

Robotic System for Strawberry Harvesting: Evaluation of CNN Architectures for Fruit Classification and Validation in Simulation

Cristian Stiven León Álvarez¹; Andrés-F Jiménez López²; Fabián-R Jiménez López³

^{1,2}Universidad de los Llanos, Facultad de Ciencias Básicas e Ingeniería, Macrypt - FarmTechnology, Colombia,
cristian.leon.alvarez@unillanos.edu.co, ajimenez@unillanos.edu.co

³Universidad Pedagógica y Tecnológica de Colombia, I2E, Colombia, fabian.jimenez02@uptc.edu.co

Abstract– This paper presents advances in the development of a robotic system for strawberry harvesting, based on computer vision and deep learning. The mechanical design of a robotic platform with multiple degrees of freedom is proposed, along with its integration into a CoppeliaSim simulation environment via ZMQ remote API, as a preliminary stage prior to physical implementation. For the perception stage, a pipeline of HSV color segmentation followed by classification with convolutional neural networks was implemented, evaluating three architectures on a dataset with eight-fold physical augmentation: custom CNN, VGG-16, and ResNet-50. VGG-16 achieved the best performance with 99.21% accuracy and $AUC = 1.0000$. However, the quantitative evaluation of HSV segmentation revealed a 45.9% fruit loss, which limits the robustness of the system for field conditions. Preliminary validation in simulation showed a reduction in classifier confidence to around 0.5 due to the domain gap between real and synthetic images. Further work is needed on physical implementation, vision-guided dynamic control, and evaluation under real crop conditions.

Keywords– Agricultural robotics; computer vision; deep learning; fruit classification; robotic simulation.

Sistema robótico para la cosecha de fresas: evaluación de arquitecturas RNC para clasificación de frutos y validación en simulación

Cristian Stiven León Álvarez¹; Andrés-F Jiménez López²; Fabián-R Jiménez López³

^{1,2}Universidad de los Llanos, Facultad de Ciencias Básicas e Ingeniería, Macrypt - FarmTechnology, Colombia,
cristian.leon.alvarez@unillanos.edu.co, ajimenez@unillanos.edu.co

³Universidad Pedagógica y Tecnológica de Colombia, I2E, Colombia, fabian.jimenez02@uptc.edu.co

Resumen— Este trabajo presenta los avances en el desarrollo de un sistema robótico para la recolección de fresas, basado en visión artificial y aprendizaje profundo. Se propone el diseño mecánico de una plataforma robótica con múltiples grados de libertad y su integración en un entorno de simulación en CoppeliaSim mediante API remota ZMQ, como etapa previa a la implementación física. Para la etapa de percepción, se implementó un pipeline de segmentación por color HSV seguido de clasificación con redes neuronales convolucionales, evaluando tres arquitecturas sobre un conjunto de datos con aumento físico de ocho veces: RNC personalizada, VGG-16 y ResNet-50. VGG-16 obtuvo el mejor desempeño con 99,21 % de exactitud y AUC = 1,0000. No obstante, la evaluación cuantitativa de la segmentación HSV reveló una pérdida del 45,9 % de frutos, lo que limita la robustez del sistema para condiciones de campo. La validación preliminar en simulación mostró una reducción de la confianza del clasificador alrededor de 0,5 debido al domain gap entre imágenes reales y las evaluadas en el entorno simulado. Se debe seguir trabajando en la implementación física, el control dinámico guiado por visión y la evaluación en condiciones reales de cultivo.

Palabras clave: aprendizaje profundo; clasificación de frutos; robótica agrícola; simulación robótica; visión artificial.

I. INTRODUCCIÓN

Con el continuo crecimiento de la población mundial, el sector agrícola enfrenta el desafío de garantizar el abastecimiento alimentario, al tiempo que lidia con la limitación de recursos disponibles y la necesidad de implementar herramientas que promuevan un desarrollo sostenible [1]. En este contexto, la automatización de labores agrícolas mediante sistemas robóticos permite mejorar la eficiencia operativa y favorecer un uso más racional de los recursos, como lo han demostrado investigaciones recientes en recolección automatizada de frutos [2]. En cultivos de cosecha manual intensiva, la mecanización de las tareas reduce la exposición de los operarios a posturas forzadas y movimientos repetitivos, mitigando la aparición de trastornos musculoesqueléticos [3].

