



UNIVERSIDAD AUTÓNOMA CHAPINGO

**POSGRADO EN INGENIERÍA AGRÍCOLA Y USO
INTEGRAL DEL AGUA.**

**MONITOREO FENOLÓGICO DE LECHUGA MEDIANTE CNN EN
FÁBRICA DE PLANTAS CON ILUMINACIÓN ARTIFICIAL (PFAL)**

TESIS

Que como requisito parcial para obtener el grado de:

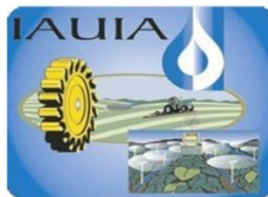
MAESTRO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA.

Presenta:

LUIS GERARDO CHÁVEZ COVARRUBIAS

Bajo la supervisión de:

DRA. RAQUEL SALAZAR MORENO



Chapingo, Estado de México, noviembre de 2025

**MONITOREO FENOLÓGICO DE LECHUGA MEDIANTE CNN EN
FÁBRICA DE PLANTAS CON ILUMINACIÓN ARTIFICIAL (PFAL)**

Tesis realizada por el C. Luis Gerardo Chávez Covarrubias, bajo la
dirección del comité asesor indicado, aprobada por el mismo y aceptada
como requisito parcial para obtener el grado de:

Maestro en Ingeniería Agrícola y uso Integral del Agua.

Director: _____


DRA. RAQUEL SALAZAR MORENO

Asesor: _____


DR. LUIS ARTURO SORIANO AVENDAÑO

Asesor: _____


DR. IRINEO LORENZO LÓPEZ CRUZ

AGRADECIMIENTOS

A la Universidad Autónoma Chapingo

Al posgrado de Ingeniería Agrícola y Uso Integral del Agua, por brindarme las herramientas para realizar esta investigación.

A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti) por el apoyo monetario durante mis estudios de maestría.

A mi comité asesor por compartirme su conocimiento y amor por la ciencia.

DATOS BIOGRÁFICOS



Datos personales:

Nombre: Luis Gerardo Chávez Covarrubias

Fecha de nacimiento: 2 enero 1998

Lugar de nacimiento: Zacatipán, San Luis Potosí

Curp: CACL980102HSPHVS03

Profesión: Ingeniero Mecánico Agrícola

Cédula Profesional: 13763626

DESARROLLO ACADEMICO:

Preparatoria: Colegio de Estudios Científicos y Tecnológicos del Estado de Hidalgo Plantel Santa Ana de Allende Hidalgo.

Licenciatura: Universidad Autónoma Chapingo.

CONTENIDO

CONTENIDO	iv
LISTA DE CUADROS	viii
LISTA DE FIGURAS	ix
RESUMEN GENERAL	xii
ABSTRACT	xiii
CAPÍTULO 1. INTRODUCCIÓN GENERAL	14
1.1 Objetivos:.....	16
1.1.1 Objetivos Generales:	16
1.1.2 Objetivos específicos:.....	17
1.1.3 Hipótesis:.....	17
1.1.4 Estructura de la tesis	17
1.2 Literatura citada	18
CAPÍTULO 2. REVISIÓN DE LITERATURA	21
2.1 Fábrica de plantas:	21
2.2 La lechuga:	21
2.2.1 La lechuga en sistemas PFALs	22

2.2.2	Lechuga variedad francesa	22
2.2.3	Etapas Fenológicas de la lechuga:.....	23
2.3	Fenotipado no destructivo	24
2.3.1	Técnicas de fenotipado no destructivo	24
2.3.2	Plataformas de fenotipado no destructivo:	25
2.4	Redes neuronales convolucionales de segmentación:	27
2.4.1	Yolo (You Only Look once)	28
2.4.2	YOLOV8n-seg y YOLOv11n-seg:	29
2.5	Investigaciones recientes de fenotipado.....	30
2.6	Referencias	33
CAPÍTULO 3. APLICACIÓN DE YOLOV8N-SEG Y YOLOV11N-SEG PARA LA SEGMENTACIÓN Y EXTRACCIÓN DE CARACTERÍSTICAS MORFOLÓGICAS EN LECHUGA CULTIVADA EN PFAL.....		39
3.1	Resumen	39
3.2	INTRODUCCIÓN.....	40
3.3	MATERIALES Y MÉTODOS	42
3.3.1	Adquisición y procesamiento de imágenes	43
3.3.2	YOLOv8n-Seg y YOLO11: Descripción General	45

3.3.3	Métricas de evaluación para las redes de segmentación:	48
3.3.4	Entrenamiento de los modelos de segmentación	51
3.4	RESULTADOS Y DISCUSIÓN	52
3.4.1	Evaluación de las funciones de pérdida.	52
3.4.2	Evaluación de las métricas de rendimiento	56
3.4.3	Evaluación (Test):	57
3.4.4	Inferencias:	61
3.4.5	Discusión:	69
3.5	Conclusiones	71
3.6	Literatura citada	73
CAPÍTULO 4. PREDICCIÓN MULTIMODAL DE BIOMASA Y ÁREA FOLIAR DE LECHUGA MEDIANTE VISIÓN RGB-D Y CNN-ML		76
4.1	Resumen	76
4.2	INTRODUCCIÓN	77
4.3	Materiales y métodos	79
4.3.1	Sitio experimental y manejo del cultivo	79
4.3.2	Adquisición y procesamiento de datos RGB-D	81
4.3.3	Flujo de procesamiento RGB-D	83

4.3.4	Cálculo del centroide y determinación de la altura precisa.....	84
4.3.5	Características morfológicas extraídas de las imágenes RGB-D	86
4.3.6	Modelos de predicción:.....	88
4.3.7	Modelos híbridos CNN + ML	88
4.3.8	Métricas de desempeño	89
4.4	Resultados y discusión	90
4.4.1	Extracción de información 2D Y 3D:.....	90
4.4.2	Análisis de correlación:.....	92
4.4.3	Rendimiento de los modelos ML basados en métricas RGB-D: ..	94
4.4.4	Resultados de predicción multimodal CNN + ML	98
4.4.5	Discusión	101
4.4.6	Conclusiones	104
4.5	Literatura citada	105

LISTA DE CUADROS

Cuadro 2.1. Etapas fenológicas de la lechuga.....	23
Cuadro 3.1. Métricas de desempeño óptimas de las arquitecturas YOLOv8n-seg y YOLOv11n-seg en el conjunto de prueba, considerando tres configuraciones de distribución de datos (50–35–15, 70–20–10 y 90–5–5).	58
Cuadro 4.1. Definición de las métricas morfológicas bidimensionales y tridimensionales utilizadas para el análisis del dosel de lechuga.....	86
Cuadro 4.2. Análisis estadístico de las variables descriptoras 2D-3D.	91
Cuadro 4.3. Desempeño de los modelos de aprendizaje supervisado en la predicción de los fenotipos de lechuga.	95
Cuadro 4.4. Resultados del desempeño de las combinaciones CNN + ML en la predicción de peso fresco, área foliar y peso seco de lechugas.....	99

LISTA DE FIGURAS

Figura 2.1. Ejemplos de plataformas de fenotipado sensor a planta, en agricultura a campo abierto como en agricultura protegida, imagen recuperada de (Abebe <i>et al.</i> , 2023).....	26
Figura 2.2. Plataforma de fenotipado tipo planta sensor se puede observar que las condiciones de iluminación y fondo se controlan, imagen recuperada de (Moon <i>et al.</i> , 2022).....	27
Figura 3.1. A) Módulo PFAL, B) Cámara instalada en el 3er nivel.	43
Figura 3.2. Imágenes en diferentes condiciones de iluminación y etapas de crecimiento.....	44
Figura 3.3. Arquitectura de la red neuronal convolucional YOLOv8n-Seg. Diagrama extraído y modificado de Sapkota & Karkee (2025).	46
Figura 3.4. Arquitectura de la red neuronal convolucional YOLOv11n-Seg. Extraído y modificado de Sapkota & Karkee (2025).	47
Figura 3.6. Evolución de las pérdidas de entrenamiento y validación para clasificación, detección y segmentación de cada distribución del dataset.	55
Figura 3.6. Desempeño de YOLOv8n-Seg-seg y YOLOv11n-Seg-seg en el conjunto de prueba.	60
Figura 3.7. Proporción de falsos positivos por clase bajo diferentes condiciones de iluminación.	62

Figura 3.8. Proporción de falsos positivos por clase bajo diferentes condiciones de iluminación.	63
Figura 3.9. Resultados de las inferencias obtenidas con YOLOv11n-seg.....	66
Figura 3.10. Resultados de las inferencias obtenidas con YOLOv11n-seg sobre el mismo conjunto de imágenes utilizadas para YOLOv8n-seg.	67
Figura 3.11. Ejemplos ilustrativos de falsos positivos (FP) y falsos negativos (FN) obtenidos en las inferencias de YOLOv8n-seg (azul) y YOLOv11n-seg (naranja).	69
Figura 4.1. Flujo experimental para la adquisición de imágenes RGB-D y la determinación de variables morfofisiológicas en <i>Lactuca sativa</i> L. variedad francesa bajo condiciones controladas en un sistema PFAL.	81
Figura 4.2. Flujo de procesamiento para la extracción de características morfológicas RGB-D.	82
Figura 4.3. Corte transversal X–Z generado a partir de la nube de puntos RGB-D de la cámara Intel RealSense D405.	85
Figura 4.4. Matriz de correlación entre las variables morfológicas 2D–3D derivadas de la nube de puntos y las variables productivas de lechuga cultivada en PFAL.	93
Figura 4.5.Desempeño del modelo Random Forest (RFR) en la predicción de fenotipos de lechuga mediante validación cruzada 5-Fold”.	97

Figura 4.6. Comparación entre valores reales y predichos utilizando EfficientNet B0 + RFR (a–c) y YOLOv8n-Seg + RFR (d–f) para la estimación de área foliar (cm²), peso fresco (g) y peso seco de lechuga (g). 101

RESUMEN GENERAL¹

Las fábricas de plantas con iluminación artificial (PFALs) permiten cultivar hortalizas en entornos controlados, pero presentan limitaciones en la automatización del fenotipado y en la precisión de la segmentación bajo condiciones lumínicas variables. En particular, la iluminación púrpura generada por diodos LED rojo-azul altera la percepción del color y dificulta la delimitación del dosel vegetal, afectando la obtención confiable de parámetros morfológicos. El objetivo de esta investigación fue diseñar e implementar un sistema automatizado de fenotipado RGB-D basado en un robot cartesiano y modelos de aprendizaje profundo para analizar rasgos morfológicos de lechuga (*Lactuca sativa* L.) en un sistema PFAL. Las imágenes RGB-D se capturaron durante el ciclo de crecimiento bajo dos condiciones: fotoperiodo activo con luz púrpura y fase oscura del fotoperiodo, manteniendo una vista cenital fija y geometría constante. Se entrenaron los modelos YOLOv8n-seg y YOLOv11n-seg bajo tres configuraciones de datos (50–35–15 %, 70–20–10 % y 90–5–5 %). YOLOv11n-seg alcanzó la mayor precisión (F1 = 0.953, Dice = 0.966, mAP50–95 > 0.96), reduciendo 40 % los falsos positivos en la fase oscura del fotoperiodo y manteniendo una precisión de segmentación superior al 96 % con luz púrpura. A partir de las máscaras se extrajeron métricas morfológicas 2D–3D área proyectada, altura, volumen aparente, circularidad y relación altura-diámetro que sirvieron como entradas para un modelo multimodal orientado a la predicción no destructiva de peso fresco, peso seco y área foliar. El sistema propuesto supera las limitaciones de segmentación bajo iluminación artificial compleja y sienta las bases de un pipeline de fenotipado RGB-D automatizado para agricultura de precisión.

Palabras clave: Fenotipado RGB-D; segmentación de instancias; YOLOv8n-seg; YOLOv11n-seg; robot cartesiano; visión por computadora; aprendizaje profundo; fábricas de plantas con iluminación artificial (PFAL).

Master's Thesis in Engineering, Universidad Autónoma Chapingo.

Author: Luis Gerardo Chávez Covarrubias.

Advisor: Dra Raquel Salazar Moreno.

ABSTRACT ²

Plant factories with artificial lighting (PFALs) enable vegetable cultivation under controlled environments; however, they still face limitations in phenotyping automation and segmentation accuracy under variable lighting conditions. The purple illumination produced by red-blue LEDs alters color perception and hampers canopy delineation, thereby affecting the reliable extraction of morphological parameters. This study aimed to design and implement an automated RGB-D phenotyping system based on a Cartesian robot and deep learning models for the morphological analysis of lettuce (*Lactuca sativa* L.) grown in a PFAL system. RGB-D images were captured throughout the growth cycle under two lighting regimes: the active photoperiod with purple illumination and the dark phase, maintaining a fixed nadir viewpoint and consistent imaging geometry. YOLOv8n-seg and YOLOv11n-seg models were trained under three data configurations (50–35–15 %, 70–20–10 %, and 90–5–5 %). The YOLOv11n-Seg-seg achieved the highest performance (F1 = 0.953, Dice = 0.966, mAP50–95 > 0.96), reducing false positives by 40 % during the dark phase of the photoperiod and maintaining segmentation accuracy above 96 % under purple illumination. From the generated masks, 2D–3D morphological metrics including projected area, height, apparent volume, circularity, and height-to-diameter ratio were extracted and used as inputs for a multimodal model aimed at the non-destructive prediction of fresh weight, dry weight, and leaf area. The proposed system overcomes segmentation limitations under complex artificial lighting conditions and establishes foundation for an automated RGB-D phenotyping pipeline in precision agriculture.

A Keywords: RGB-D phenotyping; instance segmentation; YOLOv8n-seg; YOLOv11n-seg; Cartesian robot; computer vision; deep learning; Plant Factory with Artificial Lighting (PFAL).

Master's Thesis in Engineering, Universidad Autónoma Chapingo.

Author: Luis Gerardo Chávez Covarrubias.

Advisor: Dra Raquel Salazar Moreno.

CAPÍTULO 1. INTRODUCCIÓN GENERAL

Las Plant Factories with Artificial Lighting (PFALs), conocidas también como granjas verticales, han adquirido un papel destacado en la agricultura moderna por su alta eficiencia productiva y el estricto control que ofrecen sobre el entorno de cultivo (Z. Lin *et al.*, 2022a). Estas instalaciones maximizan el uso del espacio mediante la disposición del cultivo en múltiples niveles, lo que incrementa significativamente la producción por unidad de superficie en comparación con sistemas convencionales. Como resultado, se consideran una alternativa viable para responder a desafíos globales como la seguridad alimentaria y la sostenibilidad ante el crecimiento demográfico y el cambio climático (Lee *et al.*, 2013). Bajo este esquema, la lechuga se ha convertido en uno de los cultivos más adecuados para PFALs gracias a su ciclo vegetativo corto, bajo consumo energético para alcanzar la cosecha, además de su aporte nutricional de caroteno, vitamina C y vitamina E, lo que ha impulsado un creciente interés por parte de la comunidad científica (Y. Zhang *et al.*, 2022).

El fenotipado no destructivo se refiere a la medición objetiva y no invasiva de rasgos como peso fresco, número de hojas y área foliar, mediante visión por computadora e inteligencia artificial (Petropoulou *et al.*, 2023a). En lechuga, técnicas de extracción automatizada han permitido estimaciones precisas de crecimiento (Moon *et al.*, 2022). La fusión de datos RGB-D ha mejorado la estimación no destructiva del peso fresco en tiempo real, optimizando el monitoreo en sistemas PFAL (Xu *et al.*, 2023).

Con base a lo anterior, la adquisición de datos fenotípicos precisos y a gran escala se ha convertido en un elemento fundamental para poder acelerar la mejora genética de cultivos (Cobb *et al.*, 2013; Furbank & Tester, 2011; Zhou & Nguyen). Los entornos controlados, como invernaderos o los sistemas PFALs, ofrecen la ventaja de minimizar la variabilidad ambiental, garantizando

condiciones uniformes, esto permite la implementación de sistemas de automatización complejos de alta repetibilidad (Zhou & Nguyen). Los mecanismos de captura de imágenes para el fenotipado se organizan en dos arquitecturas principales según el movimiento relativo: planta a sensor (PTS), en estos mecanismos las plantas son transportadas a cámaras fijas de medición para asegurar un control estricto de la iluminación; y sensor a planta (STP), en estos sistemas los sensores se mueven mediante rieles o brazos robóticos por encima de plantas estáticas (Zhou & Nguyen), los actuadores finales de estos mecanismos de captura se apoyan de sensores ópticos que operan en diferentes regiones del espectro electromagnético (EMS). Dichos sensores nos permiten realizar mediciones no destructivas, por ejemplo, los sensores RGB nos ayudan a analizar la morfología y color. Las cámaras multiespectrales se usan para evaluar rasgos fisiológicos y bioquímicos, como el contenido de clorofila o el estrés, las cámaras de infrarrojo térmico se usan para estimar la temperatura foliar y la transpiración, también se tienen las cámaras de profundidad equipadas con sensores de tiempo de vuelo (ToF), estas posibilitan la reconstrucción tridimensional de la arquitectura de las plantas y poder obtener información geométrica importantes (Zhou *et al.*, 2021b). Las imágenes aisladas no sirven para realizar predicciones, en este contexto el aprendizaje profundo ha transformado el análisis de imágenes en la agricultura de precisión, especialmente en entornos controlados como invernaderos y PFALs. Redes neuronales convolucionales (CNN) como U-Net, DeepLab o Mask R-CNN se han consolidado como herramientas esenciales para segmentar hojas y estructuras vegetales en diferentes condiciones lumínicas y distintos fondos. En recientes investigaciones se ha mostrado que U-net puede aplicarse con éxito en fenotipado de cultivos en invernaderos (Narisetti *et al.*, 2022), mientras que variantes ligeras de DeepLabV3+ combinadas con mobileNetV2 han permitido segmentación rápida y precisa en lechuga hidropónica bajo diferentes variables ambientales experimentales (Ojo *et al.*, 2024). Así mismo (Cardenas-Gallegos *et al.*, 2025) compararon U-net frente a otros métodos tradicionales de

segmentación (Random Forest y Otsu), y mostraron que U-Net alcanzó una segmentación significativamente mejor en imágenes de lechuga bajo condiciones variables de iluminación, la tarea de segmentación de hojas de lechuga se ha abordado con varias arquitecturas profundas, logrando consistentemente alta precisión, por ejemplo Lin *et al.* (2022b) demostró que U-Net obtuvo un rendimiento confiable en la segmentación de hojas, alcanzando un mIoU (*intersección sobre unión media*) de 0.982 y una precisión de 0.998 en un conjunto de prueba, Otro estudio reportó que el modelo pudo identificar cada planta de lechuga con una puntuación F1 superior al 97.0% (Du *et al.*, 2020) . En los trabajos que se mencionaron anteriormente no se ha utilizado YOLO (You Only Look Once (Ultralytics, 2023b)), YOLO es una familia de modelos de visión por computadora diseñada principalmente para la detección en tiempo real de objetos, que ha evolucionado desde su primera versión (Redmon *et al.*, 2016) , hasta integrar tareas más complejas como la segmentación a partir de la incorporación de YOLOv8-seg. Su aplicación en la agricultura de precisión ha demostrado ventajas frente a arquitecturas clásicas como FCN, U-Net, PSPNet o Deeplabv3, al ofrecer una segmentación más rápida y precisa en escenarios de producción controlada. Por ejemplo, Wang *et al.* (2024) reportaron que variantes modificadas de YOLOv8 superaron al modelo base en el CVPPP Leaf Segmentation Challenge, logrando un Dice del 86.4% y sobrepasando metodologías de referencia como IPK, Nottingham o Wageningen. De manera complementaria Yang *et al.* (2023) demostraron que un esquema en dos etapas con YOLOv8 y DeepLabv3+ mejorado alcanzó un mIoU del 90.8%, superando en hasta 8.4 puntos porcentuales a modelos clásicos como LR-ASPP y mostrando mayor robustez en hojas de geometría compleja.

1.1 Objetivos:

1.1.1 Objetivos Generales:

Predecir rasgos fenotípicos (área foliar, biomasa) a partir de redes neuronales convolucionales usando imágenes RGB-D.

1.1.2 Objetivos específicos:

- Desarrollar e implementar un sistema automatizado para la adquisición de imágenes.
- Capturar imágenes RGB-D destinadas al entrenamiento de redes neuronales convolucionales (CNN) tipo YOLO, a fin de segmentar imágenes cenitales de lechuga y extraer parámetros morfológicos como altura, diámetro, perímetro, área proyectada.
- Estimar de manera no destructiva el área foliar, así como el peso fresco y seco de las plantas mediante técnicas de fenotipado.

1.1.3 Hipótesis:

Las redes neuronales convolucionales enfocadas a la segmentación son capaces de aprender cualquier patrón de imagen sin importar el tipo de iluminación que tengan, solo dependerá de la calidad del de datos de entrenamiento, permitiendo tener mejores resultados que algoritmos tradicionales.

1.1.4 Estructura de la tesis

Este documento está dividido en cuatro capítulos. 1) Introducción general, 2) revisión de literatura, 3) segmentación de instancias basada en YOLOv8n-Seg y YOLOv11n-Seg para la extracción automática de características morfológicas de lechuga (*Lactuca sativa L.*) en PFALs y 4) predicción multimodal de biomasa y área foliar de lechuga mediante visión RGB-D y CNN-ML. En el capítulo 2 se describe las fábricas de plantas así como como algunos de los problemas que existen en visión por computadora, se aborda información sobre la importancia que tiene la lechuga como cultivo experimental en estos sistemas de producción, también se explica algunos sistemas de captura de imágenes para fenotipado, y un historial sobre las redes convolucionales , el capítulo 3 es un artículo científico que cumple con los objetivos planteados de desarrollar un sistema de captura de imagen, construir un conjunto de imágenes de entrenamiento con fotografías con

distinta iluminación y la segmentación , el 4 capítulo es un artículo científico que cumple con el objetivo de la predicción de biomasa y área foliar.

1.2 Literatura citada

- Cardenas-Gallegos, J. S., Severns, P. M., Klimeš, P., N. Lacerda, L., Peduzzi, A., & Soranz Ferrarezi, R. (2025). Reliable plant segmentation under variable greenhouse illumination conditions. *Computers and Electronics in Agriculture*, 229. <https://doi.org/10.1016/j.compag.2024.109711>
- Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. In *Theoretical and Applied Genetics* (Vol. 126, Issue 4, pp. 867–887). Springer Verlag. <https://doi.org/10.1007/s00122-013-2066-0>
- Du, J., Lu, X., Fan, J., Qin, Y., Yang, X., & Guo, X. (2020). Image-Based High-Throughput Detection and Phenotype Evaluation Method for Multiple Lettuce Varieties. *Frontiers in Plant Science*, 11. <https://doi.org/10.3389/fpls.2020.563386>
- Furbank, R. T., & Tester, M. (2011). Phenomics - technologies to relieve the phenotyping bottleneck. In *Trends in Plant Science* (Vol. 16, Issue 12, pp. 635–644). <https://doi.org/10.1016/j.tplants.2011.09.005>
- Lee, J. G., Choi, C. S., Jang, Y. A., Jang, S. W., Lee, S. G., & Um, Y. C. (2013). Effects of air temperature and air flow rate control on the tipburn occurrence of leaf lettuce in a closed-type plant factory system. *Horticulture Environment and Biotechnology*, 54(4), 303–310. <https://doi.org/10.1007/s13580-013-0031-0>
- Lin, Z., Fu, R., Ren, G., Zhong, R., Ying, Y., & Lin, T. (2022a). Automatic monitoring of lettuce fresh weight by multi-modal fusion based deep learning. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.980581>

- Lin, Z., Fu, R., Ren, G., Zhong, R., Ying, Y., & Lin, T. (2022b). Automatic monitoring of lettuce fresh weight by multi-modal fusion based deep learning. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.980581>
- Moon, T., Choi, W. J., Jang, S. H., Choi, D. S., & Oh, M. M. (2022). Growth Analysis of Plant Factory-Grown Lettuce by Deep Neural Networks Based on Automated Feature Extraction. *Horticulturae*, 8(12). <https://doi.org/10.3390/horticulturae8121124>
- Narisetti, N., Henke, M., Neumann, K., Stolzenburg, F., Altmann, T., & Gladilin, E. (2022). Deep Learning Based Greenhouse Image Segmentation and Shoot Phenotyping (DeepShoot). *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.906410>
- Ojo, M. O., Zahid, A., & Masabni, J. G. (2024). Estimating hydroponic lettuce phenotypic parameters for efficient resource allocation. *Computers and Electronics in Agriculture*, 218. <https://doi.org/10.1016/j.compag.2024.108642>
- Petropoulou, A. S., van Marrewijk, B., de Zwart, F., Elings, A., Bijlaard, M., van Daalen, T., Jansen, G., & Hemming, S. (2023). Lettuce Production in Intelligent Greenhouses—3D Imaging and Computer Vision for Plant Spacing Decisions. *Sensors*, 23(6). <https://doi.org/10.3390/s23062929>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. <http://arxiv.org/abs/1506.02640>
- Ultralytics. (2023, January 10). *Ultralytics Documentation*. Explore Ultralytics YOLOv8.
- Wang, P., Deng, H., Guo, J., Ji, S., Meng, D., Bao, J., & Zuo, P. (2024). Leaf Segmentation Using Modified YOLOv8-Seg Models. *Life*, 14(6). <https://doi.org/10.3390/life14060780>

- Xu, D., Chen, J., Li, B., & Ma, J. (2023). Improving Lettuce Fresh Weight Estimation Accuracy through RGB-D Fusion. *Agronomy*, 13(10). <https://doi.org/10.3390/agronomy13102617>
- Yang, T., Zhou, S., Xu, A., Ye, J., & Yin, J. (2023). An Approach for Plant Leaf Image Segmentation Based on YOLOV8 and the Improved DEEPLABV3+. *Plants*, 12(19). <https://doi.org/10.3390/plants12193438>
- Zhang, Y., Li, M., Li, G. X., Li, J., Zheng, L., Zhang, M., & Wang, M. (2022). Multi-phenotypic parameters extraction and biomass estimation for lettuce based on point clouds. *Measurement: Journal of the International Measurement Confederation*, 204. <https://doi.org/10.1016/j.measurement.2022.112094>
- Zhou, J., Zhou, J., Ye, H., & Nguyen, H. T. (2021a). High-Throughput Crop Phenotyping Systems for Controlled Environments. In J. Zhou & H. T. Nguyen (Eds.), *High-Throughput Crop Phenotyping* (pp. 183–208). Springer International Publishing. https://doi.org/10.1007/978-3-030-73734-4_9
- Zhou, J., Zhou, J., Ye, H., & Nguyen, H. T. (2021b). High-Throughput Crop Phenotyping Systems for Controlled Environments. In J. Zhou & H. T. Nguyen (Eds.), *High-Throughput Crop Phenotyping* (pp. 183–208). Springer International Publishing. https://doi.org/10.1007/978-3-030-73734-4_9

CAPÍTULO 2. REVISIÓN DE LITERATURA

2.1 Fábrica de plantas:

Las fábricas de plantas con iluminación artificial (PFALs) son sistemas de cultivo interior de alta tecnología en estos sistemas se reemplaza la luz solar completamente por LEDs, y los cultivos se disponen en capas apiladas dentro de estructuras cerradas, es muy frecuente en entornos urbanos (Stanghellini & Katzin, 2024)

Las PFALs permiten el control preciso de diversos factores micro-ambientales, como luz, temperatura, humedad, CO₂ y nutrientes, maximizando el rendimiento y calidad durante todo el año (L. Zhang *et al.*, 2022).

