Comparativa de eficacia de los modelos de redes neuronales convolucionales UNet y PSPNet en la segmentación semántica de las hojas y frutos de plantas de jitomate

RESUMEN: Las enfermedades, plagas y deficiencias nutricionales en las plantas se detectan en las hojas y frutos. La detección precisa y no invasiva usando Deep learning mediante la segmentación semántica, que clasifica cada píxel de una imagen en clases mediante algún modelo de rede neuronal convolucional. En esta investigación se implementan y comparan los modelos UNet y PSPNet para segmentar píxeles de imágenes con plantas de jitomate en las clases: hojas, frutos o fondo. Los modelos se eligieron por su capacidad para capturar detalles finos y robustez en escenarios complejos. Un aporte distintivo es la clasificación multiclase, en comparación con una segmentación binaria. El rendimiento se evaluó con métricas como Accuracy, Precision, Recall, F1 e IoU. UNet presentó un promedio del 89.66% al segmentar hojas y 88.25% en frutos, mientras PSPNet obtuvo 87.50% y 81.34%, respectivamente. Los resultados resaltan el potencial de los modelos UNet y PSPNet para aplicaciones prácticas en agricultura de precisión, permitiendo la monitorización y manejo eficiente de cultivos mediante segmentación semántica de las hojas y frutos de los cultivos de jitomate.

PALABRAS CLAVE: Agricultura de Precisión, Deep Learning, Redes Neuronales Convolucionales Segmentación Semántica, Desempeño.

Colaboración
Juan Pablo Guerra Ibarra, Tecnológico Nacional de México campus Zamora; Francisco Javier Cuevas de la Rosa, Centro de Investigaciones en Óptica A.C; Lorenzo Rene Rojas Ruiz, Tecnológico Nacional de México, campus Zamora

Fecha de recepción: 07 de noviembre 2025
Fecha de aceptación: 12 de noviembre de 2025

ABSTRACT: Diseases, pests, and nutritional deficiencies in plants are detected in leaves and fruits. Accurate and non-invasive detection using deep learning through semantic segmentation, which classifies each pixel of an image into classes using a convolutional neural network model. In this research, the UNet and PSPNet models are implemented and compared to segment pixels from images of tomato plants into the following classes: leaves, fruits, or background. The models were chosen for their ability to capture fine details and their robustness in complex scenarios. A distinctive contribution is the multi-class classification, as opposed to binary segmentation. Performance was evaluated using metrics such as Accuracy, Precision, Recall, F1, and IoU. UNet averaged 89.66% when segmenting leaves and 88.25% for fruits, while PSPNet obtained 87.50% and 81.34%, respectively. The results highlight the potential of the UNet and PSPNet models for practical applications in precision agriculture, enabling efficient crop monitoring and management through semantic segmentation of tomato crop leaves and fruits.

KEYWORDS: Precision Agriculture, Deep Learning, Convolutional Neural Networks, Semantic Segmentation, Performance.

INTRODUCCIÓN

La agricultura es una actividad fundamental para la subsistencia humana, ya que gran parte de los alimentos que se consumen son generados por este medio [1], [2]. La experiencia del agricultor es fundamental en los procesos de riego, fertilización, detección de enfermedades y plagas. Sin embargo, la naturaleza humana puede llevar a procesos ineficientes. Durante el siglo XXI, la agricultura ha experimentado un gran desarrollo tecnológico en busca de mejorar la cantidad, calidad de los alimentos producidos, reducir costos y el impacto ambiental [3].

En los últimos años el uso de la Inteligencia Artificial (IA) y otras tecnologías aplicados al cultivo de la tierra han creado el concepto de Agricultura de Precisión (AP) [4]. Entre los objetivos de la AP destaca el proporcionar a los cultivos los recursos necesarios para su adecuado desarrollo, mejorando la relación costo-beneficio.

Los algoritmos de Deep Learning (DL) tienen un lugar importante en la AP. Un enfoque particular de DL consiste en el uso de diferentes modelos de Redes Neuronales Convolucionales (RNC) con variados objetivos [5]. En AP las RNC se han usado para apoyar diversas tareas de los agricultores. Wang [6] las empleo en la detección de enfermedades presentes en las hojas de árboles de peras, Latif [7] las uso RNC para la detección de condiciones no deseadas en plantíos de arroz. Shoaib [8] implemento las RNC para la detección y segmentación de plagas en hojas de tomate. De igual manera Wei [9] las implemento para la detección de frutos maduros en cultivos de tomate.