Entre los cultivos de mayor relevancia a nivel global, la fresa destaca por su elevada demanda y su alto valor nutricional. Diversos estudios han señalado que su consumo aporta beneficios significativos para la salud, principalmente por su contenido de compuestos antioxidantes, los cuales contribuyen a la disminución del riesgo de enfermedades crónicas como el cáncer y afecciones cardiovasculares [4]. No

obstante, la recolección de fresas depende en gran medida de trabajo manual, lo que la convierte en una de las etapas más costosas del ciclo productivo y ha motivado el desarrollo de sistemas de cosecha automatizada que buscan reducir la intervención humana [5], con incrementos sostenidos en superficie cultivada y volúmenes de producción que refuerzan la necesidad de optimizar los procesos asociados a este cultivo [6]. En Colombia, la fresa tiene presencia destacada en departamentos como Cundinamarca y Boyacá, sin embargo, según el MADR, la producción depende mayoritariamente de procesos manuales, lo que implica una alta participación de la mano de obra en los costos totales del cultivo [7], a lo que se suma la naturaleza delicada del fruto, vulnerable a daños mecánicos que reducen su vida útil [8].

Ante este panorama, la automatización del proceso de recolección surge como una alternativa viable para mitigar las limitaciones de la cosecha manual. Los sistemas robóticos permiten mejorar la eficiencia mediante la identificación, clasificación y recolección precisa de los frutos, reduciendo el contacto innecesario y minimizando el daño físico [9]. El diseño de estos robots implica la integración de un sistema de percepción, un sistema de actuación generalmente conformado por un brazo robótico y una garra y una plataforma de desplazamiento [10]. En particular las redes neuronales convolucionales (RNC) han demostrado ser herramientas eficaces dentro de los sistemas robóticos de cosecha para tareas de clasificación y detección de frutos [11].

La percepción basada en visión artificial no solo resulta fundamental para la clasificación de frutos, sino que también permite detectar enfermedades y evaluar la calidad del fruto, contribuyendo a una gestión agrícola más precisa [12].

En el contexto colombiano, la adopción de tecnologías de visión artificial y aprendizaje profundo para el monitoreo y clasificación de fresas representa una oportunidad estratégica, como lo han explorado investigaciones recientes que aplican aprendizaje profundo al mapeo de rendimiento de cultivos de fresa en el país [13]. En este trabajo se presenta la evaluación y comparación de tres arquitecturas de redes neuronales convolucionales una RNC personalizada, VGG-16 y ResNet-50 para la clasificación de fresas en estados cosechable y no cosechable, integradas en un sistema robótico simulado en CoppeliaSim.

II. MATERIALES Y METODOS

Para el desarrollo del presente trabajo se utilizó un computador portátil Acer Nitro 5 con procesador Intel Core i5-10300H, 12 GB de RAM, tarjeta gráfica NVIDIA GeForce GTX 1650 (4 GB) y sistema operativo Windows 11. El prototipo mecánico se diseñó en SolidWorks, la simulación robótica se realizó en CoppeliaSim, y la implementación de los modelos RNC se llevó a cabo con Python 3.10 y TensorFlow 2.10.0 en un entorno virtual Anaconda. Las actividades se desarrollaron en el semillero FarmTechnology del grupo Macrypt, Universidad de los Llanos, Colombia.

La Fig. 1 muestra la arquitectura general del sistema robótico propuesto, conformado por un sistema de percepción, una estación central y un sistema de actuación. La arquitectura contempla la operación dentro del cultivo de fresa, donde se adquiere información sobre el estado del fruto para determinar si es apto para ser recolectado, y con base en dicha información el sistema toma decisiones que integran diferentes tecnologías para definir la acción a ejecutar. La estación central está compuesta por un sistema embebido Raspberry Pi 4 encargado del procesamiento y la toma de decisiones. Es importante destacar que esta arquitectura constituye un diseño propuesto para implementación futura; en el presente trabajo, la validación del módulo de clasificación se realizó íntegramente en simulación.

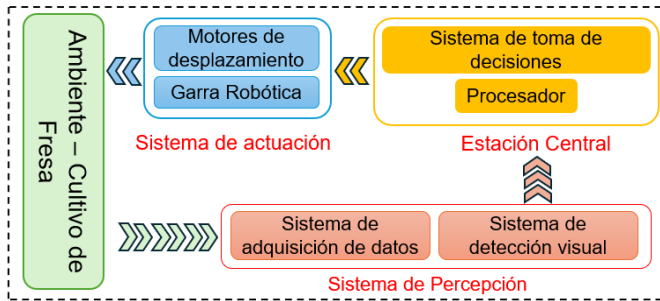


Fig. 1. Arquitectura del sistema.