Las ventajas operativas incluyen:

1. Eficiencia Hídrica
2. Mayor productividad
3. Reducción de químicos

El principal desafío es el alto consumo energético, especialmente por la iluminación y control ambiental que constituye la mayor parte de los costos operativos (Van Gerrewey *et al.*, 2021), se estima que la iluminación artificial representa entre el 75% y el 80% del consumo total de electricidad en el área de cultivo (Kozai & Japan Plant Factory Association, 2019). Esta alta demanda energética hace que la sostenibilidad ambiental de las PFALs dependa fuertemente de la disponibilidad de fuentes de energía con bajas o nulas emisiones de carbono (Cowan *et al.*, 2022)

2.2 La lechuga:

La lechuga (*Lactuca sativa* L.) es una de las hortalizas de hoja verde más cultivadas y consumidas en el mundo, tanto en sistemas a campo abierto como

en agricultura protegida (Das *et al.*, 2024; Han *et al.*, 2021). Su importancia económica se estima de alrededor de 2.93 mil millones de USD anuales, este cultivo destaca por su aporte nutricional, ya que las hojas concentran diversos compuestos bioactivos, entre los que se encuentran carotenoides, polifenoles (como ácidos hidroxicinámicos y derivados del ácido hidroxibenzoico) y vitaminas antioxidantes, principalmente vitamina C y E (Yang *et al.*, 2022) .

2.2.1 La lechuga en sistemas PFALs

La lechuga es conocida como un modelo vegetal para los sistemas en ambientes controlados (Controlled Environment Agriculture, CEA), diversos atributos fisiológicos y productivos convierten a la lechuga en un cultivo altamente eficiente en entornos CEA (Das *et al.*, 2024), por ejemplo:

- Ciclo de desarrollo corto
- Alta tasa de crecimiento vegetativo
- Posibilidad de establecer altas densidades de plantación.
- Requerimientos energéticos relativamente reducidos

2.2.2 Lechuga variedad francesa

La lechuga tipo francesa o Butterhead (*Lactuca sativa* var. *capitata*) se caracteriza por formar cabezas compactas de tamaño reducido, integradas por hojas relativamente lisas, con ligero curvado y dispuestas de manera casi esférica alrededor del cogollo. Las hojas presentan textura suave y aspecto brillante, lo que la hace especialmente apreciada para consumo en fresco (De La Rosa *et al.*, 2024). Dentro de este grupo se incluyen variedades de hoja verde, y de hoja morada ambas de particular interés para PFALs (Givens *et al.*, 2023).

Requisitos de cultivo de la lechuga variedad francesa o Butterhead

1. Requisitos de Luz (intensidad): Durante la propagación, una Integral de Luz Diaria (DLI) de 25 a 35 $\text{mol}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ (equivalente a 200–400

$\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) ha demostrado maximizar el crecimiento de variedades verdes (Givens *et al.*, 2023).

- 2. Factores adversos: Uno de los principales problemas en PFALs es el **tip burn** (necrosis marginal de hojas internas), causado por un transporte insuficiente de calcio y exceso de Luz.
- Condiciones predisponentes: Alta humedad relativa y rápido crecimiento bajo DLIs elevados, que reducen la transpiración y dificultan la distribución de Ca.
- Umbral de riesgo: DLIs superiores a $16.9\text{ mol}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ pueden incrementar la incidencia de tip burn y reducir la comerciabilidad de la lechuga.

2.2.3 Etapas Fenológicas de la lechuga:

El desarrollo fenológico de la lechuga (*Lactuca sativa*), se estructura en cuatro etapas principales: plántula, roseta, encabezamiento y fase reproductiva. La plántula inicia con la germinación y el desarrollo de hojas verdaderas, que pueden alcanzar hasta 19 hojas (Jenni & Bourgeois, 2008). La roseta se forma por hojas dispuestas en un tallo corto, posteriormente de esta etapa continua el encabezamiento es donde se alcanza la madurez de la lechuga. Finalmente, si la planta no se cosecha, entra en el estado reproductivo, elongando su tallo floral(Jenni & Bourgeois, 2008). En la Cuadro 2.1 se explica cada una de las etapas con mayor detalle.

Cuadro 2.1. Etapas fenológicas de la lechuga.

Etapa	Descripción
Plántula	Inicia desde la salida de la raíz y continua con el brote y expansión de los cotiledones.
Desarrollo vegetativo	Desarrollo foliar: Rangos desde 1 hoja verdadera, hasta 19 hojas verdaderas visibles y más largas de 2 cm

Desarrollo de cabeza	Implica que las hojas internas comienzan a cerrarse progresivamente hasta conformar una estructura compacta y firmemente agrupada.
Desarrollo de inflorescencia.	La inflorescencia surge cuando el tallo comienza a elongarse y da lugar a la formación de flores, proceso que señala la finalización del ciclo vegetativo

2.3 Fenotipado no destructivo

El fenotipado vegetal no destructivo comprende métodos que cuantifican los rasgos observables de las plantas sin alterarlas físicamente, Estos rasgos, conocidos como fenotipos, reflejan la interacción entre el genotipo y el ambiente en el que la planta se desarrolla. La medición de dichos rasgos permite analizar los cambios que ocurren durante el crecimiento y bajo diferentes condiciones climáticas o ambientales (Ye *et al.*, 2023).

2.3.1 Técnicas de fenotipado no destructivo

El fenotipado vegetal no destructivo abarca un amplio rango de tecnologías que pueden cuantificar los rasgos morfológicos, fisiológicos hasta bioquímicos de las plantas sin comprometer su integridad. Estas técnicas han evolucionado desde métodos espectroscópicos simples hacia sistemas de captura y análisis de información multiespectral, térmica y tridimensional (Abebe *et al.*, 2023), impulsando el desarrollo del fenotipado de alto rendimiento (HTPP) (Araus & Cairns, 2014; Li *et al.*, 2014).

A continuación, se presentan algunas técnicas de fenotipado:

- Espectroscopía unidimensional (1D): Se mide la reflectancia en diferentes longitudes de onda para detectar cambios fisiológicos que se asocian con el contenido de pigmentos (Li *et al.*, 2014).

- Fenotipado en 2D: Aquí se utilizan imágenes RGB, multiespectrales, fluorescentes hiperespectrales para poder estimar biomasa, colorimetría, morfología, mediante la extracción de parámetros correlacionados con estas características, extraídos por técnicas de visión por computadora (Li *et al.*, 2014)
- Fenotipado 3D: Se utiliza información espacial a partir de sensores como cámaras estéreo o también llamadas cámaras de profundidad, escáneres láser o sistemas de estructura a partir del movimiento (Structure from Motion, SfM) para poder reconstruir la geometría de la planta y calcular volumen, altura o compactación, rugosidad etc. (Paulus, 2019).
- Fenotipado temporal (T-3D): Usa mediciones tridimensionales con series de tiempo para analizar el crecimiento y los cambios dinámicos en la estructura vegetal, en agricultura, se puede utilizar para poder observar el crecimiento del cultivo a lo largo del tiempo y poder optimizarlo (Li *et al.*, 2014; Ye *et al.*, 2023).
- Fenotipado RGB-D: Fusiona información de color (RGB) con mapas de profundidad (Depth) obtenidos de sensores de luz estructurada o cámaras de tiempo de vuelo, como las Intel RealSense o Kinect. Este enfoque permite estimar rasgos morfológicos tridimensionales junto con propiedades ópticas superficiales, lo que lo convierte en una herramienta versátil para el fenotipado de precisión (Li *et al.*, 2014) .

2.3.2 Plataformas de fenotipado no destructivo:

La automatización de la adquisición de imágenes es una alternativa accesible para evaluar el crecimiento y el desarrollo vegetal. En todas las plataformas de fenotipado de alto rendimiento, los diferentes tipos de imágenes constituyen el núcleo del sistema de captura de datos. En general, las estrategias de adquisición de imágenes para fenotipado se dividen en dos formas, según la relación de movimiento ya sea el sensor a la planta o viceversa (Abebe *et al.*, 2023). La clasificación de tipo de plataformas fue tomada de (Zhou & Nguyen, 2021):

Sensor a planta:

Estos sistemas de fenotipado se aplican cuando no es conveniente o no es posible trasladar las plantas durante su crecimiento, se usan principalmente en sistemas que se desarrollan en bancos experimentales fijos o en sistemas hidropónicos donde los tratamientos se aplican en solución acuosa y mover la planta alteraría las condiciones de crecimiento alterando el experimento (Figura 2.1).

- Plataformas elevadas: Mueven a los sensores sobre el dosel vegetal mediante rieles o marcos superiores para capturar imágenes de vista aérea.
- Plataformas terrestres: Están sobre rieles, se puede ajustar la altura para la obtención de vistas laterales.
- Robots móviles



Figura 2.1. Ejemplos de plataformas de fenotipado sensor a planta, en agricultura a campo abierto como en agricultura protegida, imagen recuperada de (Abebe et al., 2023).

Los sistemas Planta a Sensor (PTS) (Figura 2.2) representan son técnicas más eficientes del fenotipado de alto rendimiento, estos sistemas permiten automatizar la captura de datos mediante el traslado de las plantas hacia estaciones fijas equipadas con sensores ópticos. Gracias a su diseño cerrado es posible controlar la iluminación reduciendo así la variabilidad ambiental y mejora la calidad de las imágenes.



Figura 2.2. Plataforma de fenotipado tipo planta sensor se puede observar que las condiciones de iluminación y fondo se controlan, imagen recuperada de (Moon *et al.*, 2022).

2.4 Redes neuronales convolucionales de segmentación:

La segmentación en lechuga (*Lactuca sativa* L.) es una herramienta esencial para la extracción automatizada de rasgos fenotípicos. La segmentación es para separar con precisión el follaje del fondo y así poder calcular métricas como el área foliar proyectada, el diámetro máximo y mínimo, la circularidad, el volumen aparente, y parámetros que describen el crecimiento y vigor vegetal, también son usadas para la detección de zonas infectadas en hojas etc. (Hu *et al.*, 2017)

Las redes neuronales convolucionales (CNN, por sus siglas en ingles) de detección de objetos se pueden clasificar en detectores de dos y una etapa. Los primeros, como R-CNN (Girshick *et al.*, 2014), Fast R-CNN, Faster R-CNN (Hu *et al.*, 2017; Ren *et al.*, 2016) y Mask R-CNN, estos algoritmos hacen la tarea de detección en dos fase: primero generan regiones de interés (RoI) y luego se

aplica una red adicional de clasificación para refinar las predicciones. Estos algoritmos ofrecen alta precisión pero suelen necesitar de mayor tiempo de inferencia, lo que limita que se puedan aplicar a sistemas en tiempo real. Por otro lado, los detectores de una etapa, como SSD (Leibe *et al.*, 2016) y la familia YOLO (You Only Look Once), realizan la detección de objetos mediante una única red neuronal, sin necesidad de generar regiones de interés ni aplicar procesos adicionales de refinamiento. Al integrar todas las operaciones en una sola arquitectura, se logra obtener una mayor eficiencia computacional, reduce el tiempo de inferencia con estos algoritmos se puede operar en tiempo real teniendo un equilibrio entre precisión y velocidad de inferencia.

2.4.1 Yolo (You Only Look once)

El modelo Yolo apareció por primera vez en 2016 fue propuesto por (Redmon *et al.*, 2016) tuvo un enfoque revolucionario en la detección en tiempo real. La principal diferencia que se notó en relación con las redes de dos etapas fue que YOLO integró en una única red neuronal convolucional los procesos de localización y clasificación, dividiendo la imagen en una malla de celdas desde la cual predice simultáneamente las cajas delimitadoras y las probabilidades de clase en una sola inferencia o sea integra la detección como un problema de clasificación.

Al paso del tiempo Yolo evolucionó a Yolo v2 (Redmon & Farhadi, 2016) y Yolo v3 estos modelos mejoraron la precisión y la exhaustividad inspirándose en Resnet, se añadió detección multiescala y capas residuales. En 2020 (Bochkovskiy *et al.*, 2020), propuso la versión Yolo V4 diseñada para mejorar la velocidad de inferencia de sus predecesores y otros algoritmos de detección de imágenes, tuvo un 10 % de mejoras en precisión y un 12 % en cuestión de velocidad de inferencia comparado con la versión 3, para lograr esto se mejoró la arquitectura, se sustituyó el backbone y se incorporó un neck más avanzado. la versión Yolov4 usa CSPDarknet53 como red base y conserva un head con anclas similar al de YOLOv3, pero añade un módulo SPP (Spatial Pyramid

Pooling) para ampliar el campo receptivo y reemplaza la tradicional FPN por PANet, lo que mejora la fusión de características entre diferentes escalas.

La tarea de segmentación de instancias se incorporó a la red neuronal YOLO en las variantes de YOLOV3, esta aportación se introdujo en la investigación llamada Poly-YOLO (Hurtik *et al.*, 2020), respecto a YOLOV3 se mejoró la precisión, velocidad y se le añadió un mecanismo de detección poligonal, que sustituye las cajas rectangulares por polígonos dinámicos definidos en coordenadas polares, lo que permite delinear con mayor exactitud la forma de los objetos.

En las versiones más recientes de YOLO, como YOLOv8n-Seg-seg (Ultralytics, 2023b) y YOLOv11n-Seg-seg(Ultralytics, 2024) , integran de forma nativa la segmentación de instancias dentro del modelo. Estas arquitecturas unifican múltiples tareas en detección, segmentación, clasificación y estimación de pose en un solo sistema de red neuronal optimizado, manteniendo la rapidez y precisión características de la familia YOLO.

2.4.2 YOLOV8n-seg y YOLOv11n-seg:

YOLOv8n-seg:

YOLOv8n-Seg, desarrollado por Ultralytics en 2023, introduce una arquitectura rediseñada, pero manteniendo la filosofía de detección en tiempo real, característica de la familia YOLO, pero con una estructura más modular y eficiente. El modelo se organiza en tres componentes principales:

- El backbone, basado en una versión optimizada de *CSPDarknet53*, utiliza bloques C2f que mejoran la transferencia de información y la reutilización de características entre capas profundas.
- El neck, implementa una combinación de redes *PAN-FPN (Path Aggregation Network–Feature Pyramid Network)*, que refuerzan la integración de características provenientes de distintas escalas,

optimizando la detección de objetos pequeños y grandes dentro de una misma imagen.

- El head desacoplado separa las tareas de clasificación, regresión de cajas y segmentación, lo que reduce las interferencias entre ellas y acelera la convergencia del entrenamiento.

(Ultralytics, 2023b).

YOLOv11n-seg:

YOLOv11n-Seg, desarrollado por Ultralytics (2024), conserva la estructura modular introducida en YOLOv8n-Seg, pero incorpora ajustes sustanciales orientados a mejorar la precisión y la eficiencia computacional. El modelo emplea un backbone más liviano, construido sobre los nuevos módulos Efficient-C2f, que optimizan la relación entre rendimiento y velocidad de procesamiento. Integra además un bloque de atención adaptativo, diseñado para resaltar las regiones más relevantes en escenas complejas, incrementando la capacidad de detección en condiciones variables de iluminación o fondo. En el área de segmentación, introduce cabezas de refinamiento de máscara (mask refinement heads), las cuales generan bordes más precisos y segmentaciones más limpias, reduciendo el ruido en los límites de los objetos. Finalmente, el soporte nativo para entrenamiento de precisión mixta (Mixed Precision Training) disminuye el uso de memoria y acelera el aprendizaje en GPU.

2.5 Investigaciones recientes de fenotipado

Las investigaciones recientes sobre segmentación de cultivos han adoptado múltiples estrategias de *deep learning* para estimar variables morfológicas y fisiológicas, diferenciándose principalmente por el tipo de red neuronal empleada, el sensor óptico utilizado y la geometría de adquisición de las imágenes.

En un estudio orientado al análisis estructural del dosel, se implementó un modelo RT-DETR de tipo *anchor-free*, complementado con módulos de refinamiento espacial (DRS e IRC), para calcular la altura y el ancho del follaje de lechuga. Las imágenes fueron capturadas mediante una cámara Intel RealSense D435i, que combina canales RGB, infrarrojo y profundidad, en una disposición lateral multimodal a 37 cm de las plantas, alcanzando un total de 18 000 imágenes para entrenamiento, validación y prueba (Lu *et al.*, 2025).

Otro trabajo, centrado en plántulas cultivadas en ambientes controlados, integró la red Mask R-CNN con un *backbone* ResNet-101 reforzado con una estructura CB-Net, mejorando la detección de hojas pequeñas y la estimación del área foliar proyectada. En este caso, se utilizó una cámara Raspberry Pi V2 en configuración cenital (*top-view*), propia de las fábricas de plantas, generando aproximadamente 1 000 imágenes anotadas (Islam *et al.*, 2024).

Para la estimación de biomasa temprana, se aplicó una U-Net clásica con información RGB-D, enfocada en la predicción del peso fresco del brote a partir de rasgos geométricos como área, perímetro o área convexa. El sistema se probó en condiciones de invernadero con cerca de 380 muestras (Lin *et al.*, 2022).

En un enfoque híbrido, se combinó YOLOv3 con una U-Net, donde el primer modelo detectó regiones de interés y el segundo realizó la segmentación precisa del contorno foliar. Este método permitió obtener indicadores de crecimiento y color (en espacios RGB y HSV) a partir de imágenes capturadas con cámaras Canon 77D y sensores Kinect v2 en una plataforma tipo *Sensor-to-Plant*, alcanzando unas 5 900 imágenes procesadas (Du *et al.*, 2020)

Otro desarrollo reciente adoptó YOLOv8n-Segm-seg, una versión ligera optimizada para dispositivos embebidos, aplicada al cálculo de peso fresco, dimensiones y volumen aparente. Se integraron sensores RGB, una cámara de profundidad Orbbec Gemini 336L y un módulo termográfico en un sistema lateral

robotizado a un metro del cultivo, obteniendo 132 imágenes multimodales (X. Zhang *et al.*, 2025a).

También se exploraron arquitecturas DCNN de doble rama con fusión tardía, que prescinden del uso de máscaras explícitas. Este enfoque permitió estimar la biomasa fresca directamente a partir de pares estéreo registrados por una cámara Intel RealSense D435i en vista cenital, con un conjunto de 3 800 plantas (Buxbaum *et al.*, 2022).

De igual forma, se combinaron YOLOv3 y una red de regresión multimodal (TMSCNet) para predecir peso fresco, peso seco, altura, diámetro y área foliar sin necesidad de segmentación previa. Las imágenes RGB-D se adquirieron desde una cámara RealSense D415 a 0.9 m de distancia, reuniendo 390 muestras (Q. Zhang *et al.*, 2022)

En otro caso, se utilizó una red CNN-284 adaptada para fusión RGB-D, con el objetivo de estimar el peso fresco de las plantas capturadas mediante una Kinect 2.0, a partir de 850 imágenes RGB y RGB-D (Xu *et al.*, 2023).

Asimismo, se aplicó la técnica de fotogrametría basada en estructura desde movimiento (SfM) utilizando el software *Agisoft Metashape*, con cámaras USB de alta resolución ubicadas a 25 cm y en ángulos de 0° y 45°, para reconstruir modelos 3D del dosel y calcular volumen y dimensiones. Este método procesó 225 plantas en 3D y 120 en 2D (Jiang *et al.*, 2024).

En cuanto a reconstrucciones volumétricas avanzadas, se propuso el método Vacuum Package, que combina cuatro cámaras 3D (una superior y tres laterales, incluyendo una RealSense D405) para estimar el volumen y la biomasa a partir de nubes de puntos multivista generadas en 402 capturas (Bloch *et al.*, 2025).

Para el seguimiento fenológico con cámaras de profundidad, se aplicó DeepLabv3+, enfocándose en la cuantificación de altura, cobertura y peso fresco.

El sistema, basado en cámaras Intel RealSense D415 operando en modo RGB-IR-Depth, capturó imágenes cada 15 minutos, de las cuales 35 fueron etiquetadas (Petropoulou *et al.*, 2023).

En un contexto espectral, se utilizó el algoritmo Spectral Angle Mapper (SAM) sobre datos hiperespectrales adquiridos con una cámara GaiaField-V10E, en un entorno oscuro y a un metro de distancia, para estimar el contenido de sólidos solubles y pH a partir de la reflectancia media de la región de interés, con un total de 387 muestras (Yu *et al.*, 2022).

Finalmente, se desarrolló un sistema de clasificación basado en CNN para imágenes NDVI obtenidas mediante un sensor montado en dron (UAV), con el propósito de contar plantas y clasificarlas según su tamaño. Este método alcanzó una resolución espacial de 3 cm/píxel y procesó más de 100 000 fotogramas (Bauer *et al.*, 2019).

2.6 Referencias

Abebe, A. M., Kim, Y., Kim, J., Kim, S. L., & Baek, J. (2023). Image-Based High-Throughput Phenotyping in Horticultural Crops. *Plants*, 12(10), 2061. <https://doi.org/10.3390/plants12102061>

Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science*, 19(1), 52-61. <https://doi.org/10.1016/j.tplants.2013.09.008>

Bauer, A., Bostrom, A. G., Ball, J., Applegate, C., Cheng, T., Laycock, S., Rojas, S. M., Kirwan, J., & Zhou, J. (2019). Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: A case study of lettuce production. *Horticulture Research*, 6(1), 70. <https://doi.org/10.1038/s41438-019-0151-5>

Bloch, V., Shapiguzov, A., Kotilainen, T., & Pastell, M. (2025). A method for phenotyping lettuce volume and structure from 3D images. *Plant Methods*, 21(1), 27. <https://doi.org/10.1186/s13007-025-01347-y>

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection* (No. arXiv:2004.10934). arXiv. <https://doi.org/10.48550/arXiv.2004.10934>

Buxbaum, N., Lieth, J. H., & Earles, M. (2022). Non-destructive Plant Biomass Monitoring With High Spatio-Temporal Resolution via Proximal RGB-D Imagery and End-to-End Deep Learning. *Frontiers in Plant Science*, 13, 758818. <https://doi.org/10.3389/fpls.2022.758818>

Cowan, N., Ferrier, L., Spears, B., Drewer, J., Reay, D., & Skiba, U. (2022). CEA Systems: The Means to Achieve Future Food Security and Environmental Sustainability? *Frontiers in Sustainable Food Systems*, 6, 891256. <https://doi.org/10.3389/fsufs.2022.891256>

Das, P. R., Del Moro, D. S., Givens, S. R., Armstrong, S. P., & Walters, K. J. (2024). Propagation light intensity influences yield, morphology, and phytochemistry of purple-leaf butterhead lettuce (*Lactuca sativa*). *Journal of Agriculture and Food Research*, 16, 101210. <https://doi.org/10.1016/j.jafr.2024.101210>

De La Rosa, L. A., Moreno-Escamilla, J. O., Martínez-Ruiz, N. D. R., Alvarez-Parrilla, E., González-Aguilar, G. A., & Rodrigo-García, J. (2024). The Preharvest Application of Stress Response Elicitors Improves the Content of Bioactive Compounds without Modifying the Sensory Attributes of Butterhead Lettuce (*Lactuca sativa* var. *Capitata*). *Foods*, 13(16), 2574. <https://doi.org/10.3390/foods13162574>

Du, J., Lu, X., Fan, J., Qin, Y., Yang, X., & Guo, X. (2020). Image-Based High-Throughput Detection and Phenotype Evaluation Method for Multiple Lettuce Varieties. *Frontiers in Plant Science*, 11, 563386. <https://doi.org/10.3389/fpls.2020.563386>

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation* (No. arXiv:1311.2524). arXiv. <https://doi.org/10.48550/arXiv.1311.2524>

Givens, S. R., Del Moro, D. S., Parker, S. E., Renny, A. G., Sams, C. E., & Walters, K. J. (2023). Light Intensity during Green-Leaf Butterhead Lettuce Propagation Influences Yield and Carotenoids at Harvest. *Horticulturae*, 9(2), 223. <https://doi.org/10.3390/horticulturae9020223>

Han, R., Truco, M. J., Lavelle, D. O., & Michelmore, R. W. (2021). A Composite Analysis of Flowering Time Regulation in Lettuce. *Frontiers in Plant Science*, 12, 632708. <https://doi.org/10.3389/fpls.2021.632708>

Hu, Q., Tian, J., & He, D. (2017). Wheat leaf lesion color image segmentation with improved multichannel selection based on the Chan–Vese model. *Computers and Electronics in Agriculture*, 135, 260-268. <https://doi.org/10.1016/j.compag.2017.01.016>

Hurtik, P., Molek, V., Hula, J., Vajgl, M., Vlasanek, P., & Nejezchleba, T. (2020). *Poly-YOLO: Higher speed, more precise detection and instance segmentation for YOLOv3* (No. arXiv:2005.13243). arXiv. <https://doi.org/10.48550/arXiv.2005.13243>

Islam, S., Reza, M. N., Chowdhury, M., Ahmed, S., Lee, K.-H., Ali, M., Cho, Y. J., Noh, D. H., & Chung, S.-O. (2024). Detection and segmentation of lettuce seedlings from seedling-growing tray imagery using an improved mask R-CNN

method. *Smart Agricultural Technology*, 8, 100455.
<https://doi.org/10.1016/j.atech.2024.100455>

Jenni, S., & Bourgeois, G. (2008). Quantifying Phenology and Maturity in Crisphead Lettuce. *HortTechnology*, 18(4), 553-558.
<https://doi.org/10.21273/HORTTECH.18.4.553>

Kozai, T. & Japan Plant Factory Association. (2019). Towards sustainable plant factories with artificial lighting (PFALs) for achieving SDGs. *International Journal of Agricultural and Biological Engineering*, 12(5), 28-37.
<https://doi.org/10.25165/j.ijabe.20191205.5177>