La Segmentación Semántica (SS) es el empleo de RNC para clasificar cada píxel de una imagen en una clase en particular [10], [11]. En la AP la SS se ha implementado por diferentes autores, por ejemplo, Kang presento un método de SS de ramas y manzanas utilizando un modelo de tipo piramidal [12]. Ni [13] trabajó con cultivos de arándanos para determinar el grado de madurez en imágenes con un fondo contrastado. Majeed [14] empleó los modelos de RNC FCN, VGG-16 y SegNet-VGG-16 para segmentar los tallos de las plantas de uva y tener un parámetro de control sobre su crecimiento.

Este trabajo se presenta y describen las etapas de entrenamiento y análisis de resultados de dos modelos de RNC UNet y PSPNet para realizar SS de las hojas y frutos de plantas de jitomates cultivadas en condiciones semihidropónicas.

La mayoría de los métodos descritos procesan imágenes adquiridas en condiciones de laboratorio y se centran en una clasificación binaria entre fondo y objeto de interés. Esto plantea interrogantes respecto al comportamiento de estos y otros métodos al aplicarse a imágenes provenientes de ambientes agrícolas reales, la viabilidad de efectuar una clasificación multiclase y cuál de todos los modelos disponibles ofrece el mejor desempeño en métricas de rendimiento como la F1-Score e IoU en las condiciones mencionadas.

En el contexto descrito anteriormente son dos las aportaciones de este trabajo. La primera es el uso de imágenes tomadas en condiciones no controladas, donde variables como iluminación, ángulos de toma y oclusiones de la lente no se controlan. Este enfoque representa un escenario más cercano a las aplicaciones reales en campo. El segundo aporte consiste en abordar la tarea de clasificación multiclase para distinguir hojas, frutos y fondo. Para potenciar el entrenamiento y resultado de los modelos de RNC UNet y PSPNet, además se implementa la técnica de Transfer Learning (TL).

La organización del documento es la siguiente: la Sección 2 describe los recursos utilizados durante el entrenamiento y la experimentación, incluyendo el software, hardware, conjunto de imágenes, su preprocesamiento y los valores de los hiperparámetros empleados. La Sección 3 presenta y analiza los resultados obtenidos por los modelos. Finalmente, la Sección 4 expone las conclusiones de la investigación.

MATERIAL Y MÉTODOS

La implementación de RNC para realizar tareas de SS requiere de elementos indispensables como lo es el conjunto de imágenes a procesar y las máscaras o anotaciones en donde se identifique lo que se desea aprenda la RNC. Existen otros componentes que es preferible contar con ellos como lo son tarjetas de video más no son indispensable.

Características del hardware y software empleados.
La implementación de los modelos de RNC se realizó con las características hardware y software listadas a continuación:

Fabricante: ASUSTeK COMPUTER INC., China.
Modelo: X510UNR
Procesador: Inte Core™ i7-8550U CPU @ 1.80GHz × 8.
RAM: 16 GB.
Tarjeta de video: NVIDIA® GeForce® 150MX.
Sistema operativo: Ubuntu 22.04.2 LTS 64 bits.
Versión Cuda tool kit: 10.1.243.
Versión Cudnn: 7.6.5.
Versión Tensorflow: 2.4.1.
Versión Keras: 2.4.3.
Versión OpenCV: 4.7.0

Conjunto de imágenes

El proceso de entrenamiento de los modelos de RNC se realizó con un conjunto de imágenes de cultivos de tomates rojos en invernaderos, el cual es de acceso público desde el link: https://www.kaggle.com/datasets/andrewmvd/tomato-detection [15]. El sitio de descarga no proporciona información referente a las condiciones de captura de las imágenes, sin embargo, se infiere por la diversidad entre ellas que no hay condiciones de captura controladas como lo es la iluminación, ángulo de captura de las imágenes y no se hace mención del tipo de dispositivo de captura empleado. El conjunto de imágenes consta de 895 elementos, de los cuales se seleccionaron aleatoriamente 300, y fueron asignas 180, 60 y 60 imágenes para formar los conjuntos de Entrenamiento, Validación y Prueba, el número de imágenes es acotado por el tiempo requerido durante el etiquetado. El tamaño de la muestra es de un 34% del total de imágenes. La distribución de las cantidades de imágenes en los conjuntos de Entrenamiento, Valida-

ción y Prueba se realizó con base en lo reportado en literatura, 60%, 20% y 20% respectivamente. La asignación de las imágenes y sus máscaras a los conjuntos se realizó empleando la función de tensorflow Split. Disponible en Google Drive [16].

