

Determinación de la madurez de mazorcas de Cacao, haciendo uso de redes neuronales convolucionales en un sistema embebido

Cocoa pods ripeness estimation, using convolutional neural networks in an embedded system

Juan F. Heredia-Gómez¹ , Juan P. Rueda-Gómez¹ , Leonardo H. Talero-Sarmiento² ,
Juan S. Ramírez-Acuña¹ , Roberto A. Coronado-Silva³ 

¹Universidad Autónoma de Bucaramanga, Bucaramanga, Colombia

²Universidad Industrial de Santander, Bucaramanga, Colombia

³Centro de Investigación La Suiza, AGROSAVIA, Río Negro, Colombia

jheredia106@unab.edu.co, jrueda405@unab.edu.co, lehetasa@correo.uis.edu.co,
jramirez89@unab.edu.co, rcoronado@agrosavia.co

(Recibido: 10 Julio 2020; aceptado: 19 Octubre 2020; Publicado en Internet: 1 Diciembre 2020)

Resumen. Una correcta cosecha Cacao implica determinar si la mazorca se encuentra en un adecuado estado de madurez. No obstante, este proceso suele darse de manera artesanal y basarse en atributos como el tamaño y color de la mazorca, características que difieren según la variedad cultivada, lo cual dificulta su estandarización. Con el fin de simplificar la cantidad de variables y presentar un método automatizado, el presente trabajo propone desarrollar una herramienta portable, de bajo costo, y hecha a medida, la cual hace uso de una red neuronal convolucional para indicar si una mazorca de cacao se encuentra en el momento oportuno para ser cosechada. Entre los principales resultados del presente trabajo se encuentran: 1) la construcción de tres conjuntos de datos etiquetados (1992 imágenes cada uno), y 2) un sistema embebido con una precisión de 34.83% mAP (*mean Average Precision*). Finalmente, se demuestra estadísticamente que el tamaño de las imágenes (4033x4033 p, 1009x1009 p y 505x505 p) no incide sobre la eficacia del entrenamiento.

Palabras clave: Cacao, Clasificación de Imágenes, Detección de objetos, Madurez, Reconocimiento de Imágenes, YOLO, Raspberry Pi.

Abstract. A correct cocoa harvest involves determining a pod maturity. However, this farm activity is usually handmade, using criteria such as Size and Color of the pod; those characteristics differ according to the cocoa variety, making it difficult to standardize. For this reason, this work proposes an automated method to simplify the number of variables to develop a portable, low-cost, and custom-made tool, which makes use of a convolutional neural network to indicate whether a cocoa pod is found it at the right time to harvest. The main results of this work are: 1) the construction of three labeled data sets (1992 images each), and 2) we developed an embedded system with a 34.83% mAP (mean Average Precision) accuracy. Finally, variance analysis demonstrates that image size (i.e., 4033x4033 p, 1009x1009 p, and 505x505 p) does not affect accuracy.

Keywords: Cocoa, Image recognition, Image classification, Object detection, Raspberry pi, Ripeness, YOLO.

Tipo de artículo: Artículo de investigación.

1 Introducción

La cosecha de un fruto es una etapa del proceso productivo igual de importante a la siembra y postcosecha; no obstante, en algunos casos a ésta no se le atribuye la misma relevancia, a pesar que una cosecha inoportuna representa pérdidas a los agricultores por la no comercialización del producto originado por el incumplimiento de los estándares mínimos de aceptación (Chamo et al., 2017; Yen & Nguyễn, 2018). En este sentido, la cosecha inoportuna del Cacao genera tres principales impactos. Primero, se disminuyen las propiedades organolépticas en el producto final (Yen & Nguyễn, 2018), con lo cual no se logra obtener el sabor y olor de un cacao de alta calidad debido a la disminución de antioxidantes, los cuales son los partícipes de garantizar que los alimentos tengan su sabor, color y olor (Machado Cuellar et al., 2018). Segundo, aumenta la presencia de oxalatos solubles, a la par de disminuir los oxalatos insolubles dentro del

grano de cacao; ya que en ocasiones se cosecha la mazorca en un estado anterior a la madurez donde se presenta la mayor cantidad de oxalatos solubles, mientras hay una baja presencia de oxalatos insolubles, por consiguiente, la comercialización de granos sub-maduros puede tener un impacto negativo en la salud humana, ya que favorecen la formación de cálculos en los riñones (Nguyễn et al., 2018). Tercero, en algunas variedades de Cacao cuyas mazorcas no han sido cosechadas a tiempo, los granos empiezan a germinar al interior del fruto generando así una pérdida completa del producto (CAOBISCO/ECA/FCC, 2015).

Por consiguiente, para evitar lo anterior es necesario realizar una cosecha oportuna. Para el caso del Cacao (*Theobroma Cacao*) los procesos de identificación del nivel de madurez de la mazorca para su posterior recolección implican retos tecnológicos, debido a que éstos son enteramente artesanales, es decir; los cacaoteros identifican mediante sus sentidos las características morfológicas (principalmente tamaño y color) para determinar si el fruto se encuentra maduro y listo para cosechar (Riskiawan et al., 2018; Tan et al., 2018). Ahora bien, es importante señalar que soportar el proceso de identificación de madurez en las capacidades sensoriales del cacaotero puede llegar a generar incongruencias durante la estandarización del proceso, ya que el color junto con las tonalidades de un producto maduro son diferentes según la variedad de cacao (Perez B & Contreras M, 2017; Ramos Ospino & Gómez Álvarez, 2019). En añadidura, en una mazorca de cualquier variedad, los colores que ésta expone no se distribuyen de manera homogénea, presentando así tonalidades diferentes o, en el peor de los casos, colores distintos (Cubillos et al., 2019), lo que dificulta generar una estandarización de procesos y/o herramientas que permitan recrear el proceso de identificación del nivel de madurez de manera global (Taiwo & Bart-Plange, 2016) y, por tanto, se suelen considerar factores adicionales como la exposición a la luz solar, humedad, altura, entre otros, con el fin de determinar el instante adecuado para cosechar, aun cuando ello implica el desarrollo de modelos más complejos (CAOBISCO/ECA/FCC, 2015).

Teniendo en cuenta lo anterior, es necesario diseñar estrategias o modelos que permitan identificar oportunamente el momento de cosechar una mazorca de Cacao (lo cual tiene el potencial de aumentar la producción de granos en los terrenos ya cultivados, incrementando con ello el rendimiento agrícola). Una posible estrategia a implementar es el desarrollo de herramientas de base tecnológica, sobre todo en el contexto colombiano, donde existe una carencia de las mismas en el sector agropecuario, y es a partir de la transferencia de las mismas que se puede impulsar un desarrollo tecnológico para reducir las brechas económicas entre el sector urbano y rural, mejorando así la calidad de vida de los campesinos (CORPOICA, 2015).

