

**MODELO DE PREDICCIÓN DEL RENDIMIENTO DE UN CULTIVO DE
CAFÉ MEDIANTE ANÁLISIS DE IMÁGENES AÉREAS
MULTIESPECTRALES**



**Universidad
del Cauca**

Julián Andrés Bolaños Ortega

Tesis de maestría en Ingeniería Telemática

Director:

PhD(C). Liseth Viviana Campo

Co Director:

PhD. Juan Carlos Corrales

Asesor:

PhD. Cristhian Figueroa

Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Telemática
Línea de Investigación e-@mbiente
Popayán, Enero de 2023

JULIAN ANDRES BOLAÑOS ORTEGA

**MODELO DE PREDICCIÓN DEL RENDIMIENTO DE UN CULTIVO DE
CAFÉ MEDIANTE ANÁLISIS DE IMÁGENES AÉREAS
MULTIESPECTRALES**

**Tesis presentada en a la facultad de Ingeniería Electrónica y
Telecomunicaciones de la Universidad del Cauca para la obtención del título
de**

Magister En:
Ingeniería Telemática

Director:
PhD(C). Liseth Viviana Campo
Co Director:
PhD. Juan Carlos Corrales
Asesor:
PhD. Cristhian Figueroa

Popayán
2023

Agradecimientos

Dedico este trabajo a mi hijo Sergio por ser la razón de ser de todo lo que hago. A mis papás por ser esas guías tan importantes en el camino y por ser los principales responsables de estar en este punto de mi vida. A mi esposa Paola por todo su amor y apoyo durante la ejecución de este trabajo y por apoyarme en todos los retos que he tenido.

Especialmente agradezco a todas las personas que brindaron desinteresadamente su valioso tiempo y espacio para la recolección de muestras manuales en este trabajo. Sin su apoyo este trabajo no hubiera sido posible, gracias y mil gracias a los profesores Janeth Narvaez, Eyder Castillo, Fernando Goyes, Doly Jannin, Juan Esteban. A los estudiantes de Ingeniería Electrónica Stiven Delgado, Alexis Valdez, Felipe Bolaños. A mi papá Marco Bolaños y mi esposa Paola por estar ahí disponibles dándolo todo para apoyar en la logística y todo lo necesario. A los dueños de la finca San Antonio don William y su ayudante Fabián, y a los administradores de la finca La Sultana de la Universidad del Cauca por el espacio y el apoyo.

Quiero agradecer a mi directora Liseth Campo por su paciencia y su guía durante el desarrollo del proyecto. Quiero agradecer al Ingeniero Juan Carlos Corrales y al Ingeniero Cristhian Figueroa por su apoyo en las etapas iniciales del proyecto que ayudaron a definir las bases de toda la investigación.

Agradezco a la Universidad del Cauca, al departamento de Telemática y a todos los docentes que pusieron un granito de arena para lograr este trabajo de grado.

Resumen Estructurado

Antecedentes: El rendimiento del cultivo es un factor importante para la evaluación de los procesos productivos y la rentabilidad de la industria cafetalera. Con frecuencia, el total de granos por unidad de área se estima utilizando mecanismos manuales como el conteo de granos, ramas, flores u otras medidas físicas. Por lo tanto, la estimación del rendimiento requiere inversión en tiempo y mano de obra; por ello, el proceso no es habitual entre los pequeños cultivos. Esta investigación estudia una alternativa no intrusiva y alcanzable para predecir el rendimiento del cultivo de café a través de imágenes aéreas multiespectrales. La propuesta está diseñada para cultivos pequeños y no técnicos que se controlan mediante fotografías aéreas capturadas por la cámara MapIR como carga útil de un vehículo aéreo no tripulado.

Objetivos: Proponer un método de predicción temprana del rendimiento de cultivos de café basado en características físicas y espectrales obtenidas desde imágenes aéreas.

Métodos: Proponemos un modelo de predicción basado en algoritmos de regresión alimentados por imágenes obtenidas desde aeronaves no tripuladas capturadas en la etapa de floración de cultivo. De esta forma el proyecto pretende calcular el rendimiento de forma temprana.

Resultados: La presente propuesta entregó como resultado un modelo de predicción del rendimiento con una exactitud del 57 % para un modelo de regresión lineal múltiple. También se tiene una aplicación web que permite la obtención del rendimiento manual de forma ágil y dinámica.

Conclusiones: El trabajo demostró que es factible ejecutar una predicción del rendimiento basado en información proveniente solo desde imágenes aéreas multiespectrales obtenidas con herramientas accesibles.

Palabras Clave: Rendimiento de los cultivos; Café; Segmentación de imágenes; multiespectral; MapaIR; Predictor; UAV

Structured Abstract

Antecedentes: The crop yield is an important factor for the evaluation of the productive processes and the profitability of the coffee industry. Frequently, the total grains per unit area is estimated using manual mechanisms such as counting grains, branches, flowers, or other physical measurements. Therefore, yield estimation requires investment in time and labor; for this reason, the process is not common among small crops. This research studies a non-intrusive and achievable alternative to predict coffee crop yield through multispectral aerial imagery. The proposal is designed for small, non-technical crops that are monitored by aerial photography captured by the MapIR camera as payload from an unmanned aerial vehicle.

Objetivos: Propose a method of early prediction of the yield of coffee crops based on physical and spectral characteristics obtained from aerial images.

Métodos: This research proposes a prediction model based on regression algorithms with images obtained from unmanned aircraft captured in the crop flowering stage. In this way, the project calculates crop yield early.

Resultados: The present proposal resulted in a crop yield prediction model with an accuracy of 57 % for a multiple linear regression model. There is also a web application that allows obtaining manual performance in an agile and dynamic way.

Conclusiones: The work demonstrated that it is feasible to execute a yield prediction based on information obtained only from multispectral aerial imagery obtained with accessible tools.

Palabras Clave: Crop Yield; Coffee; Image Segmentation; Multispectral; MapIR; Predictor; UAV

Contenido

1. Introducción	11
1.1. Planteamiento del Problema	11
1.2. Objetivos	12
1.2.1. Objetivo general	12
1.2.2. Objetivos específicos	12
1.3. Hipótesis	13
1.4. Contribuciones	13
1.5. Contenido de la monografía	13
2. Estado del arte	15
2.1. Contexto	15
2.1.1. Índices de vegetación	15
2.1.2. Agricultura de precisión	16
2.1.3. Percepción remota	16
2.1.4. La fenología del cultivo de café	16
2.1.5. Rendimiento del cultivo	17
2.2. Trabajos relacionados	17
2.2.1. Técnicas de muestreo manual	18
2.2.2. Modelos agrometeorológicos espectrales	18
2.2.3. Técnicas de aprendizaje automático	20
2.2.4. Brechas	21
3. Descriptores Espectrales	23
3.1. Obtención de descriptores espectrales	23
3.1.1. Captura de fotografías	24
3.1.2. Post Procesamiento de Imágenes	29
3.2. Obtención del rendimiento por planta	39
3.2.1. Experimento	39
3.2.2. Ejecución	42
3.3. Resultados y discusión	43
4. Modelo de Predicción de rendimiento	47
4.1. Modelos de regresión	47
4.2. Validación	51

4.2.1. Cultivos de prueba	52
4.2.2. Validación de la segmentación	52
4.2.3. Validación del modelo de regresión	54
4.2.4. Resultados	56
5. Conclusiones y trabajos futuros	59
5.1. Conclusiones	59
5.2. Trabajos Futuros	60
Referencias	62

Lista de Tablas

2.1. Brechas	22
3.1. Índices de vegetación	37
3.2. Correlación de variables con Rendimiento(Yield)	44
4.1. Variables para modelo de regresión múltiple	49
4.2. Resultados de los modelos de regresión [1]	50
4.3. Información de Manejo de los Cultivos * 1 Arroba = 11.34Kg	52
4.4. Medidas de granos de café:	56
4.5. Resumen proceso de detección	56

Nomenclatura

ANN Red Neuronal Artificial. 20

AP Agricultura de precisión. 15, 16

DNN Redes Neuronales Profundas. 20

DSM Modelo digital de superficie. 38, 44

DVI Índice de vegetación de diferencia. 12

EML Aprendizaje de máquina extremo. 20

ExG Índice de vegetación de exceso de verde. 12, 15

FAPAR Proporción de la radiación foto-sintéticamente activa absorbida. 20

GCP Puntos de control en tierra. 20

GPS Sistema de posicionamiento global. 31, 32

IAF Índice de área foliar. 18, 19

LAI Índice de área foliar. 20, 38, 44, 51

LR Regresión Lineal. 20

MAPE Error porcentual absoluto medio. 20

MSE Error absoluto medio. 20

NDVI Índice de vegetación de diferencia normalizada. 12, 15, 19–21, 28, 30, 31, 37, 44, 51

NDWI Índice de agua de diferencia normalizada. 15, 16, 19

NIR Infrarrojo Cercano. 15, 16, 28, 37, 46, 49

-
- PLS** mínimos cuadrados parciales. 20, 50
- RBF** Funcion de base radial. 51
- RF** Bosque Aleatorio. 20, 21
- RGB** Banda Roja Verde y Azul. 12, 18, 20, 24, 36
- RMSE** Raíz del error cuadrático medio. 20, 51
- SAVI** Índice de vegetación ajustado al suelo. 19
- SDG** Gradiente descendente estocástico. 49–51
- SVM** Máquinas de vectores de soporte. 20
- SVR** Vectores de Soporte de Regresión. 20, 49–51
- UAV** Vehículos aéreos no tripulados. 12, 13, 20, 21, 23, 47
- UV** Ultravioleta. 12, 21
- XGBoost** Incremento de gradiente extremo. 21

Capítulo 1

Introducción

1.1. Planteamiento del Problema

El café es uno de los productos más importantes para la economía colombiana y representa una fuente de ingresos para 540.000 familias de medianos y pequeños productores. Según el Ministerio de Hacienda de Colombia, ha sido un importante producto de exportación para la reactivación económica gracias a su valor internacional. La producción de café en Colombia fue de 12,1 millones de sacos de 60Kg en 2021, cayendo un 11 % respecto a 2020. Esta disminución se debe a condiciones climáticas desfavorables para el cultivo del café. En el departamento del Cauca, 93 mil familias son cafeteras, aportando un 9.96 % a la producción nacional, que para el 2019 alcanzó un histórico de más de 14 millones de sacos de café, consolidando este producto como uno de los más importantes en el sector agrícola. Dicho resultado histórico, fue posible gracias a que en ese año los cultivos eran en un 83 % variedades resistentes teniendo una densidad de siembra aproximada de 5200 plantas por hectárea, con un rendimiento promedio de 21.4 sacos de café [2, 3, 4]. Por esta razón, es de particular interés enfocar los esfuerzos en investigar el proceso de producción de café que permita optimizar la producción [3].

