Ingeniería Investigación y Tecnología
volumen XXIII (número 4), octubre-diciembre 2022 1-13
ISSN 2594-0732 FI-UNAM artículo arbitrado
Información del artículo: Recibido: 15 de octubre de 2021, reevaluado: 5 de mayo de 2022, aceptado: 6 de agosto de 2022

https://doi.org/10.22201/fi.25940732e.2022.23.4.031

Ingeniería Investigación y Tecnología

Integración de una cámara multiespectral y aprendizaje automático para clasificación de manzanas

Integration of a multispectral camera and machine learning for apple sorting

Gante-Díaz Saulo Abraham
Instituto Politécnico Nacional
Escuela Superior de Ingeniería Mecánica y Eléctrica unidad Zacatenco
Correo: sganted1500@alumno.ipn.mx
https://orcid.org/0000-0001-6012-4003

Lozano-Hernández Yair
Instituto Politécnico Nacional
Escuela Superior de Ingeniería Mecánica y Eléctrica unidad Zacatenco
Correo: ylozanoh@ipn.mx
https://orcid.org/0000-0001-8157-3510

Maldonado-Trinidad Marco Antonio
Escuela Superior de Ingeniería Mecánica y Eléctrica unidad Zacatenco
Instituto Politécnico Nacional
Correo: mmaldonadot1500@alumno.ipn.mx
https://orcid.org/0000-0001-7363-2904

Flores-Colunga Gerardo Ramón
Centro de Investigaciones en Óptica
Correo: gflores@cio.mx
https://orcid.org/0000-0003-2107-7405

Villegas-Piña Daniel
Escuela Superior de Ingeniería Mecánica y Eléctrica unidad Zacatenco
Instituto Politécnico Nacional
Correo: dvillegasp1501@alumno.ipn.mx
https://orcid.org/0000-0002-3237-415X

Resumen

Este trabajo presenta un sistema capaz de realizar una clasificación binaria (buena y podrida) de manzanas *red delicious*, lo cual se logra mediante el uso del Índice de Vegetación de Diferencia Normalizada (NDVI) y una Red Neural Xception. Para ello, se utiliza una cámara multiespectral para observar detalles fuera del alcance del ojo humano, reduciendo la presencia de errores durante su clasificación. La selección del NDVI es resultado de su comparación con los índices de vegetación GNDVI, GVI, NDRE, NDVIR, NG, NGRDI y RVI, aplicados a un banco de imágenes obtenido mediante la cámara multiespectral. Además, se muestran los resultados de la clasificación al utilizar redes neuronales Xception, ResNet y MobileNet, lo que justifica el uso de la red Xception. Finalmente se describe la instrumentación e iluminación empleada en un prototipo que emula el proceso de clasificación real utilizando una banda transportadora, lo que permite validar de forma experimental el sistema de clasificación propuesto, obteniéndose un 73 % de éxito en la clasificación en línea. El desarrollo de este trabajo se basa en la siguiente metodología: Se utiliza una cámara multiespectral para la creación de una base de datos, las imágenes obtenidas pasan a una etapa de procesamiento conformada por la alineación y reconstrucción RGB. Posteriormente, se utilizan dos o más bandas para el cálculo, comparación y análisis de diferentes índices de vegetación. Una vez determinado el índice de vegetación a utilizar, se procede al entrenamiento y comparativa entre distintas arquitecturas de redes neuronales. Respecto a la etapa de entrenamiento, se emplea transferencia de aprendizaje para reducir la necesidad de una gran base de datos. Por último, se realizan pruebas experimentales para validar el comportamiento.
Descriptores: Clasificación de manzanas, aprendizaje profundo, aprendizaje automático, cámara multiespectral, índice de vegetación.

Abstract

This paper presents a system capable of performing a binary classification (ripe and rotten) of red apples, which is achieved through the use of the Normalized Difference Vegetation Index (NDVI) and an Xception Neural Network. In order to carry out this process a multispectral camera is used to observe details outside the scope of the human eye, reducing the presence of errors during their classification. The selection of NDVI is the result of its comparison with the GNDVI, GVI, NDRE, NDVIR, NG, NGRDI, and RVI vegetation indices applied to an image bank obtained by means of the multispectral camera. Furthermore, classification results are displayed when using Xception, ResNet, and MobileNet neural networks, thus justifying the use of the Xception network. Finally, the instrumentation and lighting used in a prototype, which emulates the actual classification process using a conveyor belt, are described. This method allows to experimentally validate the proposed classification system, in this case, 73 % success in the online classification was achieved. The development of this work is based on the following methodology: a multispectral camera is used to create a database, the obtained images are sent to a processing stage, which involves alignment and RGB reconstruction. Subsequently, two or more bands are used for the calculation, comparison and analysis of different vegetation indices. Once the vegetation index is established, the training and comparison among different neural network architectures is carried out. Regarding the training stage, transfer learning is used to reduce the need for a large database. Finally, experimental tests are carried out to validate the behaviour.
Keywords: Apple sorting, deep learning, machine learning, multispectral camera, vegetation index.

https://doi.org/10.22201/fi.25940732e.2022.23.4.031

Introducción

Hoy en día, el almacenamiento de manzanas en cámaras de refrigeración suele presentar condiciones no favorables como el escaso mantenimiento del sistema de aire acondicionado y la recirculación de aire, lo cual permite a bacterias y hongos mantenerse y reproducirse, provocando pudrición (Paulus *et al.*, 1997). Otro problema frecuente radica en las contusiones debidas a un impacto, compresión, vibración o abrasión durante el manejo; los síntomas de estos daños (ennegrecimiento y ablandamiento del tejido) no aparecen inmediatamente (Baranowski *et al.*, 2013). Por lo cual, se requiere inspeccionar la calidad. Además, las exigencias de calidad son cada vez mayores debido a las normativas oficiales y petición de los mercados. Con base en lo anterior, existe un riesgo elevado de error humano en los procesos de clasificación manual; ya que las decisiones se ven afectadas por factores como fatiga o hábitos adquiridos.

Por lo anterior, se han realizado distintos estudios y prototipos para clasificación de manzanas, enfocándose la mayoría al diseño de algoritmos de visión artificial (Fan *et al.*, 2020; Tan *et al.*, 2018; Zhang *et al.*, 2015). En dichos sistemas, los algoritmos tienen como objetivo la clasificación y el procesamiento de imágenes mediante los cuales se detecta una zona de pudrición a través de su forma, textura o color; lo cual limita su uso en detecciones de pudriciones internas o de difícil apreciación por el ojo humano (Lu & Lu, 2018; Sofu *et al.*, 2016). Debido a la variedad de formas irregulares, colores y texturas; la clasificación de frutas y vegetales se ha convertido en un problema complejo en el área de *Machine Learning* (ML) (Hameed *et al.*, 2018); lo anterior debido a que representa un problema de naturaleza multidimensional con características hiperdimensionales.

Por ello, surge la necesidad de construir y diseñar sistemas de clasificación para control de calidad en menor tiempo, de manera automática, con menor empleo de personal y mayor exactitud. Lo cual se puede lograr mediante la implementación de Redes Neuronales (NN por sus siglas en inglés), generando mejores estándares de calidad y ahorro económico (Bhargava & Bansal, 2020; Dong & Guo, 2015).