En los antecedentes estudiados, se evidencia que las herramientas deben ser hechas a medida, lo que suele implicar la identificación de una variedad de cultivo, el levantamiento de un nuevo conjunto de datos, y la calibración de modelos. En cuanto al lenguaje de desarrollo, en concreto la mayoría de las soluciones implementadas se ejecutan en herramientas que no son de libre acceso como MATLAB, siendo una restricción respecto a los recursos disponibles. Del mismo modo, se presentan aplicaciones de escritorio desarrolladas en C#, por consiguiente, se dificulta transportar estas herramientas dentro de la unidad productiva y, por tanto, es necesario que las soluciones consideren las limitaciones tecnológicas identificadas en el sector agropecuario como conectividad y portabilidad (Alston et al., 2008; Huffman, 2009). Máxime, cuando los métodos encontrados en la revisión de la literatura suelen ser de alto costo (2,362.30 USD) comparados al método imperante en los laboratorios, que se basan en el indicador de grados Brix (basado en su pH) determinado a través de un espectrómetro, pero que trae consigo una solución invasiva y lenta, al tener que abrir todas las mazorcas. En este sentido, el uso de aplicaciones móviles o implementaciones de sistemas embebidos es recomendable, ya que permiten el desarrollo de herramientas portables, escalables, a un bajo costo.

Teniendo en cuenta lo anterior, el presente trabajo surge con el objetivo de contestar la siguiente pregunta de investigación: ¿Cómo generar una herramienta portable, de bajo costo, y hecha a medida, que permita indicar el momento oportuno para cosechar una mazorca de Cacao? Para ello, el documento se organiza de la siguiente manera: en la sección 2 se muestra la revisión de la literatura realizada, en la sección 3 se describe la variedad de Cacao a la cual se realizará el estudio, así como la construcción del conjunto de datos (3.1); en las siguientes subsecciones se expone la arquitectura propuesta (3.2), el proceso de entrenamiento de una máquina de aprendizaje (3.3), y la propuesta de un diseño experimental para determinar el efecto que tiene el tamaño de las imágenes en la eficacia del modelo (3.4). En la sección posterior se presentan los resultados de la investigación, seguido de la discusión de los resultados, y en la última sección, las conclusiones del trabajo.

2 Revisión de literatura

De manera específica, en la producción de cacao la determinación de madurez se ha realizado por medio de espectrometría, haciendo uso de sensores acústicos (Arenga et al., 2017), y la valoración por medio de inspección visual implementando técnicas de colorimetría a través de valores RGB extraídos de imágenes convencionales (Riskiawan et al., 2018). No obstante, en los antecedentes teóricos relacionados con visión por computador se encuentra que la toma de imágenes se suele realizar en ambientes controlados, es decir, con luminosidad, distancia al objeto, y sombras controladas, es por ello que el hacer uso del espacio de color RGB no es oportuno cuando las imágenes se han de tomar en el sitio del cultivo. Lo anterior, debido a que durante este proceso, pueden existir condiciones que dificulten la estandarización del proceso como intensidad lumínica (Lakens et al., 2013), distancia entre el lente y el producto (Třebický et al., 2016), y efecto de las sombras (O'Brien & Farid, 2012).

En este sentido, se encuentran antecedentes de herramientas para estandarizar el proceso de cosecha basado en la identificación de madurez. Se resalta que, en la literatura científica la mayoría de las soluciones encontradas sobre el tema son desarrolladas para otros productos agrícolas diferentes al cacao, tales como, tomates, bananas, papayas, pimentones, y manzanas. A continuación, se presenta la revisión de los siguientes trabajos asociados al cacao y a estos productos agrícolas, de los cuales se extrae información acerca del tamaño del conjunto de datos, la cantidad de categorías usadas, el uso o no de un escenario controlado, y en el caso en que lo mencionen, la distancia de captura. Arenga et al., (2017) realizaron una estimación de la madurez de las mazorcas de cacao haciendo uso de sensores acústicos, para ello disponían de 49 mazorcas, de las cuales fueron extraídas 392 muestras sensoriales, considerando dos categorías (Arenga et al., 2017). De igual manera, un año después, Riskiawan et al., (2018) realizaron una estimación de la madurez de las mazorcas de cacao, capturando imágenes a 25-30 cm del fruto en un escenario controlado, considerando tres categorías con 38 imágenes en total (Riskiawan et al., 2018).

En cuanto al cultivo de tomates, Polder et al. (2002) realizaron una estimación de madurez capturando las imágenes en un escenario controlado, considerando cuatro categorías sin especificar el tamaño del conjunto de datos (Polder et al., 2002). 12 años después, Rupanagudi et al., (2014) realizaron una estimación de madurez tomando imágenes a 46 cm de distancia en un escenario controlado, considerando tres categorías con aproximadamente 33 imágenes por categoría (Rupanagudi et al., 2014); al siguiente año, Mhaski et al., (2015) realizaron un estudio similar sin especificar el uso de un escenario controlado, considerando un total de seis categorías para un total de 100 imágenes (Mhaski et al., 2015). En ese mismo año, El-Bendary et al., (2015) realizaron una estimación de madurez en tomates haciendo uso de un escenario controlado, considerando cinco categorías con 50 imágenes por categoría (El-Bendary et al., 2015). Por su parte, 3 años después, Zhang et al., (2018) realizaron una estimación de madurez en tomates capturando imágenes *in situ*, considerando cinco categorías y 30 imágenes por categoría (Zhang et al., 2018). A la par ese mismo año, Taofik et al., (2018) realizaron una estimación de madurez en tomates *in situ* considerando cuatro categorías (Taofik et al., 2018).

Del mismo modo, Mustafa et al., (2008) realizaron una estimación de madurez en bananas capturando imágenes en un escenario controlado, considerando cuatro categorías y sin especificar el tamaño del conjunto de datos (Mustafa et al., 2008). Al siguiente año, Saadl et al., (2009) realizaron una estimación de madurez de dicho fruto capturando imágenes en un escenario controlado, considerando tres categorías con 20 imágenes cada una (Saadl et al., 2009). Posteriormente, Kipli et al., (2018) realizaron una estimación de madurez en bananas tomando 338 imágenes de un repositorio público, sin categorías específicas (Kipli et al., 2018). Continuando, Mazen & Nashat, hicieron una estimación de madurez en bananas capturando imágenes en un escenario controlado, considerando cuatro categorías para un total de 300 imágenes (Mazen & Nashat, 2019). Finalmente, Le et al., (2019) realizaron una estimación de madurez en bananas tomando 194 imágenes de un repositorio público, considerando dos categorías (Le et al., 2019).

En cuanto al cultivo de papaya se resaltan los trabajos de Saad & Hussain (2006), quienes realizaron una estimación de madurez sin especificar el uso de un espacio controlado, considerando tres categorías para un total de 18 imágenes (Saad & Hussain, 2006). Luego de 12 años, Santos Pereira et al., (2018) realizaron una estimación de madurez en papayas tomando imágenes a 17.5 cm del fruto en un escenario controlado, considerando tres categorías para un total de 114 imágenes (Santos Pereira et al., 2018). Posteriormente, Simbolon et al. (2019) realizaron una estimación de madurez en papayas tomando imágenes a 74 cm del fruto en un escenario controlado, no estableciendo categorías con un total de 300 imágenes (Simbolon et al., 2019). Asimismo, Elhariri et al., (2014) realizaron una estimación de la madurez en pimentones sin especificar el uso de un escenario controlado, considerando cinco categorías con 175

imágenes en total (Elhariri et al., 2014). Por su parte, Hamza & Chtourou (2018) realizaron una estimación de la madurez en manzanas sin especificar el uso de un escenario controlado, haciendo uso de tres categorías, con 200 imágenes en total (Hamza & Chtourou, 2018).