Un interés común entre los agricultores es la gestión de recursos para garantizar el rendimiento, es decir, la cantidad de sacos de café por unidad de área, lo que a su vez representa la productividad de un cultivo [5]. Realizar una predicción temprana del mismo puede ayudar a establecer requerimientos de forma temprana optimizando el riego y fertilizantes; permite optimizar costos en cuanto a procesos como el secado, cosecha y almacenamiento así como para el compromiso de vigencias futuras con propósitos de financiación, entre otros [6, 7]. El rendimiento del café, en general está determinado por la influencia del sistema Planta-Suelo-Ambiente, con factores determinantes en la productividad como lo son la temperatura, la pluviosidad, la disponibilidad de agua, entre otros [8]; así como también factores edáficos dependen del tipo de suelo y su composición; y los factores de la planta tal como la edad, la variedad, la densidad de siembra, la salud y la fisiología de la planta, que influyen directamente en el rendimiento. De acuerdo a los factores anteriormente mencionados, y según Idol *et al.* [9], los agricultores suelen utilizar métodos manuales para predecir el rendimiento guiados por la fisiología de la planta. Es-

pecíficamente el método manual basado en muestreo de flores o frutos, el cual demanda esfuerzos en tiempo y dinero por parte de los agricultores, causando afectación a la integridad de las plantas. Otros métodos para predecir el rendimiento están centrados en el uso de parámetros de imágenes satelitales, principalmente para el cálculo de índices de vegetación como el NDVI [10]. Sin embargo, su uso presenta altos costos de implementación y baja precisión, generando resultados a escala regional [11]. Esto ha generado interés por evaluar alternativas como el uso de vehículos aéreos no tripulados o UAVs (por sus siglas en inglés) [10], que en comparación a los sensores remotos satelitales son menos costosos. Estas tecnologías integran dispositivos tales como sensores láser, sensores de calor, cámaras infrarrojas, cámaras RGB o cámaras UV; de las cuales destacan las cámaras RGB e infrarrojas para obtener imágenes de alta resolución en diferentes bandas espectrales. Dichas imágenes son procesadas para obtener mosaicos georreferenciados y posteriormente índices de vegetación fundamentales como NDVI, DVI, ExG, entre otros; que en conjunto permiten el análisis de la salud y fisiología de un cultivo agrícola [12]. Teniendo en cuenta las anteriores consideraciones, es posible evidenciar que el rendimiento del café es estimado mediante modelos agrometeorológicos y métodos manuales. Pero dichas alternativas involucran una inversión económica significativa y una gestión de recursos no focalizada, lo cual es el principal motivo para estudiar nuevos modelos de rendimiento basados en tecnología UAV que facilita la adquisición de imágenes aéreas de alta resolución, con costos manejables.

De acuerdo a lo anterior, esta propuesta está centrada en la pregunta de investigación:

¿Cómo predecir el rendimiento de un cultivo de café mediante el entrenamiento de un modelo de aprendizaje automático basado en características espectrales obtenidas desde imágenes aéreas?

1.2. Objetivos

1.2.1. Objetivo general

Proponer un modelo de predicción del rendimiento de un cultivo de café mediante características espectrales obtenidas desde imágenes aéreas.

1.2.2. Objetivos específicos

1. Identificar descriptores espectrales relacionados con el rendimiento por planta a partir del análisis de imágenes aéreas.
2. Definir un modelo de predicción basado en análisis espectral para estimar el rendimiento de un cultivo de café.
3. Evaluar la precisión del modelo de predicción del rendimiento en un cultivo de café para un ciclo productivo controlado.

1.3. Hipótesis

La predicción del rendimiento de un cultivo de café es factible entrenando un modelo de aprendizaje automático basado en características espectrales obtenidas desde imágenes aéreas.

1.4. Contribuciones

- Diseño de un método de análisis de imágenes para la segmentación efectiva de los cultivos de café con el fin de efectuar un análisis por planta. Este proceso permite eliminar el ruido del fondo e individualizar cada planta calculando para cada una las características físicas y espectrales.
- Una herramienta web dinámica para la obtención del rendimiento manual que tiene en cuenta las plantas que fueron obtenidas en el proceso de segmentación. Esta herramienta hace uso de las coordenadas y el contorno de las plantas para dibujarlas en un mapa que permite la obtención del rendimiento manual por planta.
- Un conjunto de datos de información de imágenes multi espectrales geo-referenciadas enriquecido con características de rendimiento por planta para su uso en posteriores trabajos de investigación. Todos los recursos necesarios están alojados en el siguiente enlace: bit.ly/3h1nvdG. La estructura de los datos se basa en el uso de un motor de base de datos Mysql relacional donde cada registro corresponde a una planta como lo representa la figura 3.23 de la sección 3.1.2.4.
- Un artículo de investigación denominado: Feasibility of Early Yield Prediction per Coffee Tree based on Multispectral Aerial Imagery: Case of Arabica Coffee Crops in Cauca-Colombia Aceptado el 17 de Noviembre por la revista Remote Sensing homologada en Q1 del JCR y en A1 de publinde

1.5. Contenido de la monografía

- Capítulo 1: Este capítulo aporta una introducción que consta del planteamiento del problema junto con la exposición de los objetivos del trabajo.
- Capítulo 2: Este capítulo busca establecer una base conceptual basado en la revisión de literatura. La finalidad del capítulo dar una claridad sobre el estado actual del conocimiento del problema y las brechas más importantes encontradas.
- Capítulo 3: Este capítulo describe el proceso de obtención de descriptores espectrales basados en el procesamiento de imágenes obtenidas desde un UAV. El resultado del capítulo es el conjunto de datos físicos y espectrales por planta.

- Capítulo 4: Este capítulo describe el proceso de predicción del rendimiento del cultivo del café mediante el uso de modelos de regresión. Así como el proceso de validación del modelo con un cultivo controlado.
- Capítulo 5: Este capítulo describe las conclusiones finales del trabajo y algunos trabajos futuros propuestos.

Capítulo 2

Estado del arte

Este capítulo describe el marco conceptual del trabajo de grado. Inicialmente el capítulo establece un contexto que define términos relevantes como la percepción remota o los índices de vegetación. Posteriormente, el trabajo expone una sección de trabajos relacionados que permiten conocer el estado actual del conocimiento mediante una revisión sistemática que identifica una lista de brechas que permiten establecer de forma precisa los objetivos. Este proceso también permite abrir el espectro de técnicas y métodos hacia la predicción del rendimiento sobre cultivos de café.

2.1. Contexto

2.1.1. Índices de vegetación

Los índices de vegetación son combinaciones matemáticas entre bandas espectrales de una imagen provenientes de la reflectancia espectral de las superficies [13]. Estas pueden ser calculadas utilizando las bandas visibles e infrarrojas del espectro electromagnético que están directamente relacionadas con factores fisiológicos y físicos de las plantas, como lo son el proceso de la fotosíntesis, humedad, clorofila, el coeficiente de cultivo etc. Normalmente, la reflectancia espectral en una planta varía de acuerdo a la longitud de onda y a su estado fisiológico, y dependiendo de las bandas y la relación establecida es posible inferir distintas condiciones físicas. En general, el índice de vegetación de diferencia normalizado NDVI, que involucra la banda NIR y la banda roja del espectro es uno de los más utilizados en la agricultura de precisión (AP). Las plantas reflejan gran parte de la banda NIR y absorben la banda roja cuando están en buena condición, lo que representa un indicador del estado fisiológico de las mismas [11]. Otros índices importantes son el NDWI que puede aportar información de humedad de las plantas, o el índice de la banda visible de exceso de verde ExG que puede ser usado para definir qué es planta y que no, así como las versiones mejoradas del NDVI como el ENDVI que es utilizado también para minimizar el efecto del suelo en el resultado final [14]. Entre más bandas espectrales estén disponibles, más índices de vegetación es posible obtener y por consiguiente habrá mayor caracterización de la superficie observada.

2.1.2. Agricultura de precisión

La agricultura de precisión (AP) involucra técnicas destinadas a optimizar el rendimiento de los cultivos [15]. La AP utiliza métodos como la percepción remota para efectuar recolección y procesamiento de datos que provienen directamente del área o cultivo de interés. La AP es una herramienta para los agricultores y productores que permite prevenir posibles plagas o enfermedades. Al tiempo que monitorean el crecimiento de las plantas, y la adecuada aplicación de insumos repercutiendo así en un beneficio económico y ecológico [16]. Normalmente, la AP es descrita como un proceso repetitivo de actividades como recolección de datos, procesamiento e interpretación de los mismos. Teniendo la presente propuesta como finalidad, hacer un recorrido por dichas etapas para finalmente obtener una predicción del rendimiento [15].

2.1.3. Percepción remota

La percepción remota es uno de los métodos de recolección de datos de la AP que es usado obtener información de un cultivo sin estar en contacto con él, ya sea mediante fotografías, temperatura, humedad o luminosidad, que pueden ayudar a las labores de irrigación y mantenimiento [17, 18]. Dicha información puede ser obtenida mediante sensores satelitales, proximales o aerotransportados. Los sensores satelitales son comúnmente utilizados en la cobertura a gran escala de una determinada región, pero pueden ser inoperables en existencia de condiciones atmosféricas no deseadas. Además, cuando el área es pequeña, los costos y la imprecisión son elevadas. Los sensores aerotransportados son más ágiles, destacándose las cámaras, radares, sistemas LiDAR, entre otros. Los sensores proximales miden variables para definir el comportamiento de los parámetros biofísicos del cultivo como temperatura, agua y nutrientes [15]. Existe un interés en este caso por explorar las relaciones que existen directamente entre los parámetros biofísicos de un cultivo y la reflectancia espectral que puede ser percibida por una cámara multi espectral. Dichas relaciones pueden establecerse mediante los rangos de niveles de la reflectancia de las plantas, por ejemplo, el infrarrojo profundo que es relacionado con los niveles de humedad de las plantas índice de vegetación NDWI [14]. Así mismo, existen índices de vegetación en la banda del NIR y en el espectro visible pueden establecer relaciones importantes con el rendimiento de las plantas [11].

