

Un método no invasivo para la clasificación de manzanas

Juan Carlos Olguín-Rojas^{1,3}, J. Irving Vasquez-Gomez^{1,2},
Gilberto de Jesus Lopez-Canteñs³, Juan Carlos Herrera-Lozada¹

¹ Instituto Politécnico Nacional,
Centro de Innovación y Desarrollo Tecnológico en Cómputo,
México

² Consejo Nacional de Ciencia y Tecnología,
Ciudad de México,
México

³ Universidad Autónoma Chapingo,
Departamento de Ingeniería Mecánica Agrícola,
México

olguinr.juancarlos@gmail.com

Resumen. Actualmente, en muchas empresas agroindustriales de México, la clasificación de manzanas se hace de forma manual lo que en ocasiones genera deficiencias en la calidad del producto. Estos problemas podrían resolverse o reducirse con la implementación de equipos de visión en sitio equipados con algoritmos inteligentes. En este trabajo presentamos como objeto de estudio la clasificación de manzanas Red Delicious, Granny Smith, Golden Delicious y Gala, mediante algoritmos de inteligencia artificial. Comparamos la eficiencia de dos arquitecturas, la primera es una red neuronal convolucional (CNN) Lenet5, y la segunda es la red convolucional VGG16 a la cual se le aplicaron técnicas de transferencia de aprendizaje de entrenamientos previos con Imagenet que contiene 1.4 millones de imágenes con 1,000 clases diferentes, presentamos para nuestros experimentos un conjunto de datos de cuatro categorías de manzanas que en total suman 2,400 imágenes con dimensión de 800x600 píxeles cada una, el 70 % de las imágenes de manzanas fueron para entrenamiento, el 15 % para validación y 15 % para prueba. La red que mejor resultados entrego fue VGG16 con la técnica de entonación fina (fine-tuning) se reentrenaron 8,130,564 parámetros de un total de 15,765,828 y su precisión fue de 99 %.

Palabras clave: Clasificación, transferencia de aprendizaje, manzanas.

A Non-Invasive Method for Apple Classification

Abstract. Currently, in many agribusiness in Mexico, the classification of apples is done manually, which sometimes generates deficiencies in the quality of the product. These problems could be solved or reduced with the implementation of on-site vision equipment based on intelligent algorithms.

In this work, we present as an object of study the classification of Red Delicious, Granny Smith, Golden Delicious and Gala apples, using artificial intelligence algorithms. We compared the efficiency of two architectures, the first is a Lenet5 convolutional neural network (CNN), and the second is the CNN VGG16 network. Both networks were set using transfer learning techniques from previous trainings with Imagenet were applied that contains 1.4 million images with 1000 different classes. We present for our experiments a set of four categories of apples that in total add up to 2,400 images with a dimension of 800×600 pixels each, 70% of the images of apples were for training, 15% for validation and 15% for testing. The network that gave the best results was VGG16 with fine-tuning, 8,130,564 parameters were retrained out of a total of 15,765,828 and its precision was 99%.

Keywords: Classification, transfer learning, apples.

1. Introducción

En México la manzana se produce en más de 10 entidades federativas, siendo el estado de Chihuahua el principal productor de la zona noreste con aproximadamente 569,000 toneladas al año [9]. Dentro del proceso de la cosecha y postcosecha de manzanas, la correcta clasificación es fundamental ya que los frutos son catalogados por su grado de maduración y los precios de mercado están determinados por dichas inspecciones [4].

Actualmente, existen problemas en el campo mexicano referidos con elementos de mediciones convencionales basados en técnicas físico-químicas para determinar el grado de madurez o de inocuidad de un fruto, principalmente porque estas técnicas requieren instrumental de laboratorio; esto ocasiona que en muchos casos los productores nacionales implementen líneas de inspección visual en donde personas son entrenadas para identificar: clase, grado de madurez, texturas y tamaños de frutos principalmente, sin embargo estos métodos son subjetivos y dificultan la inspección de grandes lotes del fruto o análisis en masa.

Una mala clasificación en postcosecha de la manzana podría evitar cumplir con los estándares de calidad internacional. México presenta en la actualidad un gran déficit comercial en la exportación de manzanas, pues en 2018 se importaron 282,756 toneladas con un valor de 264 millones de dólares y se exportaron sólo 766 toneladas con un valor de 1.1 millones de dólares, además, la variación en porcentaje de las exportaciones del fruto en 2018 redujo 17.7% con respecto a las exportaciones de 2017 [9].