De la revisión se puede destacar que en cuanto a técnicas, las más comunes son técnicas de aprendizaje híbridas (Kipli et al., 2018; Mhaski et al., 2015; Polder et al., 2002; Rupanagudi et al., 2014; Simbolon et al., 2019). En segundo lugar, los autores suelen usar redes neuronales artificiales (Hamza & Chtourou, 2018; Mazen & Nashat, 2019; Saad & Hussain, 2006). En tercer lugar, máquinas de soporte vectorial SVM (Arenga et al., 2017; El-Bendary et al., 2015; Elhariri et al., 2014). En cuarto lugar, los autores emplean *K-Nearest Neighbors* (KNN) (Arenga et al., 2017; Riskiawan et al., 2018). En último lugar implementaron diversas técnicas como: *Convolutional Neural Networks* CNN (Zhang et al., 2018), *K-Means* con Fuzzy Logic (Taofik et al., 2018), *Mask R-CNN* (Le et al., 2019), *Rule Based Classification* (Kipli et al., 2018), y *Random Forest* (Santos Pereira et al., 2018). Es preciso destacar que aunque la mayoría de trabajos efectuaron la toma de imágenes en ambientes controlados, se encuentran investigaciones donde las imágenes son tomadas *in situ* (Taofik et al., 2018; Zhang et al., 2018), empleando dispositivos de bajo costo (León-Roque et al., 2016; Taofik et al., 2018).

3 Materiales y métodos

Para el presente trabajo fue empleado un genotipo de cacao tipo criollo denominado TCS 01 (*Teobroma Corpoica la Suiza*), obtenido mediante mejoramiento participativo. Esta variedad propia de la región de Santander, fue entregada a los productores de la subregión “natural montaña” en el año 2014 y presenta, entre otras, las siguientes características: Inicio de producción: 20 meses de edad, índice de grano: 3 g/grano (siendo una de las almendras de mayor tamaño entre las variedades comerciales de cacao), índice mazorca; 9 mazorcas por kg de cacao seco, número de semillas por fruto o mazorca: 38 unidades, producción de grano al cuarto año: 3,3 kg/año (CORPOICA, 2020).

3.1 Construcción del conjunto de datos

Las imágenes fueron tomadas en el centro de investigación La Suiza de AGROSAVIA, el cual se encuentra ubicado en las márgenes del río Salamagá en el municipio de Rionegro (Santander), a 34 km vía Bucaramanga a la Costa Atlántica, a una altura de 500 msnm, con 26 °C de temperatura y 85 % de humedad relativa, en una extensión de 32 hectáreas en varios relieves (Arguello Castellanos et al., 1999). A cada imagen se le añadió una etiqueta de estado de madurez. Esta categoría fue asignada por Agrónomos de AGROSAVIA expertos en el cultivo de TCS. La categoría fue verificada a muestras evaluadas en laboratorio, donde se midieron los grados Brix de las mazorcas. El proceso de construcción del conjunto de datos se expone a continuación.

Tamaño de imágenes y estandarización. El levantamiento del conjunto de datos fue realizado a partir de cuatro visitas comprendidas entre el 07/02/2020 y el 13/03/2020 al centro de investigación la Suiza. Del total de imágenes tomadas (521) se seleccionaron 498 según disponibilidad de información relacionada a la variedad y el estado de madurez, con la siguiente cantidad de elementos por categoría: Inmadura (226 imágenes), Madura (99), y Sobre-madura (173). Las imágenes fueron tomadas con una cámara Samsung SM-G975F con las características descritas a continuación: Punto F: f/1.5, Tiempo de exposición 1/193 segundos, Velocidad ISO: ISO-50, Compensación de la exposición: +0.1paso, Distancia focal: 4 mm, Apertura máxima: 1.16, Modo de medición: Foco, Modo de flash: sin flash, y longitud focal de 35 mm: 26. La dimensión de las imágenes obtenidas es de 4032x3024 pixeles, a una resolución de 72 pixeles por pulgada (ppp), en un modelo de color sRGB.

Con el propósito de estandarizar las imágenes para el entrenamiento, se organizaron las fotos procurando que las mazorcas queden orientadas de manera vertical. Posteriormente, se añaden bordes negros hasta obtener una composición cuadrada de 4033x4033 pixeles manteniendo la resolución de 72 ppp. Además, con el objetivo de aumentar la cantidad de imágenes por categoría, se realizaron modificaciones de manera sistemática a las fotos según parámetros de color. En total se generan tres transformaciones: 1) Niveles de color automáticos, contraste automático, blanco y negro automático; 2) Aumento del brillo en un 35% y de contraste en un 60%; 3) Niveles de color automáticos, contraste automático, blanco y negro automático,

Corrección de brillo -67%, corrección de contraste 11%. Con lo anterior se generó un conjunto de datos de imágenes compuesto por 1992 fotos, dividido en las categorías Inmadura (904 imágenes), Madura (396), y Sobre-madura (692).

Para finalizar, a partir del conjunto de datos original se crearon dos conjuntos más buscando determinar si el tamaño de la imagen incide sobre la capacidad de la máquina de aprendizaje para identificar las diferentes categorías. Mediante esta estrategia se cuenta con la misma cantidad de imágenes por categoría, manteniendo la resolución proporcional. Los tamaños de las imágenes de los tres conjuntos de datos finales son Original (4033x4033 p), Pequeño (1009x1009 p), y Diminuto (505x505 p). En la **Figura 1**, se indica de manera gráfica el proceso para la generación de los tres conjuntos de datos.

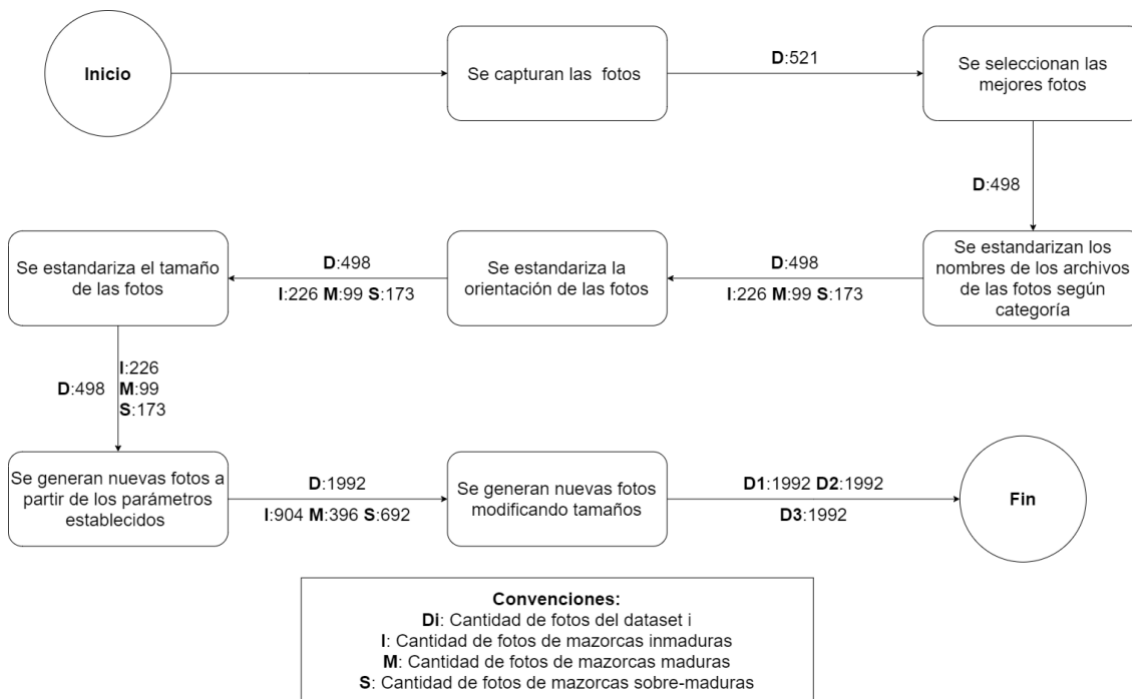


Figura 1. Secuencia de actividades para la creación de los tres conjuntos de datos