2.1.4. La fenología del cultivo de café

Uno de los aspectos más relevantes en el estudio de los cultivos es la fenología del café, que corresponde al análisis de procesos productivos periódicos en cualquier cultivo. La fenología es utilizada para la toma de decisiones en términos de producción, pues permite tener idea del plan de fertilización, el manejo de enfermedades, entre otros [19]. Existen

procesos fenológicos importantes en la producción partiendo desde la floración y el desarrollo del fruto, que deben ser tenidos en cuenta para determinar la demanda de agua y nutrientes. El proceso fenológico para el café define 4 etapas que en total suman 224 días, y va desde la floración hasta justo antes de la maduración [19]. En general, para incrementar el rendimiento, es importante que en el proceso de llenado de fruto la planta no presente déficit hídrico y esté en óptimas condiciones en su área foliar. Las plantas de café pueden llegar a una óptima producción a la edad de alrededor de 5 años donde la formación de nudos y ramas plagiotrópicas llega a su máximo.

2.1.5. Rendimiento del cultivo

El rendimiento del cultivo determina la capacidad de un cultivo de producir una cantidad de materia seca por unidad de área, aunque también puede determinarse como la cantidad de sacos de 60kg por hectárea[5]. El concepto de rendimiento que este trabajo maneja estará centrado en la cantidad de granos por planta pero en la sección 4.2.4 se analiza el peso de cada grano de café en Kg.

2.2. Trabajos relacionados

Esta sección describe los diferentes trabajos que están involucrados en el desarrollo de la presente propuesta de investigación. El proceso de revisión inicia con un mapeo sistemático centrado en el establecimiento de las técnicas actuales de predicción de rendimiento iniciando con el planteamiento de una pregunta denominada: "¿Cuál es el estado actual del conocimiento en rendimiento alrededor del café?". Se extrae la consulta de dicha pregunta la cual es definida como "TITLE-ABS-KEY ((yield* OR producti*)) AND TITLE (coffe*)" generando 5704 resultados que son introducidos en la herramienta SciMAT. Los documentos anteriores al 2018 fueron descartados. El proceso de mapeo sistemático es definido en 2 sub periodos de los cuales el primer periodo de 2018 a 2019 establece al café arábico y al rendimiento como temas importantes relacionados con la eficiencia y la calidad alimentaria. Del año 2020 al 2022 es posible notar que el cambio climático aparece como tema motor de la investigación teniendo una relación directa con el rendimiento. El rendimiento aparece estrechamente relacionado con la temperatura y las condiciones climáticas, estableciendo una influencia directa de estos factores con la productividad. Posterior a la revisión del estado actual de la ciencia, es necesario centrar los esfuerzos en determinar mediante una revisión sistemática detallada el estado actual de las investigaciones dirigidas hacia el cálculo del rendimiento del cultivo. En la ejecución y revisión de este proceso aparecen tres grandes tendencias las cuales inician con unos métodos de muestreo manuales, así como técnicas con modelos agrometeorológicos y algunos procesos que involucran el uso de algoritmos de aprendizaje automático.

2.2.1. Técnicas de muestreo manual

El conteo manual es una de las técnicas más comunes encontradas dentro de la revisión realizada. Uno de los enfoques dentro de esta técnica lo describe Rendón J *et al.* [7], quien propone realizar un conteo manual de flores y de esta forma determinar la relación de este método con el rendimiento. El documento define conceptos relevantes como la relación entre el cuajamiento de flores y el rendimiento final. Otro de los enfoques encontrados determina la cantidad de frutos en la planta, como en el trabajo Idol *et al.* [9] que presenta un método manual de estimación basado en el número de frutos por ramas y en el número de ramas por árbol; realizando un muestreo que finalmente refleja la estimación real de todo el lote. Así mismo, Castro *et al.* [20] evalúa un método rápido no destructivo que propone el cálculo y conteo de frutos por lateral ortotrópico, es decir, la estimación rápida de la cantidad de carga de café en un lateral, para posteriormente extrapolar dicha medida hacia el total de la planta y luego al total de la finca.

Uno de los enfoques cercanos a la presente propuesta lo presenta Unigarro *et al.* [21], quien realiza la revisión de cada una de las características fenotípicas del café tal como el área foliar, la altura, el diámetro, entre otras, para realizar correlaciones entre ellas y el cultivo de café. El trabajo destaca un resultado en el que el área foliar es uno de los factores relevantes en el análisis del rendimiento del cultivo de café. De forma similar Rendón J *et al.* [22] establece un estudio específico del rendimiento relacionado con el área foliar destacando la alta correlación entre las dos variables y que puede ser obtenidas mediante el uso de aeronaves no tripuladas. En ese sentido, el trabajo de Taugourdeau S *et al.* [23] muestra la relevancia del IAF dentro del manejo de la producción de los cultivos de café [24], y mientras que Dos Santos *et al.* [25] presenta el cálculo del IAF mediante imágenes aéreas RGB obteniendo variables como la altura y el diámetro [26, 27]. De otra manera, Ramos *et al.* [28] presenta un enfoque de segmentación de granos de café, calculando el rendimiento mediante visión por computador sobre imágenes RGB de las ramas del árbol. Es un enfoque que utiliza técnicas tecnológicas para facilitar la labor de conteo manual, pero sigue siendo un proceso dispendioso y complejo. La principal brecha de los trabajos anteriores radica en la dificultad para tomar manual mente las muestras de rendimiento, pues dicha tarea es dispendiosa en el tiempo. Además, a excepción del trabajo que involucra flores [7], los frutos del café ya formados son los importantes, lo que está alejado de la predicción temprana del rendimiento.

2.2.2. Modelos agrometeorológicos espectrales

Este enfoque analiza distintos modelos que relacionan variables como las características fenotípicas, la evapotranspiración, la temperatura, la radiación solar entre otras variables. En Miranda *et al.* [6] un modelo agrometeorológico es evaluado teniendo en cuenta algunas características fenotípicas como variables de entrada de 3 índices fenológicos diferentes. El trabajo destaca la alta correlación de 0.9431 obtenida por el índice fenológico 3 que simula la arquitectura de la planta. El trabajo de Lima *et al.* [29] evalúa modelo matemático ya existente, dicho modelo denominado “kiring” permite establecer relaciones entre dos variables distintas. En este caso las variables seleccionadas fueron las ramas plagiotrópicas

de la planta de café para correlacionarlas con el rendimiento. Así como el trabajo de Silva *et al.* [10] que relaciona variables ambientales como la temperatura y la evapotranspiración, introduce un modelo agrometeorológico que establece el cálculo del rendimiento con dichas entradas. Pero, además, adicionando la coordenada geográfica a cada muestra generando una distribución espacial en el resultado.

Otro de los enfoques relevantes para esta investigación es el manejo de los índices de vegetación. En este sentido, el trabajo de Rosa *et al.* [11] involucra en su modelo el parámetro NDVI que es un factor de interés en el caso de análisis de imágenes aéreas para la agricultura. Dicho modelo establece relación entre los índices de vegetación y el rendimiento, obteniendo una baja correlación en los resultados. La desventaja de esta variable podría deberse a que es obtenida a partir de datos satelitales a escala de región, lo cual es una limitante que plantea superar la presente investigación. De igual forma Nogueira *et al.* [30] involucra un modelo que incluye los índices de vegetación SAVI y NDVI para relacionarlos con el rendimiento. Este estudio establece el efecto que tiene la varianza espacial y la topografía en los índices de vegetación. Este trabajo define las etapas fenológicas importantes para llevar a cabo la estimación del rendimiento, como lo son la floración y el llenado de fruto. También el trabajo de Bolton *et al.* [14] estudia las relaciones entre los índices de vegetación basados en infrarrojo como el NDWI y el rendimiento, concluyendo que este índice representa información del estrés hídrico de las plantas.

Además de las variables hasta aquí analizadas, es destacable que Picini *et al.* [31] trabaja dos modelos agrometeorológicos que involucran la producción del año anterior y la evapotranspiración, lo que puede definir bases importantes para consolidar un modelo general que involucre todas las variables anteriormente analizadas. El trabajo de Arruda *et al.* [32] presenta el efecto que tiene la evapotranspiración sobre el rendimiento e introduce los conceptos de evapotranspiración real y evapotranspiración potencial. Y por último el trabajo de Eduardo *et al.* [33] analiza variables como la lluvia, la temperatura y el déficit hídrico mediante modelos de regresión lineal con el rendimiento, comparando diferentes modelos de regresión para establecer cuál es el más adecuado para el análisis.

Esta sección evalúa el uso de distintas variables con algunos modelos agrometeorológicos espectrales. El uso de dichos modelos destaca variables de entrada como la evapotranspiración, el NDVI, la temperatura, entre otros. Dichas características, en su mayoría, son obtenidas mediante el uso de satélites, lo que puede llegar a ser costoso. El manejo de información satelital está disponible a escala regional lo que afecta la resolución espacial de las imágenes [34].

Es importante destacar que durante la revisión de los modelos agrometeorológicos y los métodos manuales hay trabajos destacados como el de Muñoz *et al.* [21] que define el IAF como uno de los factores más representativos a la hora de establecer el rendimiento, así como variables como la altura y el dosel de la planta que según dos Santos *et al.* [25] pueden ser obtenidas mediante imágenes aérea. O el trabajo de Bolton *et al.* [14] que establece relaciones entre los índices de vegetación como el NDWI y el estado hídrico de las plantas. En general, el enfoque de la presente propuesta está enmarcado la consecución

de los parámetros que son influenciados por el sistema de factores de la Planta, Atmósfera y Suelo, que se pueden calcular a partir de imágenes aéreas [34].