Etiquetado de imágenes

El etiquetado o marcado de objetos de interés en una imagen es un proceso previo al entrenamiento y prueba de las RNC, también es usado para medir la eficiencia del SS realizado por las RNC. Los píxeles de Las imágenes seleccionadas fueron etiquetados usando la herramienta de marcado web "Computer Vision Annotation Tool" (CVAT), la cual accesible desde el link https://www.cvat.ai/ [17]. Los pixeles de las imágenes se marcaron con colores representativos a las clases de interés del estudio, verde para las hojas, rojo para los frutos y por default negro para el fondo.

El proceso de etiquetado de los píxeles de las imágenes genera una distribución de los píxeles es mostrada en la Figura 1.

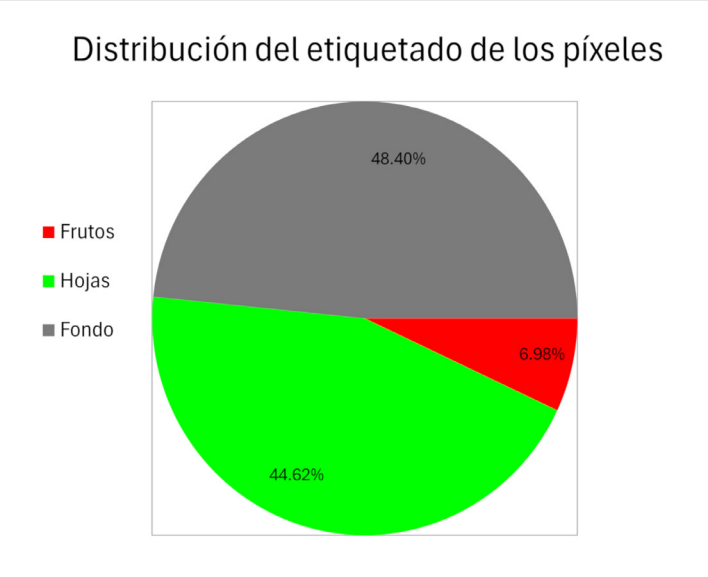

Figura 1. Distribución del etiquetado de los píxeles.
Fuente: Elaboración propia.

Preprocesamiento de imágenes

Las imágenes que conforman los conjuntos de Entrenamiento, Validación y Prueba, así como sus máscaras, tienen dimensiones de 500 × 400 píxeles y formato PNG. Para cumplir los requisitos de entrada de los modelos RNC, es requiere ajustar las dimensiones de las imágenes dependiendo del modelo, dicho ajuste de tamaño se especifica en la sección de los modelos seleccionados. Dentro del preprocesamiento se requiere para llevar a cabo el proceso entrenamiento la conversón de las máscaras a escala de grises, donde cada nivel de gris corresponde a una clase. Los procesos mencionados se realizaron con la librería OpenCV.

La Figura 2 muestra de izquierda a derecha: la Imagen original con plantas de jitomate, máscara generada con el proceso de etiquetado y la imagen resultante del preprocesamiento requerido para entrenar los modelos RNC seleccionados.

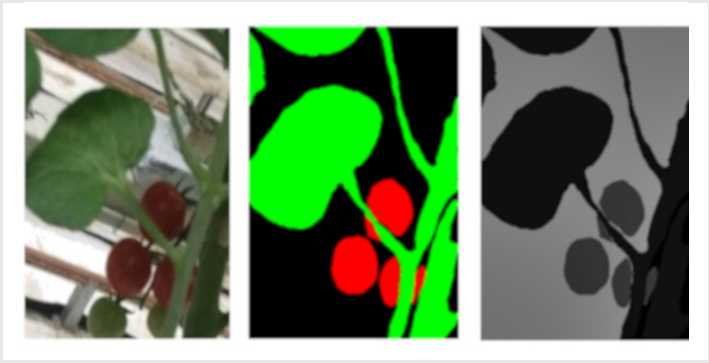

Figura 2. Ejemplos: Imagen original, máscara, preprocesada.
Fuente: Elaboración propia.

Métricas de rendimiento.

Es fundamental medir el rendimiento en la tarea de SS de las hojas y frutos realizada por los modelos RNC. Las métricas de Accuracy, Precision, Recall, F1-Score e IoU se utilizan con la finalidad de tener un parámetro cuantitativo del desempeño de las RNC.

Para definir las métricas mencionadas, es necesario definir cuatro conceptos básicos: Verdaderos Positivos (VP), Verdaderos Negativos (VN), Falsos Positivos (FP) y Falsos Negativos (FN), observe la Figura 3.

Figura 3. Matriz de confusión.
Fuente: Elaboración propia.