3.2 Arquitectura

Para el desarrollo de la herramienta portable se deben integrar diferentes elementos de software (para el entrenamiento de una máquina para segmentación y clasificación de imágenes) y hardware (para la implementación de la herramienta). Para ello y como primer paso, los conjuntos de datos son analizados mediante técnicas de inteligencia artificial, particularmente, haciendo uso de redes neuronales convolucionales (*Convolutional neural networks, CNNs*), las cuales se caracterizan por la facilidad que presentan para la extracción y análisis de imágenes (Samui et al., 2017). En este trabajo se hacen uso de unas CNNs denominadas YOLO (*You Only Look Once*), las cuales están desarrolladas en el lenguaje de programación C++, presentan una gran eficiencia en el uso de los recursos, y se caracterizan por su velocidad en la identificación de objetos, haciendo uso de 53 capas ocultas (Redmon & Farhadi, 2018). De las cuatro variantes de esta arquitectura, en el presente trabajo se hace uso de la versión YOLOv3.0-tiny que presenta menos capas convolucionales (17 capas) para mejorar la velocidad de procesamiento a cambio de disminuir la eficacia en la detección (en la **Figura 2** se presenta la arquitectura de red implementada hecha con base a detalles de la referencia). En el segundo paso, se utiliza el framework Darknet, puntualmente, la versión para dispositivos embebidos con procesadores ARM (*Advanced RISC Machine*) denominada Darknet NNPACK, la cual se puede ejecutar en sistemas embebidos y se caracteriza por presentar mejoras en el funcionamiento del reconocimiento de objetos (Park, 2020). Darknet NNPACK presenta las siguientes dos ventajas: es posible calibrar YOLOv3 en sistemas operativos Linux, además, permite realizar pruebas en tiempo real con la cámara conectada al dispositivo de entrenamiento. El sistema abordado en este trabajo es ejecutado totalmente offline en un sistema embebido Raspberry Pi 4, el cual

cuenta con 4 GB de RAM, y un procesador ARM Broadcom BCM2711, además, se hace uso de una webcam de Microsoft modelo Lifecam HD 3000, la cual captura una imagen en 1280x720 píxeles.

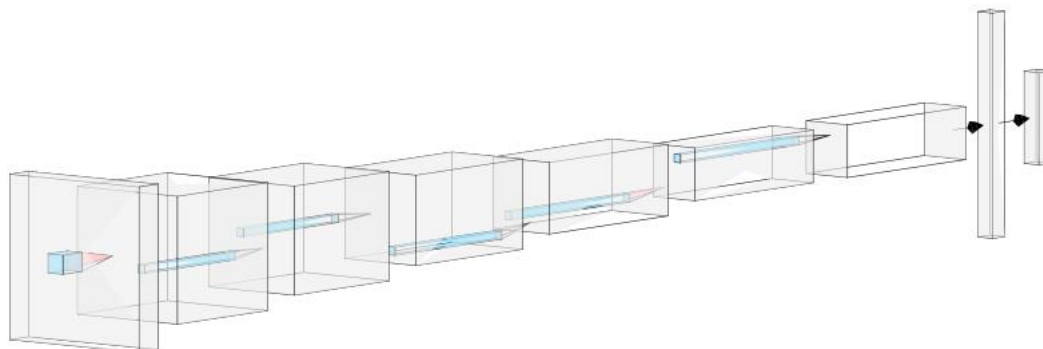


Figura 2. Estructura de la red convolucional adaptado de (Redmon & Farhadi, 2018)

3.3 Entrenamiento de la máquina de aprendizaje

El proceso de entrenamiento de la red de aprendizaje parte del proceso de etiquetado a las imágenes, el cual consiste en indicar dentro de cada foto el objeto y asignarle la categoría Inmadura, Madura, o Sobre-madura. Para ello se hace uso del programa LabelImg el cual recibe de entrada una imagen en formato JPG y entrega de salida un archivo en formato XML con la ubicación del objeto y su categoría (Lin, 2015). Inmediatamente se carga una red previamente entrenada con un dataset de 20 categorías en una GPU GTX Titan X (Redmon & Farhadi, 2018) y se asignan el conjunto de datos con sus respectivas etiquetas mediante la adaptación de archivo yolov3-tiny.cfg. Posteriormente, se definen los parámetros de entrenamiento. Para el presente trabajo se establecieron 100 *epoch* (representan un barrido desde el primer elemento al último del dataset, y viceversa), un lote (*batch*) de 16 unidades (cantidad de imágenes a procesar en cada paso de la calibración), y una tasa de aprendizaje (*learning rate*) de 0.005, la cual indica el ajuste de los pesos en cada iteración. En la **Figura 3** se indica la secuencia aplicada para la calibración de cada red neuronal.

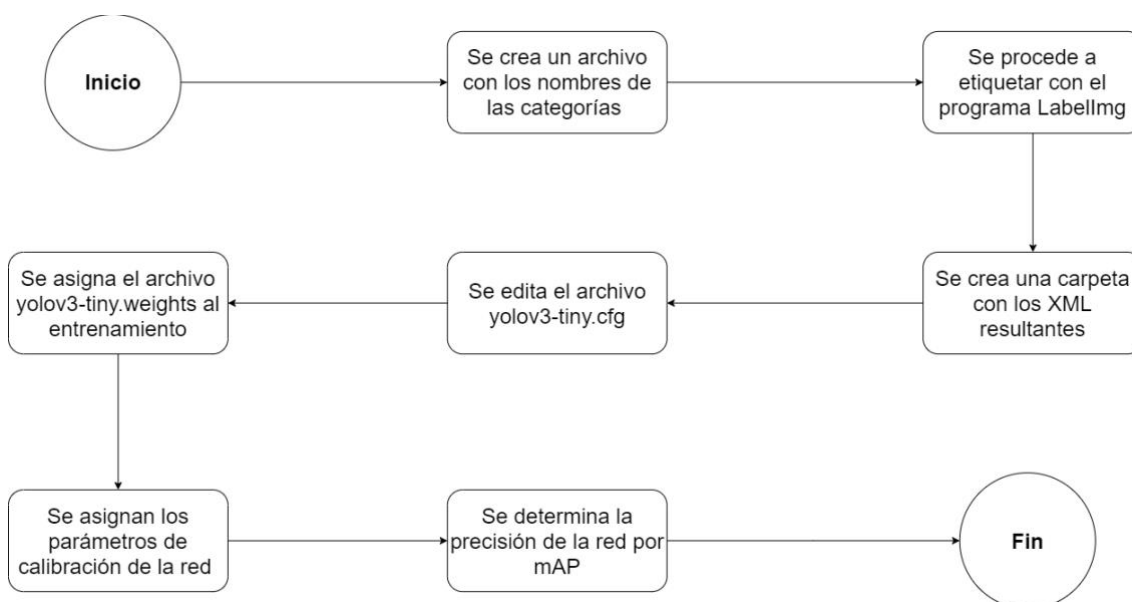


Figura 3. Diagrama de la secuencia de actividades para el entrenamiento de la detección de mazorcas maduras de cacao aplicando la arquitectura YOLO