La Fig. 2 presenta los componentes de la arquitectura propuesta, que incluye alimentación mediante panel solar de 10 W y batería de 12 V/5.5 Ah, una Raspberry Pi 4 como estación central, una cámara Raspberry Pi V2 de 8 MP con sensor infrarrojo E18-D80NK para percepción, y motores paso a paso Nema 17 con drivers TB6600 para actuación.

Se evaluaron tres arquitecturas de RNC: una personalizada entrenada desde cero, VGG-16 con aprendizaje transferido y ajuste fino, y ResNet-50 con aprendizaje transferido y ajuste fino; los hiperparámetros de entrenamiento se observan en la Tabla I. Inicialmente se exploró la segmentación por color HSV con Watershed para extraer fresas individuales de fotografías reales del cultivo, con el fin de construir el conjunto de entrenamiento; sin embargo, se evidenció una pérdida del 45,9 % de los frutos.

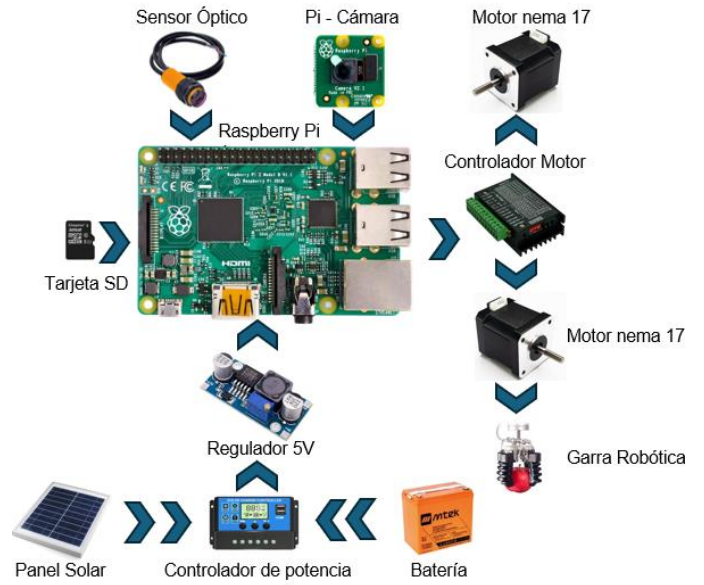


Fig. 2. Componentes del sistema propuesto.

Tabla I
HIPERPARÁMETROS DE ENTRENAMIENTO POR ARQUITECTURA.

Hiperparámetro	RNC Personalizada	VGG-16	ResNet-50
Optimizador	Adam	Adam	Adam
Función de pérdida	Binary Crossentropy	Binary Crossentropy	Binary Crossentropy
Tasa de aprendizaje fase 1	3×10^{-4}	1×10^{-5}	1×10^{-5}
Tasa de aprendizaje fase 2 – ajuste fino	-	1×10^{-5}	1×10^{-5}
Batch Size	32	32	32
Early Stopping	15	15	15
Epocas	100	100	100
Parámetros entrenables	102 K	14.9 M	24.1 M
Bloques descongelados	-	Bloques 3-5	Conv5

La Fig. 3 muestra el resultado obtenido con este enfoque. Ante estas limitaciones, se construyó el conjunto de datos a partir de tres conjuntos públicos de Kaggle [14],[15],[16], seleccionando manualmente imágenes de fresas individuales con visibilidad adecuada para clasificación binaria. El conjunto resultante comprende 840 imágenes originales (420 cosechables y 420 no cosechables) redimensionadas a $224 \times 224 \times 3$ píxeles.

Las 840 imágenes originales se dividieron de forma estratificada en 70% entrenamiento (586), 15% validación (128) y 15% prueba (126) antes de cualquier aumentación, para evitar fuga de datos. Posteriormente, se aplicó aumentación de datos únicamente al conjunto de entrenamiento con un multiplicador de 8, generando 4688 imágenes de entrenamiento mediante rotaciones aleatorias, espejo horizontal y vertical, zoom, traslación, variación de

brillo, contraste, saturación, ruido y desenfoque gaussianos ligero. Los conjuntos de validación y prueba se mantuvieron sin modificaciones.