Nuestro objetivo en este trabajo es desarrollar un sistema de visión en sitio para comparar algunas arquitecturas de aprendizaje profundo y evaluar su rendimiento en clasificación, con la finalidad de clasificar las variedades de manzana: Red Delicious, Granny Smith, Golden Delicious y Gala; que permitan como consecuencia un mejor manejo del fruto en postcosecha o incluso posiblemente en puntos de venta.



Fig. 1. Banda transportadora para captura de imágenes.

2. Trabajo relacionado

El trabajo de Mohammadi en [1] informa que desarrolló un sistema electrónico que puede clasificar las frutas en masa sobre una banda transportadora con un gramo de precisión, analiza los requisitos y desarrollos de hardware y software para sistemas de visión artificial, con énfasis en imágenes monocromas, imágenes en color e imágenes multiespectrales, además, su desarrollo clasifica por color, por hematomas, forma y densidad; informa que su clasificador permite la detección de enfermedades, defectos y contaminación en frutas y verduras.

Kondo y Ting en [5] presentan un estudio del estado del arte para describir la configuración básica de la adquisición de datos, incluidos el color, la masa y el tamaño. En este trabajo los autores consideran que los componentes de una concepción simple deberían contener un software confiable para reenviar el producto a los canales adecuados de clasificación, y reportan que existen equipos que pueden clasificar hasta 10 frutas por segundo.

Los instrumentos de clasificación mejoran los tiempos de clasificación e inspección sanitaria. Bhargava en [2] afirma que la clasificación simultánea de frutas por tamaño y color ahorraría el tiempo de inspección sanitaria, reduciendo significativamente el manejo de la fruta. En el trabajo de Sun [12] se considera que un sistema de clasificación automática de manzanas debe involucrar los aspectos de: color, peso, dimensión y defectos.



Fig. 2. Ejemplos del conjunto Gala.



Fig. 3. Ejemplos del conjunto Golden Delicious.

El autor destaca que el brillo superficial de una manzana refleja la frescura y sus defectos, considerando a esta característica sensorial entre las más importantes [12], en la literatura especializada también se reporta el trabajo de Sofu Er [11], donde se reporta que encontraron que la característica que más afecta la calidad de la clasificación de la manzana es la mancha y la descomposición.

El trabajo de Duran Melo en [3] reporta que lograron clasificar manzanas tomando en cuenta características de su superficie, basándose en un preprocesamiento y segmentación de la imagen para extraer las características de interés para la clasificación. Reporta que, extrajeron 14 características de imágenes de manzanas Gala, tales como: niveles de rojo y verde, contornos internos en la manzana, perímetro del contorno principal, y posteriormente entrenaron una red Perceptrón Multicapa (MLP) para clasificar las manzanas en tres categorías con una exactitud del 88.33 %.

El trabajo de Moallem [8], es un buen referente en cuanto a técnicas de clasificación de manzanas, pues reportan que comparan diferentes modelos de aprendizaje automático, extrajeron características estadísticas, texturales y geométricas de imágenes de manzanas Golden Delicious, y usaron clasificadores tipo SVM (Maquinas de Soporte), MLP (Perceptrón Multicapa) y KNN (K-Vecinos Cercanos) las clasificaron las manzanas en saludables y desértadas, con una precisión del 92.5 %.

El trabajo de Valdez en [13] informa que utilizó dos arquitecturas de aprendizaje profundo para clasificar y detectar daños en manzanas, una de las arquitecturas fue YOLOV3 y la otra fue un método alternativo de detección de simple disparo (SSD) propuesto por Liu et al. [7].



Fig. 4. Ejemplos del conjunto Granny Smith.



Fig. 5. Ejemplos del conjunto Red Delicious.

3. Clasificación de manzanas

Para lograr la clasificación de manzanas con métodos no invasivos, en este trabajo se propone utilizar un sistema de visión en sitio, comparando la eficiencia de dos arquitecturas de redes neuronales convolucionales; la primera es la red Lenet5 que está basada en la arquitectura propuesta por Yann LeCun et al. [6] y la segunda es la red VGG16 [10] en la cual se utilizan técnicas de transferencia de aprendizaje, específicamente, extracción de características (feature-extraction) y entonación fina (fine-tuning).

