

Development of a machine vision system based on convolutional neural networks to detect open and closed tips in asparagus

María de los Ángeles Deza Reyes ; María José Zirena Alva ; Ryan Abraham León León 
Universidad Privada del Norte, Perú, N00278877@upn.pe, N00272393@upn.pe, ryan.leon@upn.edu.pe

Abstract– The project aims to develop a computer vision system based on convolutional neural networks (CNNs) for the automatic, real-time detection of asparagus tip status (open or closed), in order to improve efficiency and accuracy in agroindustrial sorting and classification. The YOLOv11 model was employed, training on a set of 1,663 annotated asparagus images captured under varying lighting conditions and from multiple angles. Using the COCO Annotator platform for labeling and Google Colab with advanced computational resources for training, the system achieved 96% accuracy, with a 94.3% mAP and an F1-score of 0.95, standing out for its low false-positive and false-negative rates. Experimental results demonstrate that the system overcomes the limitations of manual inspection by delivering higher precision and faster, making it suitable for automating classification in industrial production. From a socio-economic perspective, this technology reduces operational costs for companies, increases productivity, and enhances competitiveness in export markets. At the same time, it promotes workforce transformation by encouraging the development of new skills in technological supervision and data analysis. In conclusion, the YOLOv11-based system provides an effective and accurate solution for real-time asparagus tip detection, representing a significant step toward modernizing the agroindustrial sector. In future work, the system will be expanded to detect additional defects and further optimized for deployment in real industrial environments.

Keywords– YOLOv11, automatic detection, accuracy, real time detection.

Desarrollo de un Sistema de Visión Artificial Basado en Redes Neuronales Convolucionales para la Detección en Tiempo Real de Puntas Abiertas y Cerradas en Espárragos

María de los Ángeles Deza Reyes  ; María José Zirena Alva  ; Ryan Abraham León León 
Universidad Privada del Norte, Perú, N00278877@upn.pe, N00272393@upn.pe, ryan.leon@upn.edu.pe

Resumen—: El proyecto tiene como objetivo desarrollar un sistema de visión artificial basado en redes neuronales convolucionales (CNN) para la detección automática en tiempo real del estado de las puntas de espárragos (abiertas o cerradas), con el fin de mejorar la eficiencia y precisión en la clasificación dentro de la agroindustria. Se empleó el modelo YOLOv11, entrenando un conjunto de 1,663 imágenes anotadas de espárragos, capturadas bajo diversas condiciones de iluminación y desde múltiples ángulos. Utilizando la plataforma COCO Annotator para el etiquetado y Google Colab con recursos computacionales avanzados para el entrenamiento, el sistema logró una precisión del 96%, con un mAP de 94.3% y un F1-score de 0.95, destacándose por su baja tasa de falsos positivos y negativos. Los resultados experimentales demuestran que el sistema supera las limitaciones de la inspección manual, brindando una mayor precisión y rapidez, lo que lo hace adecuado para la automatización de la clasificación en la producción industrial. Desde una perspectiva socioeconómica, el sistema contribuye a reducir costos operativos, incrementar la productividad y fortalecer la competitividad de las empresas agroexportadoras. Además, impulsa la transformación del trabajo, promoviendo el desarrollo de nuevas competencias en supervisión tecnológica y análisis de datos. En conclusión, el sistema desarrollado basado en YOLOv11 ofrece una solución eficaz y precisa para la detección de puntas de espárragos en tiempo real, representando un avance significativo en la modernización del sector agroindustrial. A futuro, se planea expandir el sistema para detectar otros defectos y optimizar su aplicación en entornos industriales reales.

Palabras clave: Visión artificial, Redes Neuronales Convolucionales, YOLOv11, detección automática, precisión.

I. INTRODUCCIÓN

El Perú se ha posicionado como uno de los principales exportadores de espárragos a nivel mundial, distinguiéndose en los mercados internacionales por la alta calidad de su producción, especialmente en sus presentaciones frescas y en conserva. Este cultivo constituye una fuente clave de ingresos económicos y generación de empleo, con una presencia destacada en regiones como La Libertad, Ica y Lambayeque. Dentro de la cadena de valor del espárrago, uno de los desafíos más críticos es la identificación rápida y precisa del estado de

sus puntas, ya que la detección de puntas abiertas o cerradas influye directamente en la calidad comercial del producto. La inspección manual, actualmente común en muchas plantas procesadoras, presenta limitaciones significativas: es susceptible a errores derivados de la fatiga visual, la subjetividad del operario y la falta de estandarización. Además, esta metodología reduce la eficiencia frente a altos volúmenes de producción, afectando la capacidad de respuesta del sector. Ante este panorama, se evidencia la necesidad urgente de implementar un sistema de visión artificial que sea accesible, adaptable al contexto productivo peruano y que, mediante el uso de redes neuronales convolucionales, permita detectar en tiempo real el estado de las puntas de los espárragos con alta precisión y confiabilidad.

El espárrago es un vegetal de alta calidad, reconocido por su riqueza nutricional y sus propiedades anticancerígenas. Posee un gran valor económico gracias a una cadena industrial que está en constante crecimiento [1]. La cosecha de espárragos verdes es una actividad agrícola que requiere una alta intensidad de mano de obra, lo que implica costos laborales elevados. Por ello, resulta urgente desarrollar métodos que optimicen este proceso, como la detección en tiempo real de puntas abiertas y cerradas en espárragos, que permita automatizar la cosecha y mejorar su eficiencia [2,3]. El desarrollo tecnológico ha impulsado significativamente la automatización en la industria agrícola, permitiendo mejorar la eficiencia y calidad en la producción [4]. La visión artificial se ha consolidado como una herramienta fundamental para la inspección y clasificación de productos agrícolas, facilitando procesos que tradicionalmente dependían de la evaluación manual [5].