En (Wang *et al.*, 2018), desarrollaron un algoritmo de clasificación fuera de línea para determinar si el arándano está o no en buenas condiciones; para ello, emplearon dos Redes Neuronales Convolucionales Profundas (DCNN por sus siglas en inglés) y datos de transmitancia hiperespectral. Las redes utilizadas fueron ResNet y ResNeXt, siendo la DCNN ResNeXt la que mayor precisión obtuvo (0.8952). En Steinbrener *et al.* (2019) se clasifican imágenes hiperespectrales de frutas y verduras a través de Redes Neuronales Convolucionales (CNN por sus siglas en inglés) previamente entrenadas con imágenes Rojo, Verde y Azul (RGB por sus siglas en inglés); para ello, se agregó una etapa de comprensión de datos a una CNN *ImageNet*, obteniendo una precisión fuera de línea de 92.23 %. Por su parte, en Singh & Singh (2019) se clasifican manzanas en podridas y buenas, siendo una Maquina de Soporte Vectorial (SVM por sus siglas en inglés) la de mayor rendimiento (98.9 %) al compararse con 5 diferentes clasificadores. En lo que a la clasificación de manzanas *Golden Delicious* se refiere, en el año 2017 se realizó un clasificador que indicaba si estaba sana o no; lo cual se hizo mediante una NN multicapa; el clasificador primero detecta la región del cáliz y después segmenta los defectos para indicar la pudrición, todo mediante binarización de imagen (Moallem *et al.*, 2017). Algunas herramientas que actualmente se utilizan para clasificación de manzanas son: sistemas difusos (Papageorgiou *et al.*, 2018), características espectrales (Zhang & Li, 2018), imágenes hiperespectrales (Folch *et al.*, 2016), ML (Heras, 2017), entre otras.

Los trabajos citados abordan la clasificación de frutas y verduras con base en segmentación, extracción de características y clasificación de imágenes. Los sistemas de visión RGB permiten extraer textura, forma, color y tamaño, pero algunos defectos son idénticos en textura y color, haciéndolos difíciles de detectar. Para ello se utiliza visión hiperespectral (Bhargava & Bansal, 2018). A pesar de lo anterior, la implementación en línea de sistemas basados en ML e imágenes multiespectrales sigue siendo un reto por vencer. En relación con lo anterior, en Fan *et al.* (2020) se utilizan CNNs para detectar manzanas defectuosas, la clasificación es en línea y logran una precisión de 92 %, usan 6 imágenes por cada manzana, cuentan con 2 cámaras RGB y una banda transportadora para girar las manzanas. En la práctica no siempre se cuenta con ese tipo equipo, incluso, no se tienen grandes bases de datos para el entrenamiento. Además, no existe una metodología general para el uso de imágenes multiespectrales y CNN. Por lo anterior, en este artículo se plantea una metodología para el desarrollo de un algoritmo para clasificación binaria de manzanas mediante el uso de una Cámara Multiespectal (CM). Para ello, se utiliza el Índice de Vegetación de Diferencia Normalizada (NDVI por sus siglas en inglés) y una red neuronal Xception (Xc), la cual se entrena mediante un banco de imágenes provenientes de la CM, permitiendo observar detalles fuera del alcance del ojo humano. Por último, se describe un prototipo que emula el proceso de clasificación en línea, obteniendo 73 % de éxito durante la clasificación en línea, lo anterior con solo 3 imágenes de la manzana y sin necesidad de girarla.

El trabajo se organiza de la siguiente manera: la segunda sección describe la obtención y procesamiento de imágenes provenientes de la CM, en esta misma sección se explican y comparan diversos Índices de Vegetación (VI) justificando el uso del NDVI. La sección tres presenta los criterios utilizados para el entrenamiento de diversas NNs, comparando su desempeño. La siguiente sección detalla la instrumentación de un prototipo funcional para clasificación en línea, que sirve como banco de pruebas para validar el sistema propuesto. Finalmente, la sección de conclusiones y resultados.

Obtención y procesamiento de imágenes

La presente sección, describe el proceso de obtención y procesamiento de imágenes provenientes de una CM. La cámara utilizada es la *MicasenseRededge-M*, la cual toma 5 capturas correspondientes a cada banda espectral: azul (B), verde (G), rojo (R), Infrarrojo Cercano (NIR) y Borde Rojo (RE). La comunicación entre la cámara y el computador se realiza vía ethernet y la programación mediante OpenCV (RedEdge, 2015).

Alineamiento y reconstrucción RGB

La CM presenta un desfase de posición en los lentes. Por lo tanto, se procede a realizar alineamiento de las 5 bandas, teniendo como objetivo una reconstrucción centrada de la manzana.

Alineamiento

Se realiza una transformación de perspectiva, mediante la cual se estima la orientación relativa dentro de dos imágenes. Para ello se utiliza una matriz de homografía $H \in \mathbb{R}^{3x3}$ (Escamilla, 2012).

$$\tilde{q} = sH\tilde{Q} \tag{1}$$

donde $\tilde{Q} = [X\ Y\ Z\ 1]^T$ corresponde a las coordenadas de la manzana, $\tilde{q} = [x\ y\ 1]^T$ es la representación en la imagen y s un parámetro de escala. $H = MW$, $W = [R\ t] \in \mathbb{R}^{3x3}$, contiene las rotaciones y traslaciones que relacionan el plano que se observa con el de la imagen. $M \in \mathbb{R}^{3x3}$ describe la proyección debida a los parámetros intrínsecos de la cámara (Lu & Cai, 2020).

$$H = \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

C_x y C_Y son el centro del sistema de coordenadas de la imagen, f_x y f_y corresponden a la distancia focal. Debido a que se trabaja con un objeto plano, $Z = 0$, obteniéndose:

$$\tilde{q} = sH\tilde{Q}' \tag{3}$$

con $\tilde{Q}' = [X\ Y\ 0\ 1]^T$ y $s = 1$. De (3) se observa un mapeo lineal entre los planos de la imagen, es decir, cada par de puntos proporciona dos ecuaciones independientes.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{4}$$

En resumen, la relación de las coordenadas de la manzana con su representación en la imagen está dada por (Ranganathan & Olson, 2014):

$$D_x = \frac{x}{1} = \frac{H_{11}X + H_{12}Y + H_{13}}{H_{31}X + H_{32}Y + H_{33}}$$
$$D_y = \frac{y}{1} = \frac{H_{21}X + H_{22}Y + H_{23}}{H_{31}X + H_{32}Y + H_{33}} \tag{5}$$

Para calcular la matriz de homografía se localizan 4 puntos correspondientes al contorno de la manzana, los cuales han sido establecidos de forma manual (Figura 1).

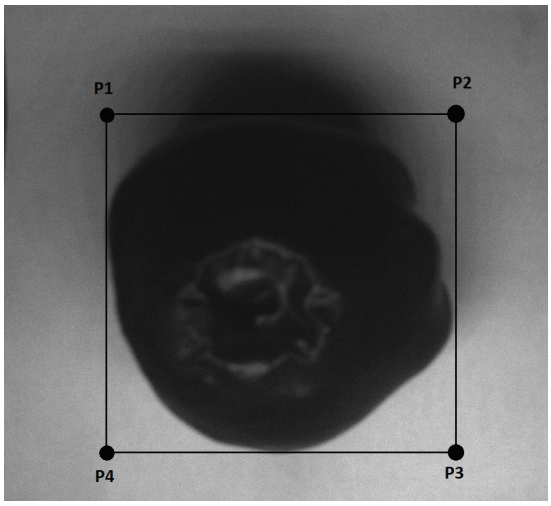

Figura 1. Selección de puntos requeridos para calcular matriz de homografía

Tomando puntos de la banda RE como puntos de destino, se obtiene la matriz de homografía respecto a una banda distinta. Posteriormente, se hace el alineamiento de la banda seleccionada con la de RE. Este procedimiento se realiza para todas las bandas.

https://doi.org/10.22201/fi.25940732e.2022.23.4.031

Reconstrucción RGB

Después del alineamiento se extraen los canales de la imagen. Cabe mencionar que la descomposición proporciona 3 canales. Una vez que se tienen imágenes de cada banda constituidas por un solo canal, se realiza la unión de las bandas B, G y R. Obteniéndose una imagen conformada por los valores de los 3 canales. La Figura 2 muestra el resultado del alineamiento y reconstrucción RGB.

