

Sistema de Visión Computarizada con Inteligencia Artificial para Extraer Información de Cultivos de Café en el Sur Occidente Colombiano

German Andres Insuasty Delgado
andresinsuasty@udenar.edu.co

Asesor: PhD. Wilson Achicanoy
 Universidad de Nariño

Resumen—This work focuses on the comparison of techniques in object detection and image segmentation for coffee crops with multispectral images, which are collected by a Parrot Bluegrass drone. Algorithms based on YOLOv5, Autoencoder, and Pix2Pix are proposed to reach automatic plant counting and zone classification in coffee crop areas. We use combinations of spectral bands such as the NDVI index as input for the training and testing of each model. Metrics associated to error, similarity, and detection success are computed, which allow to compare each model. The results indicate that the best automatic counting is accomplished by YOLOv5 with RGB and red band inputs, and segmentation is best done by both the Autoencoder and Pix2Pix with NDVI input. A web application allows to access all the information retrieved by the models, and the database and implementation code is made free.

Index Terms—yolov5, autoencoder, pix2pix, object detection, image segmentation, multispectral, NDVI, coffee crops

I. INTRODUCCIÓN

El cultivo de café crece en determinadas condiciones de temperatura, humedad y altitud. De hecho, los cafetales en su mayoría se encuentran en zonas tropicales, donde la temperatura oscila entre los 18 °C y los 24 °C durante todo el año [1]. El mantenimiento de estos cultivos juega un papel decisivo en la calidad del producto final, que se evalúa para cultivos de cafés especiales a través de un análisis de taza [2]. Con este análisis se determina el nivel de calidad por medio de un proceso de cata, donde se examina a través de diferentes sentidos las propiedades del café. Para realizar este análisis, los catadores toman en cuenta como aspectos principales de la taza el aroma, el sabor y la textura. Calificando estos tres aspectos se obtiene un valor de taza de café que cuantifica la calidad del producto. Este valor tiene relación directa con el valor comercial que los caficultores pueden pedir por su cosecha.

En un escenario ideal, un caficultor que tenga bastante experiencia seguramente obtendrá de manera seguida altos valores de taza para el tipo de café especial que cultive. Sin embargo, en un escenario real esto no siempre es frecuente, debido a que el proceso de producción del café es sensible a factores internos y externos, tanto en las etapas de crecimiento como de cosecha y beneficio. Cualquier cambio no esperado en las variables de alguna de estas etapas puede generar resultados imprevistos, y se pueden acentuar dado que la

metodología de producción se mantiene de forma manual y empírica.

Apoyando el proceso hacia la estandarización de la producción de cafés especiales, se propone un sistema de visión computarizada que analiza la información gráfica de bandas multiespectrales, adquiridas por medio de un vehículo aéreo no tripulado, para hacer un monitoreo constante del estado de cultivos de café, con la finalidad de apoyar la toma de decisiones en estrategias que apunten hacia la vigilancia y el mantenimiento de estos cultivos.

En las siguientes secciones se podrá encontrar una explicación rápida sobre los métodos de inteligencia artificial utilizados en esta investigación. El primero está enfocado en el área de la detección de arbustos de café, con el objetivo de realizar un conteo automatizado de arbustos que pueda servir de soporte para dar información aproximada del nivel de producción de las fincas. También se proponen cuatro alternativas de segmentación de imágenes, para clasificar las áreas presentes en las fincas de acuerdo con su nivel de vitalidad vegetal. Se muestra las variaciones en las implementaciones de estos algoritmos y se comparan sus resultados frente al uso de las diferentes bandas multiespectrales y el índice NDVI. También se entra en detalle sobre la instrumentación necesaria en la obtención de fotografías aéreas para los cultivos de café, generando una base de datos de acceso libre que puede ser utilizada en cualquier otra investigación o proyecto que lo requiera.

En la sección de resultados se informa de la comparación de los diferentes efectos que pueden producir las bandas multiespectrales para un modelo de detección de objetos y cómo la combinación de estas ayuda a segmentar áreas en las imágenes y que dan cuenta de la vitalidad de las plantas. La segmentación propuesta en esta investigación se hace con la comparación de dos algoritmos base, haciendo modificaciones en sus imágenes de entrada. Terminando la sección de resultados, se encuentra la descripción de un aplicativo web, que permite la interacción con los modelos utilizados y apoya la divulgación de este trabajo. Esta herramienta web es de código libre y reutilizable.

En la parte final del documento se encuentran las secciones de conclusiones, agradecimientos y referencias que apoyaron la realización de este trabajo.

II. MODELOS USADOS DE APRENDIZAJE PROFUNDO

El aprendizaje profundo utiliza modelos compuestos por varias capas de procesamiento, que pueden aprender representaciones con diferentes niveles de abstracción y así descubrir representaciones muy precisas a partir de los ejemplos de entrenamiento. Los avances en este campo han sido bastante acelerados por el incremento tecnológico en el poder de procesamiento computacional [3], [4]. Una rama del aprendizaje profundo se fundamenta en el uso de redes neuronales convolucionales CNN (*Convolutional Neural Networks*), que son el método usado por excelencia en problemas relacionados con imágenes y videos [5], [6]. El objetivo de las CNN es extraer todas las características de una imagen y luego usarlas para detectar, clasificar o segmentar los objetos en una imagen.

II-A. Detección de objetos

El algoritmo utilizado en esta investigación y relacionado a la detección de objetos fue YOLOv5 [7], que es una versión mejorada y escrita en pytorch de YOLO (*You Only Look Once*) [8], que toma el problema de detección de objetos como un problema único de regresión. En él existe solo una red convolucional que predice simultáneamente múltiples cuadros o regiones delimitadoras de los objetos en la imagen, y predice probabilidades condicionales para cada clase $p(\text{Clase}|\text{Objeto})$ en cada una de las regiones delimitadoras [9]. La arquitectura de la red neuronal YOLO puede ser vista en la Figura 1. Está compuesta por 24 capas convolucionales seguidas por 2 capas completamente conectadas; para reemplazar los módulos iniciales propuestos en GoogLeNet [10] se utilizan capas de reducción seguidas por capas convolucionales [9]. Estructuralmente, los algoritmos de YOLO mantienen la arquitectura de la red neuronal y las mejoras se notan al realizar *batch normalization* o normalización por lotes, como sucede en YOLO 9000 [11], o cambios en el *framework* de implementación, desde Darknet [12] a Pytorch [7].

II-B. Segmentación de objetos

La segmentación de imágenes trata sobre la agrupación de partes de una imagen que pertenecen a una misma clase. En el contexto de la visión computarizada se han publicado varios trabajos relacionados con la segmentación de imágenes, y dependiendo del método de segmentación o del dato que se opera, se pueden agrupar como segmentación semántica y segmentación de instancias [13], [14]. Esta investigación se orienta a la segmentación semántica utilizando *Autoencoder* [15] y una red neuronal adversaria generativa condicional cGAN (*Conditional Generative Adversarial Network*), llamada Pix2Pix [16].

La arquitectura del Autoencoder se puede ver en la Figura 2. Esta se compone de una etapa de reducción de dimensionalidad, pasando de una imagen de tamaño $256 \times 256 \times n$ hasta una representación de $16 \times 16 \times 256$, para ser de nuevo incrementada hasta las dimensiones de entrada, pero con la transformación de segmentación semántica en la imagen de salida y donde n es el número de capas de la imagen. La

arquitectura de Pix2Pix se observa en la Figura 3. Es fácil notar que la parte del generador se parece mucho a una arquitectura de autoencoder, con la diferencia de que se hace también un concatenado de las capas iniciales, conocidas como *skip connections*. Este generador, como una red U-net, es bastante sofisticado y hace parte de una red aún más grande junto con el discriminador, que a su vez consiste en una red PatchGAN generadora de una matriz de clasificación. Las comparaciones basadas en parches en la matriz de clasificación producen una mejor estructura general de las imágenes, y su evaluación en cada parche se basa en la comparativa de píxeles.