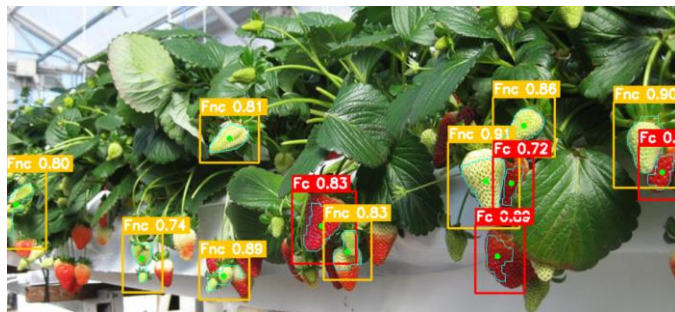


Fig. 3. Resultado segmentación.

El flujo de segmentación por color se mantiene como componente de detección para la simulación en CoppeliaSim. Este consta de: (1) conversión RGB a HSV y segmentación con máscaras independientes para cada categoría, donde Fnc utiliza dos rangos HSV para capturar frutos verdes y en transición; (2) operaciones morfológicas de apertura y cierre; (3) transformada de distancia y umbralización para marcadores de primer plano; (4) algoritmo de Watershed para separar frutos en contacto; y (5) extracción de regiones de interés individuales.

III. RESULTADOS

Como resultado del proceso de diseño, se desarrolló el prototipo mecánico mostrado en la Fig. 4, modelado en SolidWorks a partir de perfiles de aluminio de 20×20 mm que conforman el marco principal, y un perfil de 20×40 mm como riel horizontal perpendicular al surco. Sobre el riel se desliza un carro móvil con dos motores Nema 17 uno para movimiento horizontal por correa dentada y otro para movimiento vertical por tornillo sin fin, integrando una cámara Raspberry Pi V2 y un sensor infrarrojo E18-D80NK.

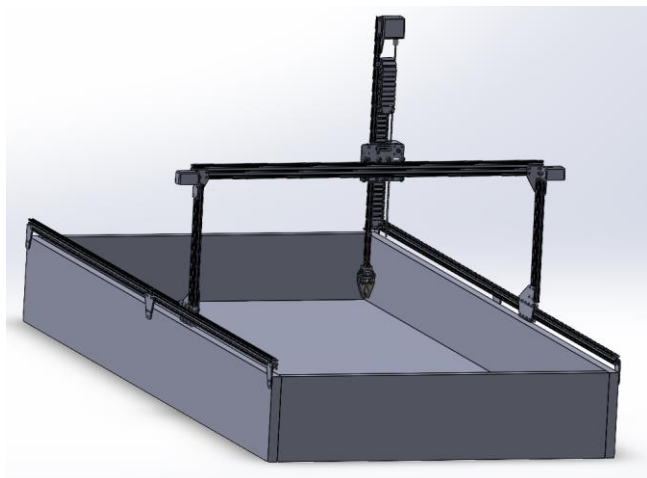


Fig. 4. Diseño plataforma robótica.

La evaluación del flujo de segmentación por color HSV sobre 26 imágenes representativas del cultivo mostró que el enfoque detecta frutos cosechables en escenarios favorables; sin embargo, presenta un desempeño deficiente en la clase Fnc, debido a la similitud cromática entre los frutos verdes y el follaje. La pérdida total de frutos alcanzó el 45,9 % y se registraron 337 detecciones falsas. Estas limitaciones motivaron la construcción del conjunto de datos a partir de imágenes públicas y constituyen la justificación para explorar segmentación instanciada (Mask R-RNC) en trabajos futuros.

Se evaluaron tres arquitecturas RNC sobre el conjunto de prueba, con aprendizaje transferido en dos fases para VGG-16 y ResNet-50. La Fig. 5 compara las métricas de los tres modelos: VGG-16 obtuvo la mayor exactitud (99,21 %) y AUC perfecto (1,0000), seguido por la RNC personalizada y ResNet-50, ambos con 98,41 %. A pesar de sus 24,1 millones de parámetros, ResNet-50 no superó al modelo personalizado de solo 102 K parámetros, lo que sugiere que la complejidad del dominio de fresas no requiere una red excesivamente profunda.