2.2.3. Técnicas de aprendizaje automático

Kouadio *et al.* [35] propuso un modelo de aprendizaje automático basado en tres distintos algoritmos: Extreme Machine Learning (EML), Random Forest (RF) y Linear Regression (LR). Todos estos modelos son entrenados utilizando variables de nutrientes del suelo que permiten efectuar el cálculo del rendimiento. Los tres modelos manejan datos de entrenamiento, de prueba y de validación. Es importante destacar que el modelo de mejor desempeño para el cálculo de rendimiento fue EML, el cual utilizaba de forma conjunta la variable de la materia orgánica, fósforo y azufre. Producto de la validación realizada mediante el uso del error cuadrático medio (RMSE), el error absoluto medio (MSE), entre otros [36].

La revisión también encontró que Barbosa *et al.* [37]. estudió modelos de regresión basados en características fisiológicas como la altura y el diámetro obtenidos de imágenes RGB utilizando UAV para la estimación del rendimiento, validando los resultados manualmente con la imagen. En esta investigación se obtuvo el total de granos del café georreferenciando cada punto con la herramienta GCP. El proceso de entrenamiento de modelos de regresión como SVM, PLS, Gradient Boosting y RF utilizó 144 registros de datos. El cálculo de altura y diámetro varió entre un 6% y un 7%, así como la medida MAPE para los modelos de regresión rondaba el 31%. Una limitante de este trabajo es que no considera el análisis espectral basado en índices de vegetación.

Del mismo modo, Thao *et al.* [38] proponen un modelo estadístico para la predicción temprana del rendimiento en cultivos de café basado en índices de vegetación a escala regional con datos de Copernicus para los predictores NDVI, FAPAR y LAI, obteniendo un $Adj R^2$ de 64 a 69% en modelos de regresión utilizando la herramienta estadística Crop Growth Monitoring System, que permite una predicción temprana hasta 6 meses antes de la cosecha, siendo un referente importante para este trabajo de grado que pretende la predicción temprana de los cultivos de café, sin embargo se basa en análisis de imágenes satelitales a escala regional pero con resultados prometedores.

El ampliar la búsqueda evidenció trabajos como el de Ashapure *et al.* [39], que propone predecir el rendimiento en cultivos de algodón mediante imágenes aéreas utilizando modelos de regresión de aprendizaje automático tales como redes neuronales artificiales (ANN), Regresión de Vectores de Soporte (SVR, Support Vector Regression) para evaluar una predicción temprana. Este trabajo destaca el uso de índices de vegetación y el uso de modelos digitales de superficie para obtener la altura del cultivo; el resultado fue una correlación de 0.85 encontrando variables de entrada óptimas mediante estudios de sensibilidad. De forma similar el trabajo de Moghimi *et al.* [40] establece el uso de imágenes aéreas y redes neuronales profundas (DNN) para predecir el rendimiento en trigo. En este caso las parcelas de trigo experimental son segmentadas con el fin de establecer el volumen y distintos índices de vegetación, entre ellos el NDVI. Este trabajo destaca la división del espacio en sub parcelas estableciendo promedios de índices de vegetación por cada una,

asociándola a un valor de rendimiento.

También el trabajo de Herrero *et al.* [41] presenta un modelo de predicción de rendimiento basado en aprendizaje automático mediante el análisis de características como la altura y el volumen provenientes del análisis de nubes de puntos desde el manejo de imágenes obtenidas por UAV en cultivos de soja. Este trabajo involucra modelos como RF y el Incremento de gradiente extremo (XGBoost) con precisiones de un 90.72 % y 91.36 % respectivamente. Dicho método establece la importancia de analizar tecnologías como el manejo de los modelos de superficie. De la misma forma, el trabajo de Kunapuli *et al.* [42] plantea un enfoque de análisis de imágenes satelitales con el fin de relacionar el NDVI y el índice de área foliar con el rendimiento en cultivos de maíz, esto mediante el uso de modelos de regresión polinomial.

A diferencia de los modelos anteriores, existen otros enfoques como el descrito en Khosla *et al.* [43] que están basados en parámetros de entrada como la pluviosidad comparando su comportamiento para predecir el rendimiento en diferentes cultivos. El trabajo de Haque *et al.* [44] involucra estudios de redes neuronales profundas con parámetros tales como la radiación UV, agua, los fertilizantes y los pesticidas. Logrando demostrar la capacidad de los modelos de redes neuronales para efectuar predicción del rendimiento teniendo en cuenta distintas variables correlacionadas.

Lo que sugieren los modelos antes estudiados es que es necesario efectuar distintas pruebas con diferentes modelos mediante las variables disponibles. Analizando la sensibilidad de los resultados de acuerdo a las diferentes combinaciones posibles de las entradas, pues en general existen algunas variables que estaban co-relacionadas más que otras, y otras que presentan información redundante [39].

La revisión de literatura Klompenburg *et al.* [45] realizó un análisis del estado actual del conocimiento, teniendo como objetivo revisar las variables utilizadas, los modelos involucrados y la cantidad de documentos por año. Estableciendo que la temperatura es una de las variables más utilizadas con 24 documentos involucrados. Las redes neuronales son las más relevantes por ser las más utilizadas en los estudios analizados. Es importante detallar que el estudio fue seccionado en dos grandes campos como lo son aprendizaje profundo y no profundo. Estableciendo que los más usados y populares son los de aprendizaje no profundo, pero aclarando que en los últimos ha existido un incremento en el enfoque de aprendizaje profundo, siendo más robusto pero más exigente en cantidad de datos, lo que en el presente trabajo es una limitante.

Considerando los trabajos anteriores, esta investigación se enfoca en obtener datos confiables que demuestren la factibilidad de la predicción temprana del rendimiento para el cultivo del café utilizando herramientas de bajo costo, facilitando el acceso a la tecnología a los pequeños y medianos productores.

2.2.4. Brechas

Según la Tabla 2.1, solo 6 de todos los métodos encontrados para la predicción del rendimiento realizan procedimientos manuales con análisis empíricos. Lo que conlleva un

Modelo	Trabajo	Tipo Imagen		Tipo de Información			Cultivo
		Satelital	Aérea	Suelo	Planta	Atmosfera	
Manual	[7]				✓		Café
	[9]				✓		Café
	[20]				✓		Café
	[21]				✓		Café
	[22]				✓		Café
	[28]					✓	Café
Agro Me- teorológico	[6]				✓		Café
	[29]				✓		Café
	[10]					✓	Café
	[11]	✓			✓		Café
	[30]	✓				✓	Café
	[14]	✓				✓	Café
	[31]				✓	✓	Café
	[32]				✓	✓	Café
[33]				✓	✓	Café	
Aprendizaje Automático	[35]			✓			Café
	[38]	✓			✓		Café
	[37]		✓		✓		Café
	[39]		✓		✓		Algodón
	[40]		✓		✓		Trigo
	[41]		✓		✓		Soja
	[42]	✓			✓		Maíz
	[43]	✓			✓		Varios
	[44]					✓	Varios
Propuesta		✓			✓	Café	

Tabla 2.1: Brechas

alto esfuerzo en términos de costo y tiempo. Así como también algunos de los trabajos que involucran modelos agrometeorológicos utilizan imágenes satelitales, que además de ser costosas, presentan baja precisión para efectuar análisis focalizado del cultivo. Por último, los trabajos encontrados en técnicas de aprendizaje automático denotan que existen muy pocos destinados al cultivo de café, de los cuales se destaca el trabajo de Barbosa et al.(2021) [37] y Thao et al. (2022) [38] que proponen unos enfoques interesantes que serán tenidos en cuenta durante toda la investigación. La presente propuesta está enfocada en el último grupo, teniendo como base de información las imágenes aéreas. Esto permitirá efectuar una segmentación para realizar análisis por planta y así obtener algunas variables tales como volumen, altura, el área entre otras como entradas del modelo de predicción.

Capítulo 3

Descriptores Espectrales

Este estudio está orientado hacia la predicción temprana del rendimiento del cultivo de café solo con datos de imágenes multiespectrales utilizando herramientas accesibles. Las imágenes obtenidas de los UAV permiten un análisis focalizado de alta precisión, lo que para esta investigación condujo a la individualización precisa y automática de las plantas, generando descriptores físicos y espectrales con potencial para predecir los rendimientos de los cultivos [46, 47, 48]. Para individualizar cada cafeto, se utilizó un algoritmo *watershed* o de cuenca hidrográfica basado en marcadores de primer plano [49]. Este algoritmo separa y define los bordes entre dos elementos de una imagen basado en la simulación de cuencas hidrográficas. Estas cuencas se rellenan en base a mínimos locales hasta definir el límite entre dos o más cuencas [49]. El capítulo inicia con la obtención de las imágenes multiespectrales siguiendo un proceso de planificación de vuelo basados en las características de las herramientas disponibles. Posteriormente el capítulo se centra en el post procesamiento de las imágenes que permite la obtención de los descriptores base para posteriormente realizar el análisis de resultados.

3.1. Obtención de descriptores espectrales

Como el primer objetivo es obtener descriptores espectrales que representan características físicas y/o fenotípicas obtenidas desde imágenes aéreas que están directamente relacionadas con el rendimiento, es importante ejecutar un proceso que permita que la información adquirida sea confiable teniendo en cuenta que es la base de todo el trabajo a realizar. Para ello en la Figura 3.1 se presenta un proceso general que conduce a la obtención de los descriptores espectrales a partir de imágenes aéreas.

El proceso de obtención de descriptores espectrales mostrado en la Figura 3.1 inicia describiendo el proceso de captura de fotografías y finaliza con el post procesamiento de información.