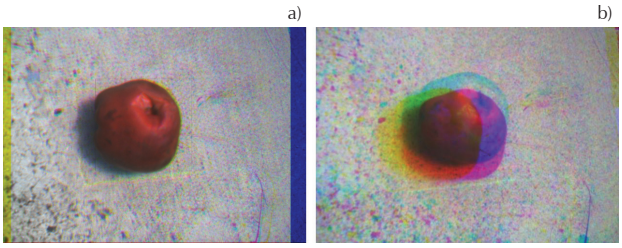

Figura 2. Resultados de la reconstrucción: a) RGB y b) Sobreposición de las imágenes

Índices de vegetación

Una vez concluido el alineamiento de imágenes se realizó la obtención del VI, el cual resulta de operaciones algebraicas entre 2 o más bandas espectrales (Ali *et al.*, 2017). Para esto, las imágenes deben estar conformadas por un canal y valores flotantes. Se experimentó con los VI descritos en la Tabla 1.

Tabla 1. Índices de vegetación (Giovos *et al.*, 2021).

Índice de vegetación	Fórmula
Diferencia Normalizada	$NDVI = NIR - R / NIR + G$
Diferencia Normalizada Verde	$GNDVI = NIR + V / NIR + G$
Proporción	$RVI = NIR / R$
Verde	$GV\ I = NIR / G$
Normalizado Rojo-Verde	$NGRDI\ G - R / G + R$

Cada píxel tiene un valor entre – 1 y 1. Para normalizar el VI de 0 a 255 se emplean las siguientes operaciones:

$$VI_2 = (VI_0)\ (255.0) \tag{6}$$

$$VI_{FIN} = \frac{VI_2 + 255.0}{2} \tag{7}$$

siendo VI_{FIN} la nueva imagen que contiene el *VI* final, VI_2 resulta de multiplicar VI_0 por 255.0 y VI_0 es el VI original.

La Figura 3 muestra los resultados de cada VI, tanto el NDVI como el GVI destacan mejor las pudriciones (Figuras 3a y 3d). Sin embargo, el GVI presenta menor contraste de la manzana, por lo cual se optó por utilizar el NDVI.

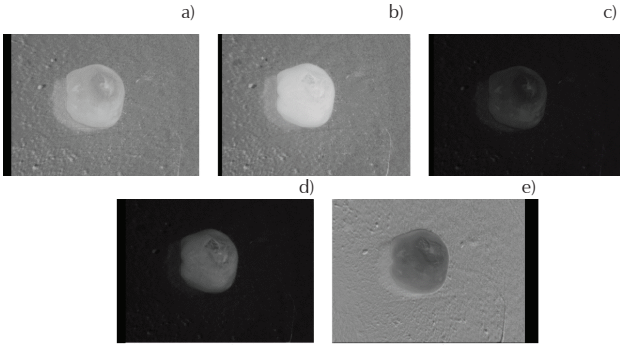

Figura 3. Resultados de la aplicación de índices de vegetación: a) NDVI, b) GNDVI, c) RVI, d) GVI, e) NGRDI

Iluminación

Para este trabajo se analizaron 4 tipos de iluminación: Luz ultravioleta (UV), LED, iluminación solar y halógena (Figura 4). La prueba con luz UV consta de un fondo color negro con el fin de contrastar la manzana; sin embargo, no se logran resaltar del todo las pudriciones (Figura 4a). La iluminación LED no presenta gran diferencia con la UV.

Por otra parte, la iluminación solar muestra mayor claridad para la obtención del NDVI, permitiendo observar con mayor facilidad las pudriciones. Lo anterior se debe a que este índice hace uso de la banda NIR, siendo esta la razón para su uso en la experimentación de los VI's y la creación de la base de datos.

La distribución espectral de la iluminación halógena se encuentra conteniendo en parte al NIR (entre 350 y 750nm) y favorece el cálculo del NDVI. Por lo anterior, en el prototipo se implementa iluminación halógena, la cual se posiciona de manera directa al objeto de estudio buscando reducir las sombras en el área de observación.

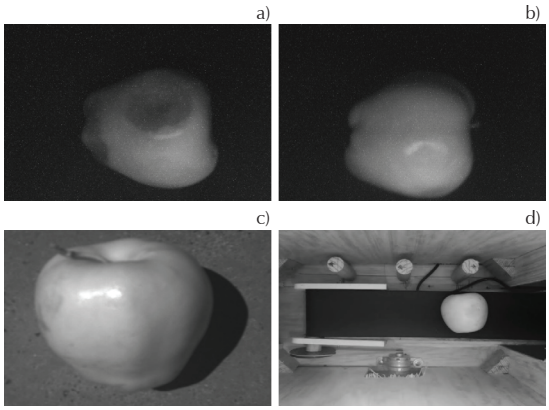

Figura 4. Pruebas de iluminación: a) luz UV, b) luz LED, c) iluminación solar d) iluminación halógena

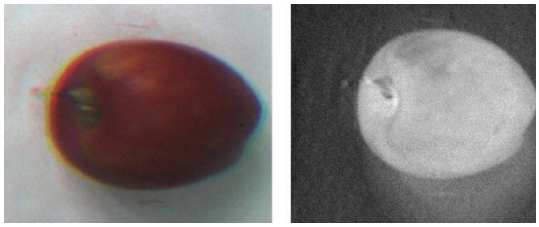
Figura 5. Manzana en buenas condiciones: a) RGB y b) NDVI

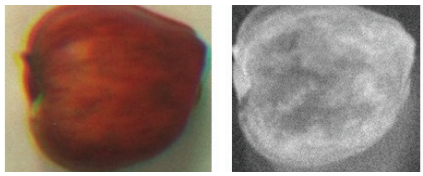
Figura 6. Manzana con pudriciones: a) RGB y b) NDVI

Obtención de base de datos

Para la obtención de la base de datos se creó un programa para calcular el NDVI y realizar reconstrucción RGB; posteriormente se guardó la información en carpetas.

A continuación se explica el algoritmo empleado:

- Lectura de archivos: Se extraen los nombres de los archivos contenidos en una carpeta (fotografías de la CM) y se ingresan a una variable.
- Ciclo for: Se encarga de repetir el proceso siguiente en función de la cantidad de archivos almacenados en la carpeta previamente inspeccionada.
- Declaración de variables: Se leen las imágenes correspondientes a las 5 bandas y se declaran 4 puntos que abarcan el contorno.
- Alineamiento de imágenes: Se realiza la alineación de las bandas B, G, R y NIR, con base en la banda RE.
- Reconstrucción RGB: Unión de las bandas B, G y R.
- Cálculo de NDVI: Se realiza la operación algebraica correspondiente.
- Almacenamiento de imagen RGB y NDVI: Una vez obtenidas las imágenes RGB y NDVI, se realiza un recorte de la imagen a fin de evitar la presencia de perturbaciones en los bordes durante el entrenamiento. Finalmente, son almacenadas.

A continuación, se muestran los resultados de alineamiento y reconstrucción RGB. La Figura 5 corresponde a una manzana sana, mientras que la manzana de la Figura 6 presenta áreas con pudrición difíciles de distinguir por el ojo humano. El algoritmo propuesto hace notar estas áreas.