Usando la matriz de confusión junto con los resultados de la SS realizada por las RNC y las máscaras de las imágenes es posible asignar cada píxel de una imagen a un cuadrante de la matriz de confusión.

La métrica Accuracy es la relación del número total de píxeles clasificados correctamente entre el total de clasificaciones realizadas por el modelo de RNC y se define de acuerdo con la Ecuación 1.

$$Accuracy = \frac{VP+VN}{VP+VN+FP+FN} . \qquad \text{Ec. (1)}$$

La métrica Precisión representa el ratio entre el número total de PV con respecto al número de positivos predichos, esta es definida por la Ecuación 2.

$$Precision = \frac{VP}{VP+FP} \quad . \qquad \text{Ec. (2)}$$

El número de VP clasificados por el modelo en relación con el número total de clasificaciones positivas es medida con la métrica Recall, la cual se expresa de acuerdo con la Ecuación 3.

$$Recall = \frac{VP}{VP+FN} \quad . \qquad \text{Ec. (3)}$$

F1-Score combina las métricas Precision con Recall y es definida por la Ecuación 4.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} . \qquad \text{Ec. (4)}$$

El ratio entre los verdaderos positivos entre el total de clasificaciones correctas se mide en la métrica IoU y es definida en la Ecuación 5.

$$IoU = \frac{VP}{VP + FP + FN} . \qquad \text{Ec (5)}$$

Transfer Learning

En los procesos de entrenamiento con RNC es preferible contar con grandes cantidades de datos. En caso de contar con una cantidad limitada de elementos para realizar el entrenamiento, es conveniente usar una alternativa para aumentar la eficiencia de las RNC. Una forma de hacerlo es implementar lo que se denomina TL, la cual toma el conocimiento adquirido de modelos de RNC entrenados con grandes volúmenes de datos y con hardware altamente especializado y usarlo como una capa adicional antes del modelo final a entrenar.

Modelos de RNC seleccionado

En DL existe una variedad de modelos de RNC, los cuales tienen gran diversidad de objetivos. En este trabajo se implementan los modelos UNet y, PSPNet con un backbone de ResNet mediante TL para mejorar la SS de las hojas y frutos de jitomate en las imágenes descrita.

Modelo de RNC UNet

En 2015 Ronneberger [18], presentó el modelo UNet para realizar la SS de imágenes médicas. Este modelo destaca debido a que el 100% de sus capas son convolucionales.

El modelo UNet consta de dos etapas: la primera se encarga de codificar la información dentro de la imagen para comprender el contexto y la segunda etapa tiene por objeto ampliar simétricamente la información codificada de la primera etapa para localizar con precisión los píxeles mediante una serie de convoluciones transpuestas. La Figura 4 muestra la estructura de RNC UNet.

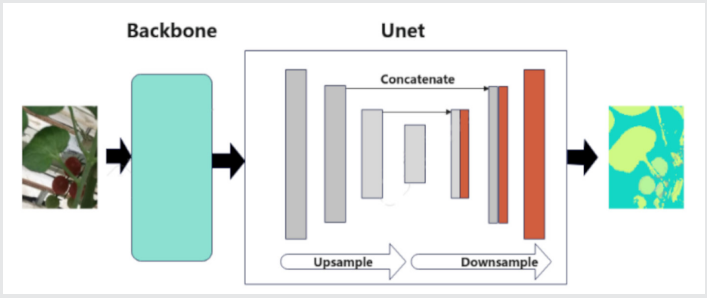

Figura 4. Estructura del modelo UNet con TL.
Fuente: Elaboración propia.

Para el modelo de RNC UNet se requiere ajustar las imágenes pertenecientes a los conjuntos de Entrenamiento, Validación y Prueba a 256 × 256 píxeles.

Modelo de RNC PSPNet

El modelo de RNC denominado PSPNet, fue presentado por Zhao [19], está compuesto por serie de capas convolucionales profundas, para ser aplicadas en la SS de imágenes de entornos urbanos. El modelo PSPNet integra un módulo de agrupación piramidal para aumentar la información contextual en el proceso de segmentación. La estructura básica del modelo se observa en la Figura 5.

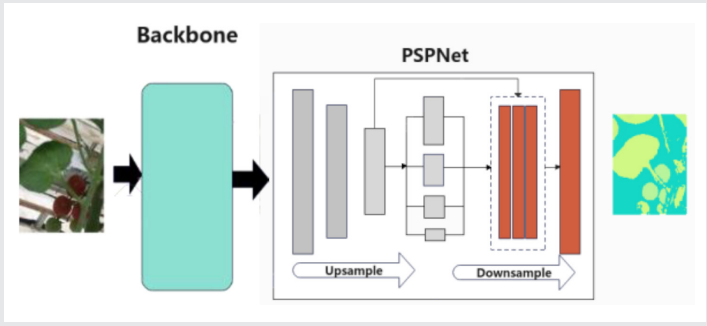

Figura 5. Estructura del modelo PSPNet con TL.
Fuente: Elaboración propia.