3.4 Entrenamiento de la máquina de aprendizaje

Con el propósito de determinar cuál es la mejor configuración para generar una efectiva arquitectura (que integre un buen nivel de exactitud en la asignación de categorías, incurriendo en un bajo tiempo computacional), se propone un diseño experimental mono factorial. Para ello se considera como predictor el tamaño de cada imagen según las categorías definidas previamente: Diminuto, Pequeño, y Original; y como métrica de calidad la variable mAP (*mean Average Precision*) la cual se mide para cada categoría de clasificación, generando así tres réplicas. Donde mAP es la media aritmética de la precisión promedio (AP) descrita en la ecuación (1) (Caragea, 2009). Por su parte, el AP (2)-(3) mide la precisión de modelos de detección de imágenes calculando cuatro ítems: 1) tipo de detección {verdadero positivo (*TP*), verdadero negativo (*TN*), falso positivo (*FP*), y falso negativo (*FN*)}, 2) Intersección sobre la unión (*IoU*) que se compara con un valor de umbral (0.5 en el presente caso), 3) *Precision* (*p*), métrica que indica que tan exitosas son las predicciones respecto a los falsos positivos (4), y 4) *Recall* (*r*), variable que representa que tan exitosas son las predicciones respecto a los falsos negativos (5) (Zhu, 2004).

$$MAP = \frac{1}{n} \sum_n AP_n \quad (1)$$

$$AP = \frac{1}{n} \sum_{r \in R} p_{interpolada}(r) \quad (2)$$

Donde

$$p_{interpolada}(r) = p(\underline{r}) \quad \underline{r}; \underline{r} > r \quad (3)$$

$$p = \frac{TP}{TP + FP} \quad (4)$$

$$r = \frac{TP}{TP + FN} \quad (5)$$

Por otra parte, se registrarán los tiempos empleados por la red neuronal para entrenar el proceso de clasificación según el tamaño de los datos. Se resalta que, debido a que no se realizan réplicas, no es posible hacer un diseño experimental.

4 Resultados

Mediante la metodología descrita en la Figura 1 se generaron tres conjuntos de datos de diferente tamaño. Su proporción se puede apreciar en la Figura 4. Dentro de cada conjunto de datos se encuentran las tres variantes realizadas a la imagen tomada. A modo de ejemplo, en la Figura 5 se aprecian los cambios en contraste, brillo, e intensidad.

En cuanto al entrenamiento y calibración de la arquitectura, fueron realizados en un Dell XPS 15 9570, procesador Intel Core i7 8700H, GPU Nvidia GTX 1050ti Max-Q y 16 GB de RAM. La duración de esta etapa presentó un lapso comprendido entre 14 a 68 horas (según el tamaño de las imágenes contenidas en cada Dataset). En cuanto a la selección de imágenes para calibrar la red, cada conjunto fue dividido en dos subconjuntos: 1) Entrenamiento, el cual contiene el 80% de elementos seleccionados de manera aleatoria; y 2) Prueba, con el 20% restante de los datos.



Figura 4. Ejemplo del tamaño relativo de los tres Datasets

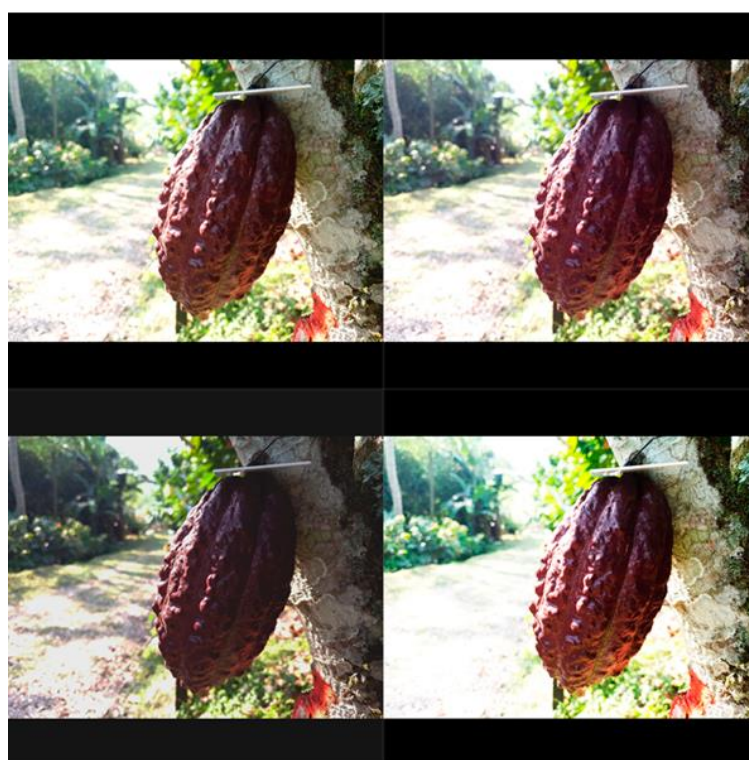


Figura 5. Ejemplo de las transformaciones de color aplicadas sobre una mazorca con categoría Inmadura

Posterior a la calibración, los datos de ajuste son contrastados mediante un análisis de varianza. Al observar los resultados de la [Tabla 1](#), donde el Valor p asociado al predictor es superior al 0.05 (nivel de significancia estadístico seleccionado durante el experimento), se concluye que el factor Dataset no incide sobre el mAP, por consiguiente, no existe diferencia estadística entre los valores de mAP para los tamaños seleccionados de imágenes. Ahora bien, como el modelo estadístico asociado al diseño ANOVA monofactorial propuesto presenta un valor de R^2 de 13.61% y los residuales que cumplen con los supuestos de normalidad (ver [Figura 6](#)), se aplica la prueba no paramétrica de Kurskall-Wallis, cuyos resultados se

registran en la [Tabla 2](#). A partir de los valores obtenidos de mediana, se rechaza con un Valor p de 0.43 que los diferentes niveles (tamaños) de las imágenes inciden sobre el mAP.

Tabla 1. Resultados del Análisis de Varianza ANOVA

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Dataset	2	19,19	9,597	0,47	0,645
Error	6	121,80	20,299		
Total	8	140,99			

Tabla 2. Resultados de medias según Dataset, para la prueba Kurskall-Wallis

Dataset	N	Mediana	Clasificación de medias	Valor Z
Diminuto	3	9,29	5,7	0,52
Original	3	10,09	6,0	0,77
Pequeño	3	7,86	3,3	-1,29
General	9		5,0	

Una vez determinado que no incide el tamaño de la imagen en la clasificación de mazorcas, se procede a hacer un nuevo entrenamiento sobre el Dataset Diminuto, en esta ocasión ampliando el tiempo de calibración a 18 horas y seleccionando aleatoriamente nuevos conjuntos de Entrenamiento y Prueba. En la [Tabla 3](#) se registran los valores de ajuste (mAP) obtenidos con el entrenamiento final.