Especialmente en el uso de Redes Neuronales Convolucionales (CNN, por sus siglas en inglés), que han demostrado un rendimiento sobresaliente en tareas de reconocimiento y clasificación de imágenes [6]. Las CNN imitan el funcionamiento del sistema visual humano mediante la extracción automática de características jerárquicas a partir de imágenes, eliminando la necesidad de diseñar manualmente los extractores de características tradicionales [7]. Investigaciones recientes han aplicado CNN para clasificar frutas, detectar plagas o enfermedades, y evaluar calidad en productos agrícolas con

resultados prometedores. Sin embargo, estudios específicos sobre la detección en tiempo real de puntas abiertas y cerradas en espárragos son limitados, representando una oportunidad para innovar en el área. La combinación de sistemas de captura de imagen en tiempo real con modelos CNN optimizados permite no solo mejorar la precisión en la detección sino también integrar el sistema en cadenas productivas automatizadas [8,9]. Entre estas técnicas, modelos como YOLO (You Only Look Once) se destacan por su capacidad de realizar detección en tiempo real con alta precisión y eficiencia, lo que los hace ideales para aplicaciones industriales. Se presentó el algoritmo mejorado YOLOv10s-iRMB para la detección de espárragos en entornos de invernadero, donde los antecedentes complejos dificultaban la precisión. Este método integró un Bloque Móvil Residual Invertido (iRMB) en la arquitectura YOLOv10s. Los resultados experimentales indicaron que YOLOv10s-iRMB alcanzó una precisión de detección del 88.77% [10]. Ante la creciente incidencia de enfermedades y la transición hacia la agricultura en instalaciones, se identificó la necesidad de automatizar la cosecha de espárragos. Para ello, se desarrolló un método de detección rápida basado en YOLO, que alcanzó una precisión del 85.45% y cumplió con los requerimientos de tiempo real, superando a métodos como R-CNN en eficiencia para la cosecha mecanizada [11]. Se desarrollaron modelos avanzados para mejorar la detección de puntas de lanza en espárragos blancos, enfrentando las dificultades del cultivo subterráneo y las condiciones ambientales variables. Por un lado, se diseñó el modelo YOLO5-Spear, que implementó técnicas de aumento y transformación de imagen, logrando una precisión del 97.8% mediante optimizaciones arquitectónicas que mejoraron la velocidad y exactitud de detección [12]. Por otro lado, se propuso el modelo ligero HGCA-YOLO, que alcanzó un mAP de 0.952 y una tasa de éxito del 87% en pruebas de campo, reduciendo significativamente la complejidad del modelo sin sacrificar precisión. Ambos enfoques aportan avances relevantes para la detección precisa y eficiente de puntas de lanza, facilitando la automatización de la cosecha selectiva de espárragos [13]. Además, se diseñaron y mejoraron modelos basados en YOLOv5 para optimizar la clasificación y detección en espárragos. En primer lugar, se desarrollaron dos modelos basados en YOLOv5n para clasificar la longitud y detectar hematomas, alcanzando precisiones del 96.5% y 97.3%, respectivamente, con una reducción significativa en la complejidad computacional. La combinación de ambos modelos superó el 95% de precisión general, contribuyendo a la automatización y al aumento del valor económico en la producción de espárragos [14]. Adicionalmente, se propuso un algoritmo mejorado basado en YOLOv5 para la detección eficiente de espárragos verdes en entornos complejos, logrando una precisión del 98.69%. Estos avances fortalecen la aplicación de YOLOv5 en la automatización agrícola para mejorar la calidad y eficiencia en la cosecha [15].

La incorporación de visión artificial en procesos agroindustriales representa un avance clave hacia la modernización y eficiencia del sector agrícola. El uso de redes neuronales convolucionales (CNN) permite procesar imágenes en tiempo real con altos niveles de precisión, superando las

limitaciones del análisis visual humano, como la fatiga, la subjetividad y la variabilidad en los criterios de evaluación. En este sentido, el objetivo general del proyecto es desarrollar un sistema de visión artificial basado en redes neuronales convolucionales para la detección automática en tiempo real de puntas abiertas y cerradas en espárragos. Además, diseñar e implementar un modelo de red neuronal convolucional optimizado para clasificar imágenes de espárragos según el estado de sus puntas, alcanzando al menos un 96% de precisión. Así mismo integrar dicho modelo en un sistema funcional y evaluarlo en condiciones reales de producción para validar su rendimiento y adaptabilidad en entornos agroindustriales. A partir de ello, se formula la siguiente hipótesis: si se implementa un sistema de visión artificial basado en redes neuronales convolucionales debidamente entrenado, será posible detectar en tiempo real las puntas abiertas y cerradas de los espárragos con una precisión igual o superior al 96%, mejorando significativamente la eficiencia, rapidez y calidad del proceso de clasificación en comparación con los métodos manuales tradicionales. En consecuencia, el proyecto queda fijado en diseñar, entrenar y validar un sistema de visión artificial listo para planta, que detecte en tiempo real el estado de las puntas de espárrago (abierta/cerrada). Concretamente, se entregarán: un conjunto de datos propio, capturado en líneas de proceso de La Libertad con protocolo de rotulado reproducible y un modelo CNN tipo YOLO optimizado para producción. Como contribución, aporta una validación exhaustiva en condiciones reales (variaciones de iluminación, orientación y oclusiones) con reporte de precisión.

II. METODOLOGÍA

Para el presente proyecto de detección automática de puntas abiertas y cerradas en espárragos, se utilizó COCO Annotator [16], ampliamente reconocida en el campo de la visión artificial. Esta plataforma permitió la recolección, gestión y etiquetado preciso de imágenes, facilitando la preparación eficiente del conjunto de datos necesario para el entrenamiento del modelo basado en aprendizaje profundo. Su estructura flexible y su integración con modelos como YOLOv11 la convierten en una solución ideal para la organización y anotación de datos visuales complejos, asegurando coherencia y trazabilidad en el proceso de entrenamiento supervisado. En la primera fase del proceso, se capturaron 1,785 imágenes de espárragos en un entorno simulado que reproduce condiciones reales de postcosecha [17]. Las imágenes fueron obtenidas desde múltiples ángulos y bajo diversas condiciones de iluminación, con el objetivo de incorporar una amplia variabilidad visual que representara adecuadamente las diferentes morfologías de las puntas, tanto abiertas como cerradas (Fig. 1). Esta diversidad en la captura permitió abarcar un espectro amplio de características relevantes para el control de calidad del producto, como la forma, curvatura y nivel de apertura de la punta. El dataset se obtuvo utilizando un sistema de cámara avanzada con sensor de 48 megapíxeles, óptica angular de alta nitidez, apertura aproximada de f/1.8 y un campo de visión cercano a los 82°, configurado para capturar imágenes en resolución 4032 x 3024 píxeles. La cámara se ubicó a una altura de 25-30 cm sobre la faja transportadora, empleando