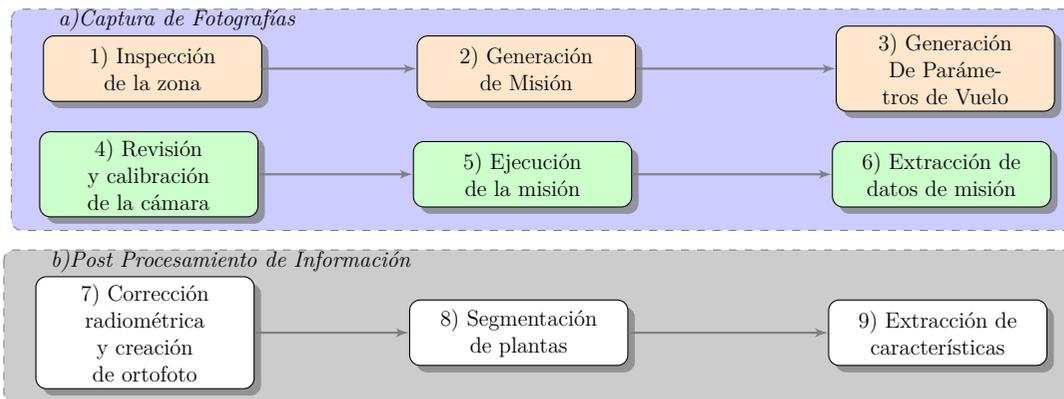


Figura 3.1: Generación de descriptores espectrales Fuente: propia

3.1.1. Captura de fotografías

La captura de fotografías aéreas es realizado utilizando una plataforma aérea y una cámara o sensor que se encarga de la recolección de la información, corresponde a la sección **a)** Figura 3.1. En dicho proceso intervienen todos los parámetros de sensor y las características de la aeronave, donde se pueden utilizar diferentes tipos de cámara ya sean multiespectrales, hiperespectrales o RGB, portados en las plataformas aéreas [46, 47]. Entre otras características, es importante destacar que la imagen que cuenta con una alta resolución espacial dada en (cm/píxel) en comparación con las obtenidas con sensores portados en satélites. Dicha resolución espacial es inversamente proporcional a la altura de vuelo, pero directamente proporcional al área cubierta por una sola fotografía, gracias a esta ventaja los sistemas aéreos no tripulados pueden abarcar áreas extensas a una baja altitud. El proceso consiste en capturar múltiples fotografías que posteriormente se pueden combinar para formar un mosaico geo referenciado.

3.1.1.1. Inspección de la zona

Es el proceso de revisión del lugar donde se va a realizar el vuelo, corresponde al paso 1 de la sección **a)** de la Figura 3.1, en este caso el experimento tuvo lugar en la finca “La Sultana” de la Universidad del Cauca, ubicada en el municipio de Timbío (Figura 3.2) latitud: 2.374733, longitud: -76.726979, departamento del Cauca, Colombia. La recolección de imágenes fue llevada a cabo en la etapa de floración entre julio (14) y agosto (29) de 2021.

Inicialmente, como el objetivo es probar la hipótesis del trabajo, se realizan vuelos sobre lotes con una orografía plana. El proceso de inspección de la zona también incluye los posibles obstáculos que pueden ser árboles, zonas de agrupación de personas, edificios, antenas de telecomunicaciones, u otros elementos que puedan afectar la realización del vuelo. En este caso, se encuentran árboles de aproximadamente 15 metros de alto teniendo que establecer la altura mínima a 18 metros. Teniendo en cuenta todas estas características se procede a establecer la misión que se ejecutará de forma automatizada.

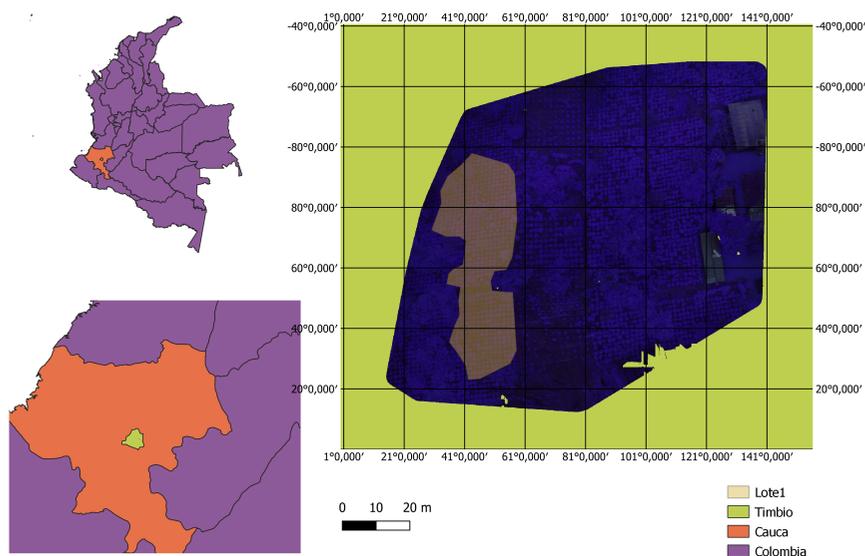


Figura 3.2: El mapa de la izquierda muestra la distribución por departamentos de Colombia, en los que se encuentra el Cauca. A la izquierda la ortofoto geolocalizada de la finca "La Sultana". Fuente: propia

3.1.1.2. Generación de la misión

La planificación y generación de la misión permite definir las características del área a cubrir, la plataforma y el sensor a utilizar con el fin de establecer criterios de configuración para realizar la captura. Corresponde al paso 2 de la sección **a)** de la Figura 3.1. Uno de los objetivos es conseguir una alta resolución de fotografía y para ello es necesario realizar múltiples tomas del terreno en diferentes lugares a una altura fija determinada. Las capturas deben tener dos tipos de superposición conocidos como lateral y longitudinal, denotados con las siglas v y u respectivamente, obteniendo cada punto del terreno en por lo menos dos fotografías asegurando la cobertura total del área objetivo.

3.1.1.3. Generación de parámetros de vuelo

Este proceso corresponde al paso 3 de la sección **a)** de la Figura 3.1. Los parámetros de vuelo tienen su base en conceptos matemáticos, estos permiten tener una aproximación de las dimensiones físicas que determinan la configuración cada vuelo, en ellos se tiene en cuenta las características del sensor y la plataforma, las variables a establecer son:

En la Figura 3.3 se representa gráficamente la planificación de vuelo propuesta.

- Altura de vuelo (h).
- Base en el aire (B) ecuación 3.1 [50].
- Separación entre líneas de vuelo (A) ecuación 3.2 [50].
- Velocidad de la aeronave

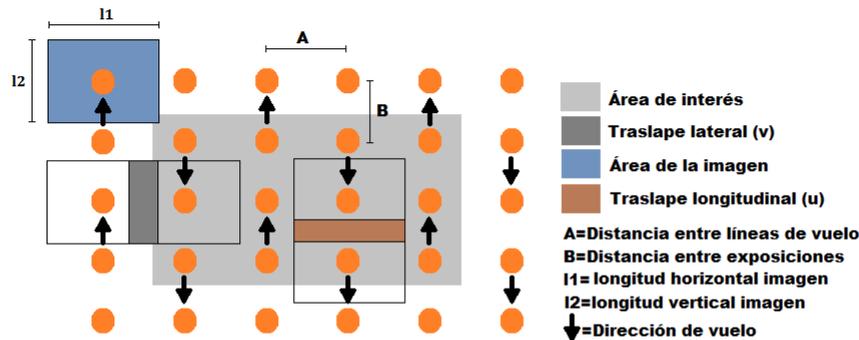


Figura 3.3: Captura de Imágenes . Fuente: propia

A continuación se presenta un resumen del cálculo de los parámetros requeridos, fundamentados en el concepto matemático expuesto en los trabajos [51, 50].

Base en el aire o distancia entre dos exposiciones (B) ecuación 3.1

$$B = l_2 * \left(1 - \frac{u}{100}\right) [m] \quad (3.1)$$

Distancia entre líneas de vuelo (A) ecuación 3.2

$$A = l_1 * \left(1 - \frac{v}{100}\right) [m] \quad (3.2)$$

Donde l_1 y l_2 se definen como las longitudes equivalentes en el terreno de acuerdo a cada lado de la fotografía.

Los parámetros conocidos son:

- Autonomía.
- Área a cubrir.
- Distancia focal (f).
- Velocidad de captura de fotografías ($V_{captura}$).
- Longitudes verticales y horizontales del sensor (dHs y dVs).
- Resolución horizontal y vertical en píxeles de la cámara ($(rHc$ y $rVc)$).

Teniendo en cuenta las variables anteriormente descritas, es posible proceder con la configuración de la misión en el software Mission Planner [52]. Para realizar la configuración es requerido conocer la altura de vuelo, los traslapes longitudinales y laterales. Para

establecer los demás parámetros solo es necesario una imagen capturada con la cámara del experimento. La Figura 3.4 muestra el resultado final de generación de puntos de referencia en tierra (*waypoints*) georreferenciados.