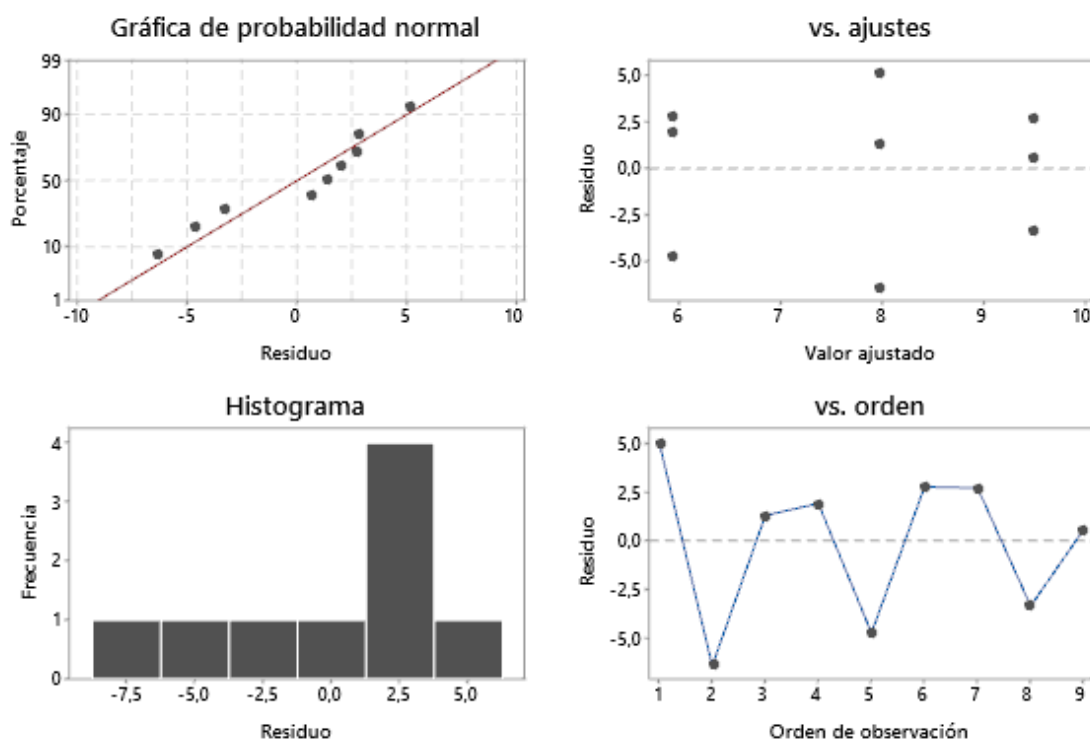


Figura 6. Gráficas de residuales para el Análisis de Varianza ANOVA

Tabla 3. Valores de mAP para el conjunto de datos Diminuto

Categoría	mAP
Inmadura	25.40%
Madura	26.41%
Sobre madura	52.67%

Por otro lado, en cuanto al tiempo empleado para entrenar la red, en la [Tabla 4](#) se registra el cambio porcentual respecto al dataset de menor tamaño (Diminuto), encontrando que, para el conjunto Pequeño, el tiempo requerido es más del doble, y más de cuatro veces para el conjunto de tamaño original.

Tabla 4. Comparación del tiempo requerido para entrenar la red según el Dataset

Dataset	Tiempo	Línea base
Diminuto	14	100,00%
Pequeño	30	214,29%
Grande	68	485,71%

Finalmente, una vez entrenada la máquina de aprendizaje, esta es implementada en la Raspberry Pi 4. En la [Figura 7](#) se muestra un ejemplo de la detección simultánea de mazorcas de cacao correspondientes a la variedad *Theobroma Cacao TCS-01*. Particularmente, se representa la capacidad de asignar múltiples categorías. Los resultados de la herramienta son supervisados por expertos de AGROSAVIA mediante imágenes de archivos, puesto que debido a la contingencia del COVID-19 no fue posible la prueba *in situ*.

**Figura 7.** Ejemplo de funcionamiento del sistema

5 Discusión

El presente trabajo de investigación se estructuró con el fin de determinar la arquitectura que permita generar una herramienta portable, de bajo costo, y hecha a medida, para indicar el instante oportuno de cosecha para una mazorca de cacao. Para ello, se generó una base de datos de imágenes que contiene tres categorías. Se resalta que, en la literatura las soluciones con más imágenes son las propuestas por Kipli et al. con 338 imágenes (Kipli et al., 2018) y Mazen & Nashat con 300 imágenes (Mazen & Nashat, 2019). Este proyecto introdujo 498 imágenes originales para el desarrollo de la solución, y 3 veces más esta cantidad en imágenes sintéticas, lo que supone mayor probabilidad de sobre ajuste, sin embargo, los resultados de precisión obtenidos no evidencian este fenómeno. No obstante, es probable que las otras investigaciones al no usar un gran número de imágenes (menos de 100 imágenes) (Riskiawan et al., 2018; Saad & Hussain, 2006; Saadl et al., 2009), presenten un ajuste insuficiente para clasificar imágenes nuevas. Además, se destaca como limitación que al ser un trabajo *in situ* las categorías de madurez no están presentes en el árbol durante un mismo lapso que las otras categorías, ya que el estado de inmadurez de una mazorca se extiende aproximadamente durante tres meses, mientras que las dos categorías restantes un mes

aproximadamente cada una, por tanto, es poco probable encontrar el mismo número de elementos en cada categoría. Por consiguiente, se recomienda para futuros trabajos incrementar la cantidad de imágenes a usar para evitar posibles sobre entrenamientos.

La anterior limitación implica que no se cuenta con la misma cantidad de elementos por categoría y esta característica puede impactar sobre el ajuste del modelo. No obstante, la decisión de capturar las imágenes en campo se toma con dos propósitos, el primero es generar un estudio no destructivo ya que no es necesario cortar los frutos para llevarlos a laboratorio, y el segundo es permitir replicar el modelo con nuevos datos directamente en campo, así como permitir que la comunidad no académica pruebe los desarrollos previos como una estrategia de prototipado rápido. Este principio es identificado en los trabajos previos (Taofik et al., 2018; Zhang et al., 2018), donde resaltan que la toma de imágenes *in situ* sin controlar variables no repercuten significativamente en el desempeño de la herramienta. Se resalta además que en comparación a los trabajos revisados que detallaron el uso de un escenario controlado para sus desarrollos (Arenga et al., 2017; El-Bendary et al., 2015; Mazen & Nashat, 2019; Mustafa et al., 2008; Polder et al., 2002; Riskiawan et al., 2018; Rupanagudi et al., 2014; Saadl et al., 2009; Santos Pereira et al., 2018), esta investigación propone una solución de bajo costo ya que no se precisó de equipos profesionales para la captura de imágenes, logrando resultados aceptables considerando una media de mAP de 34.83.

Dicho valor de mAP puede ser explicado por las siguientes razones. Primero, la diferencia de capacidad computacional, puesto que Redmon & Farhadi (2018) usan una GPU Pascal Titan X con 10.97 TFLOPS, mientras que la usada en el presente trabajo es 1050Ti Max Q con 2.18 TFLOPS, por lo cual fue necesario usar un modelo Tiny que presenta menor eficacia; Segundo, según Cubillos et al., (2019) las mazorcas de cacao no presentan madurez uniforme, por tanto, las distintas categorías presentan colores similares lo que dificulta su diferenciación de manera artificial. Tercero, durante la etapa de etiquetado se presentaron dudas respecto a la clasificación de ciertas mazorcas como consecuencia de la similitud en sus colores (error de origen antrópico). Por ello, para futuros trabajos se recomienda hacer un análisis destructivo de las mazorcas, es decir, abrir todos los frutos buscando asignar la categoría correcta según el estado de sus granos.