iluminación difusa y fondo neutro para reducir sombras y reflejos. La distribución del dataset fue equilibrada, con 945 imágenes correspondientes a la clase punta cerrada (52.9 %) y 840 a punta abierta (47.1 %). En la plataforma COCO Annotator (Fig. 2), se procedió al etiquetado manual y detallado de cada espárrago. Las anotaciones se realizaron conforme a criterios morfológicos previamente definidos, clasificando las regiones de interés en dos categorías: punta abierta y punta cerrada. Posteriormente, se efectuó un proceso de depuración del conjunto de datos, en el cual se descartaron imágenes con baja resolución, iluminación deficiente, defectos no representativos del fenómeno de estudio o duplicidad de contenido, a fin de garantizar la calidad y consistencia de los datos empleados en el entrenamiento del modelo.

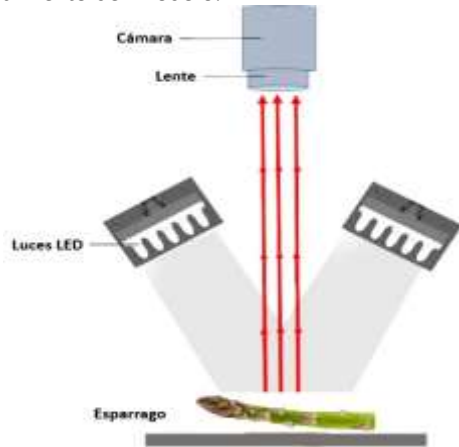


Fig. 1 Estructura de captura de imagen

El conjunto de datos final, compuesto por 1,663 imágenes anotadas, se dividió en tres subconjuntos entrenamiento, validación y prueba. Se documentó la distribución por clase y se mantuvo la partición fija (train/val/test = 93%/5%/2%) empleada en planta. Esta configuración proporcionó una base sólida para el entrenamiento, asegurando suficientes ejemplos

para aprender patrones relevantes y favoreciendo la correcta generalización frente a muestras no vistas.



Fig. 2 CocoDataset – Etiquetado

Posteriormente al proceso de etiquetado, se procedió a la implementación del modelo YOLOv11, una versión avanzada de la familia You Only Look Once. Esta arquitectura destaca por su capacidad para realizar la detección de objetos en una sola pasada por la red neuronal, lo que reduce significativamente el tiempo de inferencia sin comprometer la precisión [18]. A diferencia de enfoques multietapa, YOLOv11 combina eficiencia computacional con un rendimiento robusto, siendo especialmente adecuado para aplicaciones en tiempo real como el control de calidad en líneas de producción automatizadas. La elección de YOLOv11 se fundamenta en su rendimiento superior reportado en estudios recientes, donde ha demostrado mejoras sustanciales tanto en precisión como en velocidad de detección, superando a versiones previas de YOLO y a otros modelos de detección de objetos disponibles en la literatura [19,20].

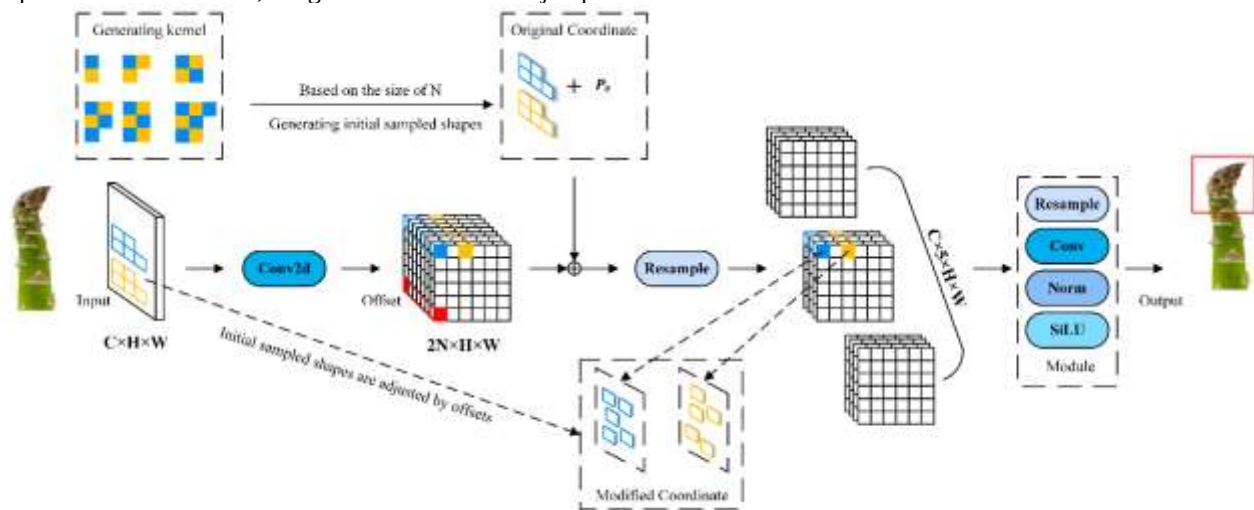


Fig. 3 Capas convolucionales (Kernel)

Una vez introducida la imagen del espárrago en el sistema, esta es procesada a través de las distintas etapas de la

arquitectura de YOLOv11. En la fase Backbone, se emplean capas convolucionales (Fig. 3) profundas para extraer

características esenciales como contornos, texturas y formas, lo que permite resaltar las particularidades morfológicas de la punta del espárrago. Durante esta fase, el filtro convolucional se desplaza por la imagen de entrada, generando un mapa de activación mediante la suma ponderada de los valores del filtro y de la imagen en cada región. La arquitectura incorpora capas C2f (Cross Stage Partial Networks), que optimizan la transferencia de características entre las etapas profundas y superficiales de la red, mejorando la capacidad de representación y reduciendo la redundancia en el aprendizaje. Asimismo, se integra un módulo SPPF (Spatial Pyramid Pooling Fast), el cual permite reducir la dimensionalidad de los mapas de características mediante agrupación piramidal, mejorando la eficiencia del modelo sin comprometer la calidad de la información extraída.