III. INSTRUMENTACIÓN Y CONSTRUCCIÓN DE BASE DE DATOS

Se ha construido una base de datos de imágenes multiespectrales pertenecientes a fincas cafeteras del departamento de Nariño, en específico al municipio de Buesaco. A continuación se explica los materiales y metodologías usadas para la creación de este conjunto de datos.

III-A. Instrumentación para la toma de datos

Actualmente, los vehículos aéreos no tripulados o drones han facilitado nuevas perspectivas en investigaciones relacionadas al estudio de las imágenes, en campos como la fotogrametría y la percepción remota, manteniendo bajos costos de operación y mostrándose como una nueva alternativa para la obtención de imágenes aéreas. Estos vehículos aéreos no tripulados tienen un amplio espectro de utilidades; entre ellas, las labores civiles de monitoreo, mediciones atmosféricas, evaluación de daños, agricultura de precisión, mapeo y cartografía, entre otras [17], [18].

Estos drones están equipados en su mayoría por una plataforma aérea, equipada con una cámara y un sistema de navegación que puede comunicarse con un centro de control en tierra, desde donde se programa y monitorea el plan de vuelo, que puede ser configurado para maniobrar manualmente o de forma automática [19]. Existen distintas clases de drones adaptados para diferentes necesidades, como por ejemplo, en una clasificación amplia se puede diferenciar a los drones de ala fija de los de ala rotatoria o multirrotores. En un nivel mayor de detalle, se puede realizar una clasificación por la tipología de número de brazos, encontrando tricópteros, cuadricópteros, hexacópteros, octocópteros y coaxiales [20].

En esta investigación se utilizó el dron Parrot Bluegrass, que se encuentra en la clasificación de ala rotatoria y cuadricóptero. Cuenta con una cámara gran angular de 14 megapixeles, vídeo full HD 1080p, con estabilización digital en los 3 ejes, GPS incorporado, procesador Dual Core GPU Quad Core y un almacenamiento interno de 32 GB de memoria. Tiene conectividad Wi-Fi 802.11a/b/g/n/ac en una red bibanda MIMO, con dos antenas dobles dipolares de 2,4 y 5 GHz, con una potencia de emisión de hasta 21 dBm, llegando a tener un alcance de señal de aproximadamente 300 m [21]. Este dron posee una cámara multiespectral Sequoia y fue adquirido por la Universidad de Nariño, en el marco del proyecto "Sistema de información integral hacia la estandarización de los procesos de producción de cafés especiales en el municipio de Buesaco" [22] (ver Figura 4).

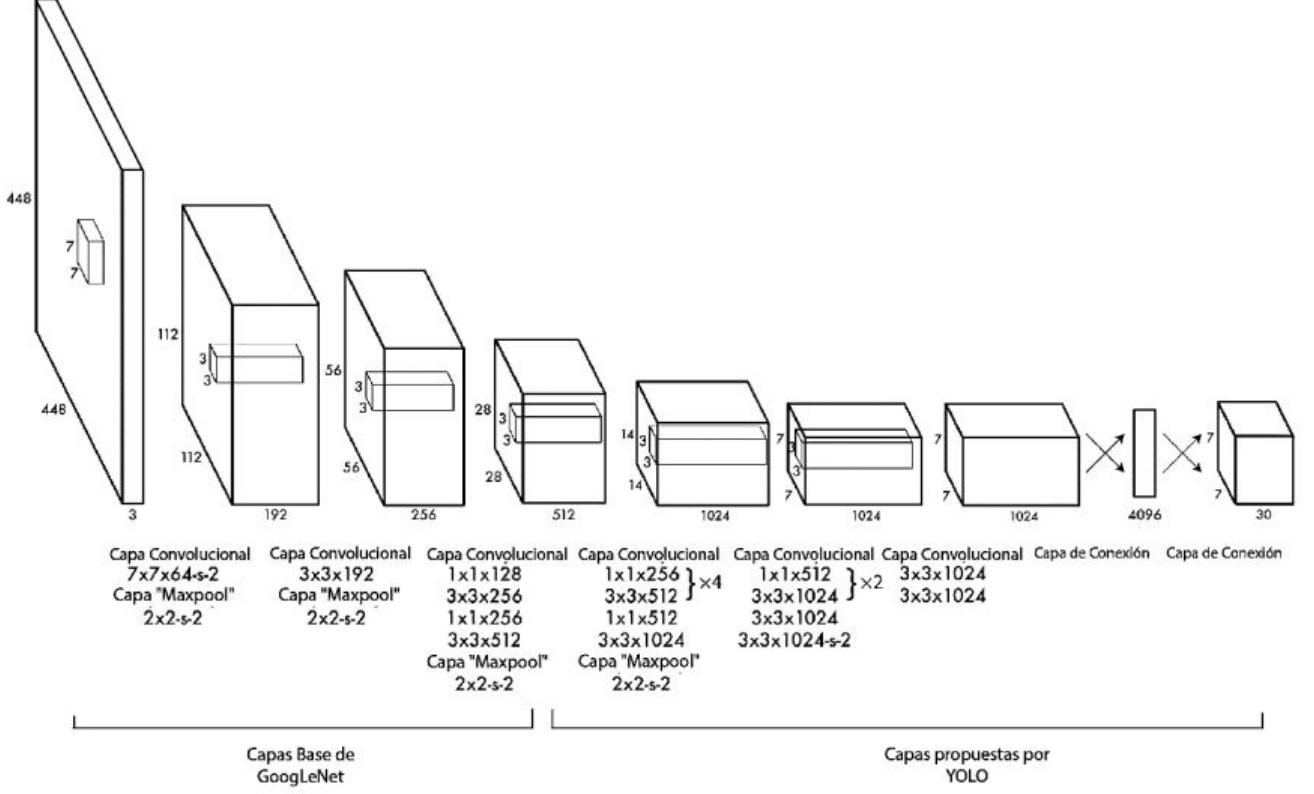


Figura 1: Arquitectura de la red neuronal YOLO. Fuente: Tomada de [9].

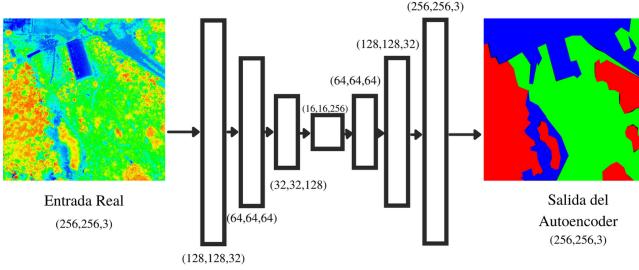


Figura 2: Arquitectura Autoencoder para la segmentación de imágenes. Sobre cada capa se observa las dimensiones de salida del procesamiento. Fuente: elaboración propia.

III-B. Software

Para la ejecución de un plan de vuelo automático se utilizó los aplicativos de CTRL+Parrot [23], que permiten la integración con Pix4Dcapture [24], haciendo posible la planificación del vuelo y el monitoreo en tiempo real cuando este se esté ejecutando. Un ejemplo de la interfaz gráfica de Pix4D puede verse en la Figura 5. En una etapa posterior a la toma de imágenes aéreas, se procedió a la creación de ortomosaicos en las diferentes bandas multiespectrales disponibles. Para este fin se utilizó el software Agisoft Metashape [25], que permite su uso de forma gratuita para propósitos académicos [26].