Se redimensionaron las imágenes que conforman los conjuntos de Entrenamiento, Validación y Prueba a 384 × 576 píxeles necesaria para cumplir con la configuración del modelo PSPNet.

La Figura 4 y Figura 5 muestran el backbone del modelo ResNet antes de las RNC seleccionadas. Este conocimiento o pesos del backbone están disponibles en la página web oficial de Keras, accesible desde el link: https://keras.io/api/applications/ [20].

Entrenamiento de los modelos

El objetivo del proceso de entrenamiento es permitir que el modelo de RNC generalice la información proporcionada por el etiquetado de las clases y posteriormente marque los píxeles de las imágenes del conjunto de Prueba, en nuestro caso, es asignar cada píxel a una de las tres clases de interés.

El proceso de entrenamiento ajusta los pesos que conectan las capas convoluciones con que está construido el modelo de RNC. El comportamiento de esta etapa controla con los valores de los hiperparámetros de entrenamiento. La configuración empleada de hiperparámetros es la siguiente:

Función de perdida: Categorical Cross-Entropy.
Optimizador usado: Adam
Épocas: 75.
Tasa de aprendizaje: 0.05.
BATCH: 16.
encoder_freeze: True.
Cantidad de elementos en los requeridos conjuntos:
Entrenamiento 180.
Validación 60
Prueba 60.

Los valores de los hiperámetros se determinaron en base a las siguientes consideraciones: la función de pérdida Categorical Cross-Entropy es adecuada para segmentación multiclase. El optimizador Adam proporciona un entrenamiento eficiente y estable. Se definieron 75 épocas para asegurar la convergencia sin sobreajuste.

La tasa de aprendizaje fue establecida con mediante experiencia, no se usó sheduler para adaptar el valor asignado. Él tamaño de batch de 16 es debido a la capacidad hardware empleado. Finalmente, el congelamiento del encoder aprovecha pesos preentrenados en ResNet, facilitando el aprendizaje ante un conjunto de datos reducido. Es importante mencionar que no se utilizó técnica alguna referente al early stopping.

RESULTADOS

El entrenamiento de los modelos UNet y PSPNet se realizó con los hiperparámetros y conjuntos de imágenes mencionados anteriormente. La Figura 6 y Figura 7 muestran el comportamiento de la función Accuracy resultado del proceso de entrenamiento con las imágenes de los conjuntos de Entrenamiento y Validación con las modelos de RNC UNet y PSPNet respectivamente.

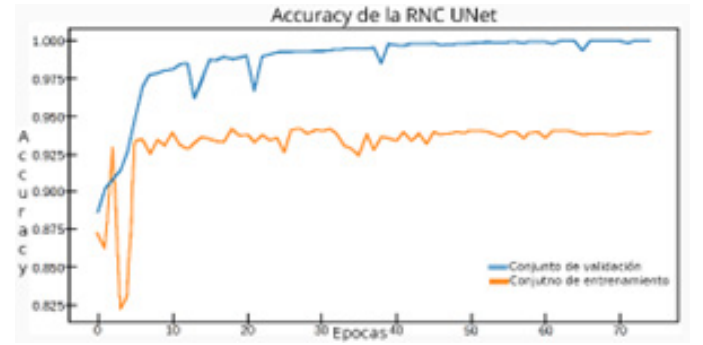

Figura 6. Comportamiento del Accuracy de la RNC PSPNet. Fuente: Elaboración propia.

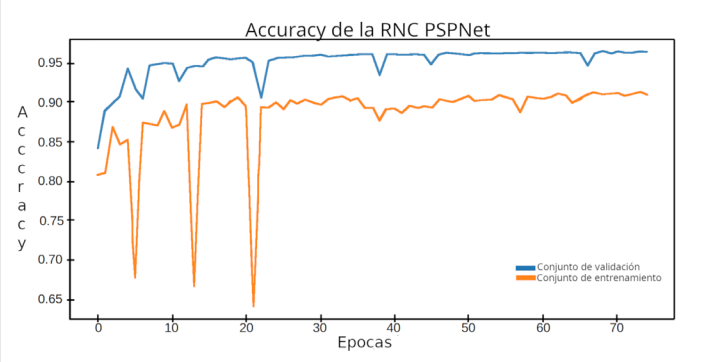

Figura 7. Comportamiento del Accuracy de la RNC PSPNet. Fuente: Elaboración propia.