Figura 3.4: Parámetros de Misión . Fuente: propia



Figura 3.5: Misión Creada . Fuente: propia

Para la plataforma Phantom 3 Standard es requerido guardar la lista de puntos a un archivo mediante la opción “Save WP File”, esto guardará los *waypoints* en un archivo extensión “.waypoints”, que tiene una estructura como la de la Figura 3.6 [53]. Es necesario darle un formato diferente al archivo generado, construyendo para ello una función en lenguaje java encargada de recibir este tipo de archivo y transformarlo a la estructura de archivos que son aceptados por la aplicación Litchi[54]. El resultado de la transformación es el mostrado en la Figura 3.7.

```

KGC WPL 110
0 1 0 16 0 0 0 0 2.447833 -76.599276 1732.885532 1
1 0 0 16 0 0.000000 0.000000 0.000000 2.447836 -76.599624 70.000000 1
2 0 0 16 0 0.000000 0.000000 0.000000 2.447836 -76.599624 70.000000 1
3 0 0 16 0 0.000000 0.000000 0.000000 2.447103 -76.599507 70.000000 1
4 0 0 16 0 0.000000 0.000000 0.000000 2.447175 -76.599382 70.000000 1
5 0 0 16 0 0.000000 0.000000 0.000000 2.447247 -76.599257 70.000000 1
6 0 0 16 0 0.000000 0.000000 0.000000 2.447319 -76.599133 70.000000 1
7 0 0 16 0 0.000000 0.000000 0.000000 2.447391 -76.599008 70.000000 1
8 0 0 16 0 0.000000 0.000000 0.000000 2.447463 -76.598883 70.000000 1
9 0 0 16 0 0.000000 0.000000 0.000000 2.447535 -76.598758 70.000000 1
10 0 0 16 0 0.000000 0.000000 0.000000 2.447547 -76.598737 70.000000 1
11 0 0 16 0 0.000000 0.000000 0.000000 2.447547 -76.598737 70.000000 1
12 0 0 16 0 0.000000 0.000000 0.000000 2.447701 -76.598850 70.000000 1
13 0 0 16 0 0.000000 0.000000 0.000000 2.447701 -76.598850 70.000000 1
14 0 0 16 0 0.000000 0.000000 0.000000 2.447633 -76.598967 70.000000 1
15 0 0 16 0 0.000000 0.000000 0.000000 2.447561 -76.599092 70.000000 1
16 0 0 16 0 0.000000 0.000000 0.000000 2.447489 -76.599217 70.000000 1
17 0 0 16 0 0.000000 0.000000 0.000000 2.447417 -76.599342 70.000000 1
18 0 0 16 0 0.000000 0.000000 0.000000 2.447345 -76.599467 70.000000 1
19 0 0 16 0 0.000000 0.000000 0.000000 2.447274 -76.599591 70.000000 1
20 0 0 16 0 0.000000 0.000000 0.000000 2.447203 -76.599715 70.000000 1
21 0 0 16 0 0.000000 0.000000 0.000000 2.447203 -76.599715 70.000000 1
22 0 0 16 0 0.000000 0.000000 0.000000 2.447369 -76.599805 70.000000 1
23 0 0 16 0 0.000000 0.000000 0.000000 2.447369 -76.599805 70.000000 1
24 0 0 16 0 0.000000 0.000000 0.000000 2.447432 -76.599696 70.000000 1
25 0 0 16 0 0.000000 0.000000 0.000000 2.447504 -76.599571 70.000000 1
26 0 0 16 0 0.000000 0.000000 0.000000 2.447576 -76.599446 70.000000 1
27 0 0 16 0 0.000000 0.000000 0.000000 2.447648 -76.599322 70.000000 1
28 0 0 16 0 0.000000 0.000000 0.000000 2.447720 -76.599197 70.000000 1
29 0 0 16 0 0.000000 0.000000 0.000000 2.447792 -76.599072 70.000000 1
30 0 0 16 0 0.000000 0.000000 0.000000 2.447854 -76.598964 70.000000 1
31 0 0 16 0 0.000000 0.000000 0.000000 2.447854 -76.598964 70.000000 1
    
```

Figura 3.6: Archivo Waypoints . Fuente: propia

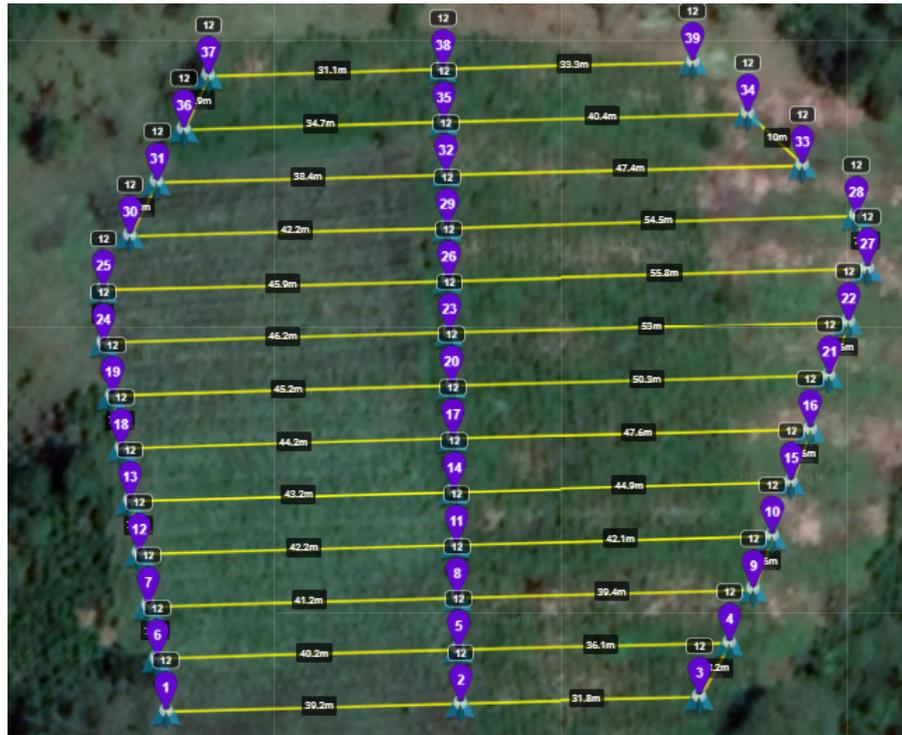


Figura 3.7: Misión Lista en Litchi App . Fuente: propia

3.1.1.4. Revisión y calibración de la cámara

Este proceso corresponde al paso 4 de la sección **a)** de la Figura 3.1 Las imágenes del trabajo fueron capturadas utilizando la cámara MAPIR Survey 3W. La cámara modificada RGN (Red-Green-NIR) está equipada con un sensor Sony Exmor R IMX117 que tiene una resolución de 12 megapíxeles utilizando los parámetros predeterminados recomendados por Mapir[55].

El montaje final del sistema de captura puede encontrarse en la Figura 3.8

3.1.1.5. Ejecución de misión y extracción de datos

Este proceso corresponde a los pasos 5 y 6 de la sección **a)** de la Figura 3.1. La ejecución de la misión se llevó a cabo en las mismas condiciones para todas las pruebas ejecutadas conservando todos los parámetros de vuelo para obtener resultados homogéneos, estos afinados durante las fases de prueba durante el año 2020 y 2021. El proceso de captura definitivo tuvo lugar durante los meses de Junio, Julio y Agosto del 2021. Las condiciones meteorológicas fueron de cielo nublado para no saturar los canales NGR de la cámara. Esto puede evidenciarse en [56], donde el autor expone que el canal rojo de la cámara puede contaminarse con reflectancia NIR. Se compararon capturas con cielos nublados y soleados para validar los mejores resultados de NDVI donde todas las configuraciones permanecieron iguales para todos los vuelos. Las imágenes resultantes se guardaron en formato JPG y RAW para su posterior procesamiento.



Figura 3.8: Montaje cámara Survey 3W y Phantom 3 Standard. Fuente: propia

3.1.2. Post Procesamiento de Imágenes

Este proceso permite obtener la información deseada a partir de las imágenes ya obtenidas en la captura de fotografías, corresponde a la sección **b)** de la Figura 3.1.

3.1.2.1. Corrección radiométrica

El procesamiento de las imágenes inicia con la calibración radiométrica llevada a cabo usando el proceso descrito en [57]. Básicamente, el proceso realizado consiste en trasladar la reflectancia de las imágenes hacia unos niveles conocidos previamente calibrados por el fabricante. Para ello existe el *target* de calibración que es fotografiado en las mismas condiciones de vuelo realizadas [58]. Este proceso corresponde a una parte del paso 1 de la sección **b)** de la Figura 3.1.

Fueron necesarias múltiples capturas en diferentes entornos entre las 10am y la 1pm para lograr determinar las mejores condiciones que permitan que la información sea fiable: Cuando las imágenes JPG de 8 bits que genera la cámara en condiciones de completa

luminosidad son procesadas (ver Figura 3.9), el resultado de NDVI es el mostrado en la Figura 3.10 evidenciando la afectación por sombra. Si efectuamos calibración radiométrica a la imagen, el resultado es el de la Figura 3.11. Puede evidenciarse claramente como, gracias a dicha calibración se mejora el efecto de la sombra en la imagen de NDVI resultante evitando la influencia de la sombra en los resultados.



Figura 3.9: Imagen Soleada. Fuente: propia

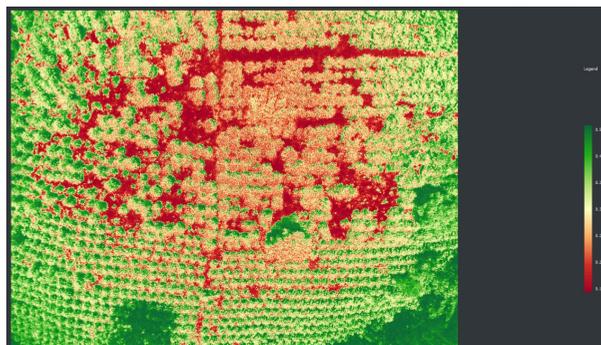


Figura 3.10: Imagen a pleno sol no calibrada.
Fuente: propia



Figura 3.11: Imagen a pleno sol calibrada.
Fuente: propia

Igualmente, el experimento en condiciones de cielo nublado arrojó una imagen como la de la Figura 3.12 donde el resultado del NDVI sin calibración radiométrica puede evidenciarse en la Figura 3.13 y el resultado con la calibración radiométrica es el de la Figura 3.14. Ambos resultados evidenciaron que los valores del índice de vegetación estaban entre 0.1 y 0.7 sin saturación en ninguno de los casos.

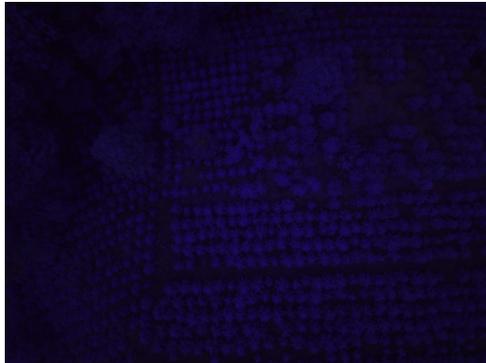


Figura 3.12: Cielo Nublado

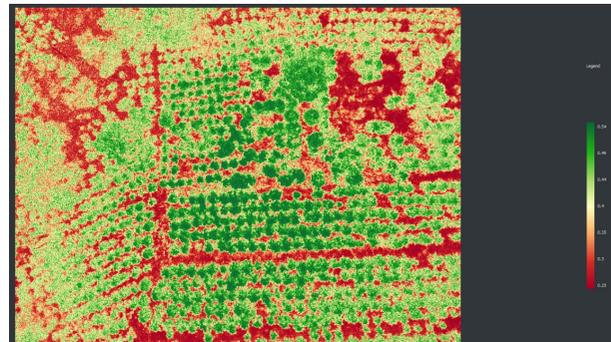


Figura 3.13: Cielo Nublado No calibrada



Figura 3.14: Cielo Nublado calibrada (fuente propia)

Basado en los anteriores resultados llegamos a la conclusión de que las mejores condiciones para llevar a cabo la captura de fotografías son con el cielo nublado. Eso concuerda con lo enunciado por Rosas *et al.* (2021) [56] donde se establece que el canal rojo de la cámara puede saturarse. Aunque las imágenes a pleno sol calibradas también pueden ser utilizadas pero estas deben ser transformadas de 16 bits a 8, pues al usar las imágenes de 16 bits ningún software de post procesamiento es capaz de manejarlas por su peso y por su formato.