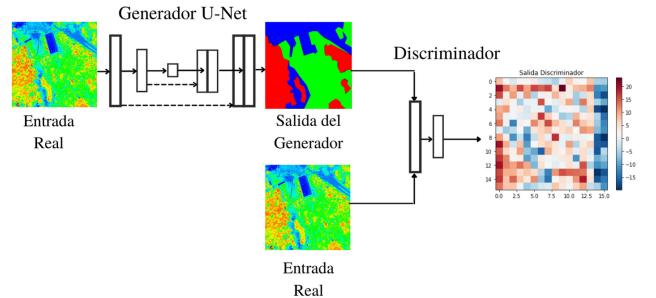


Figura 3: Arquitectura de red neuronal adversaria generativa condicional Pix2Pix. Fuente: elaboración propia a partir de [16].

III-C. Base de datos

Se compone principalmente de imágenes obtenidas a partir de la ejecución de un plan de vuelo. En cada plan de vuelo Se obtuvieron hasta 400 imágenes que se someten a una etapa de procesado, en la que se aplican funciones de alineación, optimización de puntos coincidentes, creación de nube de puntos y finalmente la obtención de un ortomosaico.

El ortomosaico [27] es un producto de imagen fotográficamente ortorrectificado, organizado como mosaico a partir de una colección de imágenes, donde se corrige la distorsión geométrica y se realiza un balance de color de las imágenes para producir un dataset de mosaico continuo. Un ejemplo de ortomosaico de la finca La Mina, en el municipio de Buesaco,

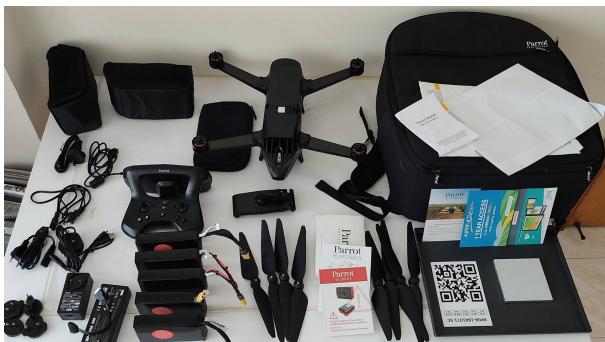


Figura 4: Dron Parrot Bluegrass adquirido por la Universidad de Nariño con todos sus elementos para la correcta ejecución del plan de vuelo. Fuente: elaboración propia.

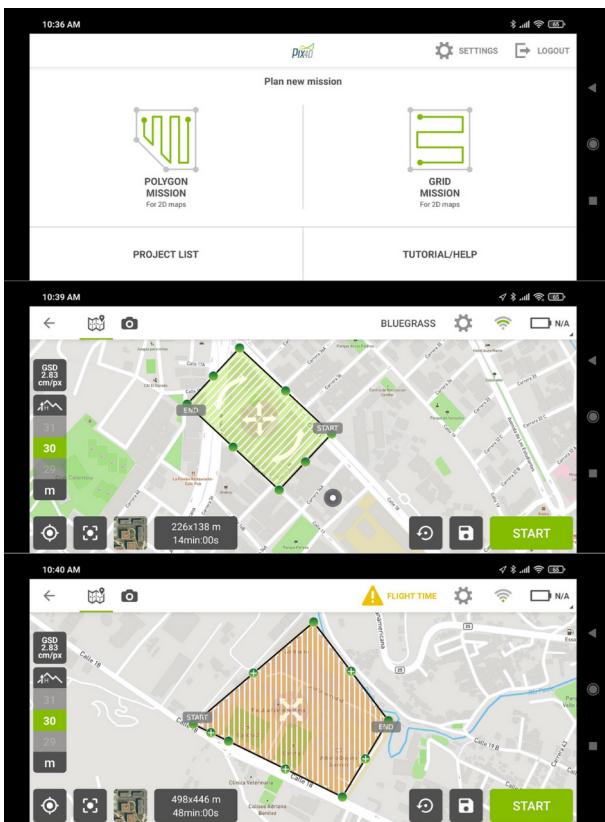


Figura 5: Interfaz Pix4D Capture para planificación de vuelo en las modalidades de polígono y malla. Fuente: elaboración propia.

se muestra en la Figura 6.

Estos ortomosaicos son generados en cuatro bandas multiespectrales que se describen a continuación:

- Banda Verde, longitud de onda de 550nm
- Banda Roja, longitud de onda de 660nm
- Banda Borde Rojo, longitud de onda de 735nm
- Banda Infrarroja, longitud de onda de 790nm

Una representación gráfica de estas bandas en escala de grises se muestra en la Figura 7, acompañada también de la vista en RGB para una mejor comprensión del terreno fotografiado.

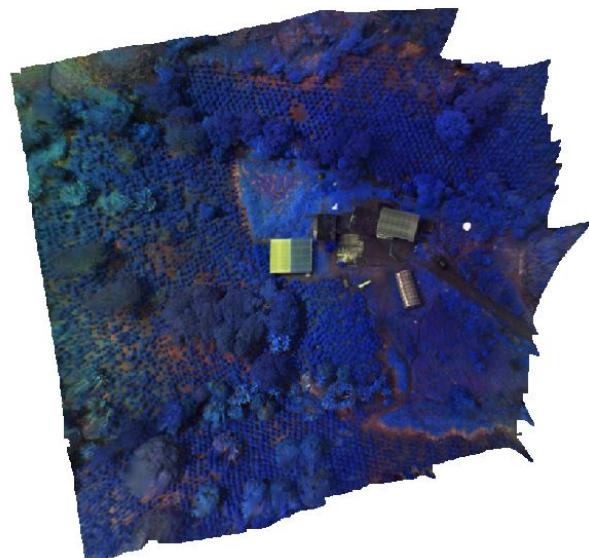


Figura 6: Ortomosaico sin procesamiento generado para la finca La Mina en el municipio de Buesaco, Nariño. Fuente: elaboración propia.

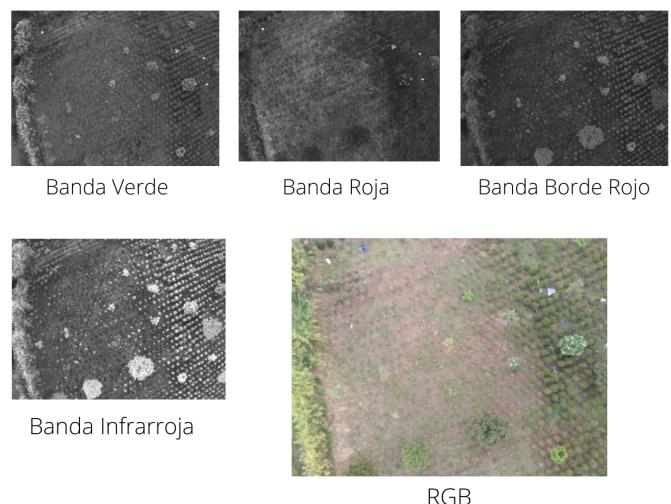


Figura 7: Fotografías en bandas multiespectrales tomadas en la Finca Piloto, ubicada en el municipio de Buesaco, Nariño. Fuente: elaboración propia.

III-C1. Mapa con índice NDVI: El índice de vegetación de diferencia normalizada o NDVI, por sus siglas en inglés, es el más usado para el análisis remoto de vegetación [28]. La reflectancia espectral de la vegetación a través de las bandas sirve como un indicador de la presencia de plantas o árboles y de su estado general. Matemáticamente, el índice NDVI esta conformado por la combinación de la banda infrarroja (NIR) y la banda roja (RED), tal como se describe a continuación

$$NDVI = \frac{NIR - RED}{NIR + RED} . \quad (1)$$

El índice NDVI ayuda a diferenciar la vegetación de otros tipos de cobertura terrestre y determinar su estado general.