El comportamiento de la función Accuracy en el proceso de entrenamiento mostrado en la Figura 6 y Figura 7 para las RNC UNet y PSPNet, se aprecia una serie de saltos en las primeras 20 épocas, estabilizándose en el resto del proceso de aprendizaje de los patrones de hojas y frutos. Cuantitativa ambos modelos alcanzan en el conjunto de entrenamiento valores superiores a 0.95, mientras que el rendimiento es inferior con el conjunto de validación, lo que es considerado como un comportamiento normal en el proceso de entrenamiento.

Un ejemplo del resultado de la SS realizada por ambos modelos se muestra en la Figura 8.

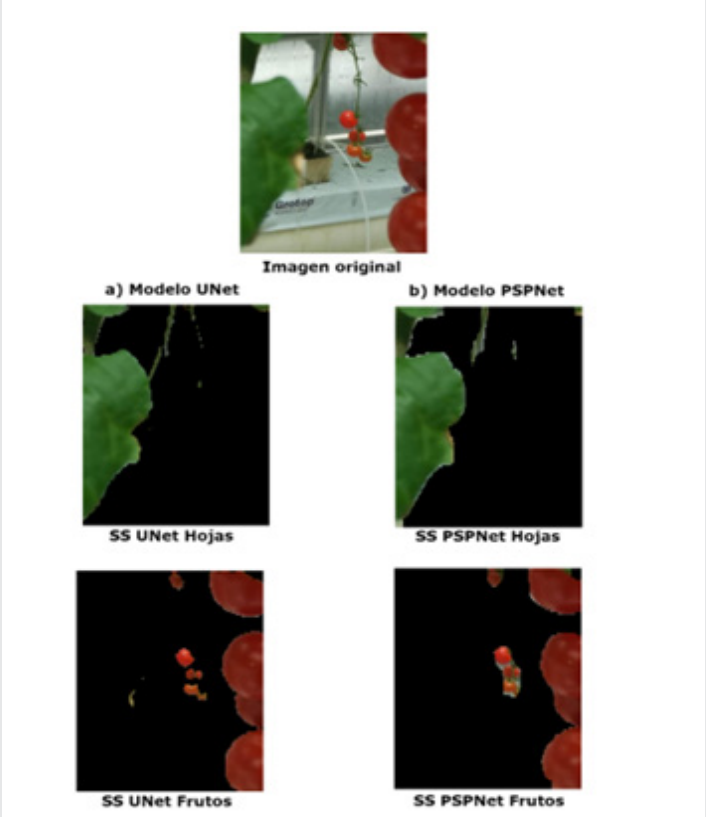

Figura 8. Ejemplos de SS realizada por los modelos. a) Resultados de la SS del modelo UNet. b) Resultado de la SS del modelo PSPNet.
Fuente: Elaboración propia

La Tabla 1 muestra los resultados cuantitativos con las métricas de desempeño en la SS de las imágenes del conjunto de Prueba.

Tabla 1 Resultados de las métricas de rendimiento al segmentar las imágenes del conjunto de prueba.

Precision	85.09%	79.21%	86.04%	84.47%
Recall	93.19%	82.84%	95.86%	91.97%
F1-score	87.59%	80.13%	89.58%	87.70%
IoU	80.19%	68.56%	83.20%	78.90%
Promedio	87.50%	81.34%	89.66%	88.25%

Fuente: Elaboración propia.

Los resultados de la Tabla 1, muestran un desempeño superior del modelo UNet en relación con el modelo PSPNet en la SS de las imágenes del conjunto de Prueba.

De la Tabla 2 a la Tabla 5 muestran datos de la prueba t de Student para muestras emparejadas de dos colas con una confiabilidad del 95% (α=0.05), realizado sobre los valores F1-Score e IoU obtenidos en la SS realizada por los modelos probados.

El análisis estadístico basado en pruebas t emparejadas de las cuatro Tablas anteriores realizadas sobre las métricas F1-Score e IoU para las clases hojas y frutos demuestra que las diferencias entre las medias los modelos UNet y PSPNet son estadísticamente significativas, dado que los valores p asociados son menores al nivel de significancia α=0.05. Lo anterior reafirma la superioridad del modelo UNet en SS de hojas y frutos en imágenes de plantas de jitomate.

Tabla 2 Prueba t Student emparejada métrica F1-Score SS Hojas.