Entrenamiento y comparativa de redes neuronales

Los métodos de segmentación y máscaras requieren que se dicten reglas para realizar una acción. Por el contrario, los métodos de ML requieren los datos y las respuestas para generar las reglas o algoritmos de relación entre estos (Xul *et al*., 2008).

Las arquitecturas de NN empleadas para clasificación de imágenes (Figura 7) están compuestas por una capa de entrada, capa de salida y capas intermedias u ocultas. Cada capa puede realizar distintas operaciones, las más comunes son:

1. Convolución: Aplica filtros convolucionales de distintos tipos para extraer características de las imágenes (*f*). El filtro convolucional se define con dimensiones (a x b) creando una matriz conocida como kernel (*k*), resultando en una imagen:

$$f \cdot k = g \tag{8}$$

   El filtro convolucional reduce las dimensiones de la imagen de entrada y extrae características según el tamaño del (*k*) utilizado.
2. Pooling: Se encarga de reducir las dimensiones de las salidas generadas. Esto se hace mediante la definición de un filtro de dimensiones n × n, el cual recorre la imagen proporcionada y evalúa subconuntos de pixeles para extraer valores máximos o promedios.

$$f \cdot pooling\ (n \times n) = \mathrm{g} \tag{9}$$

3. Fully connected: Esta operación asegura que una capa se encuentre completamente conectada con su sucesor y usualmente se emplea cuando los elementos de entrada son pocos en comparación con la entrada de la red.

https://doi.org/10.22201/fi.25940732e.2022.23.4.031

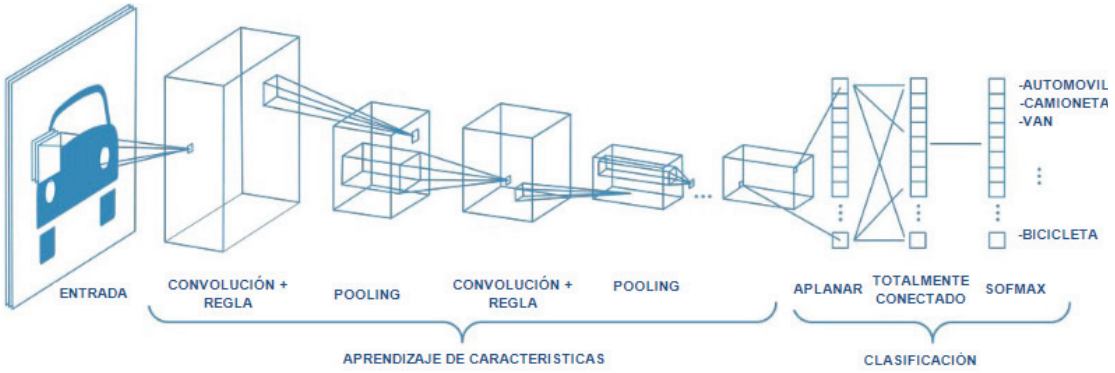

Figura 7. Ejemplo de NN con varias capas convolucionales. Se aplican filtros a cada imagen de entrenamiento con distintas resoluciones y la salida de cada imagen se emplea como entrada para la siguiente capa, tomada de (Mathworks, 2018)

Para este trabajo se empleó un conjunto de datos que contiene 816 imágenes con 425,400,3 píxeles. El banco de imágenes se dividió en dos conjuntos: entrenamiento y validación. El primer conjunto contine 705 elementos utilizados para la etapa de entrenamiento de la NN, mientras que el segundo se conforma de los 111 elementos restantes y es utilizado para la validación y pruebas en línea. Como se observa, la base de datos es pequeña, considerando trabajos como (Dargan *et al.*, 2019) y los avances en clasificación de imágenes mediante ML. En particular, *Deep Learning* (ImageNet o MNIST); realiza Transferencia de Aprendizaje (en adelante TA). Este método es utilizado cuando se cuenta con bases de datos pequeñas o es complicado/imposible recolectar datos para el entrenamiento y reconstrucción de modelos (Pan & Yang, 2009), permite utilizar el conocimiento que se ha generado para resolver una tarea que sea similar. (Hussain *et al.*, 2019) hacen uso de TA con el re-entrenamiento de una red neuronal pre entrenada con ImageNet sobre un conjunto de datos de 450 elementos, obteniendo 96.5 % de precisión en el desempeño de la nueva tarea. Shaha & Pawar, (2018) también reportan el uso de TA, lo cual les permite utilizar 80 elementos por cada clase obteniendo 88.88 % de precisión, destacando así los beneficios de esta técnica que parece no tener un límite mínimo de elementos.

La TA es un campo del aprendizaje profundo, tiene como fin el almacenamiento de conocimiento adquirido mientras se ha resuelto un problema, este desempeño se puede emplear en algún otro momento y adaptar para resolver otro problema. Se define como: sea un dominio (D_s) y tarea (T_s) fuentes, un dominio (D_t) y tarea (T_t) objetivos, TA de la función objetivo usando conocimiento en D_s y T_S, donde $D_s \neq D_S$, o $T_s \neq T_S$ (Lin & Jung, 2017), es decir, ambos dominios están relacionados.

Programación Red Neuronal

La programación y entrenamiento de las NNs se realizó mediante la librería *Keras* de Python, el algoritmo utilizado se describe en la Figura 8.

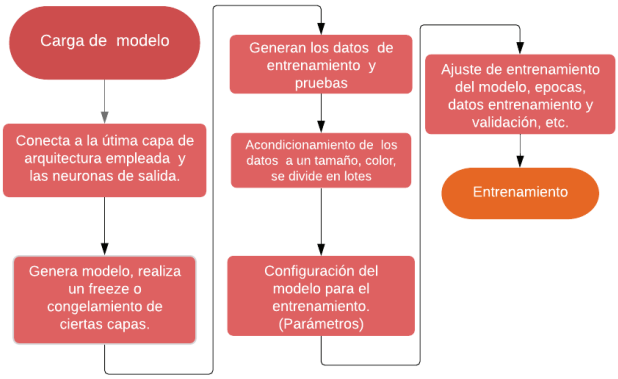

Figura 8. Etapas presentes durante el entrenamiento de las NNs

Carga y generación del modelo

Se hace uso de una metodología basada en TA, se importa el modelo a emplear, con el fin de acondicionarlo de acuerdo con las necesidades o requerimientos del problema en cuestión. Tomando en cuenta que la clasificación se dará en 2 clases, la salida de la arquitectura se conecta a un par de capas, las cuales acondicionan o reducen los datos, las características encontradas por las capas implementadas se conectan con una última capa (capa de salida) que contiene 2 neuronas. Una vez generada la arquitectura, se construye el modelo.

Congelamiento de capas

Para realizar la TA, se partió de un modelo *ImageNet* modificado para el problema en cuestión, el modelo contiene 135 capas previamente provistas con sus res-

pectivos pesos. Del modelo se "congelaron" 105 capas (pesos fijos, pre-entrenados), para la extracción de características se entrenó una cuarta parte de la red (capas 106 a 135), se realizó un *fine tuning* en las capas restantes donde no se han congelado los pesos. Así, se generan las clases requeridas para la clasificación.

Generación y acondicionamiento de los datos

Se emplean 705 imágenes para el entrenamiento y 111 para la validación. El conjunto de imágenes para el entrenamiento está conformado por 421 imágenes de manzanas buenas y 284 de manzanas podridas, mientras que el conjunto validación contiene 70 imágenes de manzanas buenas y 40 de podridas. Teniendo cerca de un 16 % respecto a los datos de entrenamiento para validar los algoritmos.