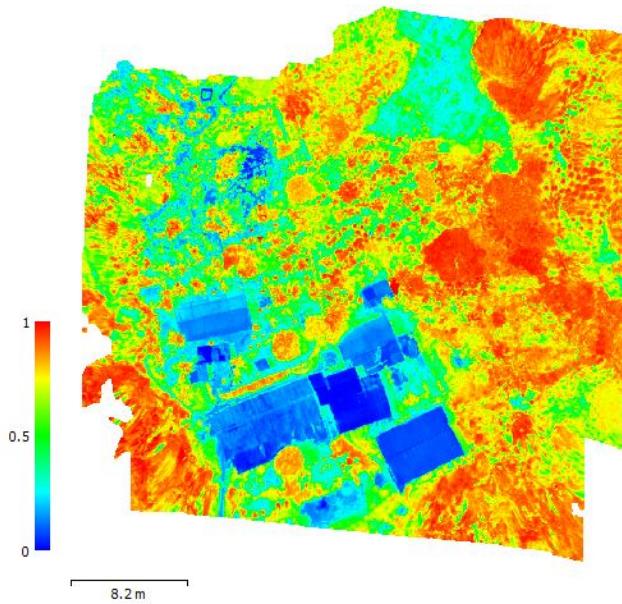


Figura 8: Ortomosaico NDVI de la finca El Arrayán, en el municipio de Buesaco, Nariño. Fuente: elaboración propia.

También permite identificar, a través de la comparación de índices con diferencias temporales, los cambios anormales en el proceso de crecimiento de las plantas. Este índice funciona comparando la cantidad de luz roja visible absorbida y la luz infrarroja cercana reflejada, dado que el pigmento de clorofila en una planta sana absorbe la mayor parte de la luz roja visible, mientras que la estructura celular refleja la mayor parte de la luz infrarroja cercana. Significa que la alta actividad fotosintética tendrá menos reflectancia en la banda roja y mayor reflectancia en el infrarrojo cercano. Al observar como se comparan estos valores entre si, es posible detectar y analizar de manera eficaz la cubierta vegetal por separado de otros tipos de cobertura natural de la tierra [29]. Un ejemplo de mapa con índice NDVI se muestra en la Figura 8, donde tiene un mapa de color normalizado entre 0 y 1 que representan el estado vital de la vegetación, siendo 0 la representación de una zona artificial o de nula actividad fotosintética y 1 una zona con una alta vitalidad.

III-C2. Modelo digital de elevación: Es una representación visual y numérica de los valores de altura en un mapa que permite identificar las formas del relieve [30]. Un ejemplo de este tipo de mapa se observa en la Figura 9. Los valores que contiene la imagen están referenciados desde la altura promedio de sobrevuelo del drone Parrot Bluegrass, que es de 30 m sobre el área de despegue, siendo los colores más cálidos las zonas con mayor altura, y los colores fríos o azules las zonas de menor altura.

IV. RESULTADOS Y ANÁLISIS

IV-A. Resultados Detección de Objetos

En una primera experimentación, se hizo el entrenamiento del modelo YOLOv5 para realizar un conteo de los arbustos de café presentes en la imagen. En este proceso se hizo una comparación del desempeño del modelo con distintos tipos

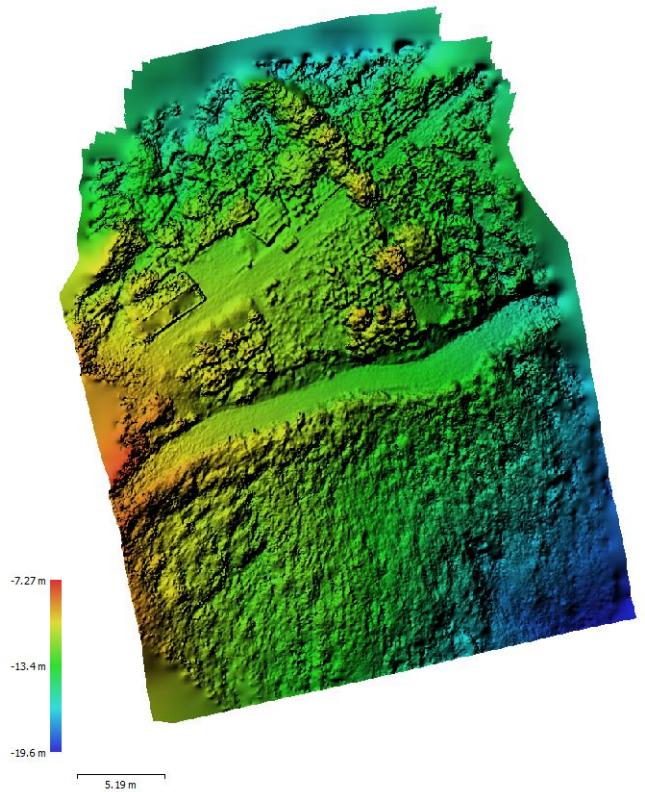


Figura 9: Modelo digital de elevación para la finca Loma Linda, en el municipio de Buesaco, Nariño. Fuente: elaboración propia.

de imagen a la entrada de la red neuronal, en específico, probando como entrada cada una de las capas de las imágenes multiespectrales (verde, roja, borde rojo, infrarroja y RGB), y obteniendo el número de arbustos de café detectados. Adicionalmente, también se obtienen las coordenadas de cada arbusto de café en la imagen.

Las imágenes utilizadas para esta experimentación son porciones de los ortomosaicos en las cuatro diferentes bandas, banda verde, roja, borde rojo e infrarrojo. Las imágenes tienen dimensiones de $416 \times 416 \times 1$ píxeles.

Para este proceso se trabajó con un *dataset* de 1100 imágenes en cada banda, dejando un 10 % de las imágenes fuera del entrenamiento para obtener métricas de desempeño. El porcentaje de acierto del experimento se calculó como

$$A = 100 \cdot \left(1 - \frac{|n_a - n_p|}{n_a} \right) \quad , \quad (2)$$

donde A el porcentaje de acierto, n_a el número real de arbustos de café presentes en la imagen, obtenido por el etiquetado previamente hecho y n_p el número de arbustos de café presentes en la inferencia de la red neuronal. Los resultados de esta experimentación se resumen en la Tabla I. En ella se observa que el mínimo valor de acierto está en 60,4 % para la capa Verde y un máximo valor de acierto esta en la entrada RGB, con un valor de 85,6 %.

En la parte de la detección espacial del arbusto de café en la imagen, se tiene como métrica de comparación al índice

Tabla I: Porcentaje de acierto A para el conteo de arbustos de café por cada banda multiespectral.

Banda	Acierto A (%)
RGB	85.6
Verde	60.4
Roja	71
Borde Rojo	64.8
Infrarrojo	64.2

IoU [31], que permite evaluar el grado de superposición de dos áreas de detección, comparando al área real o etiqueta y el área inferida por la red neuronal. El índice IoU se calcula por medio de la expresión

$$IoU = \frac{B_r \cap B_p}{B_r \cup B_p}, \quad (3)$$

siendo B_r el área real del arbusto de café en la imagen, obtenido del etiquetado hecho en la imagen y B_p el área de la inferencia del arbusto de café para la red neuronal. Este índice se encuentra en el rango de $[0, 1]$, siendo 0 cuando no existe coincidencias de las áreas y 1 cuando la coincidencia es máxima. El resultado comparativo de este índice en las diferentes bandas se observa en la Figura 10.

