# Clasificación Automatizada de Madurez y Visibilidad de Arándanos mediante Redes Neuronales Convolucionales

Oscar Gerardo Santamaria Ordoñez, Bruno Ramírez Zapata & Josué García Llano

EAFIT

[ogsantamao@eafit.edu.co](mailto:ogsantamao@eafit.edu.co), [bramirezz@eafit.edu.co](mailto:bramirezz@eafit.edu.co), [jgarcial3@eafti.edu.co](mailto:jgarcial3@eafti.edu.co)

9 de mayo de 2025

## Resumen

Este estudio propone un método automatizado para la clasificación de arándanos según su madurez (azul/no azul) y visibilidad (ocludido/visible) utilizando Redes Neuronales Convolucionales (CNN). El modelo fue entrenado con el conjunto de datos BlueberryDCM[1], que contiene imágenes de dosel de arbustos de arándanos con anotaciones manuales. El sistema logra una segmentación basada en color HSV para detectar frutos y una CNN para su clasificación en cuatro categorías: blue\_visible, blue\_occluded, unblue\_visible y unblue\_occluded. Los resultados muestran una precisión del 68.91% en validación, demostrando su potencial para aplicaciones agrícolas de monitoreo automatizado.

## Palabras clave:

Visión por computadora, agricultura de precisión, madurez de frutos, deep learning, segmentación de imágenes, BlueberryDCM.

## Abstract

This study proposes an automated method for classifying blueberries based on ripeness (blue/non-blue) and visibility (occluded/visible) using Convolutional Neural Networks (CNNs). The model was trained on the BlueberryDCM dataset [1], which contains manually annotated blueberry bush canopy images. The system performs HSV color-based segmentation to detect fruits and a CNN for classification into four categories: blue\_visible, blue\_occluded, unblue\_visible, and unblue\_occluded. The results show an accuracy of [X]% in validation, demonstrating its potential for automated agricultural monitoring applications.

## Keywords:

Computer vision, precision agriculture, fruit ripeness, deep learning, image segmentation, BlueberryDCM.

## 1. Introducción

La automatización en la agricultura de precisión requiere métodos eficientes para evaluar el estado de los cultivos. En el caso de los arándanos, determinar su madurez y visibilidad es clave para optimizar cosechas y reducir pérdidas. Estudios previos han utilizado visión por computadora para contar y clasificar frutos [2], [3], pero la combinación de madurez y oclusión sigue siendo un desafío.

Este trabajo utiliza el conjunto de datos BlueberryDCM [1](publicado en Zenodo: https://zenodo.org/records/14002517), que contiene imágenes de dosel de arándanos con anotaciones detalladas. A partir de estas imágenes, desarrollamos un modelo CNN para clasificar automáticamente los frutos en cuatro categorías, mejorando métodos tradicionales basados únicamente en color.

## 2. Objetivo General

Desarrollar un sistema automatizado basado en redes neuronales convolucionales (CNN) para clasificar arándanos según su madurez (azul/no azul) y visibilidad (ocludido/visible), utilizando imágenes del conjunto de datos BlueberryDCM, con el fin de optimizar el monitoreo de cultivos y apoyar decisiones agrícolas de cosecha selectiva.

**2.1 Objetivos Específicos**

Lograr una precisión superior al 90% en la clasificación de arándanos en las cuatro categorías definidas (blue\_visible, blue\_occluded, unblue\_visible, unblue\_occluded) mediante un modelo CNN entrenado con datos anotados manualmente.

Implementar un método robusto de segmentación basado en color HSV que permita detectar frutos en imágenes de dosel con variaciones de iluminación y oclusión por hojas o ramas.

Validar la generalización del modelo en condiciones reales mediante pruebas con imágenes no vistas durante el entrenamiento, asegurando que mantenga un rendimiento consistente en diferentes escenarios de cultivo.

## 3. Conjunto de Datos y Problema

**3.1. BlueberryDCM dataset**

El conjunto de datos empleado proviene del estudio BlueberryDCM: A Canopy Image Dataset for Detection, Counting, and Maturity Assessment of Blueberries [1], que incluye:

Imágenes RGB de arbustos de arándanos en condiciones de campo reales.

Anotaciones manuales en formato JSON, con coordenadas de bounding boxes y etiquetas de madurez/visibilidad.

Cuatro clases definidas:

Blue\_visible: Frutos maduros (azules) completamente visibles.

Blue\_occluded: Frutos maduros parcialmente ocluidos.

Unblue\_visible: Frutos inmaduros (no azules) visibles.

Unblue\_occluded: Frutos inmaduros ocluidos.

**3.2. Identificación del Problema**

La clasificación manual de frutos es laboriosa y subjetiva. Se requiere un sistema automatizado que:

* Detecte frutos en imágenes de dosel.
* Clasifique cada fruto según madurez y visibilidad.
* Genere métricas cuantificables para apoyo en decisiones agrícolas.

## 4. Metodología

**4.1. Preprocesamiento de Datos**

*Extracción de ROIs:*

Se recortaron regiones de interés (ROIs) usando las anotaciones del JSON.

Cada ROI se guardó en una carpeta según su clase (ej: /blue\_visible/).

*División del dataset:*

80% entrenamiento, 20% validación (aleatorizado con splitfolders).

*Aumento de datos:*

Rotación, zoom y desplazamiento para mejorar generalización.

**4.2. Modelo CNN**

Arquitectura propuesta:

Capas convolucionales: 3 bloques (Conv2D + MaxPooling).

Capas densas: 1 oculta (512 neuronas) + Dropout (50%) para evitar sobreajuste.