Leibe, B., Matas, J., Sebe, N., & Welling, M. (Eds.). (2016). *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I* (Vol. 9905). Springer International Publishing.
<https://doi.org/10.1007/978-3-319-46448-0>

Li, L., Zhang, Q., & Huang, D. (2014). A Review of Imaging Techniques for Plant Phenotyping. *Sensors*, 14(11), 20078-20111. <https://doi.org/10.3390/s141120078>

Lin, Z., Fu, R., Ren, G., Zhong, R., Ying, Y., & Lin, T. (2022). Automatic monitoring of lettuce fresh weight by multi-modal fusion based deep learning. *Frontiers in Plant Science*, 13, 980581. <https://doi.org/10.3389/fpls.2022.980581>

Lu, S., Lv, Y., Qian, T., Ren, W., Li, X., Li, Y., & Li, G. (2025). Lettuce architectural phenotypes extraction from multimodal images by low-light sensitivity and strong spatial perception. *Computers and Electronics in Agriculture*, 232, 109928.
<https://doi.org/10.1016/j.compag.2025.109928>

Paulus, S. (2019). Measuring crops in 3D: Using geometry for plant phenotyping. *Plant Methods*, 15(1), 103. <https://doi.org/10.1186/s13007-019-0490-0>

- Petropoulou, A. S., Van Marrewijk, B., De Zwart, F., Elings, A., Bijlaard, M., Van Daalen, T., Jansen, G., & Hemming, S. (2023). Lettuce Production in Intelligent Greenhouses—3D Imaging and Computer Vision for Plant Spacing Decisions. *Sensors*, 23(6), 2929. <https://doi.org/10.3390/s23062929>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- Redmon, J., & Farhadi, A. (2016). *YOLO9000: Better, Faster, Stronger* (No. arXiv:1612.08242). arXiv. <https://doi.org/10.48550/arXiv.1612.08242>
- Ren, S., He, K., Girshick, R., & Sun, J. (2016). *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks* (No. arXiv:1506.01497). arXiv. <https://doi.org/10.48550/arXiv.1506.01497>
- Stanghellini, C., & Katzin, D. (2024). The dark side of lighting: A critical analysis of vertical farms' environmental impact. *Journal of Cleaner Production*, 458, 142359. <https://doi.org/10.1016/j.jclepro.2024.142359>
- Van Gerrewey, T., Boon, N., & Geelen, D. (2021). Vertical Farming: The Only Way Is Up? *Agronomy*, 12(1), 2. <https://doi.org/10.3390/agronomy12010002>
- Yang, X., Gil, M. I., Yang, Q., & Tomás-Barberán, F. A. (2022). Bioactive compounds in lettuce: Highlighting the benefits to human health and impacts of preharvest and postharvest practices. *Comprehensive Reviews in Food Science and Food Safety*, 21(1), 4-45. <https://doi.org/10.1111/1541-4337.12877>
- Ye, D., Wu, L., Li, X., Atoba, T. O., Wu, W., & Weng, H. (2023). A Synthetic Review of Various Dimensions of Non-Destructive Plant Stress Phenotyping. *Plants*, 12(8), 1698. <https://doi.org/10.3390/plants12081698>

Yu, S., Fan, J., Lu, X., Wen, W., Shao, S., Guo, X., & Zhao, C. (2022). Hyperspectral Technique Combined With Deep Learning Algorithm for Prediction of Phenotyping Traits in Lettuce. *Frontiers in Plant Science*, 13, 927832. <https://doi.org/10.3389/fpls.2022.927832>

Zhang, L., Yang, X., Li, T., Gan, R., Wang, Z., Peng, J., Hu, J., Guo, J., Zhang, Y., Li, Q., & Yang, Q. (2022). Plant factory technology lights up urban horticulture in the post-coronavirus world. *Horticulture Research*, 9, uhac018. <https://doi.org/10.1093/hr/uhac018>

Zhang, Q., Zhang, X., Wu, Y., & Li, X. (2022). TMSCNet: A three-stage multi-branch self-correcting trait estimation network for RGB and depth images of lettuce. *Frontiers in Plant Science*, 13, 982562. <https://doi.org/10.3389/fpls.2022.982562>

Zhou, J., & Nguyen, H. T. (Eds.). (2021). *High-Throughput Crop Phenotyping*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-73734-4>

CAPÍTULO 3. APLICACIÓN DE YOLOV8N-SEG Y YOLOV11N-SEG PARA LA SEGMENTACIÓN Y EXTRACCIÓN DE CARACTERÍSTICAS MORFOLÓGICAS EN LECHUGA CULTIVADA EN PFAL

Luis Gerardo Chávez Covarrubias¹; Raquel Salazar Moreno^{1*}; Irineo López Cruz¹; Luis Arturo Soriano Avendaño¹

¹Posgrado de Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo. Texcoco, Estado de México. *Correo: rsalazarm@chapingo.mx

3.1 Resumen

La segmentación precisa del dosel vegetal en Plant Factory with Artificial Lighting (PFAL) es compleja debido a la variabilidad lumínica. Este estudio desarrolló un sistema de adquisición automática mediante un robot cartesiano con cámara Intel RealSense D405 para obtener imágenes RGB-D estandarizadas de lechuga (*Lactuca sativa* L.) bajo fotoperiodo activo (luz púrpura) e inactivo (luz blanca tenue). Con ellas se formó un conjunto de datos balanceado para entrenar y comparar los modelos YOLOv8n-Seg-seg y YOLOv11n-Seg-seg, evaluados con particiones 50–35–15, 70–20–10 y 90–5–5 mediante precisión, recall, F1-score, Dice y mAP@0.5–0.95. YOLOv11n-Seg-seg superó consistentemente a YOLOv8n-Seg-seg, alcanzando F1-scores >0.95, menor falsos positivos ($\leq 6\%$) y contornos más definidos. La partición 70–20–10 mostró el mejor equilibrio, reduciendo sobreajuste. En general, YOLOv11n-Seg-seg presentó mayor robustez ante variaciones espectrales, especialmente bajo luz púrpura. Los resultados confirman que la integración de visión RGB-D con modelos YOLO avanzados permite una segmentación confiable del dosel para sistemas de fenotipado automatizado en PFAL.

Palabras clave: Segmentación de instancias, YOLOv11n-Seg, YOLOv8n-Seg, mAP, Dice, RGB-D.

3.2 INTRODUCCIÓN

Las fábricas de plantas (PFAL, *Plant Factory with Artificial Lighting*) tienen la capacidad para producir de manera eficiente en lugares con condiciones ambientales adversas, usando ambientes controlados de forma precisa y automática para que el cultivo crezca en condiciones óptimas con mejores rendimientos comparado con otras técnicas de cultivo convencionales (Z. Lin *et al.*, 2022c). La lechuga (*Lactuca sativa* L) es uno de los cultivos más utilizados para experimentar en las PFALS), por su ciclo de crecimiento relativamente corto y por lo tanto menor requerimiento de energía para obtener el producto final (Lee *et al.*, 2013).

Para hacer un buen manejo del cultivo y maximizar los rendimientos, se requiere del monitoreo en tiempo real de algunas variables de crecimiento de la planta mediante métodos no destructivos (Zhang *et al.*, 2023). La fusión del aprendizaje profundo y la visión por computadora ha dado origen a los sensores virtuales (*soft sensors*), los cuales utilizan imágenes multiespectrales o RGB para estimar de manera no destructiva variables morfológicas, fisiológicas y estructurales de las plantas (Xu *et al.*, 2023). Para que estos sensores tengan buenas predicciones, resulta imprescindible contar con un sistema de adquisición de imágenes preciso y automatizado, que incluyan imágenes tomadas desde las mismas posiciones en diferentes etapas fenológicas. La automatización del sistema no solo garantiza la uniformidad en las capturas, sino que también evita alteraciones en el ambiente interno del módulo de PFAL. Esto es particularmente relevante en estudios fenotípicos, ya que cualquier cambio ambiental podría afectar el desarrollo de las plantas y comprometer los resultados experimentales. En la actualidad, se han explorado métodos de fenotipado no destructivo que incorporan sistemas de adquisición de imágenes, aplicados tanto a cultivos en invernadero como en campo abierto. Moon *et al.*, (2022) utilizó una caja oscura equipada con LEDs blancos para garantizar condiciones de iluminación consistentes durante la captura de imágenes de los cultivos, las fotografías se tomaron desde una altura

de 50 cm utilizando un teléfono inteligente Galaxy S20. Zhang *et al.* (2023) se basó en un método RGB de fenotipado no destructivo y de bajo costo, implementado en una fábrica de plantas bajo condiciones ambientales controladas. Se utilizó una cámara digital SONY SELP1650, colocada manualmente a 0.6 m sobre el dosel vegetal, con distancia focal de 100 mm y tiempo de exposición de 3 ms. Las plantas se situaron sobre una placa de espuma blanca con un cuadrado azul de 0.25 m para mejorar la segmentación.

Para la extracción de características del dosel en lechuga, es fundamental identificar individualmente cada planta y diferenciar las hojas del fondo. Este proceso requiere segmentar previamente las imágenes RGB para aislar las regiones de interés, lo cual permite posteriormente la obtención de características morfológicas y geométricas relevantes, bajo este enfoque YOLOv8n-seg incorpora una arquitectura optimizada que permite realizar detección y segmentación de instancias con alta precisión a nivel de píxel manteniendo eficiencia computacional, lo que facilita su implementación en dispositivos embebidos y plataformas robóticas (Ultralytics, 2023a). De manera complementaria, YOLOv11n-seg representa la evolución más reciente de esta línea de modelos, integrando módulos de atención, mejoras en la segmentación refinada y detección orientada, lo cual permite un mejor manejo de estructuras vegetales densas y objetos solapados, comunes en imágenes fenotípicas de cultivo. Esta versión ofrece mayor precisión y menor latencia, consolidándose como una herramienta de última generación para el monitoreo fenotípico en tiempo real

De acuerdo con Tian *et al.* (2022) la iluminación artificial predominante en los sistemas PFAL, principalmente basada en combinaciones de luces LED rojas y azules genera un ambiente visual púrpura que altera significativamente la percepción del color y reduce el contraste entre el dosel vegetal y el fondo. Estas condiciones dificultan la delimitación precisa de las hojas y provocan fallos frecuentes en métodos tradicionales de segmentación, tales como thresholding,

Otsu, watershed y K-means. El presente trabajo busca abordar las limitaciones de la segmentación de imágenes en sistemas PFAL, donde la iluminación artificial con luces LED rojas y azules genera un ambiente visual púrpura que dificulta la delimitación del dosel vegetal y afecta la extracción precisa de rasgos morfológicos. Para resolver este problema, se propone el uso de modelos de segmentación de instancias YOLOv8n-seg y YOLOv11n-seg, capaces de operar con alta precisión bajo condiciones lumínicas complejas y adaptarse a entornos de visión artificial agrícola.

Los objetivos de este trabajo son: 1) Implementar un robot cartesiano diseñado específicamente para la adquisición sistemática de imágenes dentro de un sistema de agricultura en ambiente controlado tipo PFAL (*Plant Factory with Artificial Lighting*), enfocado en el cultivo de lechuga variedad francesa.

2). Comparar el desempeño de las arquitecturas YOLOv8n-seg y YOLOv11n-seg en la tarea de segmentación, bajo diferentes condiciones de iluminación y etapas de crecimiento vegetativo del cultivo de lechuga.

3.3 MATERIALES Y MÉTODOS

La presente investigación se llevó a cabo en las instalaciones de la Universidad Autónoma Chapingo. La germinación de las semillas se realizó el 8 de noviembre de 2024 en un módulo especializado de germinación equipado con un sistema de aspersión automática, implementado con el propósito de promover un desarrollo radicular óptimo, garantizar una hidratación uniforme y favorecer un crecimiento inicial saludable. Posteriormente, el 26 de noviembre de 2024, las plántulas fueron trasplantadas al módulo PFAL, bajo un sistema hidropónico tipo NFT (Nutrient Film Technique), que permitió mantener la continuidad del crecimiento y facilitar la adaptación de las plantas a su nuevo entorno controlado. El robot cartesiano fue instalado dentro de un módulo PFAL, durante el periodo del 29 de noviembre al 26 de diciembre de 2024. La instalación se efectuó en el tercer nivel del módulo de producción PFAL (Figura 3.1, el cual opera con una

velocidad del aire equivalente al 75 % de la potencia del ventilador, y un flujo de solución nutritiva de 1.5 L s^{-1} . En cada nivel se tiene una capacidad para 27 plantas, y la radiación lumínica presenta una variación entre 190 y 230 nm desde los bordes hacia el centro del módulo, el interior del módulo está cubierto por una malla maylar reflejante, diseñada para optimizar la distribución de la luz, pero que a su vez genera reflejos y dispersión difusa. Estas características ópticas del entorno afectan la segmentación automática del dosel vegetal, al reducir el contraste espectral entre el follaje y el fondo reflectante.

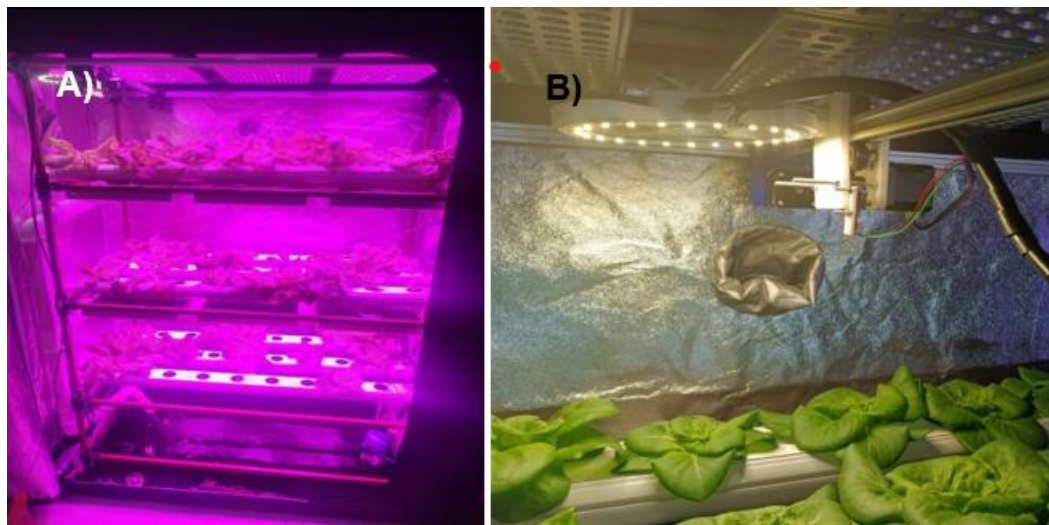


Figura 3.1. A) Módulo PFAL, B) Cámara instalada en el 3er nivel.

3.3.1 Adquisición y procesamiento de imágenes

La adquisición de imágenes se realizó mediante el sistema automático de posicionamiento XY descrito previamente, utilizando una cámara Intel D405 posicionada a 30 cm de altura con un ángulo de 90° respecto a la superficie de cultivo. Las capturas se efectuaron diariamente hasta la cosecha, cubriendo las condiciones de fotoperíodo en la PFAL (*Plant Factory with Artificial Lighting*), lo que generó un conjunto de 3049 imágenes RGB (640×480 píxeles) y sus correspondientes mapas de profundidad, almacenados en archivos CSV. De la base de datos anterior, se generó una base de datos balanceada de 800

instancias (objetos de interés). Con ayuda de la librería Albumentations (Buslaev *et al.*, 2020) realizaron giros de derecha a izquierda y de arriba hacia para aumentar la base de datos y equilibrar la cantidad de instancias en cada clase.

El etiquetado de las imágenes se efectuó en la plataforma en línea Roboflow (Dwyer *et al.*, 2024), empleando el formato de anotación YOLOv8n-Seg; las etiquetas resultantes se guardaron en archivos .txt, en la Figura 3.2 se presenta una selección de imágenes representativas tomadas durante distintas semanas de crecimiento y bajo diferentes condiciones de iluminación con fotoperiodo inactivo y activo en la PFAL.

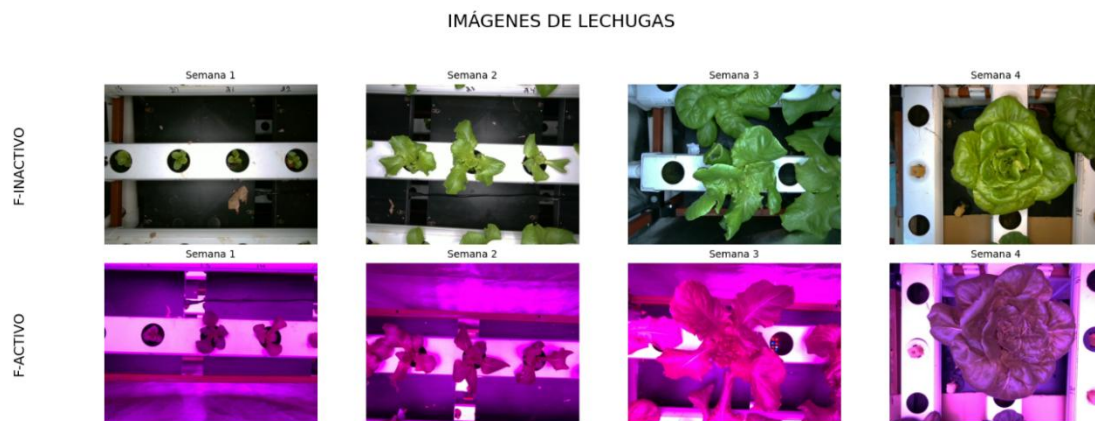


Figura 3.2. Imágenes en diferentes condiciones de iluminación y etapas de crecimiento.

Para validar la recomendación oficial de la página de YOLO (Ultralytics, 2025), el conjunto de imágenes se dividió en tres particiones para entrenamiento, validación y prueba utilizando tres configuraciones: 70–20–10 % (propuesta por Ultralytics como división estándar), 50–35–15 % y 90–5–5 %. Cada imagen está etiquetada según la clase correspondiente, donde el nombre codifica tanto la semana de crecimiento como el tipo de iluminación. Por ejemplo, F-A-S1 hace referencia a fotografías capturadas durante la semana 1 bajo condiciones de iluminación activa (horas luz) en la instalación. En contraste, F-IN-S1 identifica imágenes tomadas durante la semana 1, pero en condiciones de iluminación

inactiva. Estas imágenes capturadas por el robot servirán como entrada para una red neuronal del tipo YOLO serán las versiones YOLOv8n-seg y YOLOv11n-seg, en los procesos de entrenamiento, validación y prueba.

3.3.2 YOLOv8n-Seg y YOLO11: Descripción General

YOLOv8n-Seg y YOLOv11n-Seg son desarrollos recientes de Ultralytics (Ultralytics, 2023a, 2024), que representan el estado del arte en detección en tiempo real. A diferencia de versiones anteriores de la familia YOLO, estas versiones no solo están diseñadas para realizar detección de objetos, sino que han ampliado sus capacidades para incluir tareas como clasificación, segmentación de instancias y estimación de poses. Esta versatilidad convierte a YOLOv8n-Seg y YOLOv11n-Seg en herramientas integrales para aplicaciones complejas de visión por computadora, tanto en entornos industriales como agrícolas (Sapkota & Karkee, 2025). YOLOv8n-Seg representa una mejora considerable respecto a las versiones anteriores, en la Figura 3.3 se muestra la arquitectura de la red, así como las mejoras que se realizaron con respecto a versiones anteriores, se incorporó un diseño más eficiente y libre de anclajes (*anchor-free*), lo que simplifica el proceso de detección de objetos. Su arquitectura se apoya en una columna vertebral (*backbone*), optimizada mediante módulos CSP (Cross Stage Partial) refinados, los cuales permiten extraer características jerárquicas de forma eficaz. Estas características se organizan en diferentes niveles de abstracción: desde bordes y texturas en las primeras capas (características de bajo nivel), hasta partes de objetos y formas completas en las capas más profundas (características de alto nivel). Esta jerarquía es fundamental para que la red entienda la composición visual de los objetos en distintas escalas y contextos.

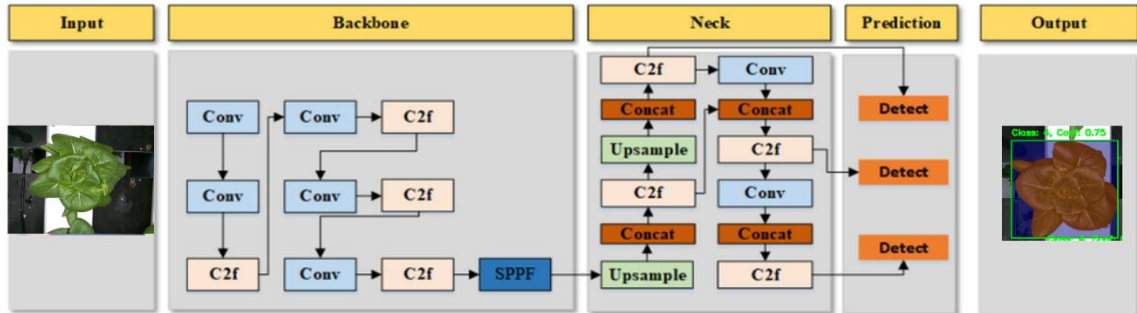


Figura 3.3. Arquitectura de la red neuronal convolucional YOLOv8n-Seg. Diagrama extraído y modificado de Sapkota & Karkee (2025).

El cuello (*neck*) de la arquitectura emplea una Red de Pirámide de Características (FPN) optimizada, que mejora la representación de objetos a múltiples escalas, permitiendo una detección más robusta tanto de objetos pequeños como grandes. Además, YOLOv8n-Seg incorpora una cabeza desacoplada, lo que significa que separa las tareas de detección en tres ramas independientes: objetividad (presencia de objeto), clasificación (tipo de objeto) y regresión (ubicación y tamaño de la caja delimitadora). Esta separación mejora la precisión sin introducir una carga computacional significativa. El bloque C2f (Cross Stage Partial with Feature Fusion) de YOLOv8n-Seg mejora la eficiencia del modelo al dividir el flujo de características en dos ramas: una procesada y otra directa, que luego se fusionan por concatenación. Este diseño optimiza la propagación del gradiente y conserva información espacial detallada, permitiendo una convergencia más rápida y estable. Gracias a ello, los modelos logran mayor precisión con menor complejidad computacional en tareas de detección y segmentación. En la Figura 3.4 se muestra la arquitectura YOLOv11n-Seg la cual incorpora avances significativos en segmentación de objetos, optimizando la precisión sin comprometer la eficiencia. Uno de sus principales cambios es la introducción del bloque C3k2 (Convolutional Block con tres convoluciones y kernel 2×2), que reemplaza al C2f anterior. Este nuevo bloque mejora el flujo de gradientes y reduce la carga computacional, favoreciendo una segmentación más precisa, especialmente en los contornos de objetos pequeños o complejos (Sapkota & Karkee, 2025). Además, se conserva el módulo SPPF (Spatial

Pyramid Pooling-Fast), que permite capturar información contextual a múltiples escalas, esencial para segmentar objetos de distintos tamaños en una misma imagen. También se incorpora el bloque C2PSA, que aplica atención espacial paralela para enfocarse en regiones clave, mejorando la segmentación en escenas con oclusiones o fondos visualmente complejos (Sapkota *et al.*, 2025).

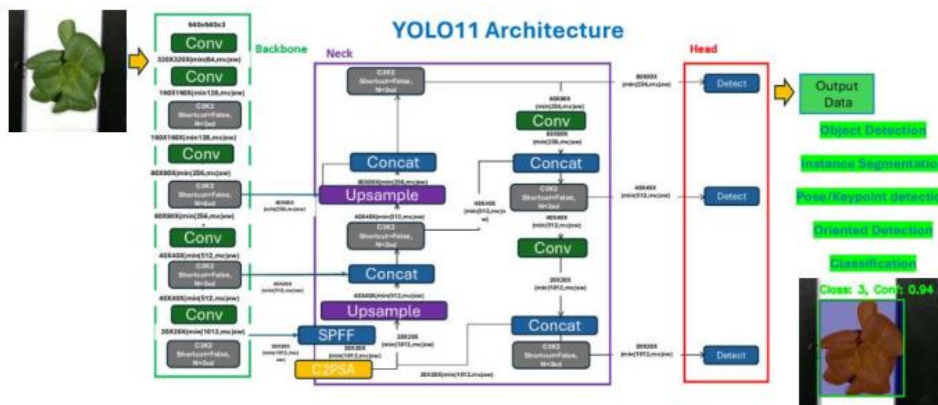


Figura 3.4. Arquitectura de la red neuronal convolucional YOLOv11n-Seg. Extraído y modificado de Sapkota & Karkee (2025).

La red recibe como entrada una imagen en formato RGB con una resolución de 640×480 píxeles (Figura 3.4). A través de sus diferentes módulos, la red procesa la imagen para identificar y segmentar los objetos presentes. La salida del modelo incluye una máscara binaria que delimita con precisión el objeto de interés, así como una caja delimitadora (bounding box) que enmarca su posición dentro de la imagen. El modelo recibe como entrada una imagen en formato RGB con una resolución de 640×480 píxeles. La arquitectura está diseñada para realizar múltiples tareas de visión por computadora, tales como detección de objetos, segmentación de instancias, detección de poses y clasificación. En este trabajo, se emplea específicamente para dos tareas: detección de objetos y segmentación semántica, aprovechando sus capacidades avanzadas para identificar y delimitar con precisión las regiones de interés dentro de la imagen.

3.3.3 Métricas de evaluación para las redes de segmentación:

Funciones de pérdida:

Las funciones de pérdida constituyen un componente esencial del entrenamiento de los modelos YOLO, ya que determinan cómo la red optimiza simultáneamente la clasificación, la localización de cajas y la segmentación de instancias. En las versiones YOLOv8n-seg y YOLOv11n-seg, la pérdida total se compone de tres términos principales: la pérdida de clasificación (L_{cls}), la pérdida de caja (L_{box}) y la pérdida de segmentación (L_{seg}).

Cálculo de Precisión, Recall, AP y mAP

El cálculo del área bajo la curva de Precisión-Recall permite obtener el valor de Average Precision (AP) para cada clase. A su vez, la mAP (mean Average Precision) se determina como el promedio de los valores de AP correspondientes a todas las clases presentes en el conjunto de evaluación. Las fórmulas utilizadas para el cálculo de estas métricas son las siguientes y se obtuvieron de (Powers & Ailab, 2011).