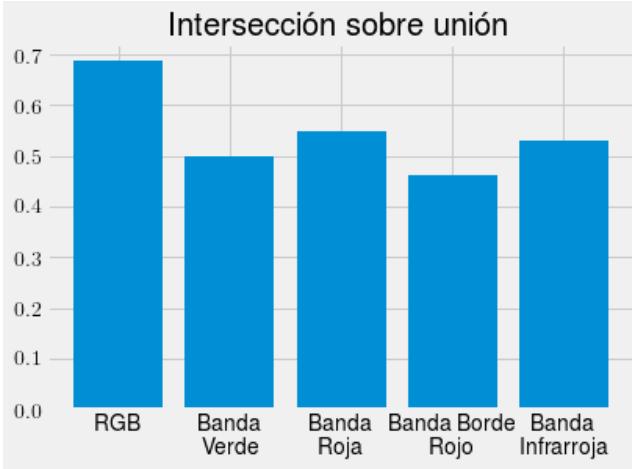


Figura 10: Métrica de desempeño para detección de arbustos de café. Intersección sobre unión

Para esta experimentación, la entrada RGB tiene el mayor índice de IoU , seguido de la banda roja, infrarroja, banda verde y por último la banda de borde rojo. Estos resultados coinciden parcialmente con los resultados del acierto en el conteo de arbustos de café. Una muestra de la inferencia del modelo YOLOv5 en la detección de arbustos de café se observa en la Figura 11.

Con los arbustos de café identificados se puede hacer un cálculo aproximado de la producción. A continuación se muestra un ejemplo de metodología de cálculo que se basa en las mediciones realizadas en campo mediante un muestreo. Este proceso es basado en el trabajo hecho en [32].

El proceso consiste en escoger una muestra del cultivo y generar estadísticas representativas. El cálculo de la producción de café se basa en los registros de floración, por medio de

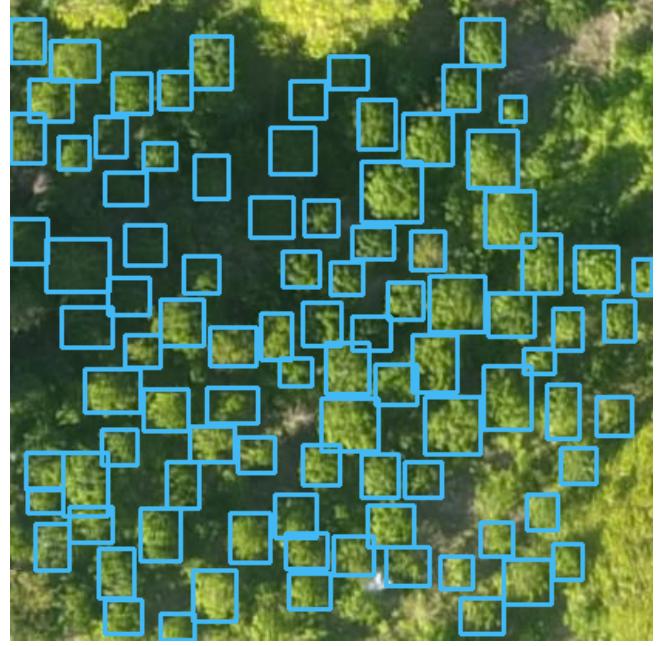


Figura 11: Inferencia gráfica del modelo YOLOv5 para la detección de arbustos de café para una imagen de entrada RGB. Fuente: elaboración propia.

$$CCT = N_F N_R N_p P_C M_c, \quad (4)$$

donde CCT es el total en gramos de café cereza a recolectar, N_F es el promedio de botones florales en cada rama para los arbustos de café, N_R es el número promedio de ramas en los arbustos de café, N_p es la cantidad de arbustos de café, P_C es el porcentaje de cuajamiento (los valores recomendados para un valor mínimo y máximo son respectivamente 50 % y 85 %). Por ultimo M_c representa el valor en gramos de un grano de café, donde el valor teórico promedio mostrado en [32] es de 1,8 g.

Para ilustrar el proceso, se puede hacer el cálculo particular con la detección realizada en la Figura 11, donde se detectó 93 arbustos de café siendo este el valor de N_p , además actuando bajo el supuesto de que los arbustos de café en promedio son iguales al mostrado en la Figura 12, se puede estimar aproximadamente $N_F = 12$ y $N_R = 20$, con estos parámetros obtendría una producción de café entre 20,1 kg y 34,1 kg.

IV-B. Resultados Segmentación

En esta experimentación se definen tres áreas generales para ser segmentadas en una imagen de cultivo de café. Estas áreas se definen de la siguiente manera.

- **Baja.** Que se refiere a la clasificación que se le da a las áreas con baja actividad fotosintética en la imagen, y que pertenecen a vegetación en mal estado o infraestructura artificial, como edificaciones o carreteras.
- **Media.** Esta clasificación se otorga a la zona con vegetación en estado medio de vitalidad.
- **Alta:** Esta clasificación se otorga a la zona con una vegetación en un estado óptimo de vitalidad, porque tiene



Figura 12: Arbusto de Café en la finca Loma Gorda del municipio de Buesaco, Nariño. Fuente: elaboración propia.

una alta actividad fotosintética y se puede inferir que corresponde a una vegetación saludable.

La identificación de estas zonas fueron hechas con la asesoría de talento humano en el área de ingeniería agroforestal teniendo como insumo el ortomosaico del mapa NDVI.

Teniendo en cuenta estas áreas de clasificación, se construyó un dataset de 3210 imágenes, de las cuales 2568 (80 %) fueron destinadas para entrenamiento y 642 (20 %) para evaluación de desempeño de los algoritmos. Las etiquetas del dataset fueron construidas manualmente con el software de Labelme [33]. Las imágenes de entrada del algoritmo corresponden a las imágenes NDVI (ver Figura 8) y a un arreglo matricial de las 4 bandas multiespectrales, cumpliendo las características explicadas en la sección II-B. Los algoritmos probados con este arreglo tendrán el sufijo *Bandas*.

En este trabajo se utiliza dos métricas o índices para medir el desempeño de los algoritmos programados, estos índices son el error medio cuadrático y el índice de similitud estructural. La medida del error medio cuadrático *MSE* [34] otorga información cuantificada sobre la diferencia entre el píxel objetivo y el estimado. Cuando dos imágenes a comparar tienen un error medio cuadrático bajo se puede considerar que las imágenes son casi idénticas, es por eso que el valor de esta métrica se considera mejor entre mas cercano a cero se encuentre. Esta métrica está definida como

$$MSE = \frac{1}{m} \sum_{i=1}^m (\hat{Y}_i - Y_i)^2 , \quad (5)$$

donde m es la cantidad de muestras en el conjunto de testeo, Y_i es el valor real del píxel i y \hat{Y}_i es el valor estimado.

Como segunda métrica se estableció el índice de similitud estructural *SSIM* [35], una métrica de calidad de imagen que evalúa el impacto visual de la luminancia, contraste y estructura en dos imágenes. Se decidió utilizar esta medida también puesto que con el uso de la esteganografía en campos de ciberseguridad se ha demostrado que el SSIM puede obviar y otorgar información sustancial con respecto al MSE, donde es aprovechado para ocultar mensajes dentro de otros objetos [36]. Por esto se ha definido tener dos métricas de medición que se complementen en su metodología de comparación de imágenes. El SSIM está definido por la ecuación

$$SSIM(x, y) = \frac{(2u_x u_y + C_1)(2\sigma_{xy} + C_2)}{(u_x^2 + u_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} , \quad (6)$$

donde x y y representan una ventana de tamaño $N \times N$ de las imágenes a comparar, u_x y u_y son los promedios de x y y , σ^2 es la varianza y σ_{xy} es la covarianza de x y y . C_1 y C_2 son valores para estabilizar la división con denominador débil o cercano a 0.