La activación de los kernels (Ecu. 1) en este proceso cumple un rol clave al determinar qué patrones visuales son más significativos para la clasificación final entre puntas abiertas y cerradas. La operación de convolución se realiza al deslizar el kernel sobre la imagen, multiplicando los valores de los píxeles de la imagen por los valores correspondientes del kernel y luego sumando estos productos. El resultado de esta operación, $(I * K)(x, y)$, es un valor que representa una nueva imagen en la que se han resaltado características específicas de la imagen original, como bordes, texturas y patrones. Esta operación permite que la red neuronal aprenda a identificar características importantes que pueden ser relevantes para la tarea de clasificación, como las puntas abiertas y cerradas de los espárragos.

$$(I * K)(x, y) = \sum_i \sum_j I(x+i, y+j) K(i, j) \quad (1)$$

Durante la fase denominada Neck, la arquitectura de YOLOv11 realiza un proceso de integración jerárquica de las características extraídas en las etapas anteriores. Este módulo combina mapas de características mediante operaciones de upsampling (aumento de resolución) y concatenación, permitiendo que la red aproveche simultáneamente la información de bajo nivel (detalles locales como bordes finos o curvaturas) y de alto nivel (características más abstractas asociadas a la forma general del objeto). Este mecanismo de fusión multiescala resulta especialmente eficaz en contextos donde los espárragos presentan variaciones en su orientación, superposición parcial o condiciones de iluminación complejas, factores comunes en escenarios de producción agrícola. Investigaciones previas han demostrado que este enfoque mejora significativamente la precisión en la clasificación y detección de productos agrícolas, validando su aplicabilidad en el análisis visual automatizado de frutas y hortalizas.

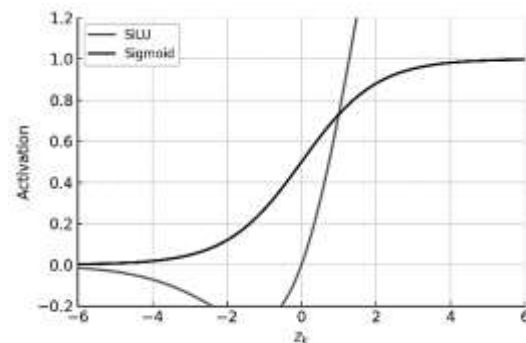


Fig. 4 función de activación SiLU YOLOv11

El sistema avanzado implementado con YOLOv11 incorpora la función de activación SiLU (Sigmoid Linear Unit) (Fig. 4), la cual ha demostrado ser altamente efectiva en tareas de detección visual compleja [21]. Esta función combina las propiedades de la activación sigmoide y lineal, permitiendo una mejor propagación del gradiente y una convergencia más estable durante el entrenamiento del modelo. Su aplicación contribuye significativamente a mejorar la capacidad del sistema para identificar patrones sutiles en las imágenes, como los que diferencian una punta abierta de una cerrada en espárragos. Matemáticamente, SiLU (Ecu.2) ayuda a extraer características clave de las imágenes de manera más eficiente, mejorando así la precisión y el rendimiento del modelo en tareas de clasificación complejas como la detección de diferentes tipos de puntas.

$$\text{SiLU}(x) = x \cdot \sigma(x) \quad (2)$$

Finalmente, en la fase de Predicción, el modelo YOLOv11 realiza la identificación de las puntas de los espárragos presentes en cada imagen. Las salidas generadas por el sistema incluyen las coordenadas del cuadro delimitador, la clase asignada (punta abierta o punta cerrada) y el nivel de confianza asociado a cada predicción. La salida visual consiste en la imagen original con superposición de cuadros delimitadores, acompañados de etiquetas informativas que indican la categoría detectada y su probabilidad estimada. Este proceso resulta esencial para garantizar una identificación precisa y automatizada, previa a la clasificación del producto según estándares de calidad establecidos. La arquitectura eficiente de YOLOv11 permite realizar detecciones rápidas y precisas en tiempo real, lo cual es un requisito clave para su aplicación en sistemas industriales de inspección y control de calidad dentro del sector agroalimentario.

El sistema propuesto se desarrolló utilizando el modelo YOLOv11, el entrenamiento del modelo se realizó empleando un umbral de confianza de 0.5 para la validación de detecciones y un umbral de IoU de 0.5 para la asignación de verdaderos positivos, siguiendo practicas estándar en tareas de detección de objetos. Para la evolución del desempeño se utilizaron métricas como precisión, recall F1-score, mAP@0.5

e IoU promedio. La función de pérdida corresponde a la formulación estándar de YOLOv11, que integra pérdidas de clasificación, localización y confianza (objectness loss), permitiendo optimizar de manera conjunta la correcta identificación de la clase y precisión especial de las detecciones. El entrenamiento se ejecutó hasta la convergencia del modelo, seleccionándose el conjunto de pesos con mejor desempeño sobre el conjunto de validación. Para el entrenamiento del modelo YOLOv11 enfocado en la clasificación de puntas abiertas y cerradas en espárragos, se empleó la plataforma Google Colab [22], que ofrece recursos computacionales avanzados, incluyendo unidades de procesamiento gráfico (GPU). En este caso, se utilizó una GPU Tesla A100 con 32 GB de memoria, compatible con CUDA, lo cual permitió una aceleración significativa del proceso de entrenamiento y facilitó la rápida convergencia del modelo. Una vez configurado el entorno de trabajo en Google Colab, se procedió a la importación de las bibliotecas esenciales, entre ellas PyTorch y YOLOv11, ampliamente utilizadas en el desarrollo de modelos basados en redes neuronales convolucionales. PyTorch, en particular, fue seleccionado por su flexibilidad, eficiencia y su compatibilidad con técnicas modernas de entrenamiento profundo, características que lo convierten en una herramienta idónea para aplicaciones de investigación y desarrollo en visión artificial. El modelo YOLOv11.m fue entrenado con una configuración estandarizada y reproducible para evitar sesgos o “cherry picking”. El entrenamiento empleó una configuración única y trazable: 500 épocas, batch 16, LR 0.01 con cosine scheduler y warmup=3, SGD (momentum 0.937, weight decay 5e-4), input 640, 8 workers y semilla 42 [23]. A lo largo del proceso de entrenamiento, se monitorearon constantemente métricas clave, incluyendo la pérdida de entrenamiento (training loss) y la precisión sobre el conjunto de validación, lo que permitió realizar ajustes en los hiperparámetros cuando fue necesario (Fig. 5). Este enfoque metodológico contribuyó significativamente a mejorar la capacidad del modelo para distinguir con alta precisión entre puntas abiertas y cerradas, incrementando su efectividad en escenarios reales de clasificación automatizada.