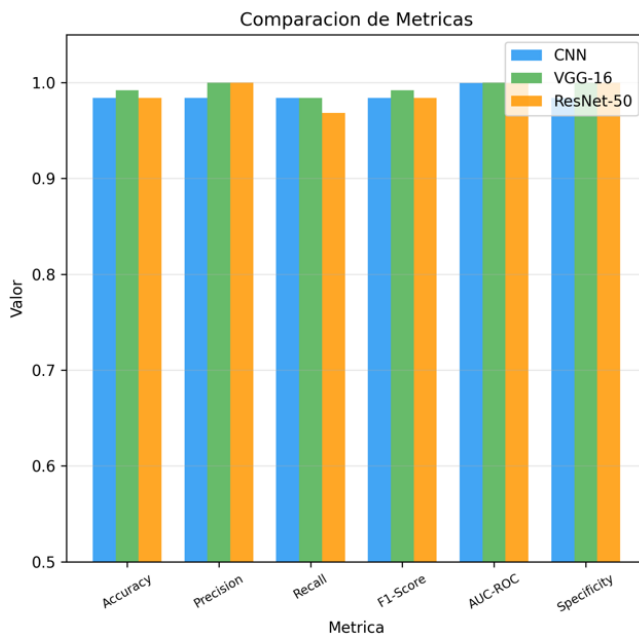


Fig. 5. Comparación de métricas de los tres modelos RNC.

La RNC personalizada alcanzó 98,41 % de exactitud en 54 épocas, cometiendo 2 errores sobre el conjunto de prueba: un falso Fc clasificado como Fnc y un falso Fnc clasificado como Fc, tal como se observa en la Fig. 6. Su ligereza la hace apta para despliegue en hardware embebido, aunque su convergencia fue la más lenta.

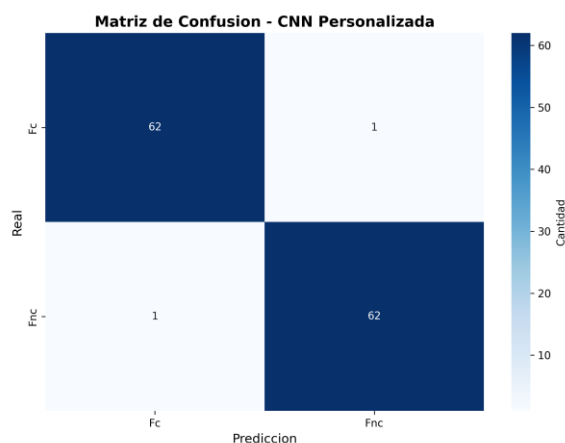


Fig. 6. Matriz de confusión RNC personalizada.

VGG-16 logró el mejor desempeño global con 99,21 % de exactitud, F1 = 99,20 % y AUC = 1,0000 en 35 épocas, tal como se observa en la Fig. 7. Cometió un único error: una fruta Fc clasificada como Fnc. El ajuste fino en los bloques 3–5 permitió adaptar características de alto nivel al dominio específico, lo que la convierte en la arquitectura seleccionada para la etapa de detección en simulación.

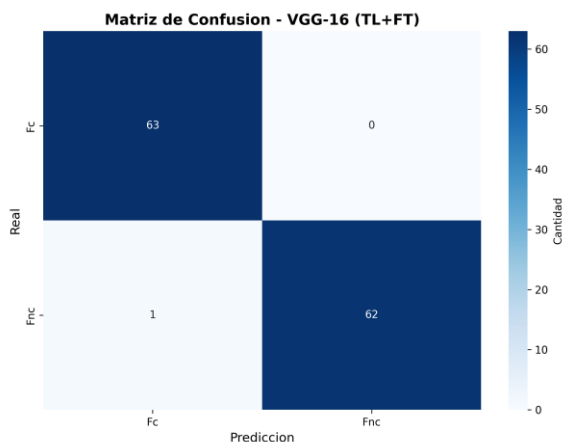


Fig. 7. Matriz de confusión VGG-16.

ResNet-50 obtuvo 98,41 % de exactitud con 2 errores de clasificación, tal como se observa en la Fig. 8, similares a los del modelo personalizado. Convergió en 32 épocas, fue la más rápida gracias al ajuste fino del bloque conv5, pero su mayor costo computacional no se tradujo en mejoras de desempeño respecto al modelo personalizado.