El número de pasos por época se determina mediante:

$$\frac{DE}{TL} = \frac{705}{32} \approx 22 \tag{10}$$

donde DE y TL corresponden al número de datos en entrenamiento y tamaño de lote, respectivamente.

$$\frac{DV}{TL} = \frac{110}{32} = 3.43 \approx 4 \tag{11}$$

siendo DV el número de datos de validación. Las ecuaciones (10) y (11) indican que la red toma 22 grupos de datos para evaluar, una vez completado el número de grupos se finaliza una época del entrenamiento y se evalúa la red con los pesos obtenidos en 4 lotes de datos propios de la validación.

Finalmente se generan los datos a emplear, estos se llevan a un tamaño de (224, 224) píxeles con 3 canales de entrada.

Compilación y entrenamiento de la red

El objetivo del entrenamiento es minimizar la diferencia entre la salida real y la calculada por la NN mediante el ajuste de los parámetros de la red, a los cuales se les pueden aportar distintas métricas según sea necesario, entre ellos se puede destacar el uso de optimizadores y métricas de pérdida. El optimizador se encarga de generar pesos (W) que ayuden a la red a converger al resultado deseado, para este caso de estudio se emplea RMSprop (Root Mean Square Propagation) que hace uso de un promedio móvil de gradientes cuadrados para normalizar el gradiente.

Por su parte, la función de pérdida (*loss)* se encarga de evaluar la desviación entre las predicciones realizadas por la red y los valores reales de observación empleados durante el aprendizaje, un resultado pequeño indica que la red es eficiente. Para nuestro caso de estudio se emplea la entropía cruzada categórica (*categorical cross entropy),* la cual es una función de pérdida utilizada en tareas de clasificación multiclase. Estas son tareas en las que un ejemplo solo puede pertenecer a una de muchas categorías posibles, y el modelo debe decidir a cuál.

Las arquitecturas empleadas han sido diseñadas para el procesamiento y clasificación de imágenes. A continuación se enlistan algunos ejemplos:

- He *et al.* (2016) presentan uno de los primeros modelos de redes profundas RN, hacen uso de filtros convolucionales en sus distintas capas (50), cada una presenta un filtro distinto (de 1x1 ó 3 × 3) y a la salida de la ultima operación se le suma el valor de entrada a las operaciones, esta conexión les permite crear arquitecturas tan profundas como se requiera.
- Respecto a MN, Howard *et al.* (2017) hacen uso de operaciones de convolución como *Depth-wise* y *Point-wise,* logrando reducir el tiempo y costo de cómputo requerido, su red se conforma de 52 capas.
- Chollet (2017) trabajan con NN Xc de 41 capas, presentan una modificación al método de convolución *Depth-wise* y realizan distintas operaciones ReLU + Convolución con las imágenes resultantes.

Es importante mencionar que Xc hace uso de *Depth wise separable convolutions*, lo cual permite un costo computacional bajo en comparación a modelos que hacen uso de convoluciones estándar (empleadas en ResNet50). Por lo cual, es computacionalmente efectiva, requiere menos parámetros y permite reducir el tiempo de entrenamiento (Chollet, 2017; Guo *et al.*, 2019).

A continuación, se enlistan los parámetros empleados para el entrenamiento de las 3 arquitecturas:

- *Optimizer:* RMSprop, con un *learning rate* = 0.001, $\rho = 0.9$, $\epsilon =$ None
- *Loss:* Categorical_crossentropy
- *Metric:* Accuracy

Entrenamiento de redes neuronales

Con la finalidad de comparar y verificar que NN es más apropiada para la tarea en cuestión, se realizó un análisis de entrenamiento de las arquitecturas antes mencionadas.

La Figura 9 muestra el desempeño de aprendizaje obtenido durante el entrenamiento de las NNs. Se pue-

https://doi.org/10.22201/fi.25940732e.2022.23.4.031

de resaltar que RN tarda más tiempo en llegar a un conocimiento o extracción de características, la red incrementa el aprendizaje de manera gradual conforme pasan las épocas obteniendo un desempeño de 0.7436. En lo que a la red MN se refiere, se observa un menor tiempo en llegar al aprendizaje en comparación con RN, el cual se mantiene con pequeñas variaciones. Los valores obtenidos en la validación comienzan con buena capacidad de predicción manteniendo ligeros cambios, logrando un desempeño máximo de 0.8462. Por su parte, Xc se mantiene oscilando cerca del 1 desde su primera época, presentando ligeras variaciones, los resultados de validación denotan un desempeño de 0.8846.

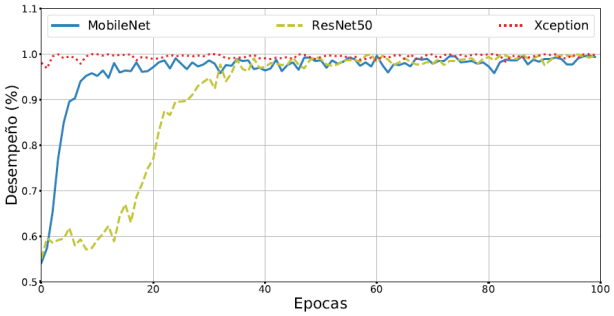

Figura 9. Desempeño de las NNs durante la etapa de entrenamiento

Los resultados mostrados en la Figura 9 no son suficientes para validar que el aprendizaje obtenido durante el entrenamiento permita realizar una correcta clasificación. Por ello, se sometió un conjunto de datos en cada época del entrenamiento de las NNs para validar su eficiencia al momento de realizar una clasificación. En la Figura 10 se presentan los resultados de validación, destacando los valores más altos (MN = 74.36 %, RN = 84.62 % y Xc = 88.46 %). Se observa que Xc obtuvo una mejor capacidad de clasificación durante el proceso de validación, lo anterior justifica su uso.

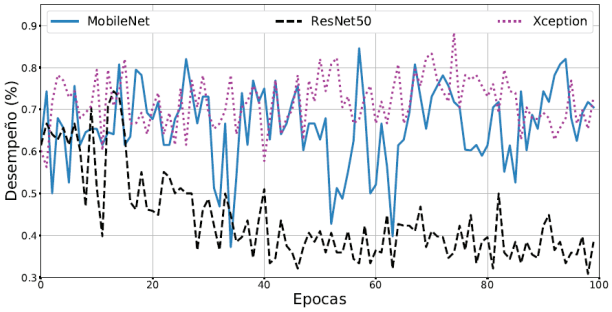

Figura 10. Gráfica de validación durante entrenamiento

Posteriormente se procedió a realizar experimentos fuera de línea. La Figura 11 muestra un ejemplo de clasificación de una manzana buena y otra con pudriciones. El experimento se realizó con 3 imágenes de