- Precisión (Precision):

$$Precision = \frac{TP}{(TP+FP)} \quad (1)$$

- Exhaustividad (Recall):

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

- Average Precision (AP):

$$AP = \int_0^1 p(r).dr \quad (3)$$

- Mean Average Precision (mAP):

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (4)$$

Donde:

- TP = Verdaderos positivos
- FP = Falsos positivos
- FN = Falsos negativos
- N = Número total de clases
- $p(r)$ = Precisión como función del recall
- AP_i = Precisión promedio de la clase i

Estas métricas permiten evaluar de forma robusta la capacidad del modelo para identificar correctamente los objetos, teniendo en cuenta tanto la calidad de las predicciones como su cobertura.

Evaluación de Segmentación Semántica

La evaluación del desempeño en tareas de segmentación semántica se basa comúnmente en diversas métricas que cuantifican la calidad de las predicciones generadas por el modelo. Una de las métricas más utilizadas es la Intersección sobre Unión (Intersection over Union, IoU) implementada por (Shelhamer *et al.*, 2017), la cual se define como la proporción entre el área de intersección de la máscara predicha y la máscara de referencia (ground truth), con respecto al área de la unión entre ambas:

$$IoU = \frac{\text{Área}(\text{Predicho} \cap \text{anotación})}{\text{Área}(\text{Predicho} \cup \text{anotación})} \quad (5)$$

A partir de la ecuación 5 se calcula el IoU promedio (mIoU) se implementó por (Shelhamer *et al.*, 2017), que representa el valor medio del IoU obtenido para todas las clases:

$$mIoU = \frac{1}{n} \times \sum_{i=1}^n IoU_i \quad (6)$$

Otra métrica relevante es la Precisión de Píxel (Pixel Accuracy, PA), que corresponde a la proporción entre el número de píxeles correctamente clasificados y el total de píxeles en la imagen:

$$PA = \frac{\text{Píxeles correctamente clasificados}}{\text{Total de píxeles}} \quad (7)$$

La Precisión de Píxel por Clase (Class Pixel Accuracy, CPA) se define como la fracción de píxeles correctamente clasificados para una clase específica en relación con el total de píxeles de esa clase:

$$CPA = \frac{\text{Píxeles correctamente clasificados de la clase } i}{\text{Total de píxeles de la clase } i} \quad (8)$$

El Promedio de Precisión de Píxel (Mean Pixel Accuracy, MPA) se obtiene al promediar las precisiones por clase:

$$MPA = \frac{1}{n} \times \sum_{i=1}^n PA_i \quad (9)$$

Finalmente, el Coeficiente de Dice (Dice Coefficient, DC) es una métrica alternativa que mide la similitud entre la máscara predicha y la de referencia se obtuvo de (Müller *et al.*, 2022), con un enfoque similar al IoU, pero ponderando la intersección con mayor peso:

$$DC = \frac{2 \times \text{Área}(\text{Máscara predicho} \cap \text{Máscara verdadera})}{(\text{Área}(\text{Máscara predicha}) + \text{Área}(\text{Máscara verdadera}))} \quad (10)$$

Estas métricas proporcionan una visión integral de la capacidad del modelo para segmentar correctamente objetos, tanto a nivel global como clase por clase.

Las redes se entrenaron de forma multitarea para realizar clasificación, detección y segmentación de manera simultánea, empleando un dataset con ocho clases que representan la semana de crecimiento vegetativo y la condición lumínica

(fotoperíodo activo o inactivo) de las plantas. Durante el entrenamiento y validación, se registraron las funciones de pérdida: *cls_loss* para clasificación, *box_loss* y *dfl_loss* para detección, y *seg_loss* para segmentación, junto con las métricas precisión (P), recall (R) y F1-score, que permiten cuantificar la capacidad del modelo para identificar correctamente los objetos detectados y clasificarlos sin errores. En la fase de prueba (test), se utilizaron los umbrales de intersección sobre unión (IoU) de 0.5 y de 0.5–0.95 para calcular las métricas mAP@0.5 y mAP@0.5–0.95, respectivamente. Estos umbrales definen el grado mínimo de superposición entre la caja predicha y la caja real para considerar una detección como correcta: un IoU = 0.5 mide la precisión general de detección y clasificación, mientras que el rango 0.5–0.95 evalúa la consistencia y robustez del modelo bajo criterios más estrictos. Finalmente, el coeficiente Dice se empleó para la segmentación, evaluando la superposición espacial entre las máscaras predichas y las reales, y cuantificando así la precisión con la que el modelo delimita las regiones de interés.

3.3.4 Entrenamiento de los modelos de segmentación

El entrenamiento de los modelos de segmentación se llevó a cabo utilizando la arquitectura YOLOv8n-seg (Ultralytics, 2023a) Y YOLOv11n-seg (Ultralytics, 2024), preentrenada sobre el conjunto COCO (Objetos Comunes en Contexto) (T.-Y. Lin *et al.*, 2014) y posteriormente ajustada con un conjunto de datos personalizado mediante la biblioteca oficial de Ultralytics, el modelo fue entrenado durante 300 épocas con early stopping activado tras 50 épocas sin mejora utilizando una GPU compatible con CUDA, las imágenes de entrada fueron de 640×640 píxeles y se empleó un tamaño de lote de 32 junto con 24 hilos de procesamiento para la carga de datos, los hiperparámetros de optimización incluyeron una tasa de aprendizaje inicial de 0.01 con un decaimiento de 0.01, el momentum de 0.937 y un término de regularización L2 de 0.0005, el warm-up se aplicó durante 3 épocas con un momentum inicial de 0.8 y una tasa de aprendizaje específica para los parámetros bias, los pesos de

pérdida fueron configurados con 7.5 para la pérdida de caja delimitadora, 0.5 para la pérdida de clasificación, y 1.5 para la pérdida Distribution Focal, además se aplicó un ratio de reducción para las máscaras de 4, el entrenamiento fue determinista con una semilla aleatoria de 42 y se utilizó entrenamiento con precisión mixta (AMP) implementado por (Narang *et al.*, 2018) para acelerar el proceso sin comprometer precisión, finalmente el modelo fue exportado en formato ONNX para facilitar su uso en entornos de inferencia, el entrenamiento se realizó en una estación de trabajo equipada con un procesador Intel Core i9-12900K de 16 núcleos y 24 hilos, una tarjeta gráfica NVIDIA GeForce RTX 3090 con 24 GB de memoria dedicada, 64 GB de memoria RAM DDR4, una unidad de almacenamiento SSD NVMe de 2 TB y sistema operativo Windows 10 Pro de 64 bits, esta configuración permitió ejecutar el entrenamiento con eficiencia en procesamiento paralelo y aprovechamiento óptimo de los recursos computacionales disponibles.

3.4 RESULTADOS Y DISCUSIÓN

3.4.1 Evaluación de las funciones de pérdida.

Entrenamiento y validación:

Las curvas de pérdida correspondientes a las tareas de clasificación, detección y segmentación se muestran en las Figuras Figura 3.5 (I-II-III), en la etapa de entrenamiento y validación, la evaluación se presenta en la sección 3.4.2 Durante las primeras 10 épocas, las pérdidas iniciales mostraron diferencias sistemáticas tanto entre modelos como entre distribuciones. En la tarea de clasificación, en la etapa de entrenamiento (Figura 3.5-I) el esquema (50 %, 35 %, 15 %) inició con pérdidas promedio de 1.141 ± 0.502 en YOLOv8n-Seg y 1.187 ± 0.591 en YOLOv11n-Seg, en (70 %, 20 %, 10 %), los valores fueron 1.235 ± 0.580 y 1.271 ± 0.605 , respectivamente y en (90 %, 5 %, 5 %), descendieron a 1.149 ± 0.552 y 1.188 ± 0.580 . En promedio, YOLOv11n-Seg inició el aprendizaje con una pérdida 3.8 % mayor y una desviación estándar 4.3 % más alta, lo que sugiere una mayor

dispersión inicial de gradientes y un ajuste menos regularizado al inicio del aprendizaje. Durante la fase de validación, las curvas siguieron una tendencia asintótica, con una diferencia media entre entrenamiento y validación $\Delta L_{val-train}$ de 0.008 para YOLOv8n-Seg y -0.001 para YOLOv11n-Seg, lo que representa una reducción del 33.3 % en la brecha de generalización en favor de este último. En general YOLOv8n-Seg converge antes que YOLOv11n-Seg debido al mecanismo de paciencia que se aplica, si en 50 épocas la métrica de pérdida deja de mejorar, el algoritmo se detiene, esto con la finalidad de evitar el sobreentrenamiento en la red.

En la tarea de detección (Figura 3.5-II), la pérdida total está compuesta por la suma ponderada de los términos de localización (*box_loss*), clasificación (*cls_loss*) y distribución de frontera (*dfl_loss*), y mostró un descenso sostenido desde las primeras épocas. Donde se observa que YOLOv11n-Seg inicia el entrenamiento con valores ligeramente menores, lo que indica una mayor estabilidad inicial, mientras que YOLOv8n-Seg presenta una tasa de reducción más pronunciada, evidenciando un aprendizaje más dinámico durante las primeras iteraciones. La configuración (70 %, 20 %, 10 %) registró el comportamiento más uniforme en ambas redes, con desviaciones estándar más bajas, lo que sugiere un ajuste más regularizado y consistente de los gradientes en las etapas iniciales. En cambio, la configuración (90 %, 5 %, 5 %) presentó una ligera dispersión mayor, posiblemente asociada con un menor volumen de datos de validación.

Durante la validación en la tarea de detección, las pérdidas siguieron un comportamiento asintótico similar al observado en el entrenamiento, manteniendo una correlación directa con las curvas de convergencia. Las pérdidas mínimas de validación fueron de 0.214, 0.203 y 0.231 para YOLOv8n-Seg, y de 0.197, 0.194 y 0.215 para YOLOv11n-Seg en las configuraciones (50 %, 35 %, 15 %), (70 %, 20 %, 10 %) y (90 %, 5 %, 5 %), respectivamente. La diferencia promedio entre validación y entrenamiento ($\Delta L_{val-train}$) fue de 0.026

para YOLOv8n-Seg y 0.017 para YOLOv11n-Seg, lo que representa una reducción relativa del 34.6 % en la brecha de generalización. Esto confirma que YOLOv11n-Seg logra una convergencia más estable y una mayor capacidad de generalización estadística, especialmente bajo la configuración 70–20–10, que mostró la menor varianza inter-época y la mejor relación entre precisión y estabilidad, se puede observar por la proximidad entre las curvas de la función de pérdida (*loss*) de entrenamiento y validación indica una mayor capacidad de generalización y una mitigación efectiva del sobreajuste.

En la tarea de segmentación (Figura 3.5-III), las curvas de pérdida en la etapa de entrenamiento mostraron un comportamiento decreciente y estable, con convergencia gradual a partir de la época 150. Durante las primeras 10 épocas, las pérdidas iniciales promedio durante en el entrenamiento fueron, para YOLOv8n-Seg, 0.508 ± 0.136 en (50 %, 35 %, 15 %), 0.514 ± 0.118 en (70 %, 20 %, 10 %) y 0.494 ± 0.110 en (90 %, 5 %, 5 %); y para YOLOv11n-Seg, 0.512 ± 0.112 en (50 %, 35 %, 15 %), 0.529 ± 0.100 en (70 %, 20 %, 10 %) y 0.507 ± 0.090 (90 %, 5 %, 5 %), respectivamente. En validación, los valores iniciales fueron ligeramente inferiores, con promedios de 0.378 ± 0.063 , 0.372 ± 0.047 y 0.370 ± 0.020 en YOLOv8n-seg, y 0.393 ± 0.058 , 0.381 ± 0.031 y 0.430 ± 0.091 en YOLOv11n-seg, reflejando una adecuada inicialización de pesos y estabilidad del gradiente.

A lo largo del entrenamiento, las pérdidas disminuyeron consistentemente en ambas redes, alcanzando valores mínimos en torno a las épocas 170–270, con ligeros desplazamientos entre entrenamiento y validación. En YOLOv8n-seg, las pérdidas finales fueron 0.263, 0.277 y 0.292 en (50 %, 35 %, 15 %), (70 %, 20 %, 10 %) y (90 %, 5 %, 5 %); mientras que las pérdidas mínimas de validación se situaron en torno a 0.239, 0.228 y 0.247, registradas entre las épocas 200–210, ligeramente posteriores al mínimo de entrenamiento. En YOLOv11n-seg, los valores finales de entrenamiento fueron 0.278, 0.270 y 0.283, y los mínimos de

validación se alcanzaron en torno a las épocas 180–280, con pérdidas de 0.238, 0.227 y 0.243, respectivamente.

La diferencia entre las pérdidas mínimas de validación y entrenamiento ($\Delta L_{\text{val-train}}$) fue de -0.024 , -0.049 y -0.045 para YOLOv8n-seg, y de -0.040 , -0.043 y -0.040 para YOLOv11n-seg, lo que indica que, en el punto de convergencia, la pérdida de validación se mantuvo ligeramente inferior a la de entrenamiento, consistente con una regularización efectiva y ausencia de sobreajuste para esta tarea. Finalmente, la configuración (70 %, 20 %, 10 %) fue la más equilibrada: produjo las pérdidas de validación más bajas (0.228 en YOLOv8n-seg y 0.227 en YOLOv11n-seg) y una convergencia más tardía pero estable. Este comportamiento sugiere que un 20 % del conjunto de datos destinado a validación permite mantener una buena capacidad de generalización sin incrementar la varianza ni penalizar la convergencia espacial del modelo.

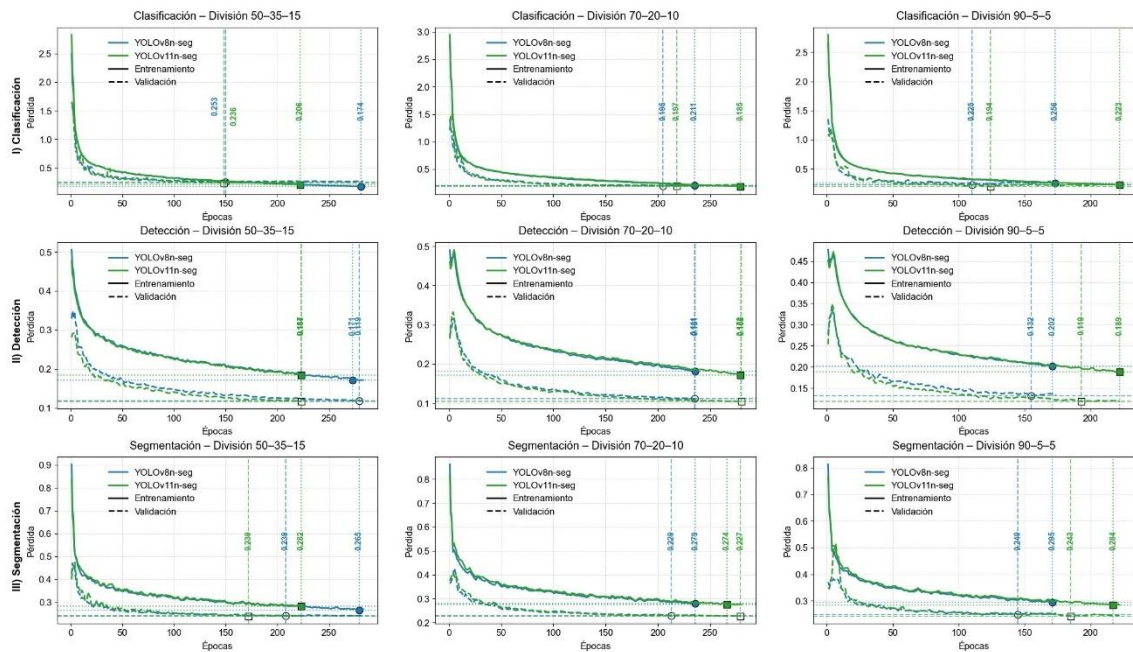


Figura 3.5. Evolución de las pérdidas de entrenamiento y validación para clasificación, detección y segmentación de cada distribución del dataset.

3.4.2 Evaluación de las métricas de rendimiento

El desempeño de los modelos YOLOv8n-Seg y YOLOv11n-Seg se evaluó mediante las métricas mAP50 y mAP50–95, que cuantifican la precisión promedio en las tareas de detección y segmentación considerando distintos umbrales de intersección sobre unión (IoU). Estas métricas permiten analizar la capacidad de generalización y la estabilidad estadística del modelo frente a las tres configuraciones del conjunto de datos: (50 %, 35 %, 15 %), (70 %, 20 %, 10 %) y (90 %, 5 %, 5 %). En todos los casos, las curvas alcanzaron más del 95 % de su valor máximo antes de la época 50, lo que refleja una convergencia temprana y una alta capacidad de aprendizaje. Al analizar la varianza inter-época, se identificó que la configuración (90 %, 5 %, 5 %) presentó una menor estabilidad estadística, con fluctuaciones métricas amplias entre épocas consecutivas, lo que indica un comportamiento más sensible al sobreajuste. En contraste, la configuración (70 %, 20 %, 10 %) mostró las varianzas más bajas, evidenciando un comportamiento más estable durante la fase post-convergencia. Esta mayor dispersión en el esquema (90 %, 5 %, 5 %) coincide con la tendencia al sobreajuste observada previamente en las curvas de pérdida, donde, si bien se mantiene un rendimiento elevado, la consistencia inter-época disminuye.

En la tarea de segmentación, las métricas mAP50 y mAP50–95 evaluaron la capacidad de los modelos para generar máscaras precisas de las regiones de interés. La evolución temporal de ambas métricas para YOLOv8n-Seg y YOLOv11n-Seg, bajo las tres configuraciones del conjunto de datos: (50 %, 35 %, 15 %), (70 %, 20 %, 10 %) y (90 %, 5 %, 5 %). En todos los casos, las curvas alcanzaron más del 95 % de su valor máximo antes de la época 50, lo que refleja una convergencia temprana y consistente con la observada en las pérdidas de entrenamiento. En la configuración (50 %, 35 %, 15 %), el desempeño inicial fue el más moderado, con $mAP50 = 0.966$, 0.9736 , $mAP50-95 = 0.9546$, 0.9617 para YOLOv8n-Seg y YOLOv11n-Seg respectivamente. Las varianzas inter-época fueron bajas, con: $\sigma^2_{mAP(50)} = 0.018 \times 10^{-4}$, $\sigma^2_{mAP(50-95)} = 0.016 \times 10^{-4}$ para

YOLOv8n-Seg, y $\sigma^2_{mAP(50)} = 0.037 \times 10^{-4}$, $\sigma^2_{mAP(50-95)} = 0.026 \times 10^{-4}$ para YOLOv11n-Seg. Este comportamiento refleja una evolución estable post-covergencia, con baja dispersión métrica y una relación directa con las pérdidas de validación. En la configuración (70 %, 20 %, 10 %), ambas redes alcanzaron el mayor equilibrio entre precisión y estabilidad estadística. Los valores máximos fueron $mAP_{50} = 0.9785$, 0.9787 , con $mAP_{50-95} = 0.9673$, 0.9681 para YOLOv8n-Seg y YOLOv11n-Seg respectivamente. Las varianzas permanecieron mínimas, con: $\sigma^2_{mAP(50)} = 0.022 \times 10^{-4}$, $\sigma^2_{mAP(50-95)} = 0.021 \times 10^{-4}$ para YOLOv8n-Seg, $\sigma^2_{mAP(50)} = 0.017 \times 10^{-4}$, $\sigma^2_{mAP(50-95)} = 0.012 \times 10^{-4}$ para YOLOv11n-Seg. Este comportamiento indica una mayor robustez temporal y menor sensibilidad a las fluctuaciones del conjunto de validación. Finalmente, la configuración (90 %, 5 %, 5 %) mostró los valores más altos de precisión, con $mAP_{50} = 0.9843$, 0.9855 , y $mAP_{50-95} = 0.9725$ y 0.976 para YOLOv8n-Seg y YOLOv11n-Seg respectivamente. Sin embargo, esta configuración presentó las varianzas más elevadas: $\sigma^2_{mAP(50)} = 0.079 \times 10^{-4}$, $\sigma^2_{mAP(50-95)} = 0.088 \times 10^{-4}$ para YOLOv8n-Seg y $\sigma^2_{mAP(50)} = 0.03 \times 10^{-4}$, $\sigma^2_{mAP(50-95)} = 0.035 \times 10^{-4}$ para YOLOv11n-Seg. Estas dispersiones son hasta cuatro veces superiores a las observadas en la configuración (70 %, 20 %, 10 %), lo que indica inestabilidad post-covergencia y mayor variabilidad métrica entre épocas, coherente con la tendencia al sobreajuste evidenciada en las curvas de pérdida. En conjunto, la configuración (70 %, 20 %, 10 %) se posiciona como la óptima para la tarea de segmentación, combinando alta precisión ($mAP_{50} > 0.97$, $mAP_{50-95} > 0.96$) con baja varianza ($< 0.022 \times 10^{-4}$). Por el contrario, aunque la configuración (90 %, 5 %, 5 %) alcanza el máximo rendimiento aparente, su alta varianza y dispersión temporal reducen su capacidad de generalización, afectando la estabilidad de la inferencia.

3.4.3 Evaluación (Test):

La evaluación de desempeño tuvo como objetivo comparar las arquitecturas YOLOv8n-seg y YOLOv11n-seg en el conjunto de datos de prueba, con el fin de

determinar su capacidad de generalización en las tareas de detección de objetos, segmentación de instancias y clasificación del estado de crecimiento de lechugas cultivadas en el sistema PFAL. Cada modelo fue evaluado mediante un barrido de confianza entre 0.05 y 0.95 con incrementos de 0.05, registrando las métricas Precisión, Sensibilidad (Recall), F1-score, mAP@0.5, mAP@0.5–0.95 y el coeficiente Dice. Las métricas mAP@0.5 y mAP@0.5–0.95 se calcularon utilizando umbrales de Intersección sobre Unión (IoU) de 0.5 y el promedio en el rango 0.5–0.95 con incrementos de 0.05, siguiendo el protocolo estándar de evaluación de COCO. El umbral de confianza óptimo se seleccionó con base al valor máximo de F1-score al representar el equilibrio entre precisión y sensibilidad, los valores asociados a lo anterior se presentan en el Cuadro 3.1.

Cuadro 3.1. Métricas de desempeño óptimas de las arquitecturas YOLOv8n-seg y YOLOv11n-seg en el conjunto de prueba, considerando tres configuraciones de distribución de datos (50–35–15, 70–20–10 y 90–5–5).

Modelo	Distribución (%)	Confianza	Precisión	Recall	F1-score	mAP@0.5	mAP@0.5–0.95	Dice
YOLO11	50–35–15	0.40	0.942	0.954	0.948	0.963	0.958	0.963
YOLO11	70–20–10	0.60	0.951	0.954	0.953	0.966	0.962	0.966
YOLO11	90–5–5	0.60	0.954	0.988	0.971	0.985	0.980	0.985
YOLO 8	50–35–15	0.05	0.927	0.960	0.943	0.964	0.960	0.964
YOLO 8	70–20–10	0.45	0.937	0.966	0.952	0.971	0.967	0.971
YOLO 8	90–5–5	0.65	0.942	0.966	0.954	0.977	0.969	0.977

En la distribución 50–35–15 (Figura 3.6-IA, IIA, IIIA), ambas redes mostraron un comportamiento similar, con valores de mAP@0.5 y Dice cercanos a 0.963–0.964, lo que refleja una segmentación precisa del dosel vegetal y una delimitación coherente de las hojas individuales. En esta configuración, YOLOv11n-Seg-seg alcanzó un F1-score de 0.948, ligeramente superior al de YOLOv8n-Seg-seg (0.943), lo que evidencia una mejor relación entre precisión y sensibilidad. La configuración 70–20–10 (Figura 3.6-IB, IIB, IIIB) presentó el mejor equilibrio general entre las métricas de desempeño. En este caso, YOLOv11n-Seg-seg alcanzó valores de F1 = 0.953 y Dice = 0.966, mientras que YOLOv8n-Seg-seg obtuvo F1 = 0.952 y Dice = 0.971. Estos resultados equivalen a una coincidencia espacial promedio del 96–97 % entre las áreas segmentadas por los modelos y las máscaras reales, lo que indica que casi toda la superficie foliar fue correctamente identificada y delimitada. La Figura 3.6-IIB muestra cómo ambas arquitecturas mantienen una tendencia ascendente del Dice hasta su punto óptimo, con curvas suaves que reflejan estabilidad en la segmentación del dosel a distintos umbrales de confianza. En la distribución 90–5–5 (Figura 3.6-IC, IIC, IIIC), donde el conjunto de validación fue reducido pero el conjunto de entrenamiento representó el mayor porcentaje de datos, se obtuvieron los valores más altos en todas las métricas. En esta configuración, YOLOv11n-seg alcanzó mAP@0.5 = 0.985, mAP@0.5–0.95 = 0.980 y Dice = 0.985, mientras que YOLOv8n-seg logró mAP@0.5 = 0.977 y Dice = 0.977. El desempeño superior en esta distribución se atribuye al mayor volumen de datos de entrenamiento, lo que permitió a los modelos aprender una mayor diversidad de patrones morfológicos y mejorar su capacidad de representación espacial del dosel vegetal. Al entrenar con un conjunto más extenso, los modelos lograron mayor generalización intra-clase, es decir, pudieron reconocer lechugas de diferentes tamaños y estados de crecimiento con alta precisión. Sin embargo, el reducido conjunto de validación (5%) limitó la detección temprana de sobreajuste, por lo que este rendimiento debe interpretarse como resultado de un aprendizaje más exhaustivo, pero menos controlado, que beneficia la exactitud en el conjunto de

prueba, aunque podría reducir la estabilidad frente a datos completamente nuevos. Las diferencias entre ambas arquitecturas en la tarea de segmentación se observan claramente en la Figura 3.6-II, donde las curvas de YOLOv11n-seg (verde) se mantienen por encima de las de YOLOv8n-seg (azul) en la mayoría de los umbrales de confianza, particularmente en las distribuciones B y C, que corresponden a las configuraciones con mayor proporción de datos de entrenamiento. Estos resultados evidencian que YOLOv11n-seg posee una mayor robustez estructural y estabilidad ante variaciones en el tamaño de los conjuntos de validación, mientras que YOLOv8n-seg mantiene un desempeño eficiente y competitivo con menor complejidad computacional.