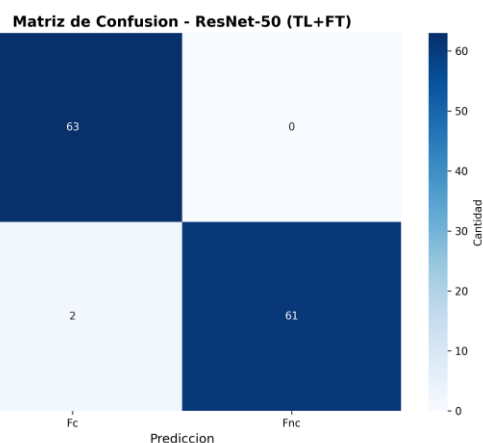


Fig. 8. Matriz de confusión ResNet-50.

La Tabla II resume las métricas comparativas de las tres arquitecturas estudiadas.

Tabla II.
DESEMPEÑO COMPARATIVO DE ARQUITECTURAS RNC EN EL CONJUNTO DE PRUEBA.

Métricas	RNC Personalizada	VGG-16	ResNet-50
Exactitud	98.41%	99.21%	98.41%
Precisión Fc	98.41%	98.44%	96.92%
Recall Fc	98.41%	100.00%	100.00%
Precisión Fnc	98.41%	100.00%	100.00%
Recall Fnc	98.41%	98.41%	96.83%
AUC	0.9997	1.000	0.9995
Epocas Entrenadas	54	35	32

La validación preliminar en simulación permitió poner a prueba el clasificador seleccionado en un escenario de detección integrado con el robot. La Fig. 9 muestra el robot posicionado sobre la planta de fresa en el entorno de CoppeliaSim.



Fig. 9. Robot posicionado sobre la planta de fresa en CoppeliaSim.

El sensor de visión capturó la escena y el *pipeline* HSV + RNC procesó la imagen, detectando y clasificando el fruto presente. La Fig. 10 presenta la captura del sensor con el resultado de la detección y la etiqueta asignada por el clasificador.



Fig. 10. Captura del sensor de visión con detección y clasificación del fruto.

Aunque el sistema identificó correctamente la presencia del fruto, la confianza promedio fue de aproximadamente 0,5, muy inferior al 99,21 % obtenido con imágenes reales. Esta reducción se atribuye al *domain gap* entre las fotografías reales de entrenamiento y las imágenes obtenidas del simulador, que difieren en texturas, iluminación y oclusiones. Los umbrales de área mínima y confianza filtraron detecciones espurias, pero también descartaron frutos válidos. Estas limitaciones motivan como trabajo futuro la generación de datos sintéticos para entrenamiento y el ajuste de umbrales específicos para simulación.

CONCLUSIONES

En este trabajo se presentan los avances en el desarrollo de un sistema robótico para la cosecha de fresas, abarcando el diseño mecánico propuesto, la evaluación del *pipeline* de segmentación HSV, la comparación de tres arquitecturas RNC para clasificación y la validación preliminar del sistema integrado en simulación. VGG-16 alcanzó el mejor desempeño con 99,21 % de exactitud y AUC = 1,0000 sobre el conjunto con aumento físico de ocho veces, superando a ResNet-50 y RNC personalizada, ambas con 98,41 %. Sin embargo, este rendimiento está condicionado por la segmentación HSV previa, cuya evaluación cuantitativa sobre 26 imágenes reales reveló una pérdida del 45,9 % de frutos y 337 detecciones falsas, confirmando que el *pipeline* no es robusto para condiciones de campo y motivando la exploración de segmentación semántica (Mask R-RNC) como trabajo futuro.

Desde el punto de vista computacional, la RNC personalizada ofrece un balance favorable para despliegue embebido, mientras que VGG-16 y ResNet-50 demandan mayores recursos sin ventaja. La validación en CoppeliaSim, donde el diseño mecánico exportado desde SolidWorks fue integrado con el clasificador mediante la API remota ZMQ, evidenció un *domain gap* que reduce la confianza del clasificador a alrededor de 0,5 respecto a las imágenes reales, limitado además por los umbrales de área mínima y confianza implementados. Se debe seguir trabajando en la

implementación física del robot, el control dinámico guiado por visión, la reducción de la brecha entre simulación y realidad, así como la evaluación del sistema completo en condiciones reales de cultivo.