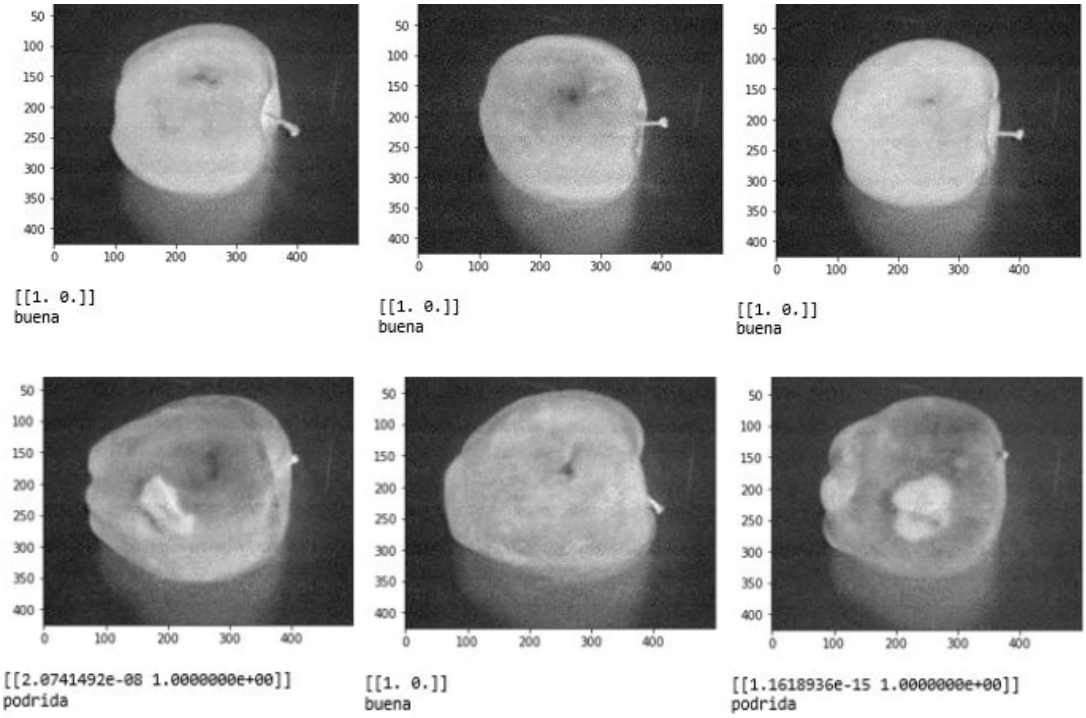

Figura 11. Resultados obtenidos fuera de línea: a) Manzana buena, no presenta pudriciones, b) Manzana mala, presenta pudriciones

diferentes ángulos de las manzanas (girándolas), logrando una mejor visualización del cuerpo del fruto. Es importante mencionar que en este experimento se clasificaron 17 manzanas mediante la red Xc, logrando un asertividad de 83 %.

Integración del sistema de clasificación

Se construyó el prototipo de la Figura 12. El cual consta de una banda transportadora de 0.09 x 0.6 m, trabaja a una velocidad de 0.075 m/s y permite procesar 4 manzanas por minuto. Además, se emplea iluminación halógena en forma directa que ofrece un flujo luminoso de 1490 lúmenes. Para detectar el paso de las manzanas se emplean 3 sensores infrarrojo LM8393 acoplados a un costado de la banda, dejando una distancia de 0.09m entre cada sensor, los cuales sirven como capturador externo, es decir; cuando un sensor detecta el paso de la manzana se manda un pulso a la CM para realizar una captura.

Clasificación en línea

Se realizaron pruebas en línea para evaluar la capacidad de clasificación real del modelo propuesto, se tomó una muestra conformada por 16 manzanas o 48 imágenes (Figura 13) divididas en 5 lotes, la distribución de estos se describe en la Tabla 2.

Tabla 2. Evaluación original (O) *vs* clasificación en línea. (C) (B=Buena, P=Podrida, L=Lote e I=Imagen, N=dato nulo)

L \ I	1	2	3	4	5	6	7	8	9	10	
1	B	B	B	B	B	B	B	B	B	B	O
	B	B	P	B	B	B	B	P	B	P	C
2	P	P	P	P	P	B	B	B	B	B	O
	B	B	B	P	B	P	B	B	B	B	C
3	P	P	P	P	P	P	P	P	P	P	O
	B	P	P	P	B	B	P	P	P	P	C
4	P	P	P	P	P	P	P	P	P	P	O
	B	P	P	P	P	B	P	P	P	P	C
5	P	P	B	B	B	B	B	B	N	N	O
	P	P	B	B	B	B	B	B	N	N	C

La red obtuvo una clasificación de 26 buenas y 22 podridas, mientras que la distribución original corresponde a 21 buenas y 27 podridas.

Los datos de la Tabla 2 se clasificaron en 4 categorías (Figura 14): Observación y predicción positivas (VP), observación positiva y predicción negativa (FN), obser-

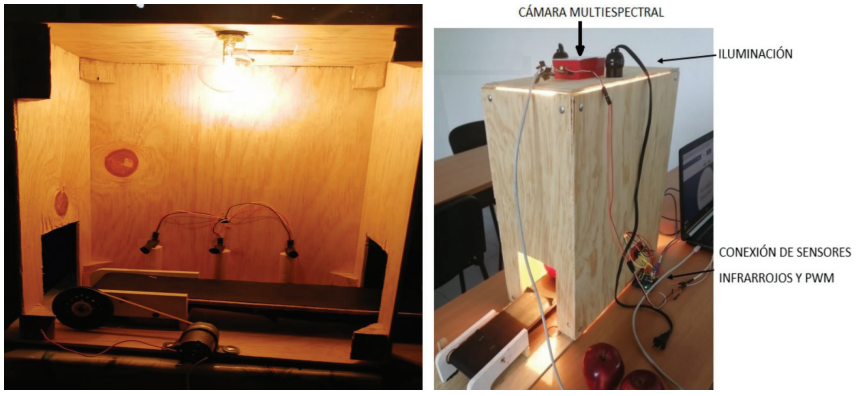

Figura 12. Prototipo funcional: a) Interior de prototipo, b) Exterior del prototipo e instrumentación

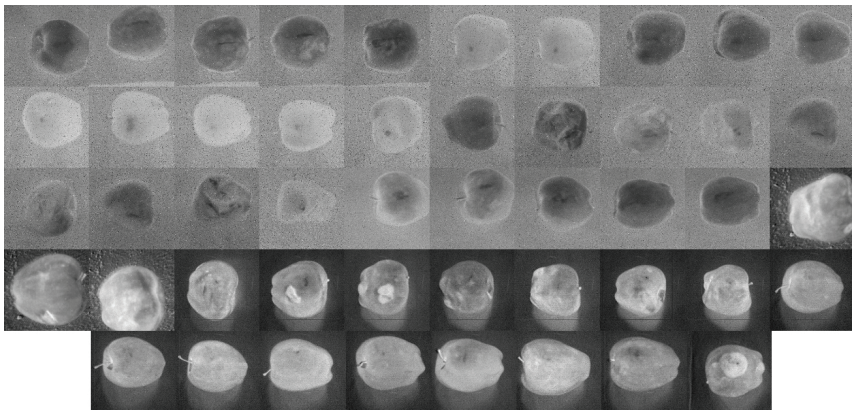

Figura 13. Banco de imágenes que conforman la matriz de confusión

vación y predicción negativas (VN) y observación negativa y predicción positiva (FP). Lo anterior para el análisis de resultados y verificación de la capacidad de asertividad del modelo (Bowes, 2012).

