres enfermedades. (a) Tizón Foliar del Norte, (b) Mancha Gris de la Hoja, y (c) Roya Común. Se observa la similitud en la coloración y forma de las lesiones que confunde al clasificador.



Figura 5. Proceso de segmentación de hoja: (Izquierda) Imagen original de campo. (Centro) Máscara de hoja generada por SAM y refinada para ser sólida y convexa. (Derecha) Hoja aislada sobre fondo negro tras recorte y redimensionamiento.

o para medir qué tan bien el modelo captura la extensión real de la enfermedad en la hoja.

La combinación de SAM con Convex Hull (Sección III-D) resolvió el problema de la segmentación *zero-shot* de forma robusta. SAM por sí solo, si bien potente, producía múltiples máscaras de ruido y a menudo discontinuidades internas (huecos). El cambio fundamental de utilizar el Convex Hull no solo como referencia geométrica, sino como base para generar una **máscara sólida** mediante relleno de polígono, junto con el refinamiento morfológico, aseguró que la máscara final fuera continua y libre de huecos, mejorando la calidad del recorte y la posterior cuantificación.

Asimismo, el preprocesamiento demostró ser crítico. El

realce condicional en HSV solo fue efectivo para la clase *Gray Leaf Spot* y resultó perjudicial al aplicarse de forma general. Esto se debe a que las lesiones de *Gray Leaf Spot* son más sutiles (tonos marrones/grises) y se benefician del realce de saturación, mientras que este mismo realce distorsionaba las características de otras enfermedades.

Si bien el *pipeline* final fue robusto, el desarrollo encontró varias limitaciones. Se exploraron múltiples arquitecturas y técnicas que no produjeron mejoras, indicando la sensibilidad del problema. Por ejemplo, el uso de *mixup-augmentation* y la eliminación del Coeficiente de Dice de la función de pérdida degradaron significativamente el rendimiento. Modelos alternativos como Swin



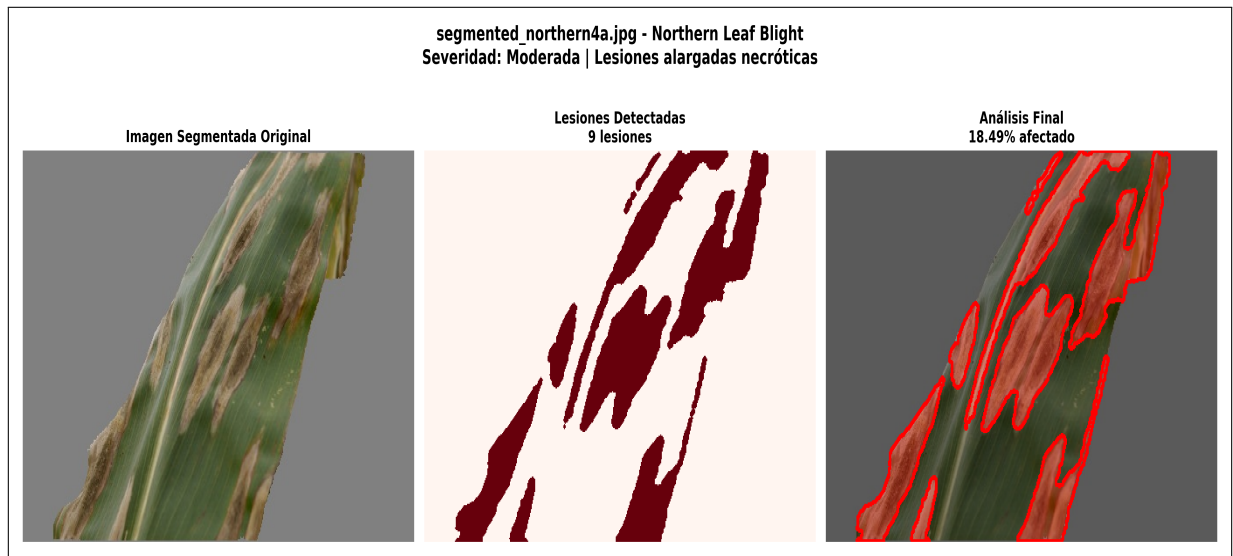


Figura 6. Resultado del pipeline de cuantificación para *Northern Leaf Blight*: (Izquierda) Hoja segmentada. (Centro) Máscara de lesión generada por umbral HSV. (Derecha) Visualización final con área afectada superpuesta (Overlay) y cálculo de severidad (20 % afectado).