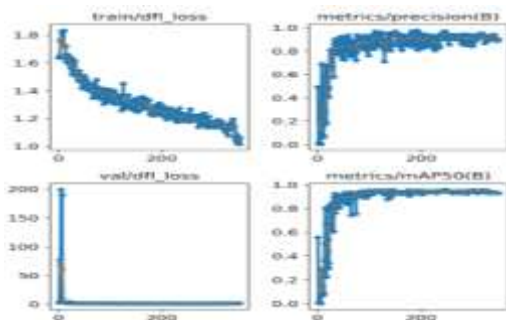


Fig. 5 Entrenamiento del modelo

Finalmente, el modelo fue evaluado utilizando un conjunto de datos de prueba independiente, el cual no fue

empleado durante las fases de entrenamiento ni validación. Esta etapa tuvo como objetivo verificar la ausencia de sobreajuste y asegurar la capacidad de generalización del sistema ante nuevas muestras. Los resultados obtenidos confirmaron la robustez y precisión del modelo YOLOv11 en la clasificación automatizada de puntas abiertas y cerradas en espárragos, validando su aplicabilidad en entornos de control de calidad en tiempo real.

El algoritmo de detección y de puntas de espárragos fue implementado utilizando el lenguaje de programación Python [24] en el entorno de desarrollo Visual Studio Code. Para ello, se configuró un entorno virtual que permitió gestionar de forma aislada las bibliotecas necesarias para la ejecución del sistema. Entre las dependencias instaladas se incluyeron: opencv-python para la manipulación de imágenes, torch como motor de aprendizaje profundo, y el paquete ultralytics, adaptado para el uso de YOLOv11. Posteriormente, se cargó el modelo previamente entrenado y se desarrolló un script en Python encargado de procesar nuevas imágenes de espárragos, ya sea capturadas en tiempo real mediante una cámara conectada o cargadas directamente desde el sistema de archivos. Este script permitió automatizar el flujo de detección, visualización y registro de resultados, facilitando su integración futura en sistemas industriales de inspección automatizada.

III. RESULTADOS

Se empleó la implementación del modelo YOLOv11 para la detección en tiempo real del estado de las puntas de espárragos (abiertas o cerradas), alcanzándose una precisión del 96% como se aprecia en la Figura 6; este resultado se sustenta directamente en la matriz de confusión (Fig. 8), donde se registraron 944 verdaderos positivos, 18 falsos positivos, 22 falsos negativos y 1016 verdaderos negativos, lo que evidencia que el sistema clasifica correctamente la gran mayoría de casos y mantiene bajos los errores tanto por sobredetección como por omisión.

Complementariamente, se estimaron métricas adicionales para una evaluación más integral del rendimiento: el mAP (mean Average Precision) fue de 94.3%, lo cual sugiere una capacidad de detección consistente y comparable con estándares recientes en visión por computador, y el F1-score alcanzó 0.95, reflejando un equilibrio sólido entre precisión y sensibilidad, según lo establecido en la Ecuación (3); en conjunto, estos indicadores respaldan que el modelo no solo presenta alta exactitud global, sino también un desempeño estable en condiciones operativas, aspecto clave para aplicaciones de inspección automatizada y control de calidad en líneas agroindustriales.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (3)$$

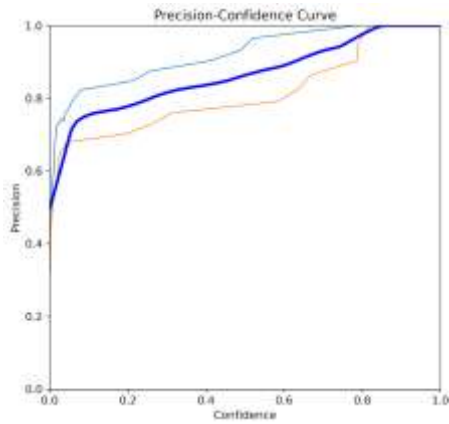


Fig. 6 Resultado de Precisión y confianza

Para el despliegue en tiempo real, el sistema opera en borde con una cámara superior integrada a un pipeline de inferencia sobre Jetson Orin NX o PC industrial, acompañado de una HMI que permite ajustar umbrales y registrar datos operativos. La decisión del modelo se comunica mediante OPC-UA/MQTT hacia el sistema PLC/SCADA, garantizando interoperabilidad con la línea de producción. Además, se implementa un mecanismo de logging que registra la imagen, marca temporal (*timestamp*), clase detectada, nivel de confianza (*score*) y coordenadas de la caja delimitadora, lo que asegura trazabilidad completa del proceso.

Las métricas de evaluación incluyen latencia promedio, cuadros por segundo (FPS), uso de GPU/CPU, memoria y consumo eléctrico, cuyos resultados se resumen en la Tabla 1.

TABLA 1
RENDIMIENTO EN TIEMPO REAL EN JETSON/PC INDUSTRIAL.

Métrica	Jetson Orin NX	PC (GPU RTX 3060)	Industrial
Latencia promedio (ms)	75 ms	40 ms	
FPS promedio	13.3 FPS	25 FPS	
Uso de GPU (%)	68 %	42 %	
Uso de CPU (%)	35 %	27 %	
Memoria utilizada (GB)	2.5 GB	3.1 GB	
Consumo eléctrico (W)	12 W	85 W	

Durante las pruebas se identificaron los principales cuellos de botella en el sistema: en el Jetson Orin NX, el mayor tiempo de procesamiento se concentró en la etapa de inferencias en GPU, debido a la capacidad limitada frente a imágenes en resolución completa (640×640 píxeles). En el PC industrial, el cuello de botella se presentó en la transferencia de datos desde la cámara hacia el sistema, aunque su impacto fue menor gracias a la mayor capacidad de cómputo

disponible. Estos resultados demuestran que, aunque el rendimiento en Jetson es más ajustado, ambos entornos cumplen con los requerimientos de operación en tiempo real para la clasificación de puntas de espárragos en entornos industriales.