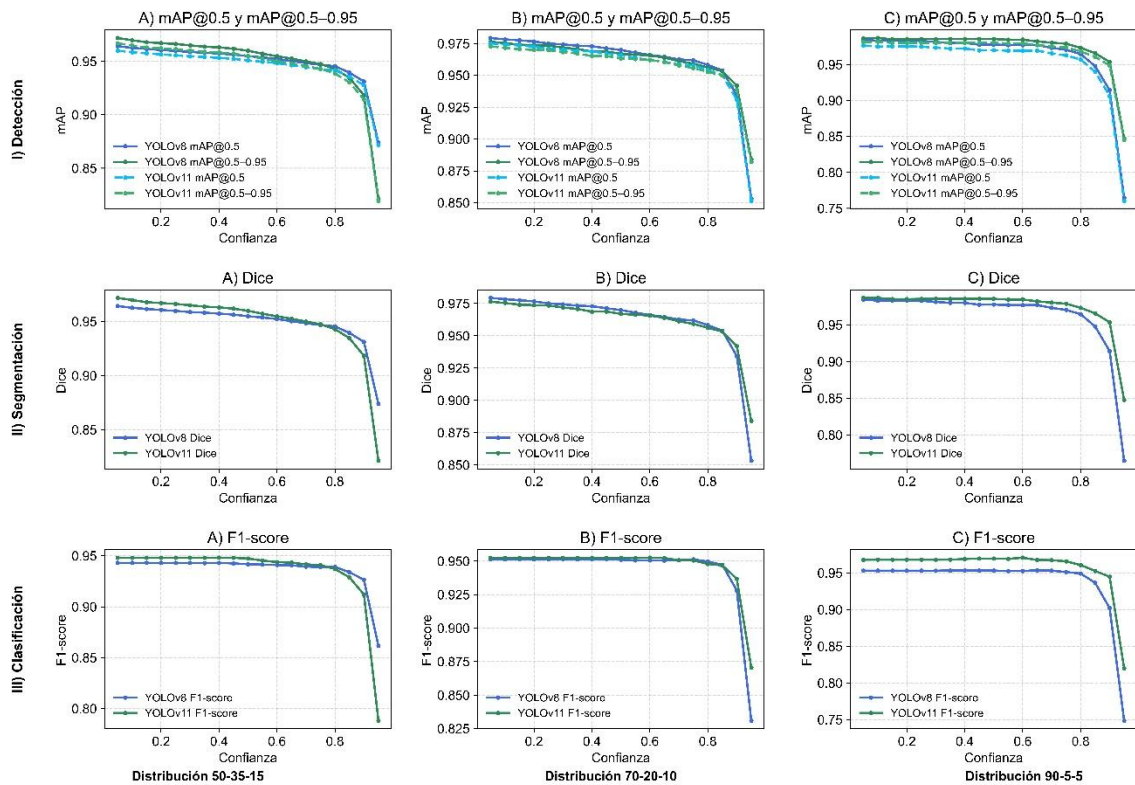


Figura 3.6. Desempeño de YOLOv8n-Seg-seg y YOLOv11n-Seg-seg en el conjunto de prueba.

3.4.4 Inferencias:

La distribución 70–20–10 (Figura 3.6-IB, IIB, IIIB) se identificó como la más estable para realizar las inferencias, al ofrecer un equilibrio adecuado entre datos de entrenamiento y validación, para las inferencias los valores correspondientes al punto de mayor equilibrio entre precisión y sensibilidad garantizan segmentaciones del dosel con una coincidencia espacial del 96–97 % respecto a las máscaras reales. Los umbrales seleccionados minimizan falsos positivos sin perder verdaderos, asegurando inferencias consistentes y de alta fidelidad geométrica. El análisis de desempeño de los modelos YOLOv8n-seg y YOLOv11n-seg se llevó a cabo bajo las dos condiciones de iluminación presentes en el sistema PFAL Y utilizando el valor de confianza óptimo descrito en el Cuadro 3.1 y un umbral IoU 0.7, el fotoperiodo activo (F-A), caracterizado por una iluminación púrpura generada mediante la combinación de diodos emisores de luz (LED) rojos y azules, y el fotoperiodo inactivo (F-IN), en el cual la iluminación del entorno es mínima y las imágenes se capturan con luz blanca tenue proporcionada por un aro de LED integrado alrededor del sistema de cámara. Este módulo óptico permite la obtención de imágenes visibles durante el periodo oscuro de las plantas sin alterar significativamente su fisiología. Los resultados mostraron diferencias significativas entre ambos modelos en cuanto a precisión, estabilidad y sensibilidad ante las condiciones de iluminación. En la Figura 3.7 se presenta la proporción de falsos positivos (FP) respecto al total de detecciones por clase. Se observa que YOLOv11n-Seg-seg mantuvo valores más bajos de FP en todas las etapas del crecimiento, mientras que YOLOv8n-Seg-seg presentó un incremento de hasta 40 % bajo el fotoperiodo inactivo. Este aumento se atribuye al menor contraste cromático generado por la luz blanca del sistema de captura, que homogeniza la luminancia del follaje y las superficies reflectantes, provocando sobre-segmentaciones en los bordes de las hojas. Bajo iluminación púrpura, ambos modelos redujeron considerablemente la proporción de FP debido al mayor contraste espectral entre la vegetación y el fondo. La relación entre falsos positivos y falsos negativos (FN) se muestra en la Figura 3.8, donde

el tamaño de los puntos representa el número de verdaderos positivos (TP). YOLOv11n-Seg-seg se concentró en la región con bajos FP y FN moderados (6–20 por clase), mostrando un comportamiento más conservador y equilibrado entre precisión y exhaustividad. Por el contrario, YOLOv8n-Seg-seg presentó un mayor número de FP, pero con FN reducidos, lo que refleja una estrategia de detección expansiva. En el contexto de las PFALs, donde la superficie de la cubierta de malla incrementa los reflejos, YOLOv11n-Seg-seg ofreció una segmentación más controlada y menos afectada por interferencias ópticas, mientras que YOLOv8n-Seg-seg tendió a incluir zonas periféricas brillantes dentro de las máscaras segmentadas.

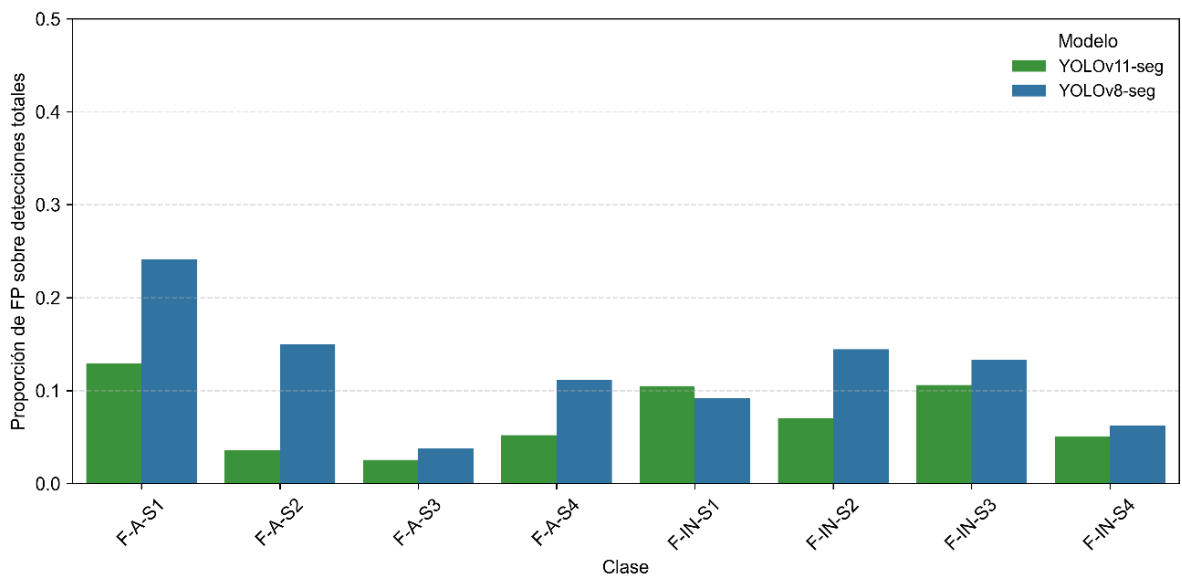


Figura 3.7. Proporción de falsos positivos por clase bajo diferentes condiciones de iluminación.

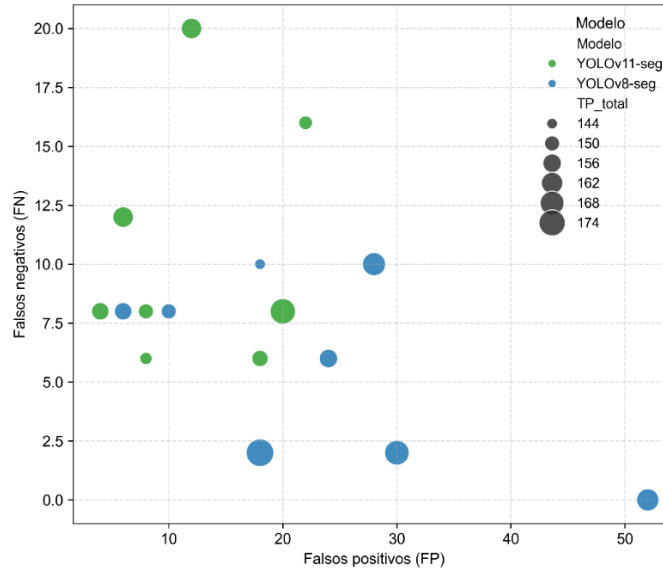


Figura 3.8. Proporción de falsos positivos por clase bajo diferentes condiciones de iluminación.

Durante las cuatro semanas de crecimiento, ambos modelos mostraron un aumento progresivo en el F1-score promedio, lo que evidencia una mejora en la capacidad de segmentación a medida que las plantas incrementan su biomasa y el dosel adquiere una morfología más definida. En las primeras semanas, cuando las plántulas presentaban un tamaño reducido y no existía superposición foliar, el rendimiento fue limitado por el bajo contraste cromático entre el follaje tierno y la superficie reflectante, que actúa como fondo dentro del sistema PFAL. Esta superficie de maylar refleja parcialmente la luz incidente, generando zonas de luminancia variable que pueden confundir los modelos durante la detección de los bordes foliares. Conforme las plantas avanzaron hacia las semanas 3 y 4, el F1-score aumentó notablemente, lo que indica una mayor estabilidad en la detección del dosel y una mejora en la coherencia espacial de las máscaras segmentadas. Este comportamiento coincide con el incremento del área foliar y la uniformidad del dosel, factores que facilitan el aprendizaje de patrones morfológicos más consistentes. Bajo iluminación púrpura (fotoperiodo activo), ambos modelos alcanzaron valores máximos de desempeño, superando 0.93 de

F1-score en la última semana, debido a la diferenciación espectral entre el follaje y el fondo reflejante, que refuerza el contraste en los canales rojo y azul del sensor RGB. En contraste, bajo el fotoperiodo inactivo (iluminación blanca del sistema de captura), el F1-score promedio fue inferior en 2–4 %, atribuible a la reducción del contraste espectral y al aumento de zonas de brillo homogéneo que dificultan la delimitación del contorno foliar. YOLOv11n-Seg-seg mantuvo una variación intersemanal inferior al 3 %, mostrando un patrón de crecimiento del desempeño casi lineal, lo que refleja una segmentación más estable y una mejor adaptación a las condiciones ópticas del entorno. Por otro lado, YOLOv8n-Seg-seg presentó un comportamiento más variable, con oscilaciones de hasta 8 % entre semanas consecutivas, indicando una mayor sensibilidad a las variaciones lumínicas y al tamaño de las plantas. En conjunto, estos resultados confirman que el aumento del área foliar y la mayor diferenciación espectral bajo luz púrpura favorecen la segmentación precisa, mientras que las reflexiones difusas y la luz blanca tenue introducen pequeñas fluctuaciones en la respuesta de ambos modelos.

La correlación entre el F1-score y la intersección sobre unión (IoU) mostró una relación positiva fuerte ($r \approx 0.93$), evidenciando que la mejora en precisión y exhaustividad se traduce en una mayor correspondencia espacial entre las máscaras segmentadas y el área real del dosel. Los resultados de YOLOv11n-seg se concentraron en la región superior del plano F1–IoU, lo que indica una segmentación más coherente y estable, mientras que YOLOv8n-Seg-seg presentó mayor dispersión, principalmente en condiciones de iluminación blanca. En conjunto, los resultados demuestran que ambos modelos alcanzaron un desempeño sobresaliente en la segmentación del dosel vegetal dentro del entorno PFAL. YOLOv8n-seg obtuvo un F1 promedio de 0.91 y una IoU media de 0.84, destacando por su alta sensibilidad y velocidad, lo que lo hace apropiado para aplicaciones de monitoreo continuo y conteo de plantas en tiempo real. Por su parte, YOLOv11n-seg alcanzó un F1 promedio de 0.93 y una IoU media de 0.88, ofreciendo mayor estabilidad frente a variaciones lumínicas y una reducción

significativa de falsos positivos. Estas características lo posicionan como el modelo más adecuado para fenotipado de precisión y estimación del área foliar en sistemas agrícolas controlados con superficies reflectantes.

Las Figura 3.9 y Figura 3.10 presentan una comparación visual de las inferencias obtenidas por los modelos YOLOv8n-seg y YOLOv11n-seg sobre el conjunto de prueba, bajo condiciones de fotoperiodo activo e inactivo. En ambos casos se muestran dieciséis imágenes representativas (dos por clase) que abarcan las diferentes etapas de crecimiento de las lechugas. Las máscaras verdes corresponden a las regiones verdaderas anotadas (Ground Truth), mientras que las máscaras azules y naranjas representan las segmentaciones predichas por YOLOv8n-seg y YOLOv11n-seg, respectivamente. Los cuadros delimitadores indican las detecciones asociadas a cada planta, acompañadas de su etiqueta de clase y nivel de confianza. En las imágenes correspondientes al modelo YOLOv8n-seg, se observa un desempeño adecuado en la detección y segmentación de las lechugas, logrando una correcta delimitación espacial en la mayoría de los casos. No obstante, en las primeras semanas de crecimiento, donde el área foliar es reducida y existe menor superposición de hojas, el modelo tiende a presentar ligeras sobre segmentaciones y falsas detecciones en los bordes de la estructura del sistema de cultivo. Estas desviaciones se asocian a la similitud de color y textura entre las hojas jóvenes y los reflejos del entorno, particularmente bajo la iluminación púrpura del fotoperiodo activo. Por su parte, el modelo YOLOv11n-seg muestra una mejor coherencia espacial en la segmentación, con bordes más ajustados al contorno foliar y una mayor estabilidad frente a variaciones lumínicas. Las inferencias obtenidas revelan una mejora en la detección de plantas parcialmente superpuestas y una reducción de los errores en regiones de alta reflectancia, lo cual coincide con los valores más altos de precisión (0.95) y F1-score (0.95) alcanzados por esta arquitectura. Además, YOLOv11n-seg logra mantener un equilibrio entre sensibilidad y especificidad en ambas condiciones de iluminación, evidenciando su capacidad para generalizar sobre distintas intensidades y espectros de luz.

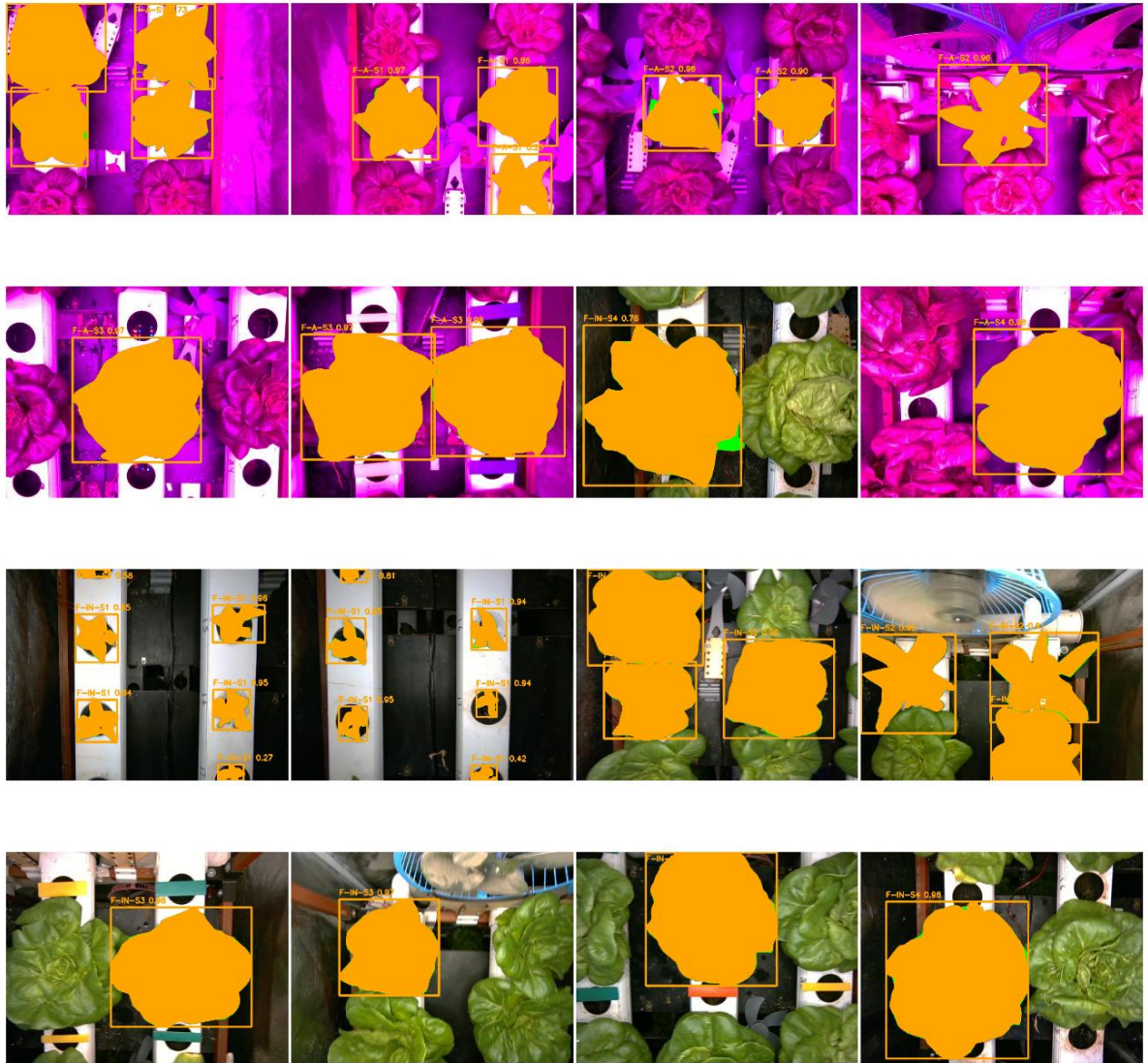


Figura 3.9. Resultados de las inferencias obtenidas con YOLOv11n-seg.

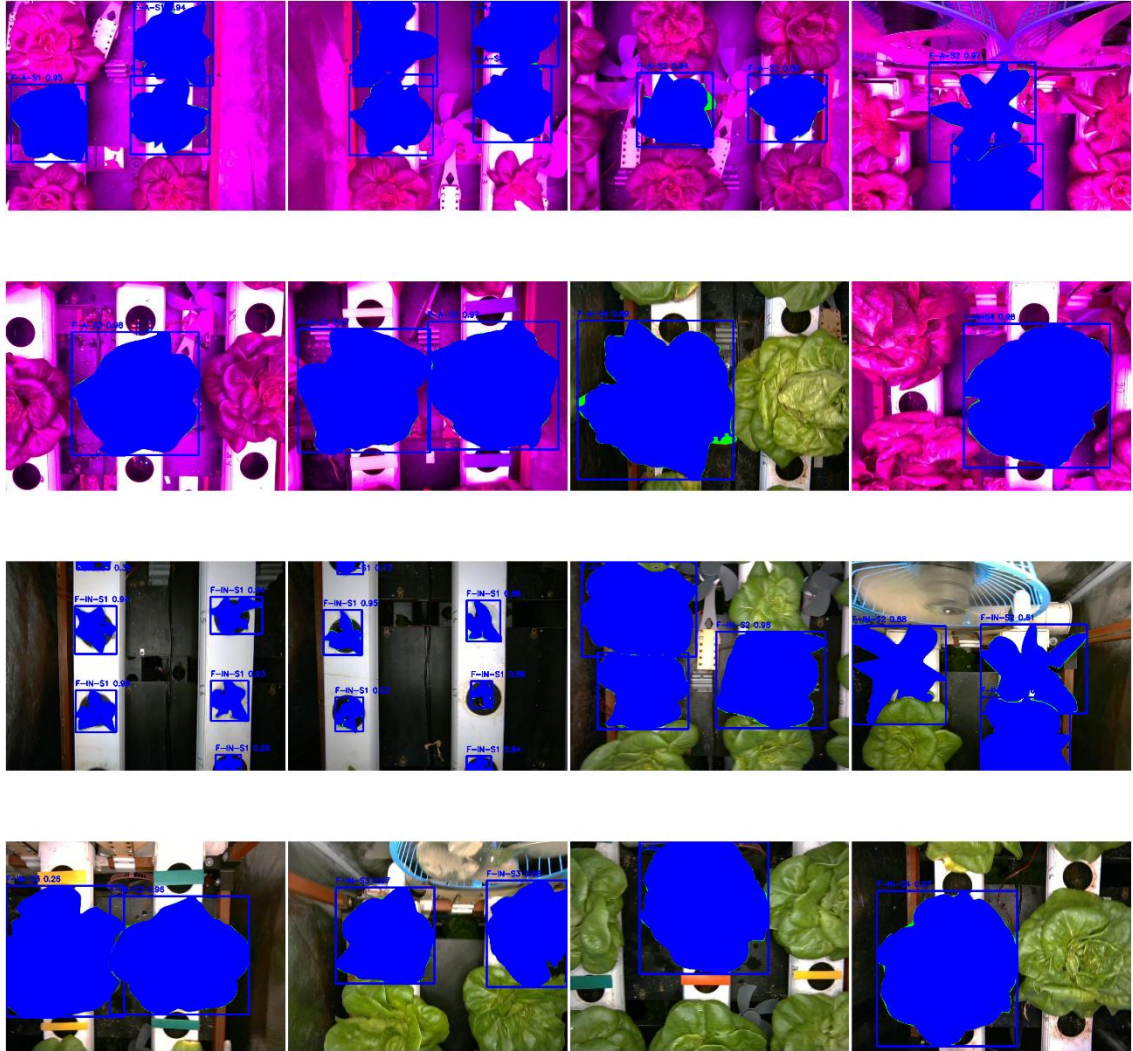


Figura 3.10. Resultados de las inferencias obtenidas con YOLOv11n-seg sobre el mismo conjunto de imágenes utilizadas para YOLOv8n-seg.

Estas observaciones cualitativas complementan los resultados cuantitativos de desempeño, confirmando la superioridad del modelo YOLOv11n-Seg-seg en precisión, consistencia morfológica y tolerancia a la variabilidad espectral de la iluminación en el entorno controlado del sistema PFAL. La Figura 23 muestra ejemplos ilustrativos de los falsos positivos (FP) y falsos negativos (FN) detectados en las inferencias de ambos modelos. En esta figura se incluyen ocho imágenes (dispuestas en una cuadrícula de 4×2), donde la primera fila corresponde a errores de YOLOv8n-seg y la segunda a YOLOv11n-seg. Las

máscaras verdes indican las etiquetas verdaderas, mientras que las máscaras azules (YOLOv8n-Seg) y naranjas (YOLOv11n-Seg) corresponden a las predicciones de los modelos. Los cuadros rojos identifican falsos positivos, es decir, detecciones erróneas de objetos inexistentes, mientras que los cuadros grises señalan falsos negativos, es decir, plantas reales no detectadas.

El análisis cualitativo de estas imágenes permite observar que los falsos positivos tienden a producirse en zonas de alto contraste o en superficies reflectantes, donde los modelos confunden brillos o texturas con estructuras foliares. En cambio, los falsos negativos se concentran en las primeras etapas de desarrollo, cuando las hojas presentan menor área proyectada y un color menos saturado. Estos errores también pueden atribuirse a las condiciones lumínicas variables dentro del sistema de cultivo vertical, particularmente en los bordes de la cámara de cultivo, donde la intensidad del LED disminuye y se generan sombras parciales. En conjunto, los resultados visuales evidencian que, aunque ambos modelos logran una segmentación robusta del dosel foliar, YOLOv11n-Seg-seg demuestra una mayor capacidad de discriminación espacial y fotométrica Figura 3.11, adaptándose mejor a la heterogeneidad del entorno y reduciendo los errores de clasificación y segmentación observados en YOLOv8n-Seg-seg. Estas observaciones cualitativas complementan los resultados cuantitativos de desempeño, confirmando la superioridad del modelo YOLOv11n-Seg-seg en precisión, consistencia morfológica y tolerancia a la variabilidad espectral de la iluminación en el entorno controlado del sistema PFAL

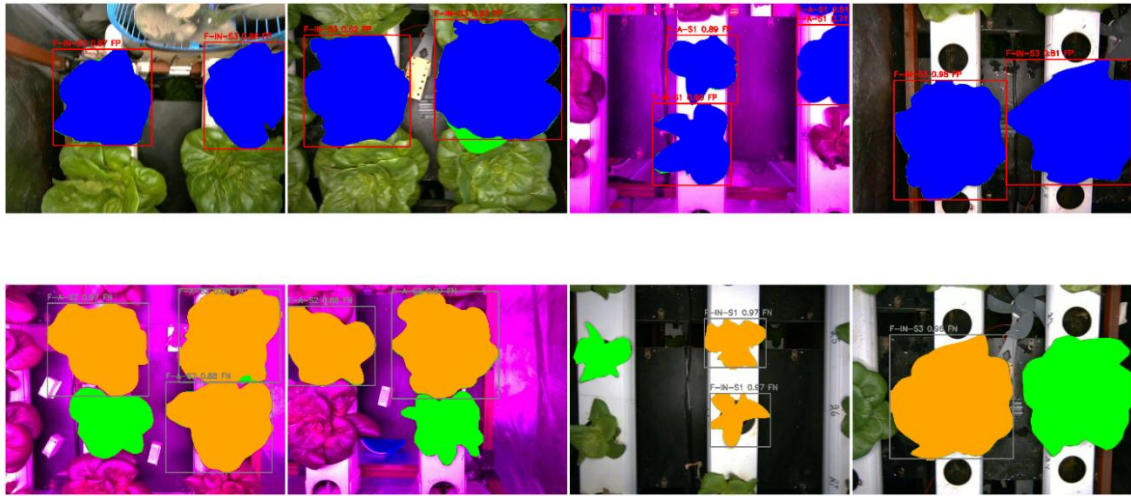


Figura 3.11. Ejemplos ilustrativos de falsos positivos (FP) y falsos negativos (FN) obtenidos en las inferencias de YOLOv8n-seg (azul) y YOLOv11n-seg (naranja).