A pesar de las limitaciones de infraestructura tecnológica y la dificultad para asignar las etiquetas, fue posible entrenar un algoritmo de clasificación e implementarlo en un sistema embebido, lo que implica que se cumplió con las recomendaciones de diseño de Alston et al., (2008) y Huffman (2009). Además, basándose en los resultados estadísticos, en los que se determinan que para el presente problema de clasificación no existe diferencia significativa entre el tamaño de las imágenes, se recomienda que para trabajos afines se utilice un tamaño de imágenes de 505x505 p, con el cual se obtienen buenos resultados a un costo computacional inferior. Este hallazgo se alinea con los resultados encontrados en trabajos previos, salvo por lo indicado en los resultados expuestos por Hamza & Chtourou (2018).

6 Conclusiones

Durante el desarrollo del presente trabajo se implementó una red neuronal convolucional en un sistema embebido generando así un prototipo de herramienta offline, portable y escalable con un costo de materiales aproximado de 112.99 USD. La arquitectura tecnológica propuesta permite la clasificación de mazorcas de cacao de la variedad TCS 01 (*Teobroma Corpoica la Suiza*) según tres estados de madurez: Inmadura, Madura, y Sobremadura. La metodología implementada durante el presente trabajo se basa en el reentrenamiento una arquitectura de redes neuronales pre-entrenadas denominada YOLOv3.0-tiny. Esta aproximación permite que el trabajo pueda ser replicable no sólo a otros cultivos agrícolas donde se busque clasificar frutos según su madurez, sino a otros problemas como son: identificación temprana de enfermedades, estimación del volumen de producción de frutos, y estimación del periodo de fermentación de los granos.

Agradecimientos

Los autores extienden sus agradecimientos al Programa de Ingeniería de Sistemas de la Universidad Autónoma de Bucaramanga, al programa Becas Bicentenario del Ministerio de Ciencia, Tecnología e

Innovación Minciencias, al programa doctoral en Ingeniería de la Universidad Autónoma de Bucaramanga, y al Centro de Investigación La Suiza, de AGROSAVIA.

Declaración de conflicto de intereses

Los autores declaran no tener conflicto de intereses con respecto a la investigación, autoría o publicación de este artículo.

Referencias

- Alston, J., Pardey, P., & Ruttan, V. (2008). Research Lags Revisited: Concepts and Evidence from U.S. Agriculture. *University of Minnesota, Department of Applied Economics, Staff Papers*.
- Arenga, D. Z. H., Dela Cruz, J. C., & Arenga, D. Z. H. (2017). Ripeness classification of cocoa through acoustic sensing and machine learning. *2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2018-Janua*, 1–6. <https://doi.org/10.1109/HNICEM.2017.8269438>
- Arguello Castellanos, O., Mejía Florez, L. A., Contreras Mayorga, N., & Toloza Ochoa, J. A. (1999). *Manual de caracterización morfoagronómica de clones elite de cacao (Theobroma cacao L.) en el noriente colombiano*. CAOBISCO/ECA/FCC. (2015). *Cocoa Beans : Chocolate & Cocoa Industry Quality Requirements*.
- Caragea, C. (2009). Mean Average Precision. En L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (p. 1703). Springer US. https://doi.org/10.1007/978-0-387-39940-9_3032
- Chamo, A., D, A., Babura, B., & Karaye, A. K. (2017). Influence of Agronomic Practices on Crop Production. *International Journal of Sciences: Basic and Applied Research (IJSBAR), Vol. 31*, 61–66. <https://gssr.org/JournalOfBasicAndApplied/article/view/6688>
- CORPOICA. (2015). *Misión para la transformación del campo. Diagnóstico*. 1–71.
- CORPOICA. (2020). *Theobroma Corpoica la Suiza*.
- Cubillos, A., García, M., S., A., R, G., & Tarazona Díaz, M. (2019). Study of the physical and chemical changes during the maturation of three cocoa clones, EET8, CCNN51 and ICS60. *Journal of the Science of Food and Agriculture, 99*. <https://doi.org/10.1002/jsfa.9882>
- El-Bendary, N., El Hariri, E., Hassanien, A. E., & Badr, A. (2015). Using machine learning techniques for evaluating tomato ripeness. *Expert Systems with Applications, 42(4)*, 1892–1905. <https://doi.org/10.1016/j.eswa.2014.09.057>
- Elhariri, E., El-Bendary, N., Hussein, A. M. M., Hassanien, A. E., & Badr, A. (2014). Bell pepper ripeness classification based on support vector machine. *2014 International Conference on Engineering and Technology (ICET)*, 1–6. <https://doi.org/10.1109/ICEngTechnol.2014.7016802>
- Hamza, R., & Chtourou, M. (2018). Apple Ripeness Estimation Using Artificial Neural Network. *2018 International Conference on High Performance Computing & Simulation (HPCS)*, 229–234. <https://doi.org/10.1109/HPCS.2018.00049>
- Huffman, W. (2009). Technology and Innovation in World Agriculture: Prospects for 2010-2019. *Iowa State University, Department of Economics, Staff General Research Papers*.
- Kipli, K., Zen, H., Sawawi, M., Mohamad Noor, M. S., Julai, N., Junaidi, N., Shafiq Mohd Razali, M. I., Chin, K. L., & Wan Masra, S. M. (2018). Image Processing Mobile Application For Banana Ripeness Evaluation. *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, 1–5. <https://doi.org/10.1109/ICASSDA.2018.8477600>
- Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J., & Midden, C. J. H. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion, 27(7)*, 1225–1246. <https://doi.org/10.1080/02699931.2013.781501>
- Le, T.-T., Lin, C.-Y., & Piedad, E. J. (2019). Deep learning for noninvasive classification of clustered horticultural crops – A case for banana fruit tiers. *Postharvest Biology and Technology, 156*, 110922. <https://doi.org/https://doi.org/10.1016/j.postharvbio.2019.05.023>
- León-Roque, N., Abderrahim, M., Nuñez-Alejos, L., Arribas, S. M., & Condezo-Hoyos, L. (2016). Prediction of fermentation index of cocoa beans (*Theobroma cacao L.*) based on color measurement and artificial neural networks. *Talanta, 161*, 31–39. <https://doi.org/10.1016/j.talanta.2016.08.022>
- Lin, T. (2015). *LabelImg*. <https://github.com/tzutalin/labelImg>
- Machado Cuellar, L., Ordoñez Espinosa, C., Katherine, Y., Cruz, L., & Suárez Salazar, J. (2018). Organoleptic quality assessment of *Theobroma cacao L.* in cocoa farms in northern Huila, Colombia. *Acta Agronómica, 67*. <https://doi.org/10.15446/acag.v67n1.66572>
- Mazen, F. M. A., & Nashat, A. A. (2019). Ripeness Classification of Bananas Using an Artificial Neural Network.