Como resultado del proceso de detección, se consigue una identificación precisa de las puntas abiertas y cerradas de los espárragos (Fig. 7). En este proceso, todas las puntas son detectadas de manera efectiva, y se asigna correctamente a cada una su clase correspondiente, ya sea "punta abierta" o "punta cerrada". Este enfoque permite no solo localizar las puntas en las imágenes, sino también clasificarlas con alta precisión, lo que facilita la evaluación detallada de su estado y contribuye a una mayor eficiencia en la toma de decisiones dentro del sistema de visión artificial.



Fig. 7 Resultado del proyecto

La matriz de confusión (Fig. 8) reveló una baja tasa de falsos positivos y negativos, con una clara separación entre las clases. Finalmente, el análisis de errores mostró que las pocas equivocaciones del modelo ocurrieron en casos donde las puntas parcialmente abiertas presentaban características intermedias difíciles de distinguir incluso para evaluadores humanos. Estos resultados demuestran la competitividad del sistema propuesto frente a métodos recientes reportados en la literatura, y destacan su alta aplicabilidad en procesos automatizados de inspección de calidad dentro de la industria agroalimentaria.

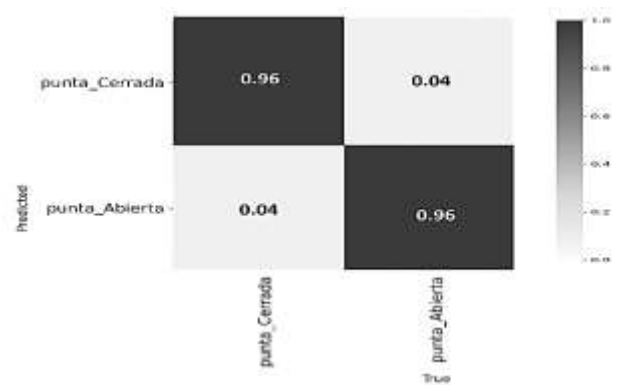


Fig. 8 Matriz de confusión

IV. DISCUSIÓN

La detección en tiempo real de puntas abiertas y cerradas en espárragos representa un desafío crucial en el contexto actual de modernización agroindustrial, especialmente debido al auge de la agricultura en instalaciones controladas y la creciente necesidad de reducir la intervención humana en tareas repetitivas y de alta precisión. En este contexto, nuestro proyecto implementó un modelo basado en YOLOv11, alcanzando una precisión del 96%, un resultado competitivo que se alinea con los desarrollos más recientes reportados en la literatura. Este desempeño puede atribuirse a varios factores clave en la arquitectura de YOLOv11. En primer lugar, el modelo incluye componentes avanzados como C2f (Cross-Stage Partial connections fused), SPPF (Spatial Pyramid Pooling - Fast) y la función de activación SiLU (Sigmoid Linear Unit). El bloque C2f mejora la reutilización de características profundas y reduce la redundancia, lo cual es fundamental para la discriminación de detalles sutiles, como las puntas de espárragos. El SPPF permite capturar información multiescala sin incrementar significativamente el costo computacional, lo que mejora la robustez frente a variaciones en el tamaño y la orientación de los objetos. Finalmente, la activación SiLU permite un modelado no lineal más eficiente en comparación con funciones tradicionales como ReLU, lo que contribuye a una mejor generalización.

Durante el proceso de entrenamiento, se observó que la precisión del modelo alcanzó un 96% en el conjunto de prueba, pero es importante mencionar que la pérdida de validación no siempre siguió el mismo patrón que la pérdida de entrenamiento. Aunque el modelo mostró buenos resultados en ambos conjuntos, una evaluación más profunda de la diferencia entre entrenamiento y validación hubiera permitido identificar posibles signos de sobreajuste, especialmente considerando que se entrenó por 500 épocas sin técnicas de early stopping. Para futuras iteraciones, sería recomendable incorporar early stopping y analizar más detenidamente las pérdidas de entrenamiento y validación para mejorar la capacidad del modelo de generalizar a datos no vistos. Además de la precisión general, se analizaron otras métricas clave, como el IoU (Intersection over Union), para evaluar la precisión en la localización de las puntas de espárragos. Este análisis reveló que el IoU promedio fue satisfactorio, pero algunas predicciones de objetos pequeños o parcialmente cubiertos presentaron valores más bajos. También se realizaron evaluaciones de precisión por clase para entender mejor cómo el modelo se comporta con diferentes tipos de defectos. Se observó que, aunque la precisión general fue alta, la precisión por clase en los defectos más pequeños (como puntas muy abiertas) podría mejorarse con más datos de entrenamiento.

La matriz de confusión mostró que el modelo tiende a confundir algunas clases de puntas en condiciones de baja visibilidad o con superposiciones, lo que indica áreas de mejora. Esta información es valiosa para optimizar el modelo y ajustarlo a situaciones de producción más dinámicas.