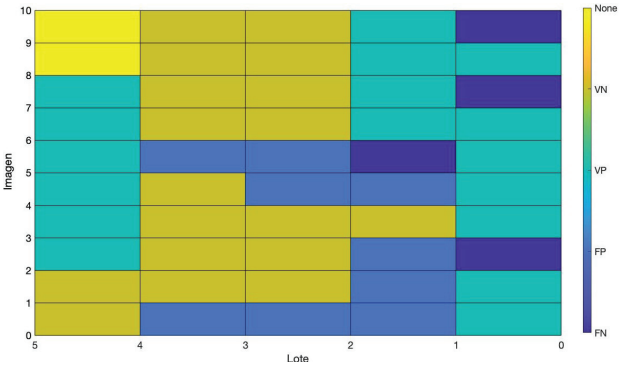

Figura 14. Comparación de resultados de evaluación

Agrupando los datos de la Figura 14; VP = 17, FN = 4, FP = 9 y VN = 18. Así, el desempeño de la clasificación está dado por:

$$A_c = \frac{VP + VN}{VP + VN + FP + FN} = \frac{35}{48} = 0.73 \quad (12)$$

Para observar en qué proporción los ejemplos positivos fueron clasificados, se utiliza:

$$R_P = \frac{VP}{VP + FN} = \frac{17}{21} = 0.81 \quad (13)$$

El número total de ejemplos correctos clasificados positivos entre el número de ejemplos positivos obtenidos, indica que los ejemplos de clase positiva en su mayoría fueron clasificados de manera correcta. Para el cálculo de la razón de ejemplos negativos se realizan cambios en (13), obteniendo:

$$R_N = \frac{VN}{VN + FP} = \frac{18}{27} = 0.66 \quad (14)$$

Lo cual indica que más de la mitad de los datos se clasifican con exactitud, resaltando que se cuenta con cierto déficit para lograr una correcta clasificación de estos.

La exactitud de los datos predichos positivos respecto a los que sí lo son se determina mediante la siguiente ecuación:

$$P = \frac{VP}{VP + FP} = \frac{17}{26} = 0.65 \quad (15)$$

El uso de la matriz de confusión muestra un desempeño a de 73 % de asertividad, permitiendo validar el desempeño de la clasificación en línea. Al analizar las imágenes se observa que algunas contienen perturbaciones que crean confusión a la red al momento de ser clasificadas. También, se resalta que los valores de predicción tienen cierta preferencia a los positivos (manzanas buenas), concluyendo que en presencia de incertidumbre la red predice el dato como positivo. Se debe resaltar que, debido a la naturaleza de la banda, no es posible girar la manzana (salvo ocasiones cuando debido a su forma, gira ligeramente).

Discusión de resultados

El desempeño de 88.46 % obtenido durante la simulación de nuestra propuesta es bajo respecto a Singh & Singh (2019), quienes clasifican fuera de línea manzanas podridas y buenas. Singh *et al.* logran 98.9 % de precisión utilizando una SVM y características de textura como: transformada Wavelet, histograma de gradientes orientados, energía de la textura de Law, etc. En un enfoque similar, Moallem *et al.* (2017) obtienen un desempeño de 92.5 % utilizando 8 características estadísticas, 5 texturales y 3 geométricas, las cuales ingresan a una SVM.

En el presente trabajo sustituimos la extracción de características por operaciones geométricas entre imágenes multiespectrales (cálculo del NDVI), reduciendo el uso de operaciones complejas y permitiendo que el CNN sea el encargado de la extracción de características.

Por su parte, Ashraf *et al.* (2019) logran un desempeño de 98 % al detectar frutas podridas o frescas entrenando una CNN con 1734 imágenes (obtenidas de internet) y empleando TA. Por nuestra parte, se destaca el empleo de 705 imágenes para el entrenamiento de la red y la base de datos obtenida mediante la CM. En un enfoque similar, en Fan *et al.*(2020) se realiza la clasificación en línea de manzanas buenas o podridas, obtuvieron 96.5 % y 92 % de desempeño en simulación e implementación, respectivamente. Utilizando una banda transportadora para rotar cada fruta y dos cámaras RGB para captar seis imágenes de cada manzana. Así, extraen 20 características de textura que entran a la CNN. El entrenamiento se hizo con 79200 imágenes y 50000 iteraciones.

Aunque el desempeño obtenido en Fan *et al.* (2020) es mayor al reportado en este trabajo, es importante considerar que se emplean 3 imágenes y que no es posible girar las manzanas. Por lo anterior, se realizaron experimentos fuera de línea con 3 imágenes de diferentes ángulos, obteniendo un desempeño de 83 %.

En resumen, el empleo del NDVI para este caso de estudio muestra un camino para la inspección de ali-

mentos sin contacto y una extracción o posible clasificación de diversos padecimientos que pueden presentarse y no son visibles al ojo humano. Sin embargo, se debe trabajar en futuras mejoras a fin de proporcionar mayor robustes al sistema y así incrementar el desempeño en la clasificación.

Conclusiones

La metodología descrita reduce la necesidad de extraer características como: textura, forma, color, etcétera. En lugar de ello, se usa el NDVI para observar pudriciones. Lo anterior se logra mediante operaciones algebraicas entre las bandas espectrales proporcionadas por la CM. Por otra parte, la arquitectura Xc se implementó en línea para clasificar manzanas buenas y podridas. El uso de TA, redujo la necesidad de una base de datos grande, obteniendo 88.46 %, 83 % y 73 % de desempeño en el entrenamiento, clasificación fuera de línea (imágenes de diferentes ángulos) y en la clasificación en línea, respectivamente.

Se debe resaltar que la mayoría de los trabajos se centra en el diseño y comparación de arquitecturas de aprendizaje (Lu & Lu, 2018; Singh & Singh, 2019), extracción de características empleando imágenes hiperespectrales (Dong, 2015; Ekramirad *et al*., 2017) y RGB (Bhargava & Bansal, 2018). Sin embargo, pocos llegan a la implementación en línea. Así, la contribución de este trabajo radica en la descripción de las etapas de un sistema de clasificación, que va desde la obtención de la base de datos, hasta el análisis de resultados de la implementación en línea. El uso de TA permitió obtener un 88.46 % de desempeño en simulación, usando 705 imágenes y 2200 iteraciones durante el entrenamiento. Lo anterior presenta un avance significativo en comparación con Fan *et al*. (2020), donde se obtuvo una precisión de 96.5 % empleando 79200 imágenes y 50000 iteraciones.

Finalmente, como trabajos futuros se tiene el diseño y modificación del algoritmo a fin de tener mayor robustes durante la clasificación, así como la creación de un prototipo que permita girar las manzanas y tomar un mayor número de imágenes. Lo anterior a fin de mejorar su desempeño durante la clasificación.

Agradecimientos

Este trabajo fue financiado a través del proyecto 292399: "Generación de estrategias científico-tecnológicas con un enfoque multidisciplinario e interinstitucional para afrontar la amenaza que representan los complejos ambrosiales en los sectores agrícola y forestal de México", del Fondo Institucional de Fomento Regional para el Desarrollo Científico, Tecnológico y de Innovación (FORDECyT) del Consejo Nacional de Ciencia y Tecnología (CONACyT) y la Secretaría de Investigación y Posgrado del Instituto Politécnico Nacional (SIP-IPN) bajo los proyectos de investigación 20210709 y 20212098.

Referencias

Ali, A. M., Darvishzadeh, R., Skidmore, A. K., & Van-Duren, I. (2017). Specific leaf area estimation from leaf and canopy reflectance through optimization and validation of vegetation indices. Agricultural and forest meteorology, 236, 162-174. https://doi.org/10.1016/j.agrformet.2017.01.015

Ashraf, S., Kadery, I., Chowdhury, M. A. A., Mahbub, T. Z., & Rahman, R. M. (2019). Fruit image classification using convolutional neural networks. *International Journal of Software Innovation (IJSI)*, 7(4), 51-70.