T Student métrica F1 SS de Hojas		
Parámetro	PSPNet	UNet
Media	87.59	89.58
Varianza	223.64	204.22
Observaciones	60	60
Diferencia Hipotética media	0	
Estadístico	**Valor**	
Estadístico t	-4.26	
Grados de libertad	59	
Valor critico t 2 colas	7.99E-05	
P-valor 2 colas	2.00	

Fuente: Elaboración propia.

Tabla 3 Prueba t Student emparejada métrica IoU SS Hojas.

T Student métrica IoU SS de Hojas		
Parámetro	PSPNet	UNet
Media	80.19	83.20
Varianza	303.69	277.04
Observaciones	60	60
Diferencia Hipotética media	0	
Estadístico	**Valor**	
Estadístico t	-4.20	
Grados de libertad	59	
Valor critico t 2 colas	9.91E-05	
P-valor 2 colas	2.00	

Fuente: Elaboración propia.

Tabla 4 Prueba t Student emparejada métrica F1-Score SS Frutos.

T Student métrica F1 SS de frutos		
Parámetro	PSPNet	UNet
Media	80.13	87.70
Varianza	167.15	61.71
Observaciones	60	60
Diferencia Hipotética media	0	
Estadístico	**Valor**	
Estadístico t	-6.07	
Grados de libertad	59	
Valor critico t 2 colas	1.03E-07	
P-valor 2 colas	2.00	

Fuente: Elaboración propia.

Tabla 5 Prueba t Student emparejada métrica IoU SS Frutos.

T Student métrica IoU SS de frutos		
Parámetro	PSPNet	UNet
Media	68.56	78.90
Varianza	263.50	137.92
Observaciones	60	60
Diferencia Hipotética media	0	
Estadístico	**Valor**	
Estadístico t	-7.01	
Grados de libertad	59	
Valor critico t 2 colas	2.86E-09	
P-valor 2 colas	2.00	

Fuente: Elaboración propia.

CONCLUSIONES

La comparación cuantitativa entre los modelos UNet y PSPNet demuestra que el modelo UNet supera a PSPNet en métricas clave. Para la SS de hojas, UNet alcanzó un promedio de 89.58% en F1-Score y 83.20% en IoU, superiores a los 87.59% y 80.19% obtenidos por el modelo PSPNet, respectivamente. En la SS de frutos, UNet reportó un promedio de 87.70% en F1-Score y 78.90% en IoU, superando al modelo PSPNet con 80.13% y 68.56% respectivamente. Además, las pruebas t de Student confirman estadísticamente la superioridad de las medias del modelo UNet por sobre PSPNet dentro del contexto manejado en la investigación.

Los resultados y el análisis presentado muestran el potencial de las RNC al realizar la SS de los pixeles de las hojas y frutos en imágenes de plantas de jitomate. Lo anterior plantea la posibilidad de desarrollar un modelo propio basado en la UNet o bien probar con diferentes backbones que mejor los resultados obtenidos.

El mejoramiento del desempeño de los modelos evaluados se abordará con múltiples enfoques. El primero es mejorar el proceso de entrenamiento aumentando el tamaño del conjunto de imágenes disponibles, ya sea etiquetando otras imágenes de la fuente o mediante la técnica de "data augmentation".

Al ser un conjunto pequeño de imágenes se plantea la posibilidad de implementar validación cruzada mediante k-fold para evaluar el desempeño de los modelos probados, permitiendo obtener estimaciones más robustas. De la misma manera se pueden usar algunas técnicas que disminuyan el riesgo de sobre ajuste, como lo es el early stopping y sheduler.

El error humano en el etiquetado de las imágenes al ser un proceso manual es un factor que afecta el entrenamiento de los modelos y al posterior cálculo de las métricas, lo que expone la necesidad de revisar a detalle este proceso de etiquetado, mediante una validación cruzada con varios etiquetadores, guiados por de expertos en fisiología vegetal.

Las imágenes resultantes de la SS de las hojas obtenidas mediante los modelos evaluados pueden ser utilizadas como entradas a otros métodos reportados en la literatura como los cuales realizan la detección e identificación de plagas y enfermedades en las plantas. Estos métodos requieren imágenes de hojas libres de otros elementos que se encuentran de manera habitual como parte del fondo de estas. Por otro lado, las imágenes segmentadas de los frutos de jitomate mediante SS pueden ser empleadas para desarrollar estimadores de cosecha o de grado de madurez de los frutos, contribuyendo así a aplicaciones prácticas en el monitoreo y gestión agrícola.