Para la transferencia de aprendizaje se utilizó la arquitectura VGG16 que fue entrenada con Imagenet y que contiene 1.4 millones de imágenes con 1000 clases diferentes. Los entrenamientos se realizaron utilizando la plataforma Google Colab [14] en Python 3, con librerías de Keras y tensorflow versión 2.4.1.

3.1. Conjunto de datos

Una de las tareas fundamentales para el desarrollo de este trabajo, fue la creación de la base de datos de imágenes de manzanas de las categorías de nuestro interés, para lograr esta importante tarea, se diseñó y construyó una banda transportadora que tiene iluminación controlada y materiales adecuados para el manejo de alimentos, ya que son de fácil limpieza y no guardan microorganismos. Véase Fig. 1.

El conjunto de datos está compuesto por cuatro categorías de manzanas que en total suman 2,400 imágenes con dimensión de 800×600 píxeles cada una, el 70 % de las imágenes de manzanas se utilizaron para entrenamiento, el 15 % para validación y 15 % para prueba, los conjuntos están organizados de la siguiente manera: 600 imágenes de manzana Gala (ver Fig. 2), 600 imágenes de manzana Golden Delicious (ver Fig. 3), 600 imágenes de manzana Granny Smith (ver Fig. 4) y 600 imágenes de manzana Red Delicious (ver Fig. 5).

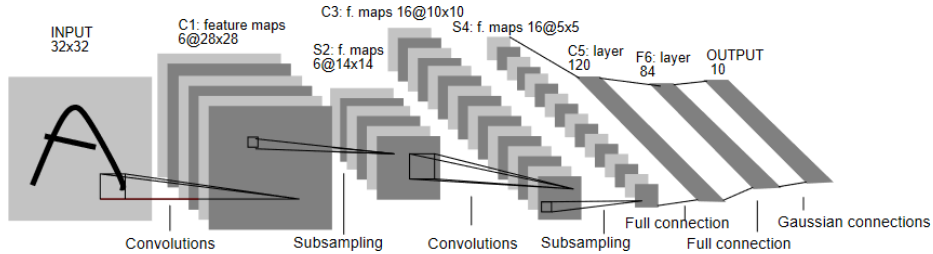


Fig. 6. Muestra los bloques que componen a la arquitectura LeNet5.

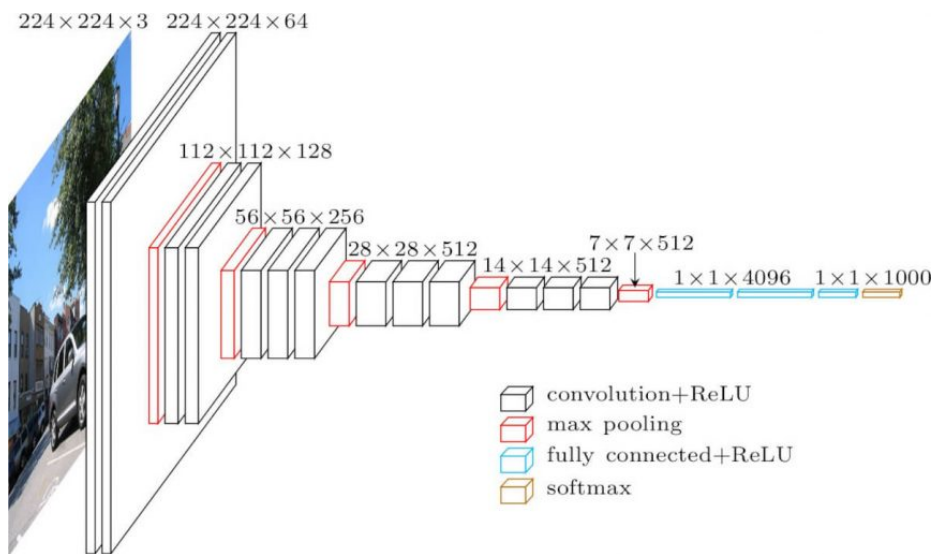


Fig. 7. Muestra la arquitectura VGG16 que participó en el ILSVRC-2014.