3.4.5 Discusión

Los resultados obtenidos muestran que el desempeño de YOLOv8n-Seg en el entorno PFAL presenta patrones que coinciden con otros estudios que emplean arquitecturas basadas en YOLOv8 para análisis de cultivos. En particular, los hallazgos de (Wolter-Salas *et al.*, 2024), quienes implementaron WS-YOLO para detectar estrés hídrico en lechuga, describen comportamientos similares: modelos compactos derivados de YOLOv8 suelen alcanzar altos niveles de sensibilidad, pero a costa de una mayor propensión a detectar regiones ambiguas y generar sobre-segmentaciones en escenarios con iluminación compleja. De forma congruente, en este estudio YOLOv8n-Seg registró un mayor número de falsos positivos, especialmente bajo iluminación blanca tenue, así como máscaras más expansivas alrededor del borde foliar. Este patrón coincide con la interpretación de (Wolter-Salas *et al.*, 2024), quienes señalan que la estructura ligera de YOLOv8 privilegia la sensibilidad sobre la precisión espacial, lo cual también se refleja en el hecho de que, aunque el modelo mantuvo altos valores de F1-score, concentró una proporción mayor de detecciones erróneas respecto a YOLOv11n-Seg.

Estos resultados también concuerdan con lo reportado por (Y. Wang *et al.*, 2024), quienes utilizaron YOLOv8 como etapa de segmentación previa en un sistema híbrido con EfficientNet. Dichos autores observaron que YOLOv8 tiende a arrastrar ruido de fondo y confundir superficies blancas, brillantes o reflectantes con vegetación cuando el entorno carece de contraste cromático. Esto coincide claramente con las observaciones realizadas en imágenes del fotoperiodo inactivo, donde YOLOv8n-Seg llegó a incrementar hasta un 40 % sus falsos positivos y presentó dificultades en zonas con reflejos de la malla de cultivo. Tal como señalan (Y. Wang *et al.*, 2024), esta limitación puede requerir filtrados adicionales para estabilizar la segmentación; sin embargo, en este estudio se comprobó que el diseño más reciente de YOLOv11n-Seg permitió mitigar estos problemas sin necesidad de aplicar etapas posteriores de refinamiento. Por su parte, el comportamiento observado en YOLOv11n-Seg coincide estrechamente con lo reportado en investigaciones recientes que utilizan variantes de YOLOv11 para fenotipado vegetal. (Y. Zhang *et al.*, 2025), mediante su modelo YOLOv11-AreaNet, destacaron que esta familia de arquitecturas produce bordes más coherentes, menor varianza inter-época y una mejor capacidad para segmentar objetos pequeños u ocluidos. Los resultados obtenidos en este estudio reproducen precisamente estas tendencias: YOLOv11n-Seg mostró menor dispersión en mAP50 y mAP50–95, una mejor segmentación de plántulas en etapas tempranas y una reducción notable de falsos positivos incluso en condiciones de brillo homogéneo. Esta estabilidad se atribuye, según (Y. Zhang *et al.*, 2025) las mejoras en la fusión de características y a una mejor representación espacial, ventajas que también se reflejaron en la mayor consistencia semanal del F1-score, con fluctuaciones inferiores al 3 %. Las condiciones lumínicas del PFAL influyeron de manera clara en la precisión geométrica de ambos modelos. Bajo iluminación púrpura del fotoperiodo activo, donde existe mayor diferenciación espectral entre el follaje y el fondo, YOLOv8n-Seg y YOLOv11n-Seg alcanzaron sus mejores valores de Dice (0.964 y 0.966, respectivamente). Estos valores son comparables con los reportados en la

literatura para métodos de segmentación supervisada y auto supervisada, (X. Lin *et al.*, 2023) desarrollaron un modelo orientado a mitigar las distorsiones cromáticas producidas por iluminación artificial en cultivos de interior, alcanzando coeficientes Dice de 0.871 bajo luz amarilla y 0.839 bajo luz púrpura en imágenes de cannabis. Aunque existe una diferencia considerable entre especies y condiciones lumínicas, la comparación permite dimensionar el avance logrado por nuestro enfoque: los modelos basados en YOLO superaron dichos valores entre 7 y 12 puntos porcentuales, incluso bajo geometrías y condiciones espectrales similares. Esto indica que, en un entorno PFAL los modelos basados en YOLO pueden capturar relaciones espaciales más finas que los métodos auto supervisados, incluso sin corrección previa del color.

En el fotoperiodo inactivo iluminación blanca tenue y fondo reflectante los valores de Dice disminuyeron ligeramente para ambas redes, con variaciones de -0.02 a -0.04 respecto al fotoperiodo activo. Aun así, el desempeño permaneció muy por encima del reportado para SSCRf (Spatial Spectral Conditional Random Field), cuyo Dice cayó hasta 0.839 en escenas púrpura y mostró sensibilidad al ruido cromático en escenarios de luminancia homogénea. En contraste, YOLOv11n-Seg mantuvo una estabilidad intersemanal superior al 97 % de coincidencia espacial, demostrando mayor tolerancia al brillo uniforme y a las reflexiones difusas del PFAL.

3.5 Conclusiones

El análisis conjunto de las métricas cuantitativas y los resultados visuales demostró que la distribución de datos 70-20-10 fue la más adecuada para el entrenamiento y validación de los modelos, ya que proporcionó un equilibrio óptimo entre volumen de entrenamiento y capacidad de generalización. Con esta configuración, ambas arquitecturas YOLOv8n-Seg-seg y YOLOv11n-Seg-seg alcanzaron un desempeño sobresaliente en la detección y segmentación del dosel foliar de lechugas cultivadas en el sistema PFAL, logrando coincidencias

espaciales superiores al 96 % respecto a las máscaras reales y una estabilidad inter-época adecuada para su aplicación en entornos controlados.

No obstante, el modelo YOLOv11n-seg evidenció una mayor precisión, con valores promedio de precisión y F1-score de 0.95, superando a YOLOv8n-Seg-seg en la delimitación de bordes y en la reducción de sobre segmentaciones bajo variaciones de iluminación. La arquitectura de YOLOv11n-Seg arquitectura basada en mecanismos de atención espacial y convoluciones eficientes, permitió una mejor respuesta ante reflejos especulares y diferencias de espectro entre los fotoperiodos activo (luz púrpura) e inactivo (luz blanca).

Por otro lado, el modelo YOLOv8n-seg mostró un menor tiempo de inferencia y un procesamiento más ligero, lo que lo posiciona como una alternativa viable para aplicaciones en tiempo real, como el monitoreo continuo de cultivos mediante robots cartesianos o sistemas embebidos en cámaras de visión RGB-D. Aunque ligeramente menos preciso, su rapidez y eficiencia computacional lo hacen adecuado para sistemas de captura automatizados en entornos agrícolas inteligentes.

En conjunto, se concluye que YOLOv11n-Seg-seg es el modelo más preciso y estable para el fenotipado automatizado del dosel vegetal en sistemas PFAL, mientras que YOLOv8n-seg resulta la opción más práctica para implementaciones en tiempo real. La combinación de ambos enfoques abre la posibilidad de desarrollar sistemas híbridos que integren precisión y velocidad, permitiendo una estimación fiable y dinámica de variables morfológicas como el área foliar proyectada, el índice de área foliar (LAI) y la cobertura del dosel, contribuyendo al monitoreo no destructivo del crecimiento vegetal en agricultura de precisión

3.6 Literatura citada

- Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information (Switzerland)*, 11(2). <https://doi.org/10.3390/info11020125>
- Dwyer *et al.*, B. N. J. H. T. R. T. (2024). *Roboflow (Version 1.0)* (1.0). Roboflow. <https://roboflow.com/>
- Lee, J. G., Choi, C. S., Jang, Y. A., Jang, S. W., Lee, S. G., & Um, Y. C. (2013). Effects of air temperature and air flow rate control on the tipburn occurrence of leaf lettuce in a closed-type plant factory system. *Horticulture Environment and Biotechnology*, 54(4), 303–310. <https://doi.org/10.1007/s13580-013-0031-0>
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). *Microsoft COCO: Common Objects in Context* (pp. 740–755). https://doi.org/10.1007/978-3-319-10602-1_48
- Lin, X., Li, C. T., Adams, S., Kouzani, A. Z., Jiang, R., He, L., Hu, Y., Vernon, M., Doeven, E., Webb, L., McClellan, T., & Guskich, A. (2023). Self-Supervised Leaf Segmentation under Complex Lighting Conditions. *Pattern Recognition*, 135. <https://doi.org/10.1016/j.patcog.2022.109021>
- Lin, Z., Fu, R., Ren, G., Zhong, R., Ying, Y., & Lin, T. (2022). Automatic monitoring of lettuce fresh weight by multi-modal fusion based deep learning. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.980581>
- Moon, T., Choi, W. J., Jang, S. H., Choi, D. S., & Oh, M. M. (2022). Growth Analysis of Plant Factory-Grown Lettuce by Deep Neural Networks Based on Automated Feature Extraction. *Horticulturae*, 8(12). <https://doi.org/10.3390/horticulturae8121124>

- Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. In *BMC Research Notes* (Vol. 15, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13104-022-06096-y>
- Narang, S., Damos, G., Elsen, E., Micikevicius, P., Alben, J., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., & Wu, H. (2018). *MIXED PRECISION TRAINING*.
- Powers, D. M. W., & Ailab. (2011). *EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION*. 2(1), 37–63. <http://www.bioinfo.in/contents.php?id=51>
- Sapkota, R., & Karkee, M. (2025). *Comparing YOLOv11 and YOLOv8 for instance segmentation of occluded and non-occluded immature green fruits in complex orchard environment*. <http://arxiv.org/abs/2410.19869>
- Sapkota, R., Meng, Z., Churuvija, M., Du, X., Ma, Z., & Karkee, M. (2025). *Comprehensive Performance Evaluation of YOLOv12, YOLO11, YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environments*. <http://arxiv.org/abs/2407.12040>
- Sebastian Wolter-Salas, P. C. R. C.-V. M. C. O. R. V. S. and D. A. (2024). WS-YOLO: An Agronomical and Computer Vision-Based Framework to Detect Drought Stress in Lettuce Seedlings Using IR Imaging and YOLOv8. In T. Guarda, F. Portela, & J. M. Diaz-Nafria (Eds.), *Advanced Research in Technologies, Information, Innovation and Sustainability* (Vol. 1935). Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-48858-0>
- Shelhamer, E., Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>

- Tian, Z., Ma, W., Yang, Q., & Duan, F. (2022). Application status and challenges of machine vision in plant factory—A review. In *Information Processing in Agriculture* (Vol. 9, Issue 2, pp. 195–211). China Agricultural University. <https://doi.org/10.1016/j.inpa.2021.06.003>
- Ultralytics. (2024, August). *Ultralytics YOLO11*. Ultralytics YOLO11. <https://docs.ultralytics.com/es/models/yolo11/>
- Ultralytics. (2023, January 10). *Explore Ultralytics YOLOv8*. <https://docs.Ultralytics.Com>. <https://docs.ultralytics.com/es/models/yolov8/>
- Ultralytics. (2025, November 17). *Técnicas de preprocesamiento de datos anotados – Ultralytics Documentation*. Ultralytics.
- Wang, Y., Wu, M., & Shen, Y. (2024). Identifying the Growth Status of Hydroponic Lettuce Based on YOLO-EfficientNet. *Plants*, 13(3). <https://doi.org/10.3390/plants13030372>
- Xu, D., Chen, J., Li, B., & Ma, J. (2023). Improving Lettuce Fresh Weight Estimation Accuracy through RGB-D Fusion. *Agronomy*, 13(10). <https://doi.org/10.3390/agronomy13102617>
- Zhang, Y., Li, Y., Cao, X., Wang, Z., Chen, J., Li, Y., Zhong, Z., Bai, R., Yang, P., Pan, F., & Fu, X. (2025). Leaf area estimation in small-seeded broccoli using a lightweight instance segmentation framework based on improved YOLOv11-AreaNet. *Frontiers in Plant Science*, 16. <https://doi.org/10.3389/fpls.2025.1622713>
- Zhang, Y., Wu, M., Li, J., Yang, S., Zheng, L., Liu, X., & Wang, M. (2023). Automatic non-destructive multiple lettuce traits prediction based on DeepLabV3 +. *Journal of Food Measurement and Characterization*, 17(1), 636–652. <https://doi.org/10.1007/s11694-022-01660-3>

CAPÍTULO 4. PREDICCIÓN MULTIMODAL DE BIOMASA Y ÁREA FOLIAR DE LECHUGA MEDIANTE VISIÓN RGB-D Y CNN-ML

Luis Gerardo Chávez Covarrubias¹; Raquel Salazar Moreno^{1*}; Irineo López Cruz¹; Luis Arturo Soriano Avendaño¹

¹Posgrado de Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo. Texcoco, Estado de México.

*Correo: rsalazarm@chapingo.mx

4.1 RESUMEN

El fenotipado vegetal no destructivo basado en visión computacional es clave para cuantificar rasgos biofísicos en sistemas PFAL. Este estudio implementó un enfoque multimodal para predecir biomasa fresca, biomasa seca y área foliar de lechuga mediante integración de imágenes RGB-D, reconstrucción 3D y modelos de aprendizaje profundo y machine learning. Usando una cámara Intel RealSense D405 montada en un robot cartesiano, se capturaron imágenes estandarizadas durante los 4 muestreos y se extrajeron métricas morfológicas 2D–3D (diámetros, área convexa, volumen aparente, altura real, etc), las cuales mostraron correlaciones elevadas con los parámetros productivos ($r = 0.70\text{--}0.93$). Se evaluaron cinco modelos de regresión, destacando Random Forest y Gradient Boosting ($R^2 = 0.85\text{--}0.91$; $MAE < 6\text{ g}$; $RMSE < 11\text{ g}$). También se incorporaron descriptores visuales profundos de EfficientNet-B0, ResNet-50 y YOLOv8n-Segm, donde las combinaciones con Random Forest mantuvieron o mejoraron el desempeño, sobresaliendo ResNet-50 + RFR ($R^2 \approx 0.91$) y YOLOv8n-Segm + RFR como opción eficiente para aplicaciones en tiempo real. En conjunto, los resultados demuestran que la fusión de geometría 3D y características visuales profundas incrementa la capacidad predictiva, mostrando el potencial de sistemas multimodales y robóticos para fenotipado preciso y escalable.

Palabras clave: RGB-D, biomasa vegetal, aprendizaje profundo, Random Forest, multimodalidad.

4.2 INTRODUCCIÓN

El rápido crecimiento poblacional y la reducción de la superficie cultivable han impulsado el desarrollo de sistemas agrícolas de alta eficiencia, como las Fábricas de plantas (Plant Factory with Artificial Lighting PFALs), que permiten el cultivo controlado y continuo a lo largo del año. En este contexto, el monitoreo fenotípico no destructivo se ha convertido en una herramienta esencial para cuantificar rasgos morfológicos y fisiológicos de las plantas con alta resolución temporal y espacial (Petropoulou *et al.*, 2023b; Yu *et al.*, 2022). Las técnicas de imagen RGB-D han emergido como una alternativa viable al fenotipado debido a su bajo costo y capacidad de reconstruir información tridimensional del dosel. A partir de imágenes RGB y mapas de profundidad, es posible estimar métricas morfológicas como área proyectada, altura, volumen aparente y excentricidad, las cuales se correlacionan con biomasa fresca y seca (Lin *et al.*, 2022a; Q. Zhang *et al.*, 2022b). Diversos trabajos recientes han empleado cámaras Intel RealSense o sistemas estéreo para obtener nubes de puntos de alta densidad que representan la estructura 3D del dosel de cultivos hortícolas como la lechuga (*Lactuca sativa* L.), un modelo ampliamente utilizado por su arquitectura regular y su relevancia comercial (Buxbaum *et al.*, 2022; Lu *et al.*, 2025). Aun así, la mayoría de los estudios se centran en el análisis 2D o en entornos estáticos, limitando la escalabilidad de las mediciones. En contraste, la integración de sensores RGB-D con sistemas robóticos cartesianos o plataformas de alto rendimiento (High-Throughput Plant Phenotyping, HTPP) permite adquirir datos espaciales de manera precisa y repetible a lo largo del ciclo de crecimiento (Ghosal *et al.*, 2018; M. Wang *et al.*, 2019). (Petropoulou *et al.*, 2023b) demostró que la automatización del monitoreo de lechugas en invernaderos inteligentes mejora la eficiencia y reduce errores humanos en la estimación de volumen y biomasa. De forma complementaria, (Yu *et al.*, 2022) combinaron imágenes hiperespectrales con redes neuronales convolucionales (2D-CNN) para predecir rasgos fenotípicos, logrando coeficientes de determinación superiores a 0.90 en la estimación de sólidos solubles y pH, evidenciando el potencial del aprendizaje

profundo en la extracción de características biofísicas sin necesidad de preprocesamiento manual. Por su parte, (Ghosal *et al.*, 2018) propusieron un marco de aprendizaje profundo (Explainable Deep Machine Vision Framework) para la identificación y cuantificación de estrés en soja, logrando clasificar múltiples tipos de estrés con alta precisión a partir de más de 25,000 imágenes foliares. Este enfoque introdujo la noción de modelos interpretables en fenotipado, al permitir visualizar las regiones de la hoja que contribuían a la decisión del modelo. Estos avances abren la puerta a arquitecturas multimodales que integran señales RGB, profundidad y métricas derivadas del dosel 3D, como la red TMSCNet desarrollada por (Q. Zhang *et al.*, 2022), capaz de fusionar información visual y estructural para predecir área foliar y biomasa de lechugas en entornos controlados. A pesar de estos progresos, persisten vacíos significativos. La mayoría de las investigaciones dependen de datasets estáticos, sin considerar la variabilidad temporal del crecimiento ni la integración directa con sistemas de adquisición automatizados. Asimismo, los modelos multimodales existentes se centran en imágenes hiperespectrales o en datos 2D, sin explotar completamente el potencial del análisis geométrico 3D derivado de nubes de puntos. En este trabajo se propone un enfoque de fenotipado RGB-D aplicado a lechugas cultivadas en un sistema PFALs utilizando una cámara Intel RealSense D405 para la captura de imágenes de alta resolución y la generación de nubes de puntos tridimensionales. El objetivo de este trabajo fue comparar el desempeño de cinco modelos de aprendizaje supervisado: Regresión Lineal Múltiple (MLR), Máquinas de Vectores de Soporte (SVR), Bosque Aleatorio (RFR), Gradient Boosting (GB) y Perceptrón Multicapa (MLP), para la estimación del área foliar, peso fresco y peso seco de lechuga. Posteriormente, se integraron características visuales profundas extraídas con los backbones EfficientNet-B0, ResNet-50 y YOLOv8n-Seg, combinadas con las métricas morfológicas dentro del esquema multimodal CNN + ML. El uso de backbones es importante porque permiten codificar información visual esencial como color, contornos y organización espacial del dosel que no se obtiene únicamente de las métricas

geométricas. EfficientNet-B0 se seleccionó por su eficiencia computacional, ResNet-50 por su estabilidad ante variaciones de iluminación, y YOLOv8n-Seg por ser un backbone ligero que extrae de manera rápida patrones visuales relevantes para la estructura foliar, este enfoque permitió evaluar el efecto de la fusión de características geométricas y visuales en el rendimiento predictivo de los modelos, demostrando la efectividad del aprendizaje profundo combinado con visión tridimensional para el fenotipado no destructivo y de bajo costo en entornos controlados.

4.3 Materiales y métodos

4.3.1 Sitio experimental y manejo del cultivo

El experimento se llevó a cabo en la Universidad Autónoma Chapingo, en los edificios adjuntos a Centros Regionales, ubicados en las coordenadas 19.491515105571892, -98.88530922208855. Se realizó un experimento durante el mes de mayo de 2025, con el objetivo de evaluar el crecimiento vegetativo de lechuga variedad francesa durante un periodo de 28 días. Se efectuaron muestreos semanales y se obtuvieron datos de manera destructiva. El experimento inició el 2 de mayo de 2025, fecha en la que las plántulas fueron trasplantadas al módulo PFALs bajo un sistema de hidroponía tipo NFT (Nutrient Film Technique), que garantiza la continuidad del crecimiento mediante un flujo constante de solución nutritiva. El módulo PFAL operó con una velocidad de aire equivalente al 75 % de la potencia del ventilador, manteniendo la radiación fotosintéticamente activa (PAR) entre 190 y 230 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. La solución nutritiva se mantuvo con una conductividad eléctrica (CE) entre 2.0 y 2.6 $\text{dS}\cdot\text{m}^{-1}$ y un pH de 6.0 a 7.0, circulando a una velocidad de flujo de 1.5 $\text{m}\cdot\text{s}^{-1}$. Cada nivel de estantería alojaba 27 plantas de lechuga, distribuidas uniformemente. Los muestreos destructivos se realizaron los días 9, 16, 23 y 30 de mayo, correspondientes a las cuatro semanas del ciclo de desarrollo vegetativo. En cada muestreo se recolectaron plantas completas para la obtención de las variables peso fresco, área foliar y peso seco, obteniendo 132 muestras.

Antes de realizar los muestreos destructivos, se efectuó un muestreo fotográfico utilizando un robot cartesiano diseñado para posicionarse exactamente en el centro del nivel del módulo PFALs, con el fin de asegurar que todas las fotografías se capturaran bajo las mismas condiciones espaciales y de iluminación. El robot cartesiano cuenta con una cámara de profundidad Intel RealSense D405 como efector final, mediante la cual se capturaron imágenes RGB y datos de profundidad de cada planta. Estas imágenes fueron almacenadas utilizando una nomenclatura estandarizada que identifica el módulo, la fecha y la hora de captura, siguiendo el formato `rgb_M {número de módulo}_AAAA-MM-DD_HH-MM-SS.png` para las imágenes RGB y `depth_M{número de módulo}_AAAA-MM-DD_HH-MM-SS.png` para los mapas de profundidad. Por ejemplo, los archivos se guardaron con nombres como `rgb_M1_2025-05-07_10-45-30.png` y `depth_M1_2025-05-07_10-45-30.png`. Todos los archivos generados fueron almacenados en la memoria interna de la Raspberry Pi 4, la cual controla el funcionamiento del robot cartesiano y gestiona la sincronización automática entre las capturas RGB-D y los registros de muestreo.

Durante los muestreos destructivos del experimento, las plantas recolectadas se pesaron inmediatamente para determinar el peso fresco, utilizando una balanza analítica OHAUS Scout SPX223 (capacidad 220 g, precisión 0.001 g). Posteriormente, las hojas se procesaron en un integrador de área foliar LICOR modelo LI-3100C para determinar el área foliar total. Finalmente, las muestras se colocaron en un horno de secado FELISA modelo FE-291D, con control digital de temperatura y rango operativo de ambiente + 5 °C a 250 °C, donde se mantuvieron a 75 °C durante 48 h hasta alcanzar peso constante, con el fin de determinar el peso seco de cada muestra (Figura 4.1).

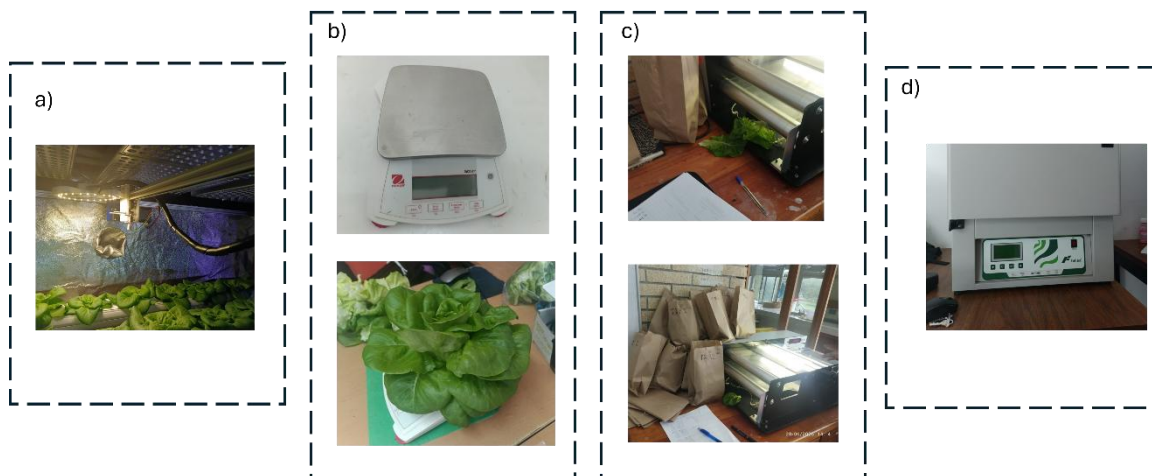


Figura 4.1. Flujo experimental para la adquisición de imágenes RGB-D y la determinación de variables morfofisiológicas en *Lactuca sativa* L. variedad francesa bajo condiciones controladas en un sistema PFAL.

4.3.2 Adquisición y procesamiento de datos RGB-D

El sistema de visión empleado estuvo compuesto por una cámara Intel RealSense D405, la cual utiliza un par estéreo calibrado y un proyector de luz estructurada para estimar la distancia de cada píxel con alta precisión. Su rango operativo va de 7 a 50 cm, aunque en este estudio se empleó únicamente dentro del intervalo de 7–36 cm, que correspondía a la distancia de interés debido a que esta distancia es el espacio que existe entre la cámara y la base del nivel del módulo PFAL (Figura 4.2, apartado III, sección a). La cámara opera con una resolución nativa de 640×480 píxeles y una frecuencia de captura de 30 fps. El dispositivo fue montado en el efector final del robot cartesiano RGB-D, asegurando que cada adquisición se realizara desde una posición fija y perpendicular al plano de cultivo. Previo a su integración en el sistema, la cámara fue calibrada geométricamente mediante el modelo de proyección pinhole, lo cual permitió obtener sus parámetros intrínsecos, coeficientes de distorsión y parámetros extrínsecos, asegurando la alineación precisa entre los sensores de color y profundidad.