Baranowski, P., Mazurek, W., & Pastuszka-Woźniak, J. (2013). Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data. *Postharvest Biology and Technology*, 86, 249-258. https://doi.org/10.1016/j.postharvbio.2013.07.005

Bhargava, A., & Bansal, A., (2018). Fruits and vegetables quality evaluation using computer vision: A review. *Journal of King Saud University-Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2018.06.002

Bhargava, A., & Bansal, A., (2020). Quality evaluation of Mono & bi-Colored Apples with computer vision and multispectral imaging. *Multimedia Tools and Applications*, 79(11), 7857-7874. https://doi.org/10.1007/s11042-019-08564-3

Bowes, D., Hall, T., & Gray, D. (2012). Comparing the performance of fault prediction models which report multiple performance measures: recomputing the confusion matrix. In Proceedings of the 8th international conference on predictive models in software engineering, 109-118. Recuperado de https://doi.org/10.1145/2365324.2365338

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, 1251-1258.

Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2019). A survey of deep learning and its applications: A new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 1-22. https://doi.org/10.1007/S11831-019-09344-W

Dong, J., & Guo, W. (2015). Nondestructive determination of apple internal qualities using near-infrared hyperspectral reflectance imaging. *Food Analytical Methods*, 8(10), 2635-2646.

Ekramirad, N., Rady, A., Adedeji, A. A., & Alimardani, R. (2017). Application of Hyperspectral imaging and acoustic emission techniques for apple quality prediction. *Transactions of the ASABE*, 60(4), 1391-1401.

Escamilla, G., & Padilla, V. (2012). Three dimensional dynamic measurements using a stereo vision system and optical flow

algorithms for high speed video applications. In CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers, IEEE, 113-117. Recuperado de http://dx.doi.org/10.1109%2FCONIELECOMP.2012.6189892

Fan, S., Li, J., Zhang, Y., Tian, X., Wang, Q., He, X., ... & Huang, W. (2020). On line detection of defective apples using computer vision system combined with deep learning methods. *Journal of Food Engineering*, 286, 110102. https://doi.org/10.1016/j.jfoodeng.2020.110102

Folch, A., Prats, J. M., Cubero, S., Blasco, J., & Ferrer, A. (2016). VIS/NIR hyperspectral imaging and N-way PLS-DA models for detection of decay lesions in citrus fruits. *Chemometrics and Intelligent Laboratory Systems*, 156, 241-248. https://doi.org/10.1016/j.chemolab.2016.05.005

Giovos, R., Tassopoulos, D., Kalivas, D., Lougkos, N., & Priovolou, A. (2021). Remote sensing vegetation indices in viticulture: A critical review. *Agriculture,* 11, 5, 457.

Guo, Y., Li, Y., Wang, L., & Rosing, T. (2019), July. Depthwise convolution is all you need for learning multiple visual domains. In Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 8368-8375. Recuperado de https://doi.org/10.1609/aaai.v33i01.33018368

Hameed, K., Chai, D., & Rassau, A. (2018). A comprehensive review of fruit and vegetable classification techniques. *Image and Vision Computing*, 80, 24-44.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

Heras, D., (2017). Clasificador de imágenes de frutas basado en inteligencia artificial Fruit image classifier based on artificial intelligence. *Revista Killkana* Técnica, 1(2).

Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, Recuperado de https://doi.org/10.48550/arXiv.1704.04861

Hussain, M., Bird, J. J., & Faria, D. R. (2019). A study on CNN transfer learning for image classification. In: Lotfi, A., Bouchachia, H., Gegov, A., Langensiepen, C., McGinnity, M. (eds) Advances in computational intelligence systems. UKCI 2018. Advances in Intelligent Systems and Computing, 840. Springer, Cham.

Lin, Y. P., & Jung, T. P. (2017). Improving EEG-based emotion classification using conditional transfer learning. *Frontiers in human neuroscience*, 11, 334.

Lu, Y., & Lu, R. (2018). Detection of surface and subsurface defects of apples using structured-illumination reflectance imaging with machine learning algorithms. *Transactions of the ASABE*, 61(6), 1831-1842.

Lu, Z., & Cai, L. (2020). Camera calibration method with focus-related intrinsic parameters based on the thin-lens model. *Optics Express*, 28(14), 20858-20878.

Mathworks (2018). Convolutional Neural Network. Recuperado de https://www.mathworks.com/content/mathworks/www/en/discovery/convolutional-neural-network.html.

Moallem, P., Serajoddin, A., & Pourghassem, H. (2017). Computer vision-based apple grading for golden delicious apples based on surface features. *Information processing in agriculture,* 4(1), 33-40. https://doi.org/10.1016/j.inpa.2016.10.003

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359.

Papageorgiou, E. I., Aggelopoulou, K., Gemtos, T. A., & Nanos, G. D. (2018). Development and evaluation of a fuzzy inference system and a neuro-fuzzy inference system for grading apple quality. *Applied Artificial Intelligence*, 32(3), 253-280. https://doi.org/10.1080/08839514.2018.1448072

Paulus, I., De Busscher, R., & Schrevens, E. (1997). Use of image analysis to investigate human quality classification of apples. *Journal of Agricultural Engineering Research*, 68(4), 341-353.

Ranganathan, P., & Olson, E. (2014). Locally-weighted homographies for calibration of imaging systems. In 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 404-409.

RedEdge, M. (2015). Multispectral camera user manual. micasense Inc.: Seattle, WA, USA, 33. Recuperado de https://support.micasense.com/hc/enus/articles/215261448rededge-user-manual-pdfdownload

Shaha, M, & Pawar, M. (2018). Transfer learning for image classification. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 656-660. Recuperado de https://doi.org/10.1109/+ICECA.2018.8474802

Singh, S., & Singh, N. P., (2019). Machine learning-based classification of good and rotten apple. In *Recent trends in communication, computing, and electronics*, 377-386. Singapore: Springer. Recuperado de https://doi.org/10.1007/978-981-13-2685-1_36

Sofu, M. M., Er, O., Kayacan, M. C., & Cetişli, B. (2016). Design of an automatic apple sorting system using machine vision. *Computers and Electronics in Agriculture*, 127, 395-405.

Steinbrener, J., Posch, K., & Leitner, R. (2019). Hyperspectral fruit and vegetable classification using convolutional neural networks. *Computers and Electronics in Agriculture,* 162, 364-372. http://dx.doi.org/10.1016/j.compag.2019.04.019

Tan, W., Sun, L., Yang, F., Che, W., Ye, D., Zhang, D., & Zou, B. (2018). Study on bruising degree classification of apples using hyperspectral imaging and GS-SVM. *Optik,* 154, 581-592.

Wang, Z., Hu, M., & Zhai, G. (2018). Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data. *Sensors,* 18(4), 1126. https://doi.org/10.3390/s18041126

Xul, Q., Zou, X., & Zhao, J. (2008). On-line detection of defects on fruit by Machinevision systems based on three-color-cameras systems. In International Conference on Computer and Computing Technologies in Agriculture, 2231-2238. Springer, Boston, MA.

Zhang, B., Huang, W., Wang, C., Gong, L., Zhao, C., Liu, C., & Huang, D. (2015). Computer vision recognition of stem and calyx in apples using near-infrared linear-array structured light and 3D reconstruction. *Biosystems engineering*, 139, 25-34. http://dx.doi.org/10.1016%2Fj.biosystemseng.2015.07.011

Zhang, M., & Li, G. (2018). Visual detection of apple bruises using AdaBoost algorithm and hyperspectral imaging. *International Journal of Food Properties*, 21(1), 1598-1607.

Cómo citar:

Gante-Díaz, S. A., Lozano-Hernández, Y., Maldonado-Trinidad, M. A., Flores-Colunga, G. R., & Villegas-Piña, D. (2022). Integración de una cámara multiespectral y aprendizaje automático para clasificación de manzanas. *Ingeniería Investigación y Tecnología*, 23 (04), 1-13. https://doi.org/10.22201/fi.25940732e.2022.23.4.031