3.2. Arquitectura de aprendizaje profundo LeNet5 y VGG16

La arquitectura de red neuronal convolucional LeNet5 de Yann Lecun [6]. Se entrenó específicamente con el conjunto de datos MNIST para el problema de reconocimiento de caracteres a mano, logrando una precisión de clasificación de aproximadamente 99.2%. Esta arquitectura consta de 2 capas convolucionales, 2 capas de agrupación máxima (max pooling) y 3 capas completamente conectadas, véase Fig. 6. Por otra parte, la red neuronal convolucional VGG16 fue propuesta por K. Simonyan y A. Zisserman [10].

Esta arquitectura participó en el ILSVRC-2014 (ImageNet Large-Scale Visual Recognition Challenge 2014) y alcanzó una precisión del 92.7% quedando entre los 5 primeros en ImageNet, que es un conjunto de datos de más de 14 millones de imágenes que pertenecen a 1000 clases. La arquitectura VGG16 tiene cinco bloques convolucionales, cada bloque convolucional es seguido por una capa de agrupación y finalmente se tienen tres capas densas.

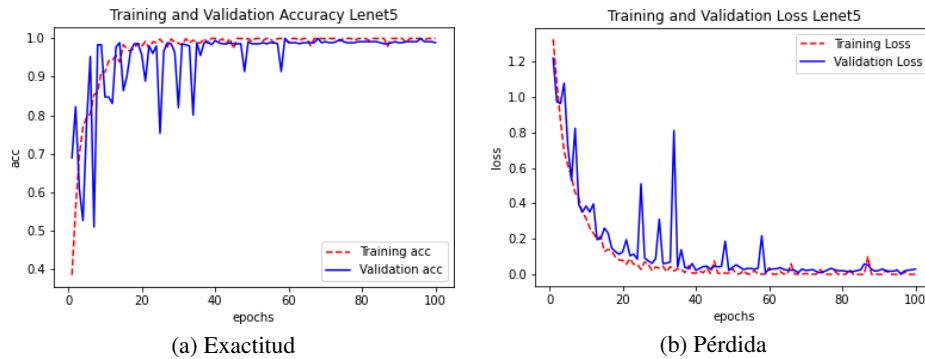


Fig. 8. Resultados de la exactitud y la pérdida usando la arquitectura LeNet5.

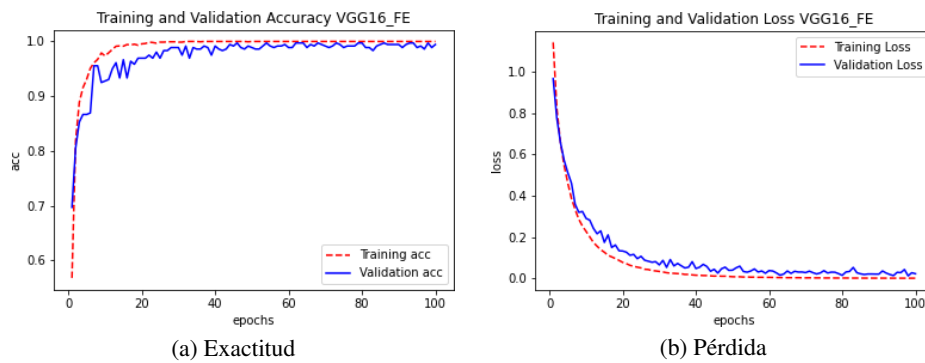


Fig. 9. Resultados de la exactitud y la pérdida usando la arquitectura VGG16 con extracción de características.

La arquitectura VGG16 que participó en el ILSVRC-2014 utilizó imágenes RGB de tamaño fijo de 224 x 224 píxeles. La imagen pasa a través de las capas convolucionales y los filtros que se usaron son de tamaño 3x3, en el caso de la agrupación máxima (max pooling) el tamaño es 2x2, véase Fig. 7.

4. Experimentación

En esta sección se describen los resultados de los entrenamientos de las arquitecturas utilizadas, la comparación de los resultados de Accuracy (exactitud) de las tres técnicas se muestra en la Fig. 11. A continuación, se presentan por separado cada una de las técnicas utilizadas.

4.1. Arquitectura de aprendizaje profundo Lenet5

Para este experimento se utiliza la arquitectura Lenet5, a la cual se le presenta el conjunto de imágenes reescaladas a 64×64 píxeles cada una. Los resultados del entrenamiento de la red Lenet5 en 100 épocas se muestran en Fig. 8.