3.1.2.2. Creación de ortofoto

Teniendo en cuenta los pasos anteriores, un punto importante dentro del trabajo es la correcta generación del ortofoto. Corresponde a la parte final del paso 1 de la sección **b)** de la Figura 3.1. Este proceso fue ejecutado mediante la herramienta WebODM[59]. Esta herramienta se encarga de procesar las imágenes usando los puntos de GPS guardados en

cada imagen para construir un mosaico geo referenciado, una nube de puntos, entre otros resultados. Uno de los inconvenientes encontrados fue la altura establecida por el GPS en cada una de las imágenes, lo que causó errores al ejecutar los procesos. El inconveniente fue resuelto usando la herramienta *exiftool* y ejecutando el siguiente comando `exiftool -m -GPSAltitude=1802m "path/to/images"`[59].

Con las imágenes analizadas y corregidas procedemos a ejecutar el proceso de generación del mosaico, en este caso se utiliza un pc con 8GB de RAM, tardando al rededor de 2 horas en completar todo el proceso. La Figura 3.15 muestra el resultado final en la interfaz del WebODM donde es posible descargar y pre visualizar todos los elementos del post procesamiento.



Figura 3.15: Resultado WebODM. Fuente: propia

3.1.2.3. Segmentación de plantas

La segmentación de plantas consiste en detectar e individualizar cada árbol de café para obtener sus características físicas y espectrales y corresponde al paso 2 de la sección b) de la Figura 3.1. La Figura 3.16 describe el proceso general que estuvo guiando toda la segmentación empezando por un pre procesamiento de las imágenes, posteriormente utilizamos el algoritmo Watershed para individualizar las plantas. Finalmente el post procesamiento consiste en generar las variables objetivo a partir plantas ya individualizadas.

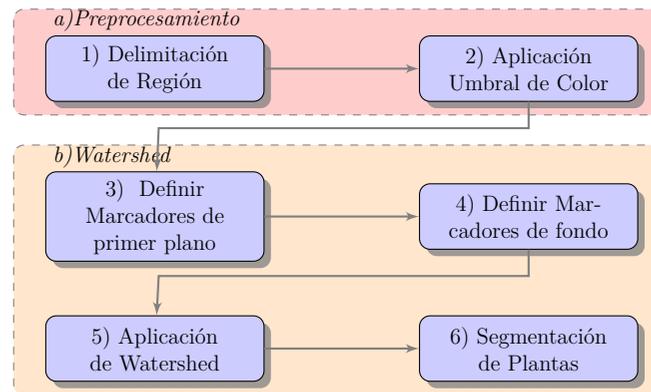


Figura 3.16: Proceso de segmentación de árboles. Fuente: propia

Pre procesamiento:

En la etapa de pre procesamiento, el cultivo a analizar es delimitado dentro de la imagen, esto ayuda a separar el entorno del área objetivo. En el paso dos, el fondo de la imagen o el suelo es eliminado usando un umbral de color basado en el espacio de color LAB definido por L(Luminosidad), a^* (valores entre verde y rojo) y b^* (valores entre azul y amarillo) o con el algoritmo kmeans que consiste en un proceso de aprendizaje no supervisado que es capaz de agrupar y separar los colores en determinados clusters, los mejores resultados fueron para la separación en 3 clusters [60]. Ambos métodos de eliminado de fondo fueron testeados y se pueden usar para la segmentación obteniendo buenos resultados. Los rangos de configuración para el método de umbral de color LAB deben establecerse para cada nueva imagen a procesar. El resultado del proceso puede verse en la Figura 3.17, este proceso corresponde a la sección **a)** de la Figura 3.16

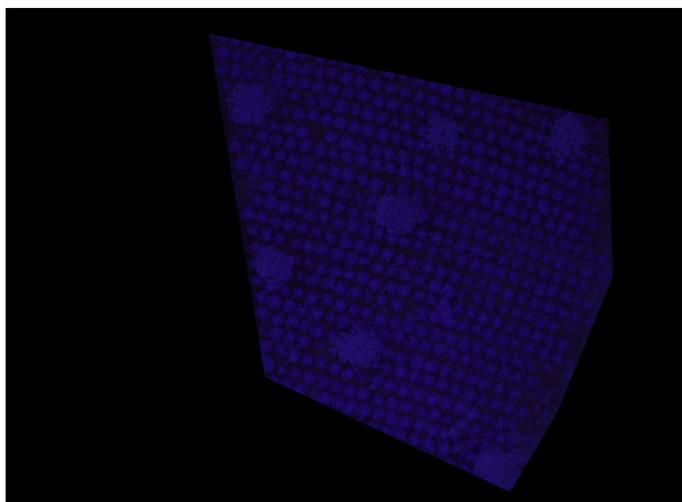


Figura 3.17: Delimitación de área y eliminación de fondo. Fuente: propia

Algoritmo de separación Watershed

Corresponde al proceso presentado en la sección **b)** de la Figura 3.16. Este algoritmo es útil para la separación y segmentación de diferentes objetos en una imagen basándose en su intensidad de color estableciendo etiquetas y límites entre distintos objetos basados en sus máximos de intensidad. Para ayudar al algoritmo a mejorar la segmentación, una de las herramientas es el uso de marcadores de primer plano y de fondo. Teniendo en cuenta las características de la imagen, múltiples ejercicios de prueba fueron realizados para encontrar una correcta definición de los marcadores de primer plano, llegando a una aplicación sucesiva de operaciones morfológicas. El resultado definitivo de la definición de marcadores de primer plano se puede ver en la Figura 3.18

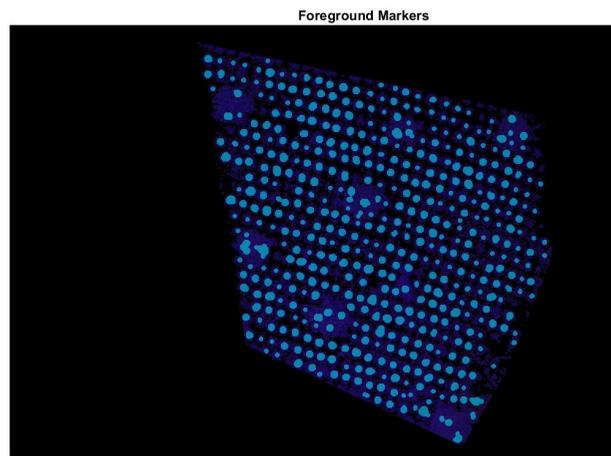


Figura 3.18: Marcadores de primer plano. Fuente: propia

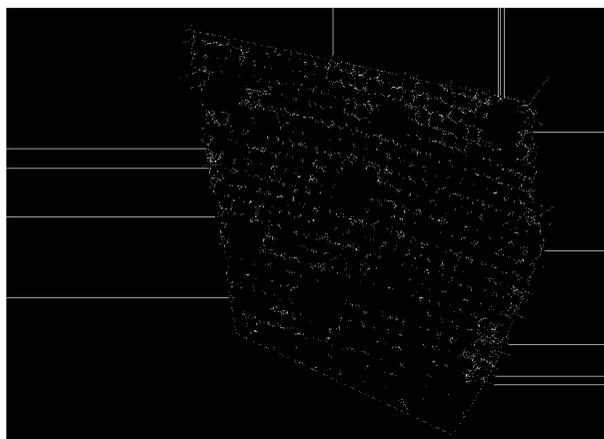


Figura 3.19: Marcadores de fondo. Fuente: propia

Los marcadores de fondo definen lo que efectivamente no es planta, para ello se obtiene

un negativo proveniente del paso 1 de la sección **a)** y se define el límite de las plantas que puede evidenciarse en la Figura 3.19 . Al ejecutar el algoritmo *watershed* con los datos anteriores se generan las etiquetas representadas en la Figura 3.20 para posteriormente evidenciar la segmentación realizada como se muestra en la Figura 3.21 .