AGRADECIMIENTOS

Sin el apoyo de las Instituciones que nos permiten contribuir a generar conocimiento no sería posible la presentación de los resultados de esta investigación, por lo anterior se agradece a:
Secretaría de Ciencia, Humanidades, Tecnología e Innovación.
Tecnológico Nacional de México.
Centro de Investigaciones en Óptica A.C

BIBLIOGRAFÍA

[1] Y. Awasthi, "Press 'A' for Artificial Intelligence in Agriculture: A Review," Int. J. Informatics Vis., vol. 4, no. 3, pp. 112–116, 2020, doi: 10.30630/joiv.4.3.387.

[2] P. P. Ray, "Internet of things for smart agriculture: Technologies, practices and future direction," J. Ambient Intell. Smart Environ., vol. 9, no. 4, pp. 395–420, 2017, doi: 10.3233/AIS-170440.

[3] G. Mohyuddin, M. A. Khan, A. Haseeb, S. Mahpara, M. Waseem, and A. M. Saleh, "Evaluation of Machine Learning Approaches for Precision Farming in Smart Agriculture System: A Comprehensive Review," IEEE Access, vol. 12, no. May, pp. 60155–60184, 2024, doi: 10.1109/ACCESS.2024.3390581.

[4] R. Bongiovanni, E. C. Mantovani, S. Best, and A. Roel, Introducción a la Agricultura de precisión. Procisur/IICA San José, Costa Rica, 2006.

[5] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Comput. Electron. Agric., vol. 147, no. July 2017, pp. 70–90, Apr. 2018, doi: 10.1016/j.compag.2018.02.016.

[6] H. Wang, J. Ding, and S. He, "MFBP-UNet: A Network for Pear Leaf Disease Segmentation in Natural Agricultural Environments," Plants, vol. 12, no. 18, p. 3209, Sep. 2023, doi: 10.3390/plants12183209.

[7] G. Latif, S. E. Abdelhamid, R. E. Mallouhy, J. Alghazo, and Z. A. Kazimi, "Deep Learning Utilization in Agriculture: Detection of Rice Plant Diseases Using an Improved CNN Model," Plants, vol. 11, no. 17, p. 2230, Aug. 2022, doi: 10.3390/plants11172230.

[8] M. Shoaib, T. Hussain, and B. Shah, "Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease," Front. Plant Sci., vol. 13, no. October, pp. 1–18, Oct. 2022, doi: 10.3389/fpls.2022.1031748.

[9] J. Wei et al., "Tomato ripeness detection and fruit segmentation based on instance segmenta-

tion," Front. Plant Sci., vol. 16, no. May, pp. 1–19, 2025, doi: 10.3389/fpls.2025.1503256.

[10] S. Asgari Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: a review," Artif. Intell. Rev., vol. 54, no. 1, pp. 137–178, Jan. 2021, doi: 10.1007/s10462-020-09854-1.

[11] S. Hao, Y. Zhou, and Y. Guo, "A Brief Survey on Semantic Segmentation with Deep Learning," Neurocomputing, vol. 406, pp. 302–321, Sep. 2020, doi: 10.1016/j.neucom.2019.11.118.

[12] H. Kang and C. Chen, "Fruit Detection and Segmentation for Apple Harvesting Using Visual Sensor in Orchards," Sensors, vol. 19, no. 20, p. 4599, Oct. 2019, doi: 10.3390/s19204599.

[13] X. Ni, C. Li, H. Jiang, and F. Takeda, "Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield," Hortic. Res., vol. 7, no. 1, p. 110, Dec. 2020, doi: 10.1038/s41438-020-0323-3.

[14] Y. Majeed, M. Karkee, Q. Zhang, L. Fu, and M. D. Whiting, "Determining grapevine cordon shape for automated green shoot thinning using semantic segmentation-based deep learning networks," Comput. Electron. Agric., vol. 171, no. November 2019, p. 105308, Apr. 2020, doi: 10.1016/j.compag.2020.105308.

[15] "Tomato Detection | Kaggle." https://www.kaggle.com/datasets/andrewmvd/tomato-detection (accessed Oct. 21, 2022).

[16] "DataSet - Google Drive." https://drive.google.com/drive/folders/1a16xRsQaEyVVe3g-b5K46aRBSjRJhJ3iv (accessed Nov. 06, 2025).
[17] "CVAT." https://www.cvat.ai/ (accessed Oct. 21, 2022).

[18] W. Weng and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," IEEE Access, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.

[19] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, vol. 2017-Janua, pp. 6230–6239, doi: 10.1109/CVPR.2017.660.

[20] "Keras: the Python deep learning API." https://keras.io/ (accessed Mar. 09, 2022).