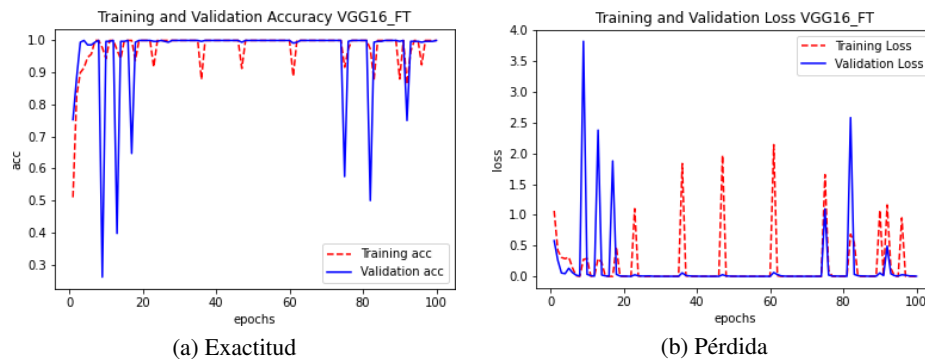


Fig. 10. Resultados de la exactitud y la pérdida usando la arquitectura VGG16 con ajuste fino.

4.2. Transferencia de aprendizaje con extracción de características (feature extraction) en VGG16

Para este experimento se utiliza la arquitectura VGG16 que fue entrenada con Imagenet y que contiene 1.4 millones de imágenes con 1000 clases diferentes, se usa transferencia de aprendizaje con la técnica de extracción de características (feature extraction) para presentar el conjunto de imágenes de manzanas, esto significa, que sólo se reentrenó la parte final de la arquitectura. Los resultados del entrenamiento de la red VGG16 con el método de extracción de características, entrenando 100 épocas se muestran en Fig. 9.

4.3. Transferencia de aprendizaje con entonación fina (fine tuning) en VGG16

Para este experimento se continúa utilizando la arquitectura VGG16 que fue entrenada con Imagenet y que contiene 1.4 millones de imágenes con 1000 clases diferentes, se usa transferencia de aprendizaje con la técnica de entonación fina (fine tuning) para presentar el conjunto de imágenes de manzanas, esto significa que se reentrenó el bloque 5 de la arquitectura interna de VGG16 además de la parte final de la arquitectura. Los resultados del entrenamiento de la red VGG16 con el método de extracción de características, entrenando 100 épocas se muestran en Fig. 10.

4.4. Prueba del rendimiento en clasificación de las arquitecturas

Para evaluar el rendimiento en clasificación se usa la matriz de confusión, el conjunto de la prueba de validación está compuesto por un total de 360 manzanas de las 4 clases, es decir, 90 manzanas por clase. Los resultados de la matriz de confusión de la arquitectura Lenet5 se presentan en la Fig. 12(a).

Por otra parte, los resultados de la matriz de confusión de la arquitectura VGG16 con la técnica de extracción de características (FE) se presentan en la Fig. 12(b). Finalmente, los resultados de la matriz de confusión de la arquitectura VGG16 con la técnica de entonación fina (FT) se presentan en la Fig.12(c).