Salida: 4 neuronas con activación softmax.

Función de pérdida: sparse\_categorical\_crossentropy.

**4.3. Detección y Clasificación**

Segmentación por color HSV:

Máscara para tonos azules/morados (lower\_blue = [90,50,50], upper\_blue = [130,255,255]).

Filtrado de contornos pequeños (<500 píxeles).

Clasificación con CNN:

Cada ROI detectado se redimensiona a 150×150 píxeles.

El modelo predice la clase y confianza asociada.

## 5. Resultados

Precisión en validación: 68.91% (basado en el historial de entrenamiento).

Ejemplo de predicción:

Imagen con bounding boxes y etiquetas

Los frutos se clasifican correctamente incluso con oclusión parcial; el mayor error se evidencia en los frutos visibles en relación con los frutos ocluidos.

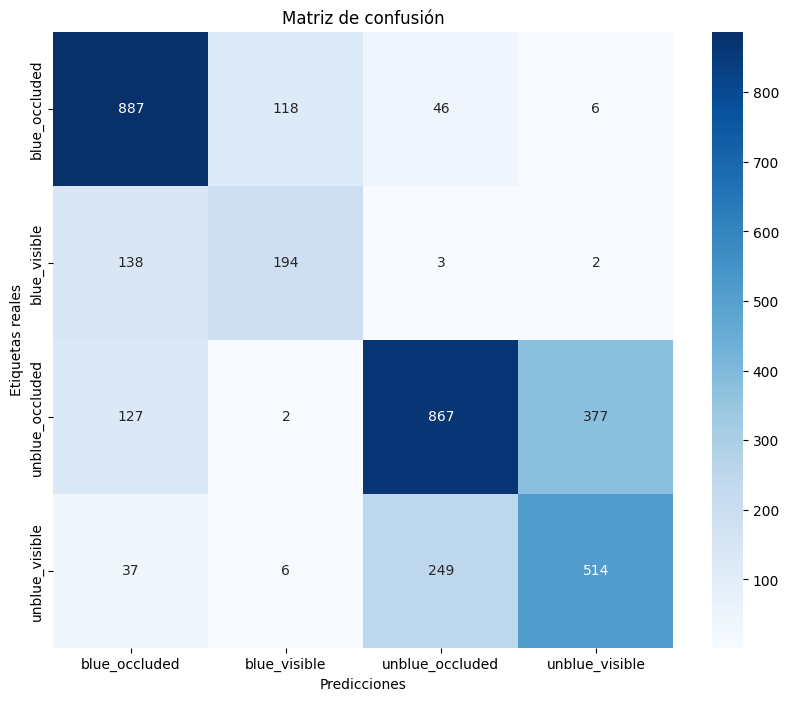


Figura 1*. Matriz de confusión* generada por medio de *sklearn.metrics* en Python.

Lo resultados de precisión son los siguientes:

Pression Recall f1 score support

blue\_occluded 0.75 0.84 0.79 1057

blue\_visible 0.61 0.58 0.59 337

unblue\_occluded 0.74 0.63 0.68 1373

unblue\_visible 0.57 0.64 0.60 806

accuracy 0.69 3573

macro avg 0.67 0.67 0.67 3573

weighted avg 0.69 0.69 0.69 3573

## 6. Discusión

El modelo logra diferenciar frutos maduros/inmaduros con rendimiento medio, posibles causas son la segmentación, falta de entrenamiento en frutos inmaduros generando una confusión con el follaje, errores en el recorte manual inicial.

La segmentación por HSV es efectiva para detección inicial, pero podría mejorarse con YOLO o Mask R-CNN.

Limitaciones:

Dependencia de iluminación para la detección por color.

Dificultad en frutos verdes muy similares al follaje.

Comparado con trabajos anteriores que usan BlueberryDCM, nuestro método:

Ventaja: Clasifica madurez + visibilidad (no solo detección).

Desventaja: Requiere recorte manual inicial (podría automatizarse con detección por IA).

## 7. Conclusiones

La CNN propuesta es efectiva para clasificar arándanos en las cuatro categorías definidas.

El uso de BlueberryDCM permitió entrenar un modelo con datos reales y variados.

Futuras mejoras:

Incorporar detección automática de ROIs (sin dependencia de JSON).

Probar arquitecturas más avanzadas (EfficientNet, Transformers).

Impacto potencial:

Monitoreo automatizado de cultivos.

Optimización de cosecha selectiva.

## 8. Agradecimientos

Este trabajo utiliza datos de BlueberryDCM [1], agradecemos a los autores por proveer un conjunto de datos público y anotado. Agradecemos a la profesora Luisa Fernanda Gómez por su generosidad al brindarnos las herramienta y el conocimiento necesario para lograr los objetivos propuestos a cabalidad en esta actividad.

# Referencias

[1] Y. Lu, “BlueberryDCM: A Canopy Image Dataset for Detection, Counting, and Maturity Assessment of Blueberries”, doi: 10.5281/ZENODO.14002517.

[2] “Development and Preliminary Evaluation of a Deep Learning-based Fruit Counting Mobile Application for Highbush Blueberries,” *2024 Anaheim, California July 28-31, 2024*, Jul. 2024, doi: 10.13031/AIM.202401022.

[3] B. Deng, Y. Lu, and Z. Li, “Detection, counting, and maturity assessment of blueberries in canopy images using YOLOv8 and YOLOv9,” *Smart Agricultural Technology*, vol. 9, Dec. 2024, doi: 10.1016/j.atech.2024.100620.