- Arabian Journal for Science and Engineering*, 44(8), 6901–6910. <https://doi.org/10.1007/s13369-018-03695-5>
- Mhaski, R. R., Chopade, P. B., & Dale, M. P. (2015). Determination of ripeness and grading of tomato using image analysis on Raspberry Pi. *2015 Communication, Control and Intelligent Systems (CCIS)*, 214–220. <https://doi.org/10.1109/CCIntelS.2015.7437911>
- Mustafa, N. B. A., Fuad, N. A., Ahmed, S. K., Abidin, A. A. Z., Ali, Z., Yit, W. B., & Sharrif, Z. A. M. (2008). Image processing of an agriculture produce: Determination of size and ripeness of a banana. *2008 International Symposium on Information Technology*, 1–7. <https://doi.org/10.1109/ITSIM.2008.4631636>
- Nguyễn, H. V. H., Lê, H. M., & Savage, G. P. (2018). Effects of maturity at harvesting and primary processing of cocoa beans on oxalate contents of cocoa powder. *Journal of Food Composition and Analysis*, 67, 86–90. <https://doi.org/https://doi.org/10.1016/j.jfca.2018.01.007>
- O'Brien, J. F., & Farid, H. (2012). Exposing photo manipulation with inconsistent reflections. *ACM Transactions on Graphics*, 31(1), 1–11. <https://doi.org/10.1145/2077341.2077345>
- Park, T. (2020). *Darknet with NNPACK*. <https://github.com/digitalbrain79/darknet-nnpack>
- Perez B, M. A., & Contreras M, J. D. (2017). Instructivo de buenas prácticas de cosecha y pos-cosecha. En *Swisscontact Colombia*.
- Polder, G., van der Heijden, G. W. A. M., & Young, I. T. (2002). Spectral Image Analysis for Measuring Ripeness of Tomatoes. *Transactions of the ASAE*, 45(4), 1155–1161.
- Ramos Ospino, A. del C., & Gómez Álvarez, M. S. (2019). *Caracterización fenotípica y genotípica de aislados de cacao (Theobroma Cacao L.) de Dibulla, Guajira* (Vol. 8, Número 5).
- Redmon, J., & Farhadi, A. (2018). *YOLOv3: An Incremental Improvement*.
- Riskiawan, H. Y., Puspitasari, T. D., Hasanah, F. I., Wahyono, N. D., & Kurnianto, M. F. (2018). Identifying Cocoa ripeness using K-Nearest Neighbor (KNN) Method. *2018 International Conference on Applied Science and Technology (iCAST)*, 354–357. <https://doi.org/10.1109/iCAST1.2018.8751633>
- Rupanagudi, S. R., Ranjani, B. S., Nagaraj, P., & Bhat, V. G. (2014). A cost effective tomato maturity grading system using image processing for farmers. *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, 7–12. <https://doi.org/10.1109/IC3I.2014.7019591>
- Saad, H., & Hussain, A. (2006). Classification for the Ripeness of Papayas Using Artificial Neural Network (ANN) and Threshold Rule. *2006 4th Student Conference on Research and Development*, 132–136. <https://doi.org/10.1109/SCORED.2006.4339325>
- Saadl, H., Ismaie, A. P., Othmanl, N., Jusohl, M. H., Naim, N. F., & Ahmad, N. A. (2009). Recognizing the ripeness of bananas using artificial neural network based on histogram approach. *ICSIPA09 - 2009 IEEE International Conference on Signal and Image Processing Applications, Conference Proceedings*, 536–541. <https://doi.org/10.1109/ICSIPA.2009.5478715>
- Samui, P., Roy, S., & Balas, V. (2017). *Handbook of Neural Computation 1st Edition*.
- Santos Pereira, L. F., Barbon, S., Valous, N. A., & Barbin, D. F. (2018). Predicting the ripening of papaya fruit with digital imaging and random forests. *Computers and Electronics in Agriculture*, 145, 76–82. <https://doi.org/10.1016/j.compag.2017.12.029>
- Simbolon, Z. K., Syakry, S. A., Mulyadi, & Syahrini, M. (2019). Separation of the Mature Level of Papaya Callina Fruit Automatically Based on Color (RGB) uses Digital Image Processing. *IOP Conference Series: Materials Science and Engineering*, 536, 12127. <https://doi.org/10.1088/1757-899X/536/1/012127>
- Taiwo, A., & Bart-Plange, A. (2016). Factors Responsible For Post-Harvest Losses And Their Effects On Rice Producing Farmers: A Case Study Of Afife And Aveyime Rice Projects in The Volta Region Of Ghana. *International Research Journal of Engineering and Technology (IRJET)*, 3, 1014–1022.
- Tan, D. S., Leong, R. N., Laguna, A. F., Ngo, C. A., Lao, A., Amalin, D. M., & Alvindia, D. G. (2018). AuToDiDAC: Automated Tool for Disease Detection and Assessment for Cacao Black Pod Rot. *Crop Protection*, 103, 98–102. <https://doi.org/https://doi.org/10.1016/j.cropro.2017.09.017>
- Taofik, A., Ismail, N., Gerhana, Y. A., Komarujaman, K., & Ramdhani, M. A. (2018). Design of Smart System to Detect Ripeness of Tomato and Chili with New Approach in Data Acquisition. *IOP Conference Series: Materials Science and Engineering*, 288, 12018. <https://doi.org/10.1088/1757-899X/288/1/012018>
- Třebický, V., Fialová, J., Kleisner, K., & Havlíček, J. (2016). Focal Length Affects Depicted Shape and Perception of Facial Images. *PLOS ONE*, 11(2), e0149313. <https://doi.org/10.1371/journal.pone.0149313>
- Yen, D., & Nguyễn, H. (2018). Effects of maturity stages and fermentation of cocoa beans on total phenolic contents and antioxidant capacities in raw cocoa powder. *Vietnam Journal of Biotechnology*, 14, 743–752. <https://doi.org/10.15625/1811-4989/14/4/12309>
- Zhang, L., Jia, J., Gui, G., Hao, X., Gao, W., & Wang, M. (2018). Deep Learning Based Improved Classification System for Designing Tomato Harvesting Robot. *IEEE Access*, 6, 67940–67950. <https://doi.org/10.1109/ACCESS.2018.2879324>
- Zhu, M. (2004). *Recall, precision and average precision*.

Sobre los autores

Juan F. Heredia-Gómez

Es estudiante de Ingeniería de Sistemas de la Universidad Autónoma de Bucaramanga. Sus áreas de investigación incluyen Visión por Computador, *Deep Learning*, *Machine Learning*, e Inteligencia Artificial.

Juan P. Rueda-Gómez

Es estudiante de Ingeniería de Sistemas de la Universidad Autónoma de Bucaramanga, coordinador del semillero de investigación de aplicaciones móviles. Sus áreas de investigación se centran primordialmente en desarrollo de software, construcción de modelos de información, y analítica de datos.

Leonardo H. Talero-Sarmiento

Es estudiante de Doctorado en Ingeniería de la Universidad Autónoma de Bucaramanga, Profesor Asistente, Magíster en Ingeniería Industrial, e Investigador de la Escuela de Estudios Industriales y Empresariales de la Universidad Industrial de Santander (grupo OPALO). Sus áreas de investigación incluyen Analítica de Datos, Diseño de Sistemas Productivos, Gestión de la Producción, y Finanzas. Con publicaciones en INGE CUC, Apuntes del Cenes, Contaduría y Administración, Estudios Gerenciales, y Suma de Negocios.

Juan S. Ramírez-Acuña

Es un estudiante de Ingeniería de Sistemas de la Universidad Autónoma de Bucaramanga. Sus áreas de investigación se enfocan en desarrollo de software e ingeniería del software.

Roberto A. Coronado-Silva

En la especie Cacao, Líder del proyecto: Recolección, caracterización, morfoagronómica y molecular de materiales criollo y del alto rendimiento en Colombia. Unión temporal Colombia 1, FEDECACAO CORPOICA (2006). Estudio de los atributos fisicoquímicos y organolépticos del cacao en Santander (2012). Investigador principal del proyecto: Evaluación de genotipos seleccionados por atributos de valor en diferentes regiones de Colombia (agenda 2013-2015). Además, ha liderado investigaciones en las áreas de suelos, plagas y poscosecha.