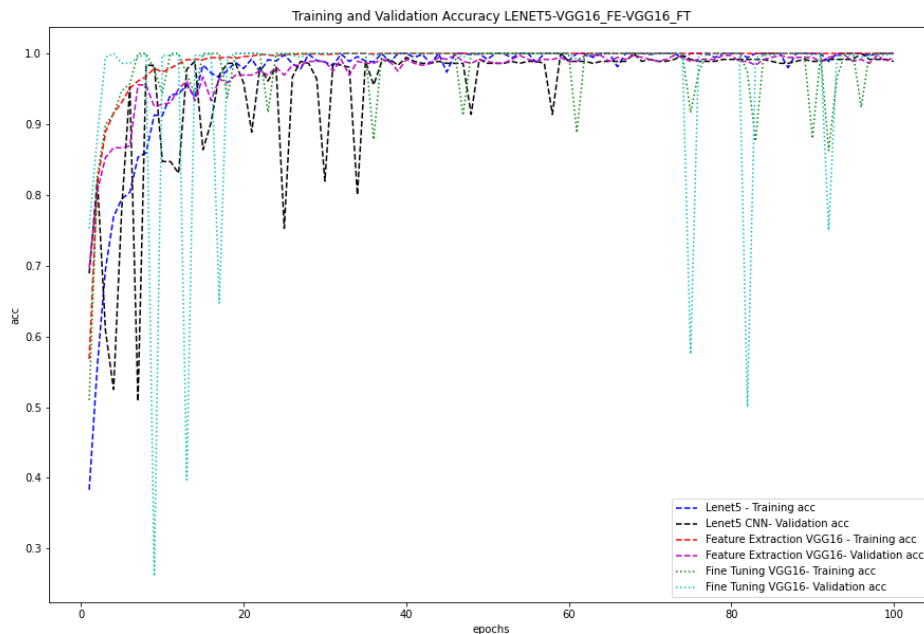


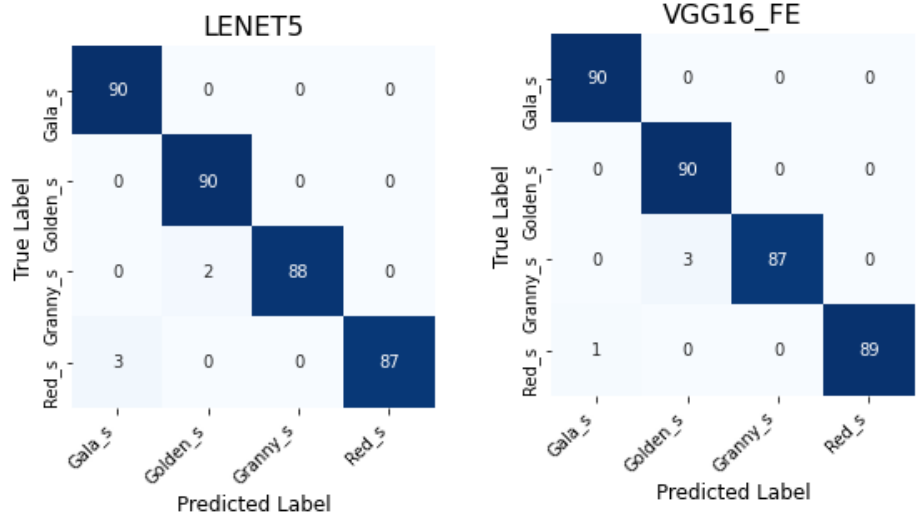
Fig. 11. Muestra los resultados de Accuracy (exactitud) en entrenamiento y en validación de las tres técnicas.

4.5. Análisis de resultados

La Fig. 18, muestra que la arquitectura VGG16 con la técnica de entonación fina (fine-tuning) tiene el mejor resultado para el rendimiento en clasificación, en la matriz de confusión se obtuvo una exactitud del 99 %.

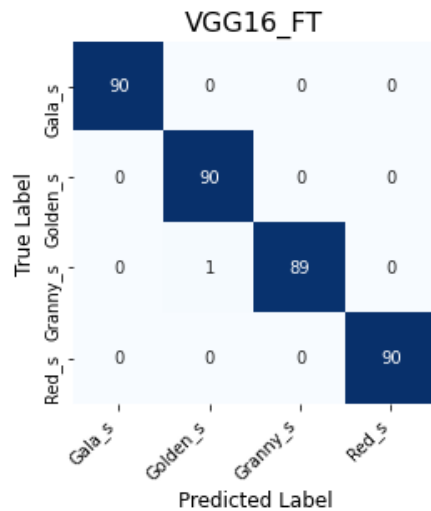
Se considera que este resultado es debido a que con esta técnica se hace un ajuste más fino y también se entrenan algunas de las capas finales de la base convolucional que se utiliza para la extracción de características, esto implica que con esta técnica se está realizando una entonación fina de las representaciones más abstractas del modelo que se está utilizando como base, esto porque las primeras capas aprenden características generales y gradualmente las capas sucesivas aprenden características más abstractas, por esta razón con esta técnica no se utiliza toda la base convolucional del entrenamiento previo, ya que, si el nuevo conjunto de datos (manzanas) difiere del conjunto de datos con el que se entrenó originalmente (Imagenet) la técnica solo utiliza las primeras capas, las cuales son congeladas para no ser entrenadas (freeze layers).