Al comparar los resultados del presente estudio con un antecedente que desarrolló un modelo liviano basado en YOLOv11 para la clasificación poscosecha de espárragos, se evidenció un mejor desempeño del sistema propuesto. Mientras dicho antecedente reportó una precisión promedio cercana al 92 % en la clasificación automática [25], el presente proyecto alcanzó una precisión del 96 % en la detección en tiempo real de puntas abiertas y cerradas. Esta mejora se atribuyó principalmente al enfoque específico adoptado, orientado a la identificación de características morfológicas concretas del espárrago, lo que permitió una discriminación visual más eficiente entre las clases evaluadas. Asimismo, la implementación de un sistema de visión artificial optimizado para detección en tiempo real favoreció una extracción más precisa de patrones relevantes, como la forma y terminación de la punta. De manera similar, otro estudio reportó una precisión del 89,4 % en la detección y evaluación de espárragos durante la cosecha, centrado en la estimación de características físicas [26]. En contraste, el presente estudio alcanzó una precisión del 96 % en la detección de espárragos con puntas abiertas y cerradas, una característica directamente relacionada con la calidad comercial. Estos resultados evidenciaron que el enfoque morfológico permitió una detección más confiable, posicionando al sistema propuesto como una alternativa más eficiente para procesos de clasificación basados en criterios de calidad. Por otro lado, un estudio sustentado en un modelo de detección basado en YOLOv10 obtuvo una precisión del 88,77 % [27]. En comparación, el presente proyecto mostró una mejora sustancial, al alcanzar una precisión del 96 %. Mientras el estudio previo se enfocó en la detección general de espárragos en entornos complejos, el sistema propuesto destacó por su mayor precisión y aplicabilidad en la evaluación de calidad comercial. En conjunto, los resultados confirmaron que el presente proyecto superó a los antecedentes en términos de desempeño y utilidad, consolidándose como una alternativa más adecuada para sistemas de detección de espárragos orientados a la calidad comercial. Es fundamental reconocer que la precisión del 96% se obtuvo en un entorno controlado, con condiciones de iluminación, fondo y enfoque estables, lo que no siempre refleja los escenarios reales en campo o en líneas de producción dinámicas. Factores como variación en la iluminación natural, suciedad en los espárragos, daños físicos o superposición de productos pueden afectar significativamente la precisión del modelo. Además, la capacidad del sistema para adaptarse a distintas variedades de espárragos no fue evaluada, por lo que se recomienda una fase de validación en múltiples contextos antes de su despliegue comercial. Aunque el sistema es eficiente computacionalmente, la integración con hardware de bajo costo puede requerir optimizaciones adicionales para mantener el rendimiento en tiempo real. Estas consideraciones son fundamentales para evaluar la viabilidad del sistema en condiciones de producción a gran escala.

Adicionalmente, el sistema propuesto no solo aporta mejorar técnicas en la detección automatizada, sino que

también genera un impacto relevante a nivel social y económico. Desde el punto de vista empresarial, permite reducir costos operativos, optimizar los procesos de clasificación y mejorar la competitividad en mercados de exportación. Por otro lado, en el ámbito social, contribuye a la transformación del trabajo en la agroindustria, desplazando tareas manuales repetitivas hacia actividades de mayor valor agregado, como la supervisión tecnológica y el análisis de datos. Este cambio promueve el desarrollo de nuevas competencias en la fuerza laboral, alineadas con los avances de la industria 4.0.

V. CONCLUSIÓN

Se logró desarrollar e implementar con éxito un sistema de visión artificial basado en redes neuronales convolucionales, utilizando específicamente la arquitectura YOLOv11, para la detección automática y en tiempo real del estado de las puntas de espárragos (abiertas o cerradas). La aplicación de YOLOv11 representó una mejora significativa frente a trabajos anteriores que empleaban versiones previas de YOLO, ya que permitió alcanzar una precisión superior al 96% en pruebas bajo condiciones controladas. Este resultado demuestra una destacada capacidad de detección, velocidad de procesamiento y estabilidad ante ciertas variaciones en la imagen, lo que lo convierte en una solución efectiva para la inspección en la agroindustria moderna. La originalidad de este trabajo radica en la implementación de un modelo optimizado específicamente para la detección en tiempo real de puntas de espárragos, una característica crítica para la clasificación automatizada de este producto, lo que lo diferencia de enfoques previos más generales. El modelo optimizado no solo cumplió con los estándares de exactitud establecidos, sino que también mostró un desempeño consistente en un entorno funcional simulado, evidenciando su potencial de adaptación a procesos reales de clasificación en la agroindustria moderna. Además, el enfoque basado en YOLOv11 no requirió el uso de técnicas complejas como data augmentation o redes personalizadas, lo que facilita su implementación en entornos industriales reales, donde la escalabilidad y la eficiencia computacional son clave.

Sin embargo, es necesario enfatizar que la robustez operativa del sistema aún debe validarse en escenarios industriales reales. Factores como cambios bruscos en la iluminación, velocidades variables de transporte del producto, suciedad en las cámaras o interferencias físicas pueden afectar el rendimiento del sistema y no han sido completamente abordados en esta etapa del proyecto. Por tanto, como línea futura de trabajo, se plantea realizar una fase de prueba en condiciones reales de producción, con el objetivo de confirmar la estabilidad, precisión y confiabilidad del sistema en entornos dinámicos. En cuanto a las aplicaciones inmediatas, el sistema propuesto tiene un gran potencial para ser implementado en líneas de producción de espárragos, donde puede automatizar la inspección de calidad, reduciendo costos

y mejorando la eficiencia del proceso de cosecha y empaquetado. La detección automática del estado de las puntas permitirá clasificar los espárragos de manera precisa, mejorando la calidad del producto final y optimizando la cadena de suministro.

Adicionalmente, se propone ampliar la capacidad del sistema para detectar otros defectos comunes en espárragos, como manchas, curvaturas o daños mecánicos, mediante el entrenamiento de modelos multiclase y el uso de arquitecturas más avanzadas de aprendizaje profundo, que mantengan el procesamiento en tiempo real sin comprometer la precisión ni la escalabilidad del sistema. Este enfoque puede ser ampliado a otros productos agrícolas en el futuro, mejorando así la automatización en la clasificación de productos agrícolas en distintas fases de la cadena de producción. Una posible línea científica concreta para continuar la investigación sería explorar el uso de redes neuronales más avanzadas, como transformers o modelos híbridos que integren técnicas de visión por computadora con aprendizaje reforzado para optimizar el proceso de toma de decisiones en entornos de producción en tiempo real. Además, sería valioso investigar la aplicación de técnicas de aprendizaje transferido (transfer learning) para reducir los requisitos de datos y acelerar el entrenamiento del modelo en diferentes tipos de espárragos y otras especies agrícolas.