Como tercera métrica se definió el porcentaje promedio de acierto en la segmentación de las áreas A_s , definida como

$$A_s = \frac{100}{m} \sum_{i=1}^m \left(1 - \frac{|z - \hat{z}|}{z} \right) , \quad (7)$$

donde el subíndice s representa el área para la cual se calcula, m representa la cantidad de muestras en el conjunto de testeo, z es el valor real del área segmentada y \hat{z} es el valor estimado.

Los resultados comparativos para el error medio cuadrático *MSE* en el conjunto de testeo se muestran en la Figura 13. En ella se dibuja el valor del error para cada muestra en cada uno de los cuatro algoritmos implementados. En el lado derecho se observa unos marcadores que representan el desempeño promedio para cada algoritmo, encontrando que el modelo de Pix2Pix, que es el que usa como imagen de entrada el índice NDVI, es el que presenta menos error en la estimación, siendo 10,2 veces menor que el algoritmo con mayor error, el Autoencoder-Bandas. Sin embargo, el algoritmo de Autoencoder, con entrada NDVI, se encuentra a 5,1 veces el error promedio mínimo de Pix2Pix. Los algoritmos que tienen como entrada el arreglo matricial de las bandas multiespectrales, Pix2Pix-Bandas y Autoencoder-Bandas, presentan un menor

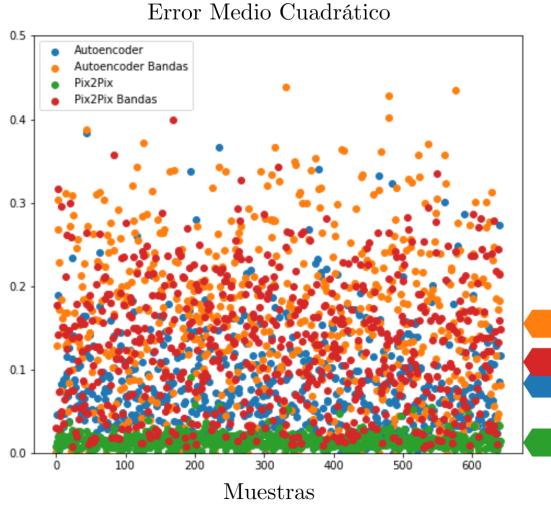


Figura 13: Valor de MSE para la segmentación de imágenes en el conjunto de testeo. En el lado derecho de la figura se ubican marcadores del valor promedio de cada algoritmo. Fuente: elaboración propia.

desempeño, pero aun así sus valores son cercanos a cero, mostrando que tienen también un buen resultado. Estos valores pueden ser vistos en la Tabla II

Tabla II: Comparativa de desempeño para algoritmos con el valor promedio MSE y desviación estándar σ para el conjunto de prueba.

Algoritmo	MSE	σ
Autoencoder	0.0781	0.0633
Autoencoder-Bandas	0.1566	0.0952
Pix2Pix	0.0153	0.0087
Pix2Pix-Bandas	0.1284	0.0761

Es interesante también notar que el algoritmo de Pix2Pix es el que presenta menor desviación estándar, siendo 10,9 veces menor que Autoencoder-Bandas, 8,7 veces menor que Pix2Pix-Bandas y 7,2 veces menor que Autoencoder. Esta dispersión en los datos puede ser explicada por la arquitectura de los algoritmos implementados y el insumo de entrada que poseen. Es decir, los algoritmos Pix2Pix y Autoencoder que poseen como entrada las imágenes NDVI son los de menor desviación estándar, explicándose a que estos tienen como entrada una imagen preprocesada que resalta las áreas de interés para este trabajo. Por su parte los algoritmos Autoencoder-Bandas y Pix2Pix-Bandas en su proceso de entrenamiento deben encontrar una combinación o procesamiento de los valores de entrada que les permita inferir información similar a la mostrada en las imágenes NDVI. Si bien la entrada de bandas posee mucha mas información, dado que es el estado original de las imágenes multiespectrales, la combinación correcta de esta información puede no ser fácilmente encontrado por el proceso de entrenamiento.

Los resultados comparativos para el índice de similitud estructural $SSIM$ se muestran en la Figura 14. Este índice toma valores en el rango $[-1, 1]$, siendo -1 el valor que

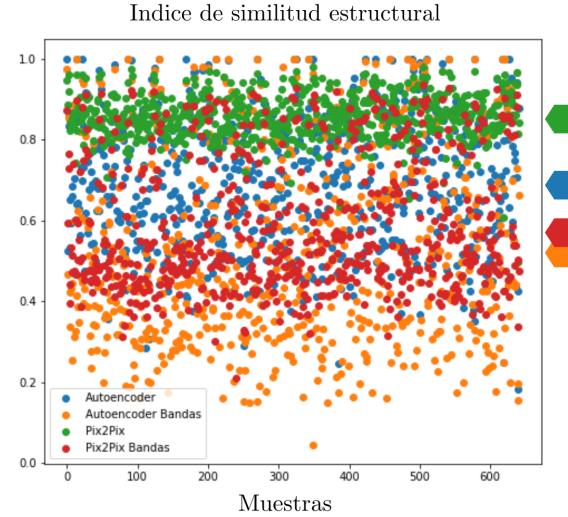


Figura 14: Valor de $SSIM$ para la segmentación de imágenes en el conjunto de testeo. En el lado derecho de la figura se ubican marcadores del valor promedio de cada algoritmo. Fuente: elaboración propia.

indica que no hay similitud entre las imágenes y 1 el valor que indica que las imágenes son idénticas. Al lado derecho de la figura se ubican marcadores en el valor promedio para las muestras del conjunto de testeo. Se observa que el algoritmo Pix2Pix tiene el mejor desempeño con respecto a este índice, seguido por el Autoencoder y con valores cercanos los algoritmos con entrada del arreglo matricial de las bandas multiespectrales. Los valores promedio y desviación estándar para las experimentaciones pueden ser vistos en la Tabla III. Los datos son consistentes con lo mostrado en los resultados de MSE , puesto que los algoritmos que tienen entrada NDVI presentan mejores resultados para este índice de similitud estructural.

Tabla III: Comparativa de desempeño para algoritmos con el valor promedio $SSIM$ y desviación estándar σ para el conjunto de prueba.

Algoritmo	$SSIM$	σ
Autoencoder	0.6921	0.1622
Autoencoder-Bandas	0.5359	0.2333
Pix2Pix	0.8590	0.0553
Pix2Pix-Bandas	0.5831	0.1603

Los resultados comparativos para el porcentaje de acierto A_s , por cada área de segmentación, se resumen en la Tabla IV. Se observa que en promedio el porcentaje para el algoritmo Pix2Pix-Bandas (58,9 %) resulta ser el más bajo, mientras que el algoritmo Pix2Pix con entrada NDVI es el más alto (86,8 %). Con respecto a esta métrica, se puede decir también que el acierto en la segmentación de la zona rotulada como baja y media es más alto en comparación con el resultado logrado para la zona alta. Analizando este hallazgo, es posible explicarlo debido a que en las imágenes recolectadas esta es la zona de menor continuidad, presenta varias interrupciones. Es decir, dentro de la zona alta es posible que exista característi-

cas de zona media o baja, ya que los espacios entre arbustos o cualquier otra vegetación por lo general son áreas de tierra descubierta sin vegetación. Sin embargo, lo mas conveniente para toda esta área es segmentar como clase alta, puesto que no interesa generar alertas para un proceso de beneficio sobre el terreno entre arbustos debido a que eso afectaría al arbusto de café que ya se encuentra en una buena clasificación.