REFERENCIAS

- [1] C. B. Barrett, "Overcoming global food security challenges through science and solidarity," *Amer. J. Agric. Econ.*, vol. 103, no. 2, pp. 422–447, 2021. <https://doi.org/10.1111/ajae.12160>
- [2] T. Yoshida, Y. Onishi, T. Kawahara, and T. Fukao, "Automated harvesting by a dual-arm fruit harvesting robot," *Robomech J.*, vol. 9, 2022. <https://doi.org/10.1186/s40648-022-00233-9>
- [3] F. A. Fathallah, "Musculoskeletal disorders in labor-intensive agriculture," *Appl. Ergon.*, vol. 41, no. 6, pp. 738–743, 2010. <https://doi.org/10.1016/j.apergo.2010.03.003>
- [4] F. Giampieri, J. M. Alvarez-Suarez, and M. Battino, "Strawberry and human health: Effects beyond antioxidant activity," *J. Agric. Food Chem.*, vol. 62, no. 18, pp. 3867–3876, 2014. <https://doi.org/10.1021/jf405455n>
- [5] S. Yamamoto, S. Hayashi, H. Yoshida, and K. Kobayashi, "Development of a stationary robotic strawberry harvester with a picking mechanism that approaches the target fruit from below," *JARQ*, vol. 48, no. 3, pp. 261–269, 2014. <https://doi.org/10.6090/jarq.48.261>
- [6] N. Hernandez-Martinez, C. Blanchard, D. Wells, and M. Salazar-Gutierrez, "Current state and future perspectives of commercial strawberry production: A review," *Sci. Hortic.*, vol. 312, art. 111893, 2023. <https://doi.org/10.1016/j.scienta.2023.111893>
- [7] Ministerio de Agricultura y Desarrollo Rural, "Subsector Productivo de la Fresa," Dirección de Cadenas Agrícolas y Forestales, Bogotá, Colombia, Mar. 2021. [Online]. Available: <https://sioc.minagricultura.gov.co/Fresa/Documentos/2021-03-31%20Cifras%20Sectoriales.pdf>
- [8] S. Aliasgarian, M. Moghaddam, G. Ghanbarian, y H. Saryazdi, "Mechanical damage of strawberry during harvest and postharvest operations," *Acta Technol. Agric.*, vol. 18, no. 3, pp. 65–70, 2015, doi: <https://doi.org/10.1515/ata-2015-0001>
- [9] R. Rahmadian and M. Widyartono, "Harvesting system for autonomous robotic in agriculture," *INAJEEE*, vol. 2, no. 1, pp. 1–6, 2019. <https://doi.org/10.26740/inajeee.v2n1.p1-6>
- [10] E. Navas, R. Shamshiri, V. Dworak, C. Weltzien, and R. Fernandez, "Soft gripper for small fruits harvesting and pick and place operations," *Front. Robot. AI*, vol. 10, 2024. <https://doi.org/10.3389/frobt.2023.1330496>
- [11] T. U. Sane and T. U. Sane, "Artificial intelligence and deep learning applications in crop harvesting robots: A survey," in *Proc. Int. Conf. Electr., Commun., Comput. Eng. (ICECCE)*, 2021, pp. 1–6. <https://doi.org/10.1109/ICECCE52056.2021.9514232>
- [12] J. M. Pearce, "Optimizing strawberry disease and quality detection with vision transformers and attention-based convolutional neural networks," *Foods*, vol. 13, no. 12, art. no. 1869, Jun. 2024, doi: <https://doi.org/10.3390/foods13121869>
- [13] C. Pardo, C. Parra, and L. Solaque, "Mapeo del rendimiento de cosecha mediante sistema de detección y conteo de fresas," in *Proc. EIEI ACOFI 2024*, 2024. <https://doi.org/10.26507/paper.3524>
- [14] "Strawberry Maturity Dataset," *Kaggle*, 2023. [Online]. Available: <https://www.kaggle.com/datasets/yihfeng/strawberry-maturity>. [Accessed: Apr. 21, 2026].
- [15] "Strawberry Dataset," *Kaggle*, 2023. [Online]. Available: <https://www.kaggle.com/datasets/abdulbasit31/strawberry-dataset>. [Accessed: Apr. 21, 2026].
- [16] "Fruits and Vegetables Dataset," *Kaggle*, 2024. [Online]. Available: <https://www.kaggle.com/datasets/muhriddinmukhammad/fruit-and-vegetables-dataset>. [Accessed: Apr. 21, 2026].