Transformer y EfficientNetV2S no superaron al ResNet50V2, obteniendo un Dice mucho menor (0.552), lo que también invalidó las pruebas de *ensembling*. Además, técnicas como *Test-Time Augmentation* (TTA) y el uso de imágenes pre-segmentadas por SAM para el entrenamiento no mostraron beneficios y fueron descartadas.

Finalmente, el modelo ResNet50V2 presentado no fue una elección inicial, sino el resultado de una extensa exploración de arquitecturas e hiperparámetros. El proceso iterativo, fue fundamental para encontrar un equilibrio entre capacidad y generalización. Se partió de un modelo que sufría de alto *overfitting*, lo cual se corrigió exitosamente al reducir el *Learning Rate*, balancear la función de pérdida (50 % CE + 50 % Dice) e incrementar la regularización ( $L2=0.02$ ), logrando una mejora en la métrica Dice. El modelo final representa el balance óptimo encontrado, aumentando las capas de *fine-tuning* a 40 y ajustando la paciencia del *Early Stopping* a 55 épocas para lograr el Coeficiente Dice máximo de 0.8264.

## VI. CONCLUSIÓN Y TRABAJOS FUTUROS

En este trabajo se ha diseñado, implementado y validado un pipeline integral para el diagnóstico de enfermedades foliares en el maíz, capaz no solo de clasificar la enfermedad, sino también de segmentar la lesión y cuantificar su severidad. El sistema completo abarca desde un preprocesamiento de imagen multi-espacio (HSV,  $L^*a^*b^*$ ) y un robusto Aumento de Datos, hasta un módulo de clasificación basado en ResNet50V2, un componente de segmentación *zero-shot* avanzado (SAM + Convex Hull Sólido) y, finalmente, un sistema de cuantificación de severidad adaptativo.

### VI-A. Contribuciones y Logros Clave

Los resultados confirman la viabilidad del pipeline. La estrategia de **Aumento de Datos** fue clave para que el modelo ResNet50V2 pudiera generalizar y clasificar imágenes de

campo, cuyas condiciones difieren significativamente de las del dataset de entrenamiento.

La contribución metodológica principal reside en la exitosa integración de tres componentes clave. Primero, un **preprocesamiento multi-espacio** (combinando realce HSV condicional para *Gray Leaf Spot* con CLAHE y AMF en el espacio  $L^*a^*b^*$ ), que fue crítico para realzar las patologías. Segundo, la exitosa integración de este clasificador con un módulo de **segmentación zero-shot**, mejorado mediante la generación de máscaras convexas sólidas y refinamiento morfológico, que resolvió robustamente el aislamiento de la hoja sin huecos internos. Tercero, la **cuantificación adaptativa**, que utiliza la clasificación previa para seleccionar el perfil de detección de lesiones, cerrando eficazmente el ciclo de diagnóstico.

Finalmente, el rendimiento del sistema (superando un Coeficiente de Dice de 0.80) se alcanzó tras una **exploración iterativa** de arquitecturas. El modelo ResNet50V2 presentado demostró ser superior a otras alternativas exploradas como EfficientNetV2S y Swin Transformer, así como a técnicas descartadas como mixup-augmentation y TTA, confirmando el balance óptimo del pipeline final.