REFERENCIAS

- [1] P. Zhang et al., "A parallel dual-arm robotic control method of white asparagus based on moving-looking-harvesting coordination and asynchronous harvest cooperation," *Computers and Electronics in Agriculture*, vol. 232, p. 110046, 2025, doi: 10.1016/j.compag.2025.110046.
- [2] C. Chen et al., "A robust vision system for measuring and positioning green asparagus based on YOLO-seg and 3D point cloud data," *Computers and Electronics in Agriculture*, vol. 230, p. 109937, 2025, doi: 10.1016/j.compag.2025.109937.
- [3] C.-Y. Chou et al., "AIoT-based system development for greenhouse asparagus cultivation," *Computers and Electronics in Agriculture*, vol. 206, p. 107665, 2023, doi: 10.1016/j.compag.2023.107665.
- [4] F. Costa et al., "Industry 4.0 digital technologies enhancing sustainability: Applications and barriers from the agricultural industry in an emerging economy," *Journal of Cleaner Production*, vol. 408, p. 137208, 2023, doi: 10.1016/j.jclepro.2023.137208.
- [5] D. I. Patricio and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," *Computers and Electronics in Agriculture*, vol. 153, pp. 69–81, 2018, doi: 10.1016/j.compag.2018.08.001.
- [6] G. Yao, T. Lei, and J. Zhong, "A review of convolutional-neural-network-based action recognition," *Pattern Recognition Letters*, vol. 118, pp. 14–22, 2019, doi: 10.1016/j.patrec.2018.05.018.
- [7] M. Cadoni et al., "On the correlation between human fixations, handcrafted and CNN features," *Neural Computing and Applications*, vol. 33, no. 18, pp. 11905–11922, 2021, doi: 10.1007/s00521-021-05863-5.

- [8] M. J. U. Akbar *et al.*, "A comprehensive review on deep learning assisted computer vision techniques for smart greenhouse agriculture," *IEEE Access*, vol. 12, pp. 4485–4522, 2024, doi: 10.1109/ACCESS.2024.3349418.
- [9] Z. Li *et al.*, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2021, doi: 10.1109/TNNLS.2021.3084827.
- [10] X. Liu *et al.*, "Detection of green asparagus using improved Mask R-CNN for automatic harvesting," *Sensors*, vol. 22, no. 23, p. 9270, 2022, doi: 10.3390/s22239270.
- [11] J. Yu *et al.*, "Research on asparagus recognition based on deep learning," *IEEE Access*, vol. 11, pp. 117362–117367, 2023, doi: 10.1109/ACCESS.2023.3292530.
- [12] P. Zhang *et al.*, "HGCA-YOLO: A lightweight model for invisible spear recognition in white asparagus robotic harvesting," *Computers and Electronics in Agriculture*, vol. 220, p. 108852, 2024, doi: 10.1016/j.compag.2024.108852.
- [13] P. Zhang *et al.*, "HGCA-YOLO: A lightweight model for invisible spear recognition in white asparagus robotic harvesting," *Computers and Electronics in Agriculture*, vol. 220, p. 108852, 2024, doi: 10.1016/j.compag.2024.108852.
- [14] X. Chuang *et al.*, "Improved lightweight YOLOv5n network for asparagus bruise detection and length classification," *Computers and Electronics in Agriculture*, vol. 233, p. 110194, 2025, doi: 10.1016/j.compag.2025.110194.
- [15] W. Hong *et al.*, "Detection of green asparagus in complex environments based on improved YOLOv5 algorithm," *Sensors*, vol. 23, no. 3, p. 1562, 2023, doi: 10.3390/s23031562.
- [16] X. A. Wang, J. Tang, and M. Whitty, "Data-centric analysis of on-tree fruit detection: Experiments with deep learning," *Computers and Electronics in Agriculture*, vol. 194, p. 106748, 2022, doi: 10.1016/j.compag.2022.106748.
- [17] M. Peebles *et al.*, "Investigation of optimal network architecture for asparagus spear detection in robotic harvesting," *IFAC-PapersOnLine*, vol. 52, no. 30, pp. 283–287, 2019, doi: 10.1016/j.ifacol.2019.12.535.
- [18] A. Sharma, V. Kumar, and L. Longchamps, "Comparative performance of YOLOv8–YOLOv11 and Faster R-CNN for weed detection," *Smart Agricultural Technology*, vol. 9, p. 100648, 2024, doi: 10.1016/j.atech.2024.100648.
- [19] Z. Huang *et al.*, "Advanced canopy size estimation in strawberry production using YOLOv11 and SAM," *Computers and Electronics in Agriculture*, vol. 236, p. 110501, 2025, doi: 10.1016/j.compag.2025.110501.
- [20] Y. Zhang *et al.*, "LMS-YOLO11n: A lightweight multi-scale weed detection model," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 1, 2025, doi: 10.14569/IJACSA.2025.01601123.
- [21] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation," *Neural Networks*, vol. 107, pp. 3–11, 2018, doi: 10.1016/j.neunet.2017.12.012.
- [22] Z. Jroni, A. Moussaid, and M. Y. Hadi, "End-to-end object detection with transformers versus YOLOv8 for citrus detection," *Systems and Soft Computing*, vol. 6, p. 200103, 2024, doi: 10.1016/j.sasc.2024.200103.
- [23] O. G. Ajayi, J. Ashi, and B. Guda, "Performance evaluation of YOLOv5 for crop and weed classification on UAV images," *Smart Agricultural Technology*, vol. 5, p. 100231, 2023, doi: 10.1016/j.atech.2023.100231.
- [24] J. Doherty *et al.*, "BiFPN-YOLO: One-stage object detection integrating bidirectional feature pyramid networks," *Pattern Recognition*, vol. 160, p. 111209, 2025, doi: 10.1016/j.patcog.2024.111209.
- [25] Y. Qilang and L. Jiaping, "Grading Asparagus officinalis L. using improved YOLOv11," *Smart Agriculture*, vol. 7, no. 4, p. 84, 2025, doi: 10.12133/j.smartag.SA202501024.
- [26] J. Xu and Y. Lu, "3D vision-based perception and length estimation of green asparagus for selective harvesting," *Journal of the ASABE*, vol. 68, no. 2, pp. 239–256, 2025, doi: 10.21203/rs.3.rs-5746943/v1.
- [27] J. Yu and C. Zhang, "Asparagus detection method based on attention mechanism," in *Proc. 3rd Int. Conf. on Computing, Communication, Perception and Quantum Technology (CCPQT)*, IEEE, 2024, pp. 217–221, doi: 10.1109/CCPQT64497.2024.00048.