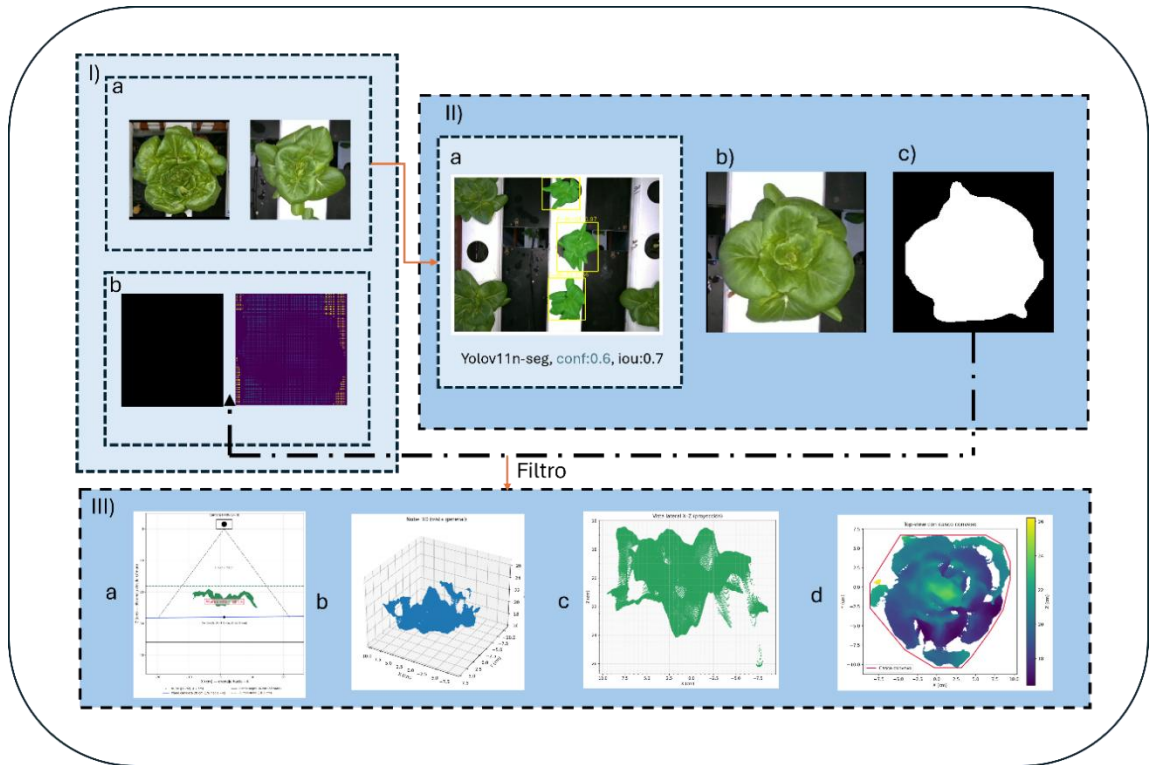


Figura 4.2. Flujo de procesamiento para la extracción de características morfológicas RGB-D.

Parámetros intrínsecos

Los parámetros intrínsecos describen las propiedades ópticas internas de la cámara y definen cómo las coordenadas métricas tridimensionales de un punto (X, Y, Z) en el espacio son proyectadas en el plano de imagen (u, v) . Esta relación se expresa mediante la matriz de calibración K :

$$K = \begin{bmatrix} fx & 0 & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 431.48 & 0 & 323.42 \\ 0 & 430.92 & 241.79 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

En esta matriz

- $fx = 431.48 \text{ px}$, $fy = 430.92 \text{ px}$ representan las distancias focales horizontales y verticales, respectivamente, que escalan la proyección de la escena.

- $cx = 431.48$ px , $cy = 430.92$ corresponden al punto principal, es decir, el centro óptico del sensor.

Los coeficientes de distorsión del lente fueron:

$$k1 = -0.0518, k2 = 0.0608, k3 = -0.0201, p1 = -0.0004, p2 = 0.0014$$

empleados para corregir distorsión radial y tangencial, eliminando curvaturas en los bordes de la imagen y garantizando una proyección métrica precisa.

Parámetros extrínsecos

Los parámetros extrínsecos describen la relación geométrica entre los sensores de profundidad y color, representada por una matriz de rotación R y un vector de traslación T . Estos se emplean en la ecuación:

$$P_{color} = R \times P_{depth} + T \quad (2)$$

$$\text{Donde } R = \begin{bmatrix} 0.999996 & -0.002494 & -0.001358 \\ 0.002495 & 0.999997 & 0.000858 \\ 0.001356 & -0.000862 & 0.999999 \end{bmatrix}, T = \begin{bmatrix} 5.89 \times 10^{-6} \\ -5.67 \times 10^{-6} \\ 7.29 \times 10^{-5} \end{bmatrix}$$

Estos parámetros permitieron transformar y alinear milimétricamente los puntos del sensor de profundidad hacia el sistema de coordenadas del sensor RGB, generando una nube de puntos 3D a color (x, y, z, R, G, B) .

4.3.3 Flujo de procesamiento RGB-D

El procesamiento de las imágenes (Figura 4.2) se estructuró en tres fases: (i) alineación y limpieza de datos, donde se transformaron las coordenadas de profundidad al marco RGB aplicando los parámetros R y T ; (ii) detección y segmentación de la planta central mediante el modelo YOLOv11n-seg (conf. = 0.6, IoU = 0.7), generando una máscara binaria precisa; y (iii) reconstrucción y análisis tridimensional, donde se filtró el rango $Z \in [0.07, 0.36]$ m, se aplicó

DBSCAN ($\epsilon = 0.5$ cm, $\text{min_samples} = 20$) para eliminar ruido, y se calculó el casco convexo simplificado por el método Douglas–Peucker. A partir de la nube de puntos resultante se extrajeron métricas morfológicas tridimensionales: diámetro mayor (D_{mayor}), diámetro menor (D_{menor}), área convexa (A_c), volumen aparente (V_a), altura real (H_r), circularidad (C), relación H/D y coeficiente de variación vertical (CV_z). Finalmente, se generó un dataset RGB-D normalizado (320×320 px) que incluyó las imágenes RGB, máscaras binarias, mapas de profundidad y un archivo meta.json con la información métrica asociada, se muestra el algoritmo 1 donde se ve el flujo de cada actividad realizada.

4.3.4 Cálculo del centroide y determinación de la altura precisa

En la fase de reconstrucción tridimensional (Figura 26 IIIa), se estimó el centroide del casco convexo de la planta a partir de la distribución de los puntos en el plano X–Y, calculado como:

$$C_x = \frac{1}{n} \sum_{i=1}^n x_i, C_y = \frac{1}{n} \sum_{i=1}^n y_i \quad (3)$$

Donde:

- x_i y y_i son las coordenadas en el plano horizontal de cada punto (X–Y) de cada punto i perteneciente a la nube de puntos 3D de la planta.
- n es el número total de puntos considerados dentro del casco convexo (la superficie externa que envuelve el contorno de la planta)

El centroide representa la proyección media del volumen foliar y constituye el referente geométrico para determinar la altura real h_d del dosel. A partir de esta coordenada se proyectó una línea perpendicular al plano inclinado de la canaleta, cuya pendiente fue del 1 % en dirección negativa del eje X, correspondiente al flujo gravitacional de la solución nutritiva. De esta forma, el punto de intersección entre el centroide y el plano de referencia

(a 28 cm de altura desde la base de la cámara) permitió ajustar la altura estimada de la planta corrigiendo el sesgo geométrico introducido por la inclinación del sustrato (Figura 4.3 se muestra la Figura 4.2-III-(a, en una escala más grande). Esta corrección incrementó la precisión del cálculo de la altura en comparación con una proyección ortogonal plana. Para fines prácticos de este experimento la planta se colocó en la parte central de la canaleta esto para tener las condiciones de iluminación adecuadas para todas las 132 muestras se encuentran en la misma posición sobre la canaleta inclinada por ello calculamos las alturas restando 28- $z_{\min}$, pero generalizamos para cuando sea en real time las imágenes deben ser tomadas bajo esta condición de inclinación.

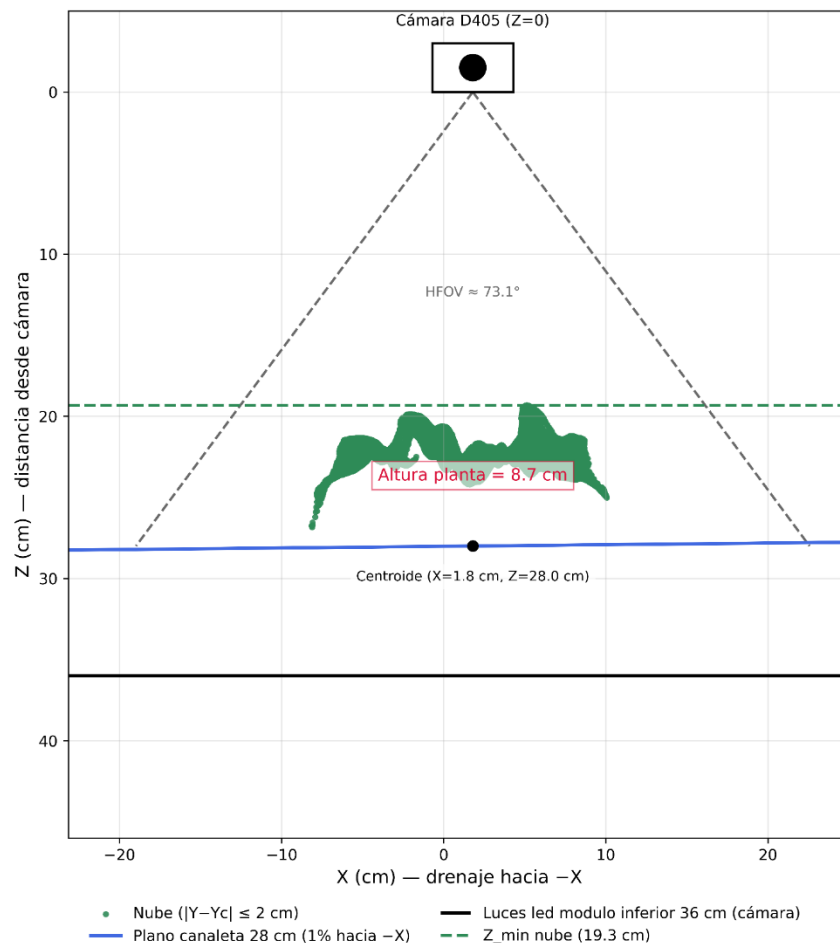


Figura 4.3. Corte transversal X-Z generado a partir de la nube de puntos RGB-D de la cámara Intel RealSense D405.

4.3.5 Características morfológicas extraídas de las imágenes RGB-D

A partir de la segmentación RGB-D y reconstrucción tridimensional, se extrajeron diez características morfológicas que describen la geometría del dosel de la lechuga en términos de expansión horizontal, compacidad y estructura vertical (Cuadro 4.1). Estas métricas, ampliamente reportadas en estudios de fenotipado de lechuga mediante cámaras RGB-D (Bloch *et al.*, 2025; Islam *et al.*, 2024; Z. Lin *et al.*, 2022c; Paulus, 2019; X. Zhang *et al.*, 2025a; Y. Zhang *et al.*, 2025b), permiten relacionar la forma y volumen de la planta con parámetros fisiológicos como el peso fresco (PF), peso seco (PS) y área foliar (AF) .

Cuadro 4.1. Definición de las métricas morfológicas bidimensionales y tridimensionales utilizadas para el análisis del dosel de lechuga.

Métrica – Abrev.	Ecuación – Unidad – Descripción funcional
Dmax — D_{mayor}	D_{mayor} — cm — Longitud máxima del casco convexo; representa la expansión lateral principal.
Dmin — D_{menor}	D_{menor} — cm — Longitud mínima ortogonal a Dmax; refleja la simetría de la roseta.
AR — Relación de aspecto	$AR = D_{\text{mayor}}/D_{\text{menor}}$ — relación — Proporción entre ejes; valores cercanos a 1 indican forma circular.
CIRC — Circularidad	$C = 4\pi A_c/P_c^2$ — sin unidad — Mide la compacidad del contorno; ≈ 1 indica plantas redondeadas.

Aconv — A_c	A_c — cm ² — Área proyectada del casco convexo; indica cobertura foliar horizontal.
Pconv — P_c	P_c — cm — Longitud del borde convexo; describe la irregularidad del contorno.
Vol — V_a	$V_a = 0.5 \times A_c \times H_r$ — cm ³ — Estima el volumen aparente y la densidad espacial del follaje.
H/D — Relación altura–diámetro	$H/D = H_r / (0.5 \times (D_{mayor} + D_{menor}))$ — sin unidad — Describe hábito vertical vs. horizontal.
CVh — CV_h	$CV_h = 100 \times (\sigma_z / \mu_z)$ — % — Variabilidad vertical del dosel; indica rugosidad.
Ziqr — IQR_z	$IQR_z = Q75(Z) - Q25(Z)$ — cm — Compactación del follaje según el rango intercuartílico.
Hreal — H_r	$H_r = Z28 - Z_{min}$ — cm — Altura real respecto al plano de referencia.
Ha — Altura media	Ha — cm — Altura media del dosel obtenida del modelo de elevación.
Wa — Ancho medio	Wa — cm — Ancho promedio del follaje en vista cenital.

IDF_B / IDF_M / IDF_A — Índices derivados	Funciones RGB-D — sin unidad — Describen textura superficial, variaciones de reflectancia y rugosidad.
--	---

4.3.6 Modelos de predicción:

Para la estimación de los rasgos fenotípicos (peso fresco, área foliar y peso seco), se entrenaron y evaluaron cinco modelos supervisados de regresión: Regresión Lineal Múltiple (MLR), Máquinas de Vectores de Soporte (SVR) (Cortes *et al.*, 1995) , Bosques Aleatorios (RFR) (Breiman, 2001), Gradient Boosting (GB) (Friedman, 2001) y Perceptrón Multicapa (MLP) (Rosenblatt, 1958). El modelo MLR asume relaciones lineales entre las variables predictoras y la variable objetivo, proporcionando un punto de referencia base para los modelos más complejos. El SVR (Support Vector Regression) utiliza una función de pérdida basada en el margen ε y un núcleo radial (RBF) para modelar relaciones no lineales mediante proyecciones en espacios de mayor dimensión. El modelo RFR (Random Forest Regressor) (Breiman, 2001) combina múltiples árboles de decisión entrenados con subconjuntos aleatorios de datos y variables, lo que mejora la robustez y reduce el sobreajuste. Por su parte, el GB (Gradient Boosting) (Friedman, 2001) optimiza secuencialmente un conjunto de árboles débiles mediante descenso del gradiente, corrigiendo los errores residuales de las iteraciones previas y aumentando la precisión global. Finalmente, el MLP (Multilayer Perceptron) (Rosenblatt, 1958) es una red neuronal densa con funciones de activación no lineales (ReLU) y entrenamiento mediante retro-propagación, capaz de capturar interacciones complejas entre las variables.

4.3.7 Modelos híbridos CNN + ML

Para fortalecer la capacidad predictiva, los modelos anteriores también se entrenaron con características visuales extraídas mediante *backbones* de redes convolucionales profundas (CNN): EfficientNet-B0, ResNet-50 y YOLOv8n-Segn.

Estas redes se utilizaron como extractores de características (sin capas de clasificación) para obtener descriptores visuales de alto nivel que representaran la textura, color, geometría y estructura foliar de las plantas.

- EfficientNet-B0: emplea un escalado compuesto que optimiza la relación entre precisión y eficiencia computacional, ideal para sistemas embebidos (Tan & Le, 2019).
- ResNet-50: incorpora conexiones residuales (*skip connections*) que permiten entrenar redes profundas evitando la degradación del gradiente (He *et al.*, 2016).
- YOLOv8n-Segm: diseñado para detección y segmentación en tiempo real, combina precisión aceptable con bajo consumo de recursos, siendo óptimo para fenotipado automatizado (Ultralytics, 2023a).

Las salidas de estas CNN se concatenaron con las métricas morfológicas y de profundidad (2D–3D), formando vectores multimodales que alimentaron los modelos MLR, SVR, RFR, GB y MLP. Esta integración permitió evaluar la sinergia entre información geométrica (RGB-D) y características visuales profundas, analizando su impacto sobre la precisión predictiva.

4.3.8 Métricas de desempeño

El rendimiento de los modelos se evaluó mediante tres métricas estándar de regresión: el coeficiente de determinación (R^2), el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE). Estas métricas cuantifican la precisión y estabilidad del modelo, y fueron calculadas bajo un esquema de validación cruzada (K-Fold = 5).

Coeficiente de determinación (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

Evalúa la proporción de la variabilidad explicada por el modelo; valores cercanos a 1 indican alto poder predictivo.

Error absoluto medio (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

Indica el error promedio absoluto entre los valores observados y los estimados (Chai & Draxler, 2014).

Raíz del error cuadrático medio (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

Mide la magnitud promedio del error, penalizando los errores grandes (Chai & Draxler, 2014).

4.4 Resultados y discusión

4.4.1 Extracción de información 2D Y 3D:

Se calcularon los estadísticos descriptivos de las 16 métricas morfológicas 2D–3D obtenidas a partir de las imágenes RGB-D. Los valores presentaron una amplia variabilidad, reflejando diferencias marcadas en la arquitectura del dosel entre planta y reflejando el crecimiento semanal, entre las variables extraídas que mejor representan el aumento de biomasa y área foliar son Diámetro menor, diámetro mayor, área convexa proyectada, volumen aparente y la altura, en el Cuadro 4.2 se muestra el comportamiento de cada una de estas variables durante la etapa de desarrollo vegetativo de la planta de lechuga.

El diámetro mayor y el diámetro menor presentan coeficientes de variación de 29.27 % y 31.01 %, lo que indica una variabilidad moderada en la expansión horizontal del dosel vegetal, estrechamente relacionada con el área proyectada y, por tanto, con el desarrollo foliar.

El área convexa, con una media de 116.86 cm² y un rango que va de 23.18 a 313.79 cm², mostró un coeficiente de variación del 59.07 %, reflejando diferencias notables en la cobertura superficial de las plantas. Este parámetro es un indicador directo del área foliar estimada, ya que representa la superficie ocupada por la planta vista desde arriba.

El volumen aparente registró la mayor variabilidad (CV = 101.36 %), con valores entre 30.17 y 1948.69 cm³, evidenciando grandes diferencias en el desarrollo tridimensional y la acumulación de biomasa aérea entre individuos. Por lo tanto, este parámetro se asocia de forma directa con el crecimiento volumétrico y el peso fresco.

Cuadro 4.2. Análisis estadístico de las variables descriptoras 2D-3D.

Variable	Media	Mediana	Desv_Est	Mínimo	Máximo	Rango	CV (%)
D mayor (cm)	12.161	11.35	3.56	5.8	20.9	15.1	29.272
D menor (cm)	10.962	10.681	3.4	4.6	19.684	15.084	31.014
area convexa (cm²)	116.863	98.642	69.035	23.175	313.79	290.615	59.073

volumen aparente (cm³)	425.73 8	222.253	431.534	30.171	1948.69 1	1918.52 1	101.36 1
altura real (cm)	5.82	4.829	2.881	1.795	16.944	15.149	49.512

4.4.2 Análisis de correlación:

La Figura 4.4 muestra la correlación entre las variables morfológicas derivadas de la nube de puntos y las variables productivas de las plantas de lechuga. Se observa que el peso fresco (Peso), el peso seco (Psec) y el área foliar (Af) presentan correlaciones altas ($r = 0.82-0.93$) con las métricas estructurales del dosel, particularmente con el diámetro mayor (Dmax), diámetro menor (Dmin), área convexa (Aconv), perímetro convexo (Pconv) y volumen aparente (Vol). Esto evidencia que el tamaño, volumen y expansión del dosel son factores determinantes del aumento de biomasa y área foliar.

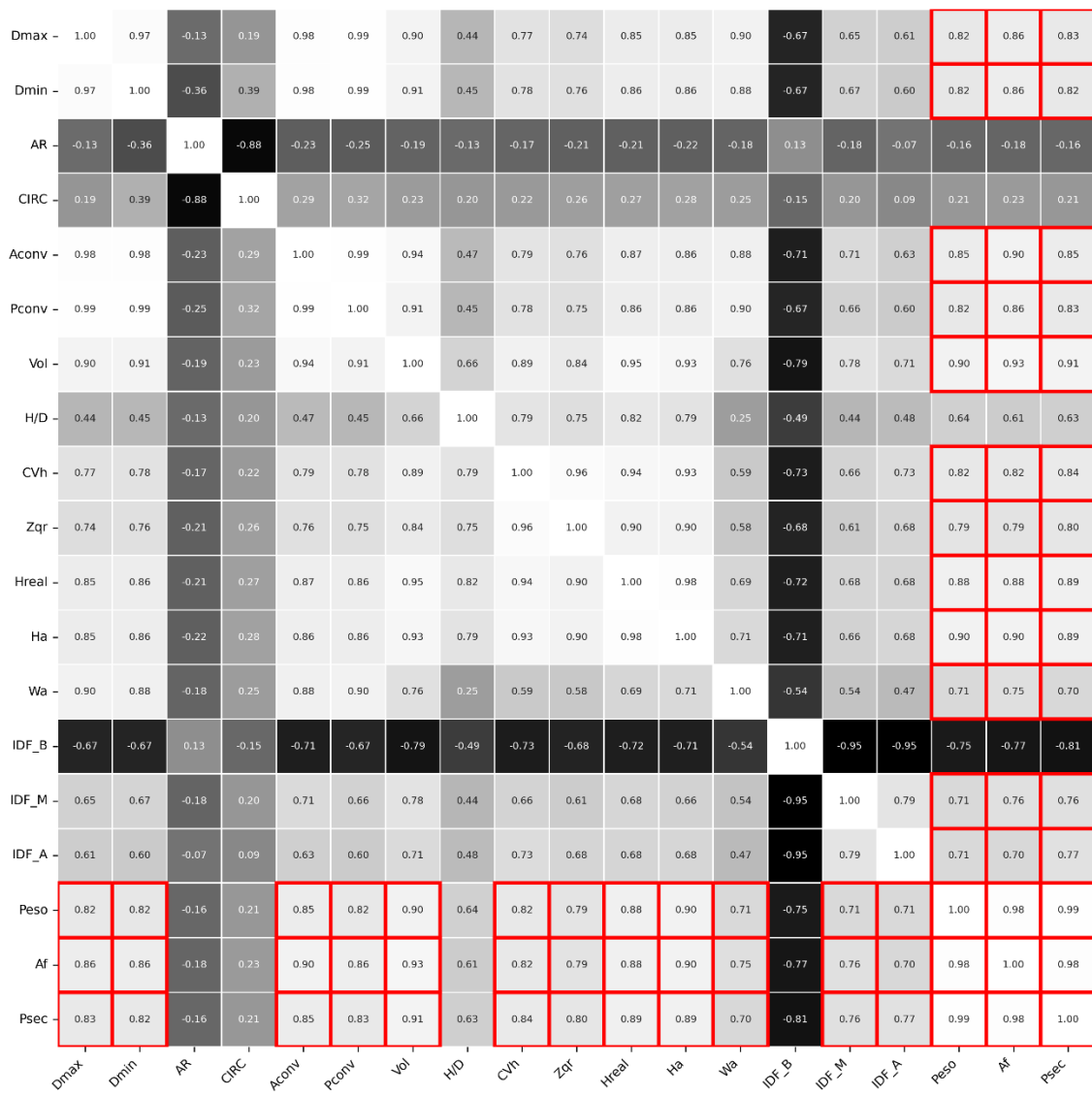


Figura 4.4. Matriz de correlación entre las variables morfológicas 2D–3D derivadas de la nube de puntos y las variables productivas de lechuga cultivada en PFAL.

De igual forma, la altura real (Hreal) y las dimensiones proyectadas del dosel (Ha y Wa) mantienen correlaciones positivas ($r = 0.70\text{--}0.89$) con las variables productivas, lo que sugiere que el crecimiento vertical complementa la expansión lateral en el desarrollo de la planta. Por el contrario, los índices de forma AR (aspect ratio) y CIRC (circularidad) muestran correlaciones bajas o negativas ($r < 0.3$), indicando que describen características de forma o simetría, pero no reflejan

directamente el incremento de biomasa o área foliar. Las variables seleccionadas para el entrenamiento del modelo CNN+ML Dmax, Dmin, Aconv, Pconv, Vol, Hreal, Ha, Wa, IDF_M, IDF_A, CVh y Ziqr, mostraron correlaciones positivas con las variables objetivo área foliar, peso fresco, peso seco (Af, Peso y Psec), con valores de r entre 0.70 y 0.93. Estas métricas describen la expansión foliar, el volumen aparente y la variación vertical del dosel, y representan los predictores más robustos del crecimiento y acumulación de biomasa en lechugas bajo condiciones PFAL.

4.4.3 Rendimiento de los modelos ML basados en métricas RGB-D:

Los modelos de aprendizaje automático fueron entrenados utilizando las 12 variables morfológicas obtenidas a partir del procesamiento RGB-D, las cuales describen características estructurales como diámetros, área y perímetro convexo, volumen aparente y altura real de las plantas. Estas variables se escalaron mediante *StandardScaler* y se aplicó validación cruzada de 5 particiones (K -Fold = 5) para evaluar la estabilidad de los modelos. Se entrenaron 5 modelos de regresión supervisada: Regresión Lineal (MLR), Máquinas de Vectores de Soporte (SVR-RBF), Bosques Aleatorios (RFR), Gradient Boosting (GB), y una red neuronal multicapa optimizada con función *ReLU* y algoritmo *Adam*.

Los resultados del desempeño promedio (R^2 , MAE y RMSE) se presentan en el Cuadro 4.3. Los modelos de regresión supervisada mostraron un desempeño consistente al predecir los fenotipos derivados del procesamiento RGB-D. El coeficiente de determinación validado (R^2) evidenció que los algoritmos basados en ensambles Random Forest (RFR) y Gradient Boosting (GB)— alcanzaron los mejores resultados generales en las tres variables analizadas (peso fresco, área foliar y peso seco). Para el peso fresco, el modelo RFR presentó el mayor ajuste con un $R^2 = 0.85$, un MAE de 5.64 g y un RMSE de 10.26 g, seguido de SVR (RBF) con $R^2 = 0.80$. Estos valores indican una adecuada capacidad de

predicción y una baja dispersión del error entre los valores medidos y estimados, lo que demuestra que los patrones de crecimiento pueden ser modelados con precisión a partir de las variables morfológicas RGB-D.

En la predicción del área foliar, se observó el mejor desempeño global, alcanzando un $R^2 = 0.91$ con el modelo Random Forest, acompañado de un MAE de 91.3 cm² y un RMSE de 156.3 cm², lo que refleja una alta correlación entre los datos observados y los estimados. El modelo Gradient Boosting también mostró resultados competitivos $R^2 = 0.87$, aunque presenta errores ligeramente superiores.