### VI-B. Trabajo Futuro

Si bien los resultados son alentadores, se identifican claras líneas de investigación futura para mejorar y extender este trabajo:

- **Aumentar el dataset con imágenes naturales clasificadas:** Aumentar la cantidad de imágenes con las cuales se entrena el dataset con imágenes reales que fueron segmentadas y clasificadas de forma correcta, para sumar un conjunto de imágenes con condiciones más cercanas a la realidad y no tanto a condiciones de laboratorio.
- **Optimización del Clasificador:** Explorar arquitecturas más recientes (ej. *Vision Transformers*) o realizar una

búsqueda de hiperparámetros más exhaustiva para el ResNet50V2.

- **Refinamiento de la Segmentación:** Investigar métodos adicionales para filtrar las máscaras generadas por SAM, incorporando información de color o textura para mejorar la precisión en lesiones complejas.
- **Validación Cuantitativa de la Severidad:** Comparar los porcentajes de área afectada calculados por el sistema con evaluaciones manuales realizadas por expertos agrónomos, estableciendo una métrica de error para la cuantificación.
- **Extensión a Otras Enfermedades y Cultivos:** Evaluar la adaptabilidad del pipeline para el diagnóstico en otras enfermedades del maíz o, incluso, en diferentes cultivos de importancia agrícola.
- **Despliegue Práctico:** Desarrollar una aplicación (móvil o web) que implemente este pipeline para su uso directo por agricultores o técnicos, optimizando los modelos para la inferencia en dispositivos con recursos limitados.

Este proyecto sienta las bases para futuros sistemas de diagnóstico agrícola que no solo identifiquen la presencia de una enfermedad, sino que también proporcionen una medida objetiva y automatizada de su impacto, facilitando así la toma de decisiones de manejo más eficientes e informadas. A su vez, permitiría generar informes de enfermedades de forma completa y simple a distancia, realizado por personas que no necesariamente tienen conocimiento del área.

#### REFERENCIAS

- [1] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, C. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, y R. Girshick, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.
- [2] D. P. Hughes y M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv preprint arXiv:1511.08060*, 2015.
- [3] S. P. Mohanty, D. P. Hughes, y M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 1419, 2016.
- [4] K. He, X. Zhang, S. Ren, y J. Sun, "Deep residual learning for image recognition," en *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [5] G. Huang, Z. Liu, L. Van Der Maaten, y K. Q. Weinberger, "Densely connected convolutional networks," en *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 4700–4708, 2017.
- [6] M. Tan y Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," en *Proc. Int. Conf. Machine Learning (ICML)*, pp. 6105–6114, 2019.
- [7] A. H. Ali, A. Youssef, M. Abdelal y M. A. Raja, "An ensemble of deep learning architectures for accurate plant disease classification," *Ecological Informatics*, vol. 81, 102553, 2024.
- [8] F. Khan, N. Zafar, M. N. Tahir, M. Aqib, H. Waheed, y Z. Haroon, "A mobile-based system for maize plant leaf disease detection and classification using deep learning," *Frontiers in Plant Science*, vol. 14, 2023.
- [9] I. Pacal, "Enhancing crop productivity and sustainability through disease identification in maize leaves: exploiting a large dataset with an advanced vision transformer model," *Expert Systems with Applications*, vol. 238, 2024.
- [10] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," en *Graphics Gems IV*, P. S. Heckbert, Ed. Boston, MA: Academic Press Professional, pp. 474–485, 1994.
- [11] O. Ronneberger, P. Fischer, y T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," en *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.
- [12] A. Kayad, J. Fernández, J. Suástegui, J. Trejo, y L. Hernández, "Segmentación y Extracción de Características de Color, Textura y Forma para la Detección de Enfermedades en Hojas de Maíz," *Research in Computing Science*, vol. 148, no. 7, pp. 195–206, 2019.
- [13] R. de Rossi, L. Couretot, P. Rago, G. Mercado Cárdenas, y A. Pou, "Principales enfermedades del maíz. Evaluación de híbridos en el norte de Buenos Aires," en *Actas Jornadas Técnicas de Maíz*, AAPRESID-INTA Pergamino, 2012.
- [14] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," *Biosystems Engineering*, vol. 172, pp. 84–91, 2018.