Figura 3.20: Labeling. Límites de plantas detectadas por el algoritmo combinadas con los marcadores de primer plano en color azul Fuente: propia

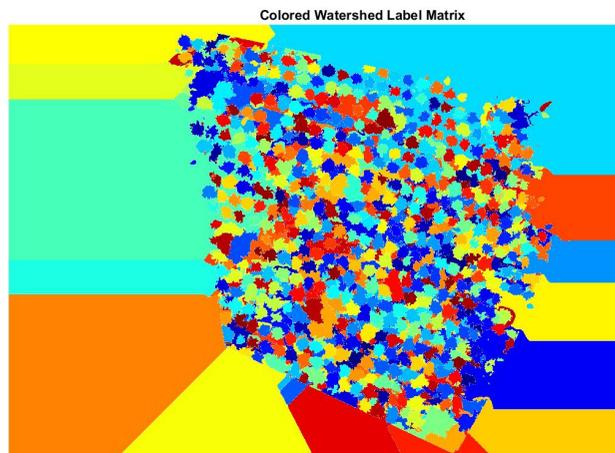


Figura 3.21: Resultados de cuenca hidrográfica: muestra todas las etiquetas detectadas en un color diferente aplicando un espacio de trabajo de color LUT. Fuente: propia

3.1.2.4. Extracción de características

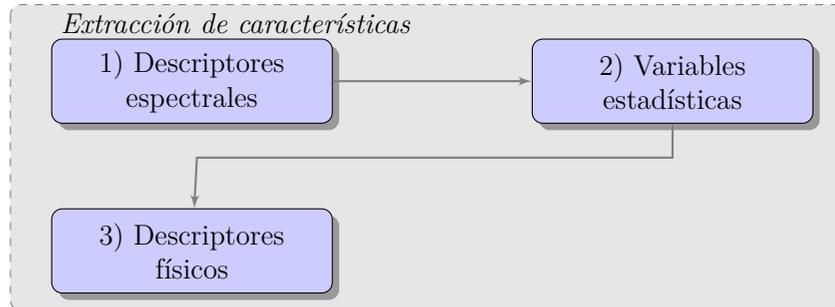


Figura 3.22: Proceso de extracción de características. Fuente: propia

Una vez completada la segmentación de los árboles, el siguiente paso es el proceso de extracción de variables (Corresponde al paso 3 de la sección **b**) de la Figura 3.1). Este proceso comienza con el análisis de ruido revisando el tamaño de cada árbol con valores previamente determinados por límites inferior y superior. El contorno de la máscara de la Figura 3.23 se procesa cambiando los puntos (x,y) en sus pares (lat, long). Este proceso está enfocado a calcular las coordenadas WGS84 utilizando la información georreferenciada de la ortofoto a través de un proceso escrito en lenguaje Python y MatLab[61]. Con esta información es posible obtener medidas físicas. Recursos adicionales de la aplicación software en el anexo 3

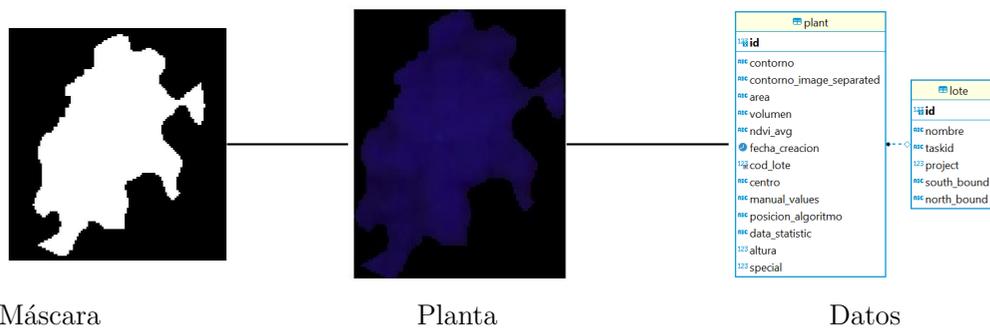


Figura 3.23: Flujo de extracción de variables: La máscara corresponde a la detección del árbol. La máscara se aplica a la imagen RGB de tres bandas para extraer los datos y almacenarlos en una base de datos relacional Mysql que fue necesaria para la persistencia de datos [62]. Fuente: propia

Descriptores Espectrales y variables estadísticas

Los descriptores espectrales corresponden a los obtenidos a través de imágenes aéreas, que serán las variables de entrada de los modelos de predicción. Estos se obtuvieron teniendo en cuenta la segmentación de la sección 3.2. Con cada árbol se procesaron los

índices de vegetación, obteniendo valores medios, máximos y mínimos. Corresponde a los pasos 1 y 2 de la Figura 3.22.

Los índices de vegetación son combinaciones matemáticas entre bandas espectrales de una imagen que representan la reflectancia espectral de la superficie [13]. Estos se pueden calcular utilizando diferentes bandas del espectro electromagnético. Las bandas están relacionadas con los factores fisiológicos y físicos de los árboles. Estos factores pueden ser el proceso de fotosíntesis, la humedad, la clorofila, el coeficiente de cultivo, entre otros. Normalmente, la reflectancia espectral de un árbol cambia según la longitud de onda y su estado fisiológico [57]. En este trabajo se recolectó el valor de los índices de vegetación y los valores máximo, mínimo y promedio de las bandas Roja, Verde, NIR y el histograma. Mediante el uso de índices de vegetación, es posible inferir diferentes condiciones físicas. Los árboles en su mayoría reflejan la banda NIR y absorben la banda roja cuando están en buenas condiciones, lo que representa un indicador de su estado fisiológico. Esta condición se evidencia utilizando el índice de vegetación de diferencia normalizada NDVI [11].

La Tabla 3.1 presenta el resumen de los índices de vegetación involucrados en este trabajo. Esta investigación basó los índices de vegetación en los resultados obtenidos por Rosas *et al.* (2021) [56] quien Realizó análisis espectral con la cámara Survey3W. Todos los índices de vegetación de la Tabla 3.1 están disponibles usando las bandas RGN de la cámara Survey3 [55].

Tabla 3.1: Índices de vegetación

Índice	Ecuación	Descripción	Ref
CRI	$\frac{R}{R_{mean}} * 100$	Índice de madurez del café	[56]
GNDVI	$\frac{NIR-G}{NIR+G}$	Índice de vegetación de diferencia normalizado verde	[56]
MCARI1	$1,2 [2,5(NIR - R) - 1,3(NIR - G)]$	Índice de relación de absorción de clorofila modificado mejorado 1	[48]
MTVI1	$1,2 [1,2(NIR - G) - 2,5(R - G)]$	Índice de vegetación triangular modificado 1	[48]
NGRDI	$\frac{G-R}{G+R}$	Índice de diferencia verde-rojo normalizado	[63]
NDVI	$\frac{NIR-R}{NIR+R}$	Índice de vegetación de diferencia normalizado	[56]
RVI	$\frac{R}{NIR}$	Índice de proporción de las radiaciones del infrarrojo cercano (NIR) y el rojo	[56]
NRVI	$\frac{RVI-1}{RVI+1}$	Normalización de RVI	[56]

Las variables estadísticas como la media de cada banda y su varianza, así como los índices de vegetación basados en la Tabla 3.1 fueron calculados recorriendo cada píxel de la planta individualizada.

Descriptores Físicos

Este proceso corresponde al paso 3 de la Figura 3.22. Para este proceso, los descriptores físicos del cafeto son medidas de características que pueden ser altura, índice de área foliar LAI o área. Este trabajo utiliza la ortofoto georreferenciada y el Modelo Digital de Superficie (DSM) para generar esta información[37]. La altura del árbol se mide restando el valor mínimo al máximo calculado con las coordenadas geográficas en el DSM. Para hacer esto, se construyó un método que itera a través de los píxeles de la imagen. Este método muestrea las alturas presentes en el DSM cada 10 píxeles en la imagen del árbol individualizado. La imagen del árbol se dilata con un radio de 20 para incluir una parte del suelo. Finalmente, se comparan todos los puntos para extraer el máximo y el mínimo, y luego su diferencia. Con esta información se calculan los valores de volumen y área lateral tomando en cuenta el trabajo de Favarin *et al.* (2002) [27], obteniendo sus valores a partir de las ecuaciones (3.3) y (3.5)

Cada planta es procesada de tal manera que es posible obtener su contorno representado en coordenadas geográficas WGS84[64]. Con este contorno en formato GEOJson es posible obtener el área en unidades de metros cuadrados utilizando el paquete *area* de python [65].

El volumen también estuvo dentro de los planes para ser calculado mediante el API que presenta WebODM para la medida de altura, mas sin embargo no encontramos una forma confiable para realizar la medida, pues con una pequeña variación de los puntos de contorno, cambia drásticamente el valor obtenido en m^2 .

Los valores de volumen finales fueron obtenidos a partir del área de contorno y la altura como ecuación (3.3), así como el área lateral que se calculó con (3.5).

$$volImage = area * \frac{h}{200} * \frac{4}{3} \quad (3.3)$$

$$diametro = 2 * \sqrt{\left(\frac{area}{\pi}\right)} \quad (3.4)$$

$$Al = \pi * \frac{diametro}{4} * \sqrt{(4 * h^2 + diametro^2)} \quad (3.5)$$

$$laiArea = -0,5786 + 0,7896Al \quad (3.6)$$

Donde el área y la altura (h) son valores previamente conocidos.

3.2. Obtención del rendimiento por planta

3.2.1. Experimento

El rendimiento de un cultivo está determinado por la cantidad de materia seca por unidad de terreno. La recopilación manual de datos es un proceso costoso, complejo y lento. Para el presente trabajo no se cuenta con información por planta que permita realizar el análisis propuesto. Las variables disponibles corresponden a las características fenotípicas y espectrales de cada una de las plantas que hacen parte del cultivo. Por ello, un método para la obtención del rendimiento de forma manual fue ejecutado siguiendo el flujo de la Figura 3.24.



Figura 3.24: Muestreo de Plantas: Proceso para ejecutar el muestreo manual del rendimiento. Fuente: propia

Por ello, desarrollamos una herramienta web que puede agilizar el proceso de obtención de muestras de rendimiento de cultivos de manera precisa y enfocada. La arquitectura puede encontrarse en la Figura 3.25. La interfaz de la aplicación web permite guardar la información automáticamente para evitar la pérdida de datos. La aplicación web integra el servicio *tiles* de WEBOdm para la carga dinámica, georreferenciada y ágil de ortofoto[66]. Esto combinado con la representación GEOJson del borde de las plantas que fue almacenado en la extracción de variables a modo de contorno, permite el uso de la librería LeafletJS que integra toda la información en un mismo mapa[67].

El resultado final es una experiencia interactiva fácil de ejecutar por personas sin conocimientos avanzados del proceso de recolección de muestras.

La segmentación de las plantas permite agilizar la recolección de datos manual en cada una de las plantas, este proceso puede ser realizado por personas no expertas de forma masiva y automática con poca supervisión.

Según el flujo propuesto en la Figura 3.24, el primer paso es la ubicación de la planta a muestrear, este paso se facilita con el reconocimiento del entorno del cultivo y la planta objetivo con la ayuda del mapa de la aplicación. Las plantas con información completa cambian de color para evitar traslapes y ayudar a la referencia en campo. En este proceso es posible al usar la segmentación masiva realizada de la sección 3.1.2.3 como lo muestra la Figura 3.26 que expone la interfaz desarrollada para la obtención de variables manuales.