Tabla IV: Porcentaje de acierto A_s para la segmentación de zonas.

Algoritmo	Acierto por zona A_s (%)			
	Baja	Media	Alta	Promedio
Autoencoder	85.2	77.3	68.8	77.1
Autoencoder-Bandas	68.4	80.1	36.2	61.6
Pix2Pix	96.8	95.3	82.1	91.4
Pix2Pix-Bandas	73.2	74.9	34.1	60.7

Una imagen comparativa de los resultados obtenidos con los diferentes algoritmos se muestra en la Figura 15. En ella se observan, en orden de izquierda a derecha, la segmentación lograda con Autoencoder, Autoencoder-Bandas, Pix2Pix y Pix2Pix-Bandas, seguido de la imagen etiqueta, el procesado NDVI y una muestra en escala de grises de la banda Infrarroja. Visualmente, el desempeño de los algoritmos con respecto a la imagen de la etiqueta resulta ser muy aproximado y podría catalogarse como acertado. Sin embargo, para un análisis más exacto se deben tener en cuenta las métricas calculadas. Se observa también, que los resultados más exactos se obtienen en general para los algoritmos con entrada igual al índice NDVI (Autoencoder y Pix2Pix).

IV-C. Aplicativo Web

Para la divulgación de esta investigación se construyó un aplicativo web interactivo en python, con el framework streamlit [37], que permite la interacción con el dataset y los modelos implementados. Una muestra visual de este aplicativo web puede ser visto en la Figura 16. El código de este aplicativo web es de libre uso y esta alojado en [38]. También existe una página web para el apoyo en la divulgación de este trabajo disponible en [39]. Este aplicativo y los resultados obtenidos en la investigación hacen parte también del Sistema de Información Integral hacia la Estandarización de los Procesos de Producción de Cafés Especiales en el Municipio de Buesaco - CAFIOT [40].

V. CONCLUSIONES

Esta investigación apoya el proceso de estandarización para los cultivos de café en el municipio de Buesaco, Nariño. Brinda herramientas tecnológicas que soportan la toma de decisiones en la gestión de los cultivos, con el objetivo de monitorear y estimar el rendimiento de los cultivos, a través de instrumentos no tradicionales de adquisición y procesamiento de información, como lo son las imágenes aéreas multiespectrales.

Por medio de la aplicación de técnicas de identificación de arbustos de café, se puede hacer inferencia sobre la cantidad de café estimada o esperada en producción. Y por medio de la aplicación de técnicas de segmentación, se puede establecer

un monitoreo automático del estado general de los cultivos, a través de las clases Baja, Media y Alta, que fueron definidas en esta investigación. Con esta información, los agricultores pueden establecer estrategias de mejora de la producción y cuidado focalizado por áreas, según el estado de vitalidad de las plantas. A futuro se espera que estas ayudas contribuyan al planteamiento de estrategias de optimización de recursos e insumos.

Con respecto a los algoritmos aquí estudiados, se nota las fortalezas logradas en aplicaciones agrícolas y específicamente para los cultivos de café de la región de Buesaco, en el departamento de Nariño. Los resultados obtenidos en la detección de objetos muestra que, para reconocimiento de arbustos de café, es más eficiente trabajar con imágenes RGB, con un acierto promedio de 85,6 %, seguido de la banda roja con un valor de 71 %; mientras que para la segmentación de imágenes clasificando áreas en niveles de calidad vegetal, los mejores resultados se obtuvieron con el algoritmo Pix2Pix usando las imágenes NDVI, llegando a ser hasta 10 veces mejor que los otros algoritmos en las métricas de rendimiento.

Como aporte adicional a la investigación, se construyó una banco de datos con toda la información de imágenes recolectada en las fincas prototipo, en los cuales también se tiene información geográfica y visual de los planes de vuelo realizados durante la investigación. Todo el trabajo aquí realizado, se encuentra documentado, versionado y de libre uso en [41]. Esto permitirá replicar y rehusar materiales para mejoras y proyectos futuros en este campo de investigación, para quienes quieran tomar este trabajo como referencia.

AGRADECIMIENTOS

Los autores agradecen a la Universidad de Nariño, al Grupo de Investigación en Ingeniería Eléctrica y Electrónica - GIIIE y al programa de Maestría en Ingeniería Electrónica - MaIE, por el apoyo académico y financiero recibido. Esta investigación fue financiada por el proyecto “Sistema de Información Integral hacia la Estandarización de los Procesos de Producción de Cafés Especiales en el Municipio de Buesaco”, Contrato de Financiamiento de Recuperación Contingente No. 80740-214-2019, de la Convocatoria 818 de Colciencias, Código 65472.

También se extiende un agradecimiento especial a los caficultores del municipio de Buesaco, departamento de Nariño, quienes han colaborado de manera incondicional en las etapas de recolección de las imágenes aéreas y han brindado todo su conocimiento y experiencia en beneficio de esta investigación, sin su valioso aporte, este trabajo no hubiera podido ser posible.

REFERENCIAS

- [1] C. F. Carvalho, S. M. Carvalho, and B. Souza, “Coffee,” in *Natural Enemies of Insect Pests in Neotropical Agroecosystems*. Springer, 2019, pp. 277–291.
- [2] G. Puerta, “Calidad en taza de las variedades de cafe arabica cultivadas en colombia,” 1998.
- [3] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose, and W. An, “Detecting non-hardhat-use by a deep learning method from far-field surveillance videos,” *Automation in Construction*, vol. 85, pp. 1–9, 2018.
- [4] J. Torres, *DEEP LEARNING Introducción práctica con Keras*. Lulu.com, 2018.

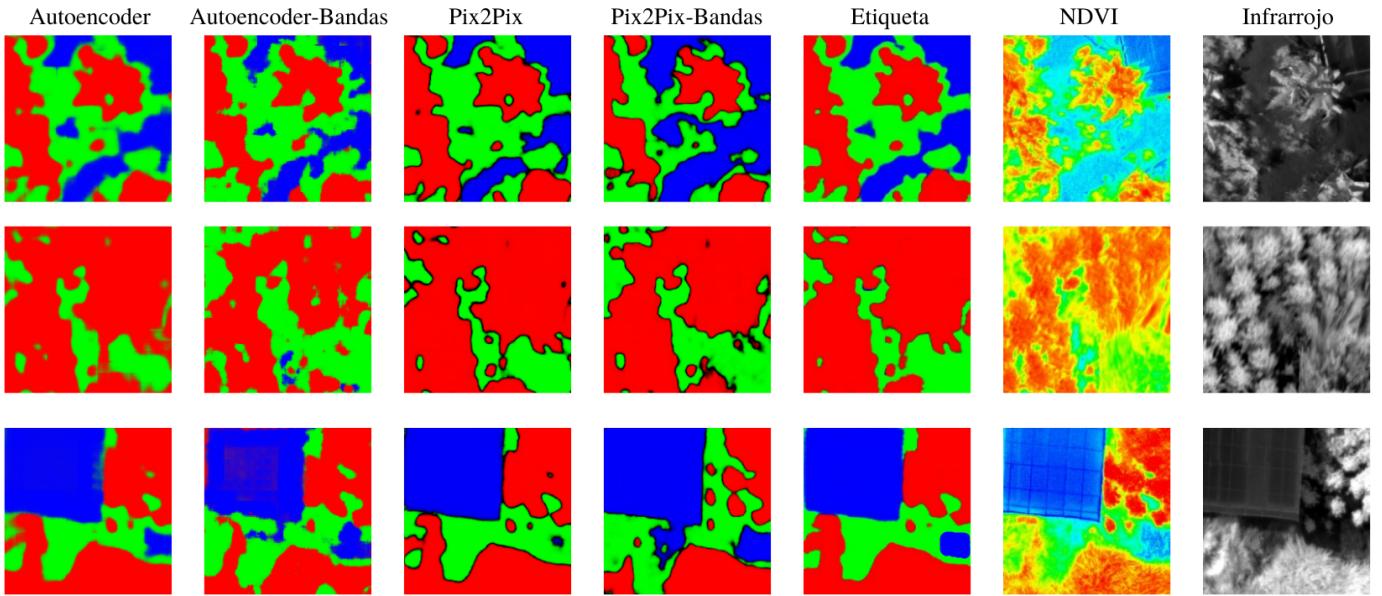


Figura 15: Resultados de algoritmos para segmentación en zona Baja (color azul), zona Media (color verde) y zona Alta (color rojo). Fuente: elaboración propia.



Figura 16: Muestra de aplicativo web para la interacción con los modelos de inteligencia artificial en la detección y segmentación de imágenes, escrito en python con streamlit. Fuente: elaboración propia.

- [5] J. Durán Suárez, "Redes neuronales convolucionales en r. reconocimiento de caracteres escritos a mano," 2017.
- [6] E. M. Bodero, M. P. Lopez, A. E. Congacha, E. E. Cajamarca, and C. H. Morales, "Google colaboratory como alternativa para el procesamiento de una red neuronal convolucional," *Revista Espacios*, vol. 41, no. 07, 2020.
- [7] G. Jocher, A. Stoken, J. Borovec, A. Chaurasia, T. Xie, C. Liu, V. Abhiram, T. Laughing *et al.*, "ultralytics/yolov5: v5. 0-yolov5-p6 1280 models," *AWS, Supervise.ly and YouTube integrations*, 2021.
- [8] N. Ketkar, "Introduction to pytorch," in *Deep learning with python*. Springer, 2017, pp. 195–208.
- [9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [10] P. Ballester and R. M. Araujo, "On the performance of googlenet and alexnet applied to sketches," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [11] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.
- [12] R.-C. Chen *et al.*, "Automatic license plate recognition via sliding-

- window darknet-yolo deep learning,” *Image and Vision Computing*, vol. 87, pp. 47–56, 2019.
- [13] M. Thoma, “A survey of semantic segmentation,” *arXiv preprint arXiv:1602.06541*, 2016.
- [14] D. De Geus, P. Meletis, and G. Dubbelman, “Panoptic segmentation with a joint semantic and instance segmentation network,” *arXiv preprint arXiv:1809.02110*, 2018.
- [15] J. Feng and Z.-H. Zhou, “Autoencoder by forest,” in *Thirty-second AAAI conference on artificial intelligence*, 2018.
- [16] Y. Qu, Y. Chen, J. Huang, and Y. Xie, “Enhanced pix2pix dehazing network,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8160–8168.
- [17] H. Bendea, P. Boccardo, S. Dequal, F. G. Tonolo, D. Marenchino, and M. Piras, “Low cost uav for post-disaster assessment,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, no. B8, pp. 1373–1379, 2008.
- [18] F. Chiabrandi, F. Nex, D. Piatti, and F. Rinaudo, “Uav and rpv systems for photogrammetric surveys in archaeological areas: two tests in the piedmont region (italy),” *Journal of Archaeological Science*, vol. 38, no. 3, pp. 697–710, 2011.
- [19] W. Aguilar, R. Costa Castelló, C. Angulo Bahón, and L. Molina, “Control autónomo de cuadricópteros para seguimiento de trayectorias,” in *Revista Digital Congreso de Ciencia y Tecnología: Memorias. Sesiones Técnicas*, 2014, pp. 140–144.
- [20] J. L. C. Villalobos, A. E. J. Menchaca, A. L. Terrazas, and F. M. Vázquez, “El mundo de los drones: tipos de drones y sus principales usos,” *FINGUACH. Revista de Investigación Científica de la Facultad de Ingeniería de la Universidad Autónoma de Chihuahua*, vol. 4, no. 14, pp. 3–5, 2018.
- [21] Parrot bluegrass fields. [Online]. Available: <https://tycgis.com/parrot-bluegrass/>
- [22] Minciencias. Sistema de información integral hacia la estandarización de los procesos de producción de cafés especiales en el municipio de buesaco. [Online]. Available: <http://www.sednarino.gov.co/SEDNARINO12/index.php/es/55-sednarino/noticias/4050-sistema-de-informacion-integral-hacia-la-estandardizacion-de-los-procesos-de-produccion-de-cafes-especiales-en-el-municipio-de-buesaco>
- [23] Pix4D. Ctrl+parrot. [Online]. Available: <https://play.google.com/store/apps/details?id=com.pix4d.pluginparrot>
- [24] —. Pix4dcapture. [Online]. Available: <https://www.pix4d.com/es/producto/pix4dcapture>
- [25] W. T. Tinkham and N. C. Swayze, “Influence of agisoft metashape parameters on uas structure from motion individual tree detection from canopy height models,” *Forests*, vol. 12, no. 2, p. 250, 2021.
- [26] Agisoft. Agisoft metashape licesing. [Online]. Available: <https://www.agisoft.com/buy/licensing-options/>
- [27] J. O. Escalante Torrado, H. Porras Díaz *et al.*, “Ortomasicos y modelos digitales de elevación generados a partir de imágenes tomadas con sistemas uav,” *Tecnura*, vol. 20, no. 50, pp. 119–140, 2016.
- [28] R. S. DeFries and J. Townshend, “Ndvi-derived land cover classifications at a global scale,” *International Journal of Remote Sensing*, vol. 15, no. 17, pp. 3567–3586, 1994.
- [29] C. M. D. Rafael, “Monitoreo de bosques utilizando ndvi rededge de rapideye,” *Instrucciones Generales para los Autores*, vol. 10, pp. 58–71, 2013.
- [30] J. Fallas, “Modelos digitales de elevación: Teoría, métodos de interpolación y aplicaciones,” *Escuela de Ciencias Ambientales. Universidad Nacional, Costa Rica*, 2007.
- [31] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 658–666.
- [32] J. Rendón, J. Arcila, and E. Montoya, “Estimación de la producción de café con base en los registros de floración,” 2008.
- [33] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, “Labelme: a database and web-based tool for image annotation,” *International journal of computer vision*, vol. 77, no. 1-3, pp. 157–173, 2008.
- [34] T. Chai and R. R. Draxler, “Root mean square error (rmse) or mean absolute error (mae),” *Geoscientific Model Development Discussions*, vol. 7, no. 1, pp. 1525–1534, 2014.
- [35] A. Hore and D. Ziou, “Image quality metrics: Psnr vs. ssim,” in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 2366–2369.
- [36] C. A. Angulo, S. M. Ocampo, and L. H. Blandon, “Una mirada a la esteganografía,” *Scientia et technica*, vol. 13, no. 37, pp. 421–426, 2007.
- [37] P. Singh, *Deploy machine learning models to production: with flask, streamlit, docker, and kubernetes on google cloud platform*. Apress, 2021.
- [38] A. Insuasty, “vision-computarizada-cultivos-cafe,” <https://github.com/AndresInsuasty/vision-computarizada-cultivos-cafe.git>, 2021.
- [39] —, “Sistema de visión computarizada con inteligencia artificial para cultivos de café en el sur occidente colombiano,” <https://andresinsuasty.com/tesis.html>, 2021.
- [40] U. Nariño, “Cafiot udifar,” <http://cafiot.udifar.edu.co:81/>, 2021.
- [41] A. Insuasty, “Repositorio de código,” <https://github.com/AndresInsuasty/tesis-cafes-especiales.git>, 2021.