Para el peso seco, las diferencias entre modelos fueron menores, pero el modelo Random Forest nuevamente destacó con un $R^2 = 0.85$, MAE de 0.34 g y RMSE de 0.61 g, seguido de la Regresión Lineal (MLR) con $R^2 = 0.83$. Esto sugiere que las relaciones lineales entre las variables morfológicas y la biomasa seca son bien captadas por modelos tanto lineales como no lineales, mientras que la red neuronal mostró un desempeño moderado con un valor de $R^2 = 0.80 - 0.81$.

En conjunto, los resultados evidencian que los modelos de ensamble basados en árboles, en particular Random Forest, fueron los más robustos y generalizables para el conjunto de datos analizado, con un equilibrio adecuado entre precisión y estabilidad (bajo MAE y RMSE). Este comportamiento se atribuye a la capacidad de estos modelos para manejar relaciones no lineales y reducir el sobreajuste mediante promediado de múltiples árboles de decisión.

Cuadro 4.3. Desempeño de los modelos de aprendizaje supervisado en la predicción de los fenotipos de lechuga.

Variable objetivo	Modelo	R^2	MAE	RMSE
-------------------	--------	-------	-----	------

Peso fresco (g)	SVR (RBF)	0.805	5.631	11.536
	Random Forest (RFR)	0.846	5.646	10.259
	Gradient Boosting (GB)	0.788	6.073	12.015
Área foliar (cm ²)	SVR (RBF)	0.760	131.577	256.121
	Random Forest (RFR)	0.911	91.309	156.254
	Gradient Boosting (GB)	0.868	109.038	189.974
Peso seco (g)	SVR (RBF)	0.737	0.433	0.813
	Random Forest (RFR)	0.853	0.338	0.609
	Gradient Boosting (GB)	0.828	0.353	0.657

Las gráficas de dispersión medido vs. predicho (Figura 4.5) muestran visualmente el ajuste obtenido por el modelo Random Forest (RFR) para cada variable fenotípica. En las tres figuras, los puntos se distribuyen próximos a la línea diagonal $y=x$, lo cual indica una alta concordancia entre los valores observados y los estimados por el modelo. En la gráfica correspondiente al peso fresco Figura 4.5-1, se observa que la mayoría de los puntos se concentran de

manera uniforme alrededor de la diagonal, lo que sugiere una predicción estable en todo el rango de valores medidos. Las ligeras dispersiones hacia los extremos indican que el modelo mantiene su precisión incluso para plantas con pesos mayores, sin presentar sesgos sistemáticos de sobrestimación o subestimación. En el caso del área foliar Figura 4.5-2, la nube de puntos sigue una trayectoria muy cercana a la línea de referencia, con una distribución compacta y sin tendencias asimétricas. Esto demuestra que el modelo logra capturar adecuadamente las variaciones estructurales del dosel foliar, tanto en plantas jóvenes como desarrolladas, reproduciendo de manera fiel la amplitud de las áreas medidas a partir de las métricas RGB-D. Finalmente, en la gráfica del peso seco Figura 4.5-3, la dispersión es mínima, y la mayoría de los puntos se ubican dentro de un margen estrecho alrededor de la diagonal, lo que confirma una excelente capacidad del modelo para estimar con precisión pequeñas variaciones en la biomasa acumulada. Este comportamiento indica una alta sensibilidad del modelo a cambios sutiles en los rasgos morfológicos que influyen en el contenido de materia seca.

En conjunto, las tres representaciones gráficas evidencian que el modelo Random Forest logra predicciones consistentes, precisas y sin sesgos, demostrando un excelente desempeño general para estimar el crecimiento y desarrollo de las plantas de lechuga a partir de información fenotípica obtenida por visión RGB-D.

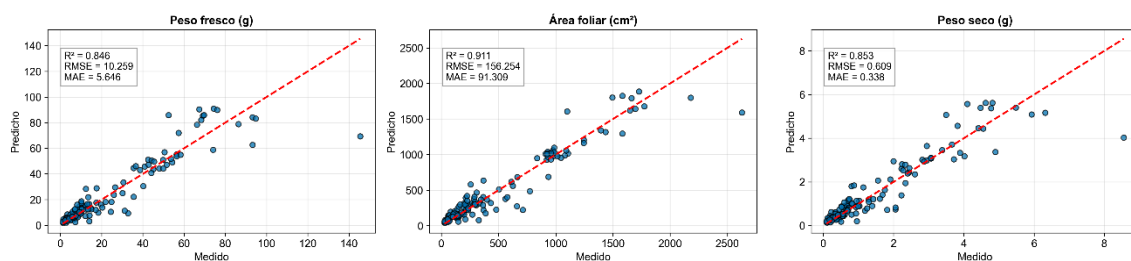


Figura 4.5.Desempeño del modelo Random Forest (RFR) en la predicción de fenotipos de lechuga mediante validación cruzada 5-Fold”.

4.4.4 Resultados de predicción multimodal CNN + ML

Al incorporar las características visuales extraídas por las CNN (*EfficientNet-B0*, *ResNet-50* y *YOLOv8n-Segm*), el desempeño de los modelos no mejoró de forma generalizada. En particular, los modelos lineales (MLR) y de vectores de soporte (SVR) empeoraron significativamente, con reducciones notables en el coeficiente de determinación. En algunos casos, incluso se obtuvieron valores negativos (por ejemplo, SVR = -0.15 con *EfficientNet* en la predicción de área foliar), en el Cuadro 4.4 se puede visualizar lo anterior.

Esto se debe principalmente a que las representaciones generadas por las CNN son de alta dimensionalidad y naturaleza no lineal, lo que incrementa la varianza y complejidad del espacio de atributos. Los modelos lineales y con pocos parámetros de regularización no son capaces de capturar las interacciones complejas entre las características, lo que conduce a sobreajuste o ruido y afecta la capacidad de generalización. Por otro lado, los algoritmos de ensamble como Random Forest (RFR) y Gradient Boosting (GB) mantuvieron o incluso mejoraron su rendimiento al incorporar las características extraídas por las CNN. Estos modelos son más robustos frente al ruido y la redundancia de variables, ya que dividen el espacio de decisión en múltiples subconjuntos y combinan los resultados de árboles individuales, lo que reduce la varianza del modelo.

En términos comparativos, las combinaciones ResNet-50 + RFR ($R^2 = 0.86\text{--}0.91$) y EfficientNet-B0 + RFR ($R^2 = 0.85\text{--}0.90$) conservaron la mejor capacidad predictiva global, alcanzando MAE $\approx 5.4\text{--}5.6$ g para peso fresco, ≈ 92 cm² para área foliar y ≈ 0.34 g para peso seco. Esto demuestra que la fusión de descriptores visuales, de profundidad y métricos fortalece la predicción cuando se emplean modelos capaces de manejar interacciones no lineales complejas.

En el caso de YOLOv8n-Segn, a pesar de ser una arquitectura más ligera y orientada a detección en tiempo real, su rendimiento fue altamente competitivo, con $R^2 = 0.85$ para peso fresco, $R^2 = 0.90$ para área foliar y $R^2 = 0.85$ para peso seco, y errores similares a los de ResNet y EfficientNet. Si bien presenta una ligera reducción en precisión (≈ 0.02 – 0.03 unidades de R^2 respecto a ResNet), su principal ventaja radica en su eficiencia computacional y bajo costo de inferencia, ya que requiere menos parámetros y menor tiempo de procesamiento. Por tanto, aunque ResNet-50 + RFR ofreció los valores de R^2 más altos, la combinación YOLOv8n-Segn + RFR se perfila como la opción más adecuada para sistemas de predicción en tiempo real, donde el equilibrio entre precisión, rapidez es esencial. Su capacidad para extraer información estructural relevante con menor carga computacional la convierte en una alternativa ideal para integración en hardware embebido (por ejemplo, *Raspberry Pi* o *NVIDIA Jetson Nano*) y aplicaciones de fenotipado automatizado en sistemas PFAL.

Cuadro 4.4.: Resultados del desempeño de las combinaciones CNN + ML en la predicción de peso fresco, área foliar y peso seco de lechugas.

Backbone	EfficientNet		ResNet 50		YOLOv11n	
Peso Fresco						
Modelo	R ²	MAE	R ²	MAE	R ²	MAE
RFR	0.87	5.42	0.86	5.52	0.85	5.64
GB	0.84	5.97	0.83	6.04	0.82	6.11
Área foliar						
Modelo	R ²	MAE	R ²	MAE	R ²	MAE
RFR	0.90	92.48	0.91	96.83	0.90	96.92
GB	0.89	103.16	0.88	112.60	0.90	103.39
Peso Seco						

Modelo	R ²	MAE	R ²	MAE	R ²	MAE
RFR	0.86	0.34	0.87	0.34	0.85	0.36
GB	0.86	0.35	0.89	0.33	0.85	0.37

En las gráficas de la Figura 4.6 correspondientes a EfficientNet-B0 + RFR del inciso (a-c), se observa una alineación cercana de los puntos respecto a la línea de identidad ($y = x$), lo que indica una alta concordancia entre los valores reales y los predichos para las tres variables. La ligera dispersión en los extremos superiores sugiere que el modelo tiende a subestimar levemente las muestras de mayor magnitud, aunque los errores se mantienen bajos. El MAE refleja la desviación promedio entre las mediciones reales y las estimaciones en este caso, de apenas unos pocos gramos o centímetros, mientras que el RMSE, ligeramente superior al MAE, indica que no existen desviaciones extremas, confirmando una predicción estable y con bajo error cuadrático. En las gráficas de YOLOv8n-Segm + RFR Figura 4.6 incisos (d-g), se aprecia un comportamiento muy similar: los puntos siguen la tendencia lineal esperada, aunque con una dispersión un poco mayor respecto a EfficientNet, principalmente en valores altos de área foliar. Esto sugiere una menor precisión en muestras grandes, pero un desempeño consistente en el rango medio de datos. Los valores de MAE y RMSE mantienen una diferencia estrecha, lo que demuestra que los errores son homogéneos y no se concentran en casos aislados. En conjunto, ambos modelos presentan errores pequeños y distribuidos uniformemente, lo que evidencia una buena capacidad de generalización. Sin embargo, YOLOv8n-seg logra resultados comparables con menor complejidad computacional, confirmando su potencial para tareas de predicción en tiempo real.

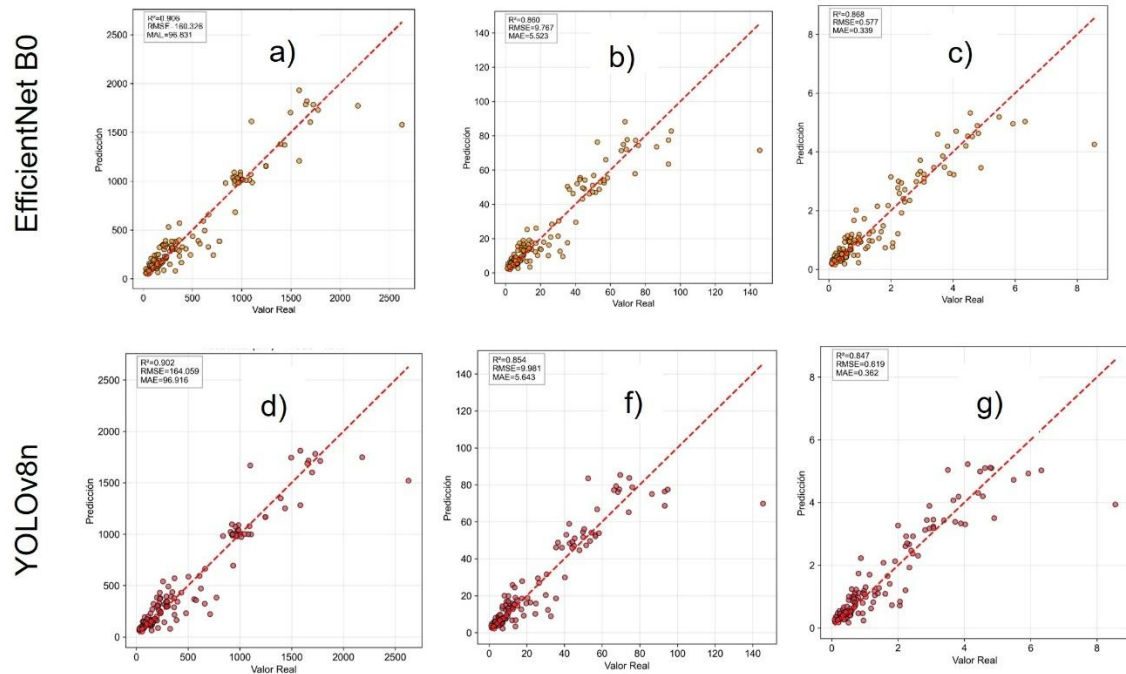


Figura 4.6. Comparación entre valores reales y predichos utilizando EfficientNet B0 + RFR (a–c) y YOLOv8n-Seg + RFR (d–f) para la estimación de área foliar (cm²), peso fresco (g) y peso seco de lechuga (g).

4.4.5 Discusión

Los resultados obtenidos en este estudio confirman la tendencia observada en investigaciones recientes sobre estimación de rasgos fenotípicos y biomasa en especies hortícolas (Dhawi *et al.*, 2025; Liu *et al.*, 2021; X. Zhang *et al.*, 2025c). En primer lugar, los modelos tradicionales de regresión (MLR, SVR) mostraron un rendimiento limitado al intentar predecir variables complejas como el peso fresco, área foliar y peso seco de lechuga a partir de rasgos morfológicos derivados de imágenes RGB-D. La Regresión Lineal Múltiple (MLR), aunque interpretable, no logró capturar la no linealidad inherente de los datos, con valores de R^2 entre 0.78 y 0.88, similares a los reportados (X. Zhang *et al.*, 2025c), donde el modelo alcanzó $R^2 = 0.61$ para peso fresco al usar parámetros morfológicos (MP) y un rendimiento muy pobre con índices de vegetación (VI) ($R^2 = -0.19$).

De manera similar, el SVR tuvo dificultades para ajustarse a la complejidad de las relaciones entre los rasgos fenotípicos, alcanzando valores de $R^2 \leq 0.80$, en concordancia con los hallazgos de (Dhawi *et al.*, 2025), donde el modelo mostró $R^2 = 0.82 - 0.83$ en la predicción de biomasa aérea (AGB) del mijo perla. Este comportamiento se explica por su sensibilidad a los valores atípicos y la dependencia de transformaciones lineales, que limitan su capacidad para manejar datos con alta varianza y no linealidad estructural.

En contraste, los modelos de conjunto basados en árboles de decisión particularmente el Random Forest (RFR) mostraron el mejor desempeño general tanto en este trabajo como en la literatura revisada. En la presente investigación, RFR alcanzó $R^2 = 0.855 - 0.91$ para las tres variables evaluadas, con errores bajos ($MAE < 6$ g, $RMSE < 11$ g), confirmando su robustez frente al sobreajuste y capacidad para modelar interacciones no lineales. Esto concuerda con los estudios de (Dhawi *et al.*, 2025) donde el RFR logró $R^2 = 0.94$ y $RMSE = 0.20$ al utilizar solo las dos características más relevantes. Asimismo, en *Toona sinensis*, (Liu *et al.*, 2021) reportaron rendimientos sólidos $R^2 = 0.77 - 0.80$ para biomasa seca y fresca, lo que subraya la consistencia de este enfoque en diferentes especies.

El modelo Gradient Boosting (GB) presentó un comportamiento similar al RFR, aunque con ligeras variaciones en el error. Su mecanismo de corrección secuencial le permitió alcanzar un $R^2 = 0.86 - 0.90$ en este estudio, comparable con el rendimiento del XGBoost ($R^2 = 0.98$, $RMSE = 0.26$) reportado por (Dhawi *et al.*, 2025). La diferencia principal radica en la mayor estabilidad del RFR con conjuntos de datos pequeños o con ruido, mientras que el GB es más eficiente cuando se dispone de una gran cantidad de características discriminativas.

Las redes neuronales mostraron un comportamiento intermedio. El modelo BPNN alcanzó $R^2 \approx 0.76 - 0.80$, mientras que la versión DE-BP mejoró hasta $R^2 = 0.81$, confirmando que el aumento de dimensionalidad mejora la capacidad de

generalización. Este hallazgo coincide con (X. Zhang *et al.*, 2025c), quienes reportaron que DE-BP superó a MLR, RFR y BPNN con $R^2 = 0.84$ y RMSE = 8 g para peso fresco. Sin embargo, la mejora no fue suficiente para superar de forma consistente a los métodos de conjunto, lo que indica que las redes multicapa tradicionales todavía requieren ajustes arquitectónicos o regularización adicional cuando el tamaño del conjunto de entrenamiento es limitado.

Al integrar características profundas extraídas mediante redes convolucionales (CNN) EfficientNet-B0, ResNet-50 y YOLOv8n-Seg con modelos ML (Machine Learning), el rendimiento se modificó sustancialmente. Los modelos lineales y SVR empeoraron debido a la alta dimensionalidad y no linealidad de las “características profundas”, generando incluso valores negativos de R^2 (por ejemplo, SVR = -0.15 con EfficientNet en área foliar). En contraste, los modelos de conjunto (RFR y GB) mantuvieron o mejoraron su desempeño, alcanzando $R^2 = 0.86 - 0.91$ con ResNet y EfficientNet, evidenciando su habilidad para explotar la riqueza informativa de las características visuales jerárquicas.

Estos resultados son coherentes con los reportes de (Dhawi *et al.*, 2025) donde ResNet y EfficientNet mostraron una gran competencia en la predicción de biomasa y nitrógeno foliar, al capturar información morfológica, cromática y estructural directamente de las imágenes. De hecho, la combinación de EfficientNet-B0 + RFR ($R^2 = 0.85 - 0.90$) y ResNet-50 + RFR ($R^2 = 0.86 - 0.91$) sobresalió como la más precisa, confirmando que la fusión de descriptores visuales, métricos y de profundidad potencia la estimación de parámetros fenotípicos complejos.

Por su parte, YOLOv8n-Seg + RFR alcanzó un rendimiento competitivo ($R^2 = 0.85 - 0.90$), con errores similares a los de las redes más profundas, aunque con una ligera dispersión adicional en valores altos de área foliar. Su principal ventaja radica en la eficiencia computacional, al ser un modelo ligero optimizado para detección y segmentación en tiempo real. A diferencia de ResNet o EfficientNet,

YOLOv8n-Seg requiere menos parámetros y tiempo de inferencia, lo que lo convierte en la opción más adecuada para aplicaciones embebidas en plataformas como Raspberry Pi o Jetson Nano.

En conjunto, los resultados de este estudio demuestran que:

1. Los modelos lineales son insuficientes para describir las relaciones no lineales entre rasgos morfológicos y biomasa.
2. Los modelos de conjunto (RFR y GB) ofrecen un equilibrio óptimo entre precisión y estabilidad, especialmente con datos tabulares y multimodales.
3. Las redes neuronales mejoradas y las características extraídas de CNNs aportan ventajas notables cuando se busca capturar información estructural y visual compleja.
4. Finalmente, YOLOv8n-Seg + RFR representa una solución práctica y eficiente para predicción fenotípica en tiempo real, mientras que ResNet-50 + RFR es la opción preferente para análisis de alta precisión en entornos controlados y aplicaciones de real time.

4.4.6 Conclusiones

El presente estudio demuestra la eficacia de la visión computacional RGB-D y los modelos de aprendizaje automático para la estimación no destructiva de rasgos fenotípicos clave en lechuga cultivada bajo condiciones controladas de fábrica de plantas (PFAL). Las métricas morfológicas 2D–3D derivadas de las imágenes RGB-D como los diámetros mayor y menor, área convexa, volumen aparente y altura real mostraron una alta correlación ($R^2 = 0.82 - 0.93$) con los parámetros productivos (peso fresco, peso seco y área foliar), confirmando su validez como descriptores estructurales del crecimiento vegetal. Entre los modelos de aprendizaje automático evaluados, los algoritmos de conjunto basados en árboles, Random Forest (RFR) y Gradient Boosting (GB), demostraron el mejor

desempeño general ($R^2 = 0.85 - 0.91$, $MAE < 6$ g), superando a los métodos lineales (MLR, SVR) y a las redes neuronales convencionales. Su robustez frente al sobreajuste y capacidad para modelar interacciones no lineales los posicionan como herramientas óptimas para predecir biomasa y área foliar a partir de rasgos morfológicos RGB-D.

La integración multimodal de características profundas extraídas por CNNs (EfficientNet-B0, ResNet-50 y YOLOv8n-Segm) con las variables morfológicas mejoró o mantuvo la precisión de los modelos de conjunto, alcanzando con combinaciones como ResNet-50 + RFR y EfficientNet-B0 + RFR. Entre ellas, YOLOv8n-Segm + RFR ofreció un equilibrio sobresaliente entre precisión, eficiencia computacional y tiempo de inferencia, consolidándose como una opción viable para aplicaciones de predicción fenotípica en tiempo real sobre plataformas embebidas (Raspberry Pi, Jetson Nano).

En conjunto, los resultados confirman que la fusión de información métrica, estructural y visual proveniente de imágenes RGB-D y redes profundas potencia la estimación precisa y automatizada de parámetros de crecimiento vegetal. Este enfoque constituye un avance hacia sistemas inteligentes de fenotipado de alta precisión y bajo costo, con potencial aplicación en la agricultura de ambiente controlado, optimización del manejo hídrico-nutricional y monitoreo dinámico del desarrollo de cultivos.

4.5 Literatura citada

Bloch, V., Shapiguzov, A., Kotilainen, T., & Pastell, M. (2025). A method for phenotyping lettuce volume and structure from 3D images. *Plant Methods*, 21(1). <https://doi.org/10.1186/s13007-025-01347-y>

- Breiman, L. (2001). *Random Forests* (Vol. 45).
- Buxbaum, N., Lieth, J. H., & Earles, M. (2022). Non-destructive Plant Biomass Monitoring With High Spatio-Temporal Resolution via Proximal RGB-D Imagery and End-to-End Deep Learning. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.758818>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Cortes, C., Vapnik, V., & Saitta, L. (1995). Support-Vector Networks Editor. In *Machine Learning* (Vol. 20). Kluwer Academic Publishers.
- Dhawi, F., Ghafoor, A., Almousa, N., Ali, S., & Alqanbar, S. (2025). Predictive modelling employing machine learning, convolutional neural networks (CNNs), and smartphone RGB images for non-destructive biomass estimation of pearl millet (*Pennisetum glaucum*). *Frontiers in Plant Science*, 16. <https://doi.org/10.3389/fpls.2025.1594728>
- Friedman, J. H. (2001). 999 REITZ LECTURE GREEDY FUNCTION APPROXIMATION: A GRADIENT BOOSTING MACHINE 1. In *The Annals of Statistics* (Vol. 29, Issue 5).
- Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences of the United States of America*, 115(18), 4613–4618. <https://doi.org/10.1073/pnas.1716999115>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>

- Islam, S., Reza, M. N., Chowdhury, M., Ahmed, S., Lee, K. H., Ali, M., Cho, Y. J., Noh, D. H., & Chung, S. O. K. (2024). Detection and segmentation of lettuce seedlings from seedling-growing tray imagery using an improved mask R-CNN method. *Smart Agricultural Technology*, 8. <https://doi.org/10.1016/j.atech.2024.100455>
- Lin, Z., Fu, R., Ren, G., Zhong, R., Ying, Y., & Lin, T. (2022a). Automatic monitoring of lettuce fresh weight by multi-modal fusion based deep learning. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.980581>
- Lin, Z., Fu, R., Ren, G., Zhong, R., Ying, Y., & Lin, T. (2022b). Automatic monitoring of lettuce fresh weight by multi-modal fusion based deep learning. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.980581>
- Liu, W., Li, Y., Liu, J., & Jiang, J. (2021). Estimation of plant height and aboveground biomass of toona sinensis under drought stress using rgb-d imaging. *Forests*, 12(12). <https://doi.org/10.3390/f12121747>
- Lu, S., Lv, Y., Qian, T., Ren, W., Li, X., Li, Y., & Li, G. (2025). Lettuce architectural phenotypes extraction from multimodal images by low-light sensitivity and strong spatial perception. *Computers and Electronics in Agriculture*, 232. <https://doi.org/10.1016/j.compag.2025.109928>
- Paulus, S. (2019). Measuring crops in 3D: Using geometry for plant phenotyping. In *Plant Methods* (Vol. 15, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13007-019-0490-0>
- Petropoulou, A. S., van Marrewijk, B., de Zwart, F., Elings, A., Bijlaard, M., van Daalen, T., Jansen, G., & Hemming, S. (2023). Lettuce Production in Intelligent Greenhouses—3D Imaging and Computer Vision for Plant Spacing Decisions. *Sensors*, 23(6). <https://doi.org/10.3390/s23062929>

- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
- Tan, M., & Le, Q. V. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*.
- Ultralytics. (2023, January 10). *Explore Ultralytics YOLOv8*. <https://docs.ultralytics.com/es/models/yolov8/>
- Wang, M., Cheng, Y., Guo, B., & Jin, S. (2019). Parameters determination and sensor correction method based on virtual CMOS with distortion for the GaoFen6 WFV camera. *ISPRS Journal of Photogrammetry and Remote Sensing*, 156, 51–62. <https://doi.org/10.1016/j.isprsjprs.2019.08.001>
- Yu, S., Fan, J., Lu, X., Wen, W., Shao, S., Guo, X., & Zhao, C. (2022). Hyperspectral Technique Combined With Deep Learning Algorithm for Prediction of Phenotyping Traits in Lettuce. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.927832>
- Zhang, Q., Zhang, X., Wu, Y., & Li, X. (2022a). TMSCNet: A three-stage multi-branch self-correcting trait estimation network for RGB and depth images of lettuce. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.982562>
- Zhang, Q., Zhang, X., Wu, Y., & Li, X. (2022b). TMSCNet: A three-stage multi-branch self-correcting trait estimation network for RGB and depth images of lettuce. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.982562>
- Zhang, X., Han, X., Zhang, Y., Hu, L., & Li, T. (2025a). Multi-Trait Phenotypic Extraction and Fresh Weight Estimation of Greenhouse Lettuce Based on Inspection Robot. *Agriculture (Switzerland)*, 15(18). <https://doi.org/10.3390/agriculture15181929>

Zhang, Y., Li, Y., Cao, X., Wang, Z., Chen, J., Li, Y., Zhong, Z., Bai, R., Yang, P., Pan, F., & Fu, X. (2025). Leaf area estimation in small-seeded broccoli using a lightweight instance segmentation framework based on improved YOLOv11-AreaNet. *Frontiers in Plant Science*, 16. <https://doi.org/10.3389/fpls.2025.1622713>