Aunque Lenet5 y VGG16 tienen la misma exactitud en las matrices de confusión (98 %), se debe considerar que el tamaño de las arquitecturas en cuestión de parámetros es muy distinto 2,471,512 y 15,765,828 respectivamente, esto quiere decir que si para la clasificación de manzanas, las condiciones de iluminación en el sistema de visión en sitio no cambian, entonces se recomienda utilizar la arquitectura Lenet5, ya que con esta arquitectura se podría tener un buen desempeño respecto al tiempo de clasificación incluso si se tratara de un sistema embebido con características de bajo costo computacional.



(a) Arquitectura LeNet5.

(b) Arquitectura VGG16 usando extracción de características.



(c) Arquitectura VGG16 usando ajuste fino.

Fig. 12. Matrices de confusión para cada una de las arquitecturas probadas.

5. Conclusiones y trabajo futuro

El rendimiento en clasificación fue de 98 % para Lenet5, 98 % para VGG16 con extracción de características y 99 % para VGG16 con entonación fina, sin embargo, se debe tomar en cuenta que, en la construcción del conjunto de datos de 2,400 manzanas, se utilizó condiciones de iluminación controlada, esto implica que para escenarios con iluminación variable los porcentajes de precisión alcanzados pueden variar.

Si las condiciones de iluminación en el sistema de visión en sitio no cambian, entonces se recomienda utilizar para la implementación la arquitectura Lenet5, esto porque la cantidad de parámetros de la red es mucho menor a la VGG16, pues se podría tener un buen desempeño respecto al tiempo de clasificación incluso si se tratara de un sistema embebido con características de bajo costo computacional.

Como trabajo a futuro, se analizarán daños y enfermedades propias de estas variedades de manzanas, además del análisis foliar de cultivares de manzanas para desarrollar propuestas de solución a los problemas que enfrenta el sector agrícola nacional.

Referencias

1. Baneh, N. M., Navid, H., Kafashan, J.: Mechatronic components in apple sorting machines with computer vision. *Journal of Food Measurement and Characterization*, vol. 12, no. 2, pp. 1135–1155 (2018) doi: 10.1007/s11694-018-9728-1
2. Bhargava, A., Bansal, A.: Fruits and vegetables quality evaluation using computer vision: A review. *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 3, pp. 243–257 (2018) doi: 10.1016/j.jksuci.2018.06.002
3. Durán, J., González, F.: Sistema automático clasificador de manzanas (2012)
4. Gutiérrez Rico, V.: Cosecha, postcosecha y comercialización de la manzana. Technical report, Consorcio de Coordinación en Salud Integral y el Centro de Multiservicios Educativos (2012)
5. Kondo, N., Ting, K.: Robotics in bioproduction systems. *The Society for Engineering in Agricultural, Food, and Biological Systems*, vol. 33, no. 29, pp. 1–11 (2000) doi: 10.1016/s1474-6670(17)36744-7
6. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. vol. 86, pp. 2278–2324 (1998) doi: 10.1109/5.726791
7. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A. C.: SSD: Single shot MultiBox detector. *Computer Vision-ECCV*, pp. 21–37 (2016) doi: 10.1007/978-3-319-46448-0_2
8. Moallem, P., Serajoddin, A., Pourghassem, H.: Computer vision-based apple grading for golden delicious apples based on surface features. *Information Processing in Agriculture*, vol. 4, no. 1, pp. 33–40 (2017) doi: 10.1016/j.inpa.2016.10.003
9. Servicio de Información Agroalimentaria y Pesquera: Panorama agroalimentario 2019. Technical report, SIAP (2018)
10. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014) doi: 10.48550/ARXIV.1409.1556
11. Sofu, M., Er, O., Kayacan, M., Cetişli, B.: Design of an automatic apple sorting system using machine vision. *Computers and Electronics in Agriculture*, vol. 127, pp. 395–405 (2016) doi: 10.1016/j.compag.2016.06.030
12. Sun, K., Li, Y., Peng, J., Tu, K., Pan, L.: Surface gloss evaluation of apples based on computer vision and support vector machine method. *Food Analytical Methods*, vol. 10, no. 8, pp. 2800–2806 (2017) doi: 10.1007/s12161-017-0849-7
13. Valdez, P.: Apple defect detection using deep learning based object detection for better post harvest handling (2020) doi: 10.48550/ARXIV.2005.06089
14. VVAA: Google colaboryatory (2021) <http://colab.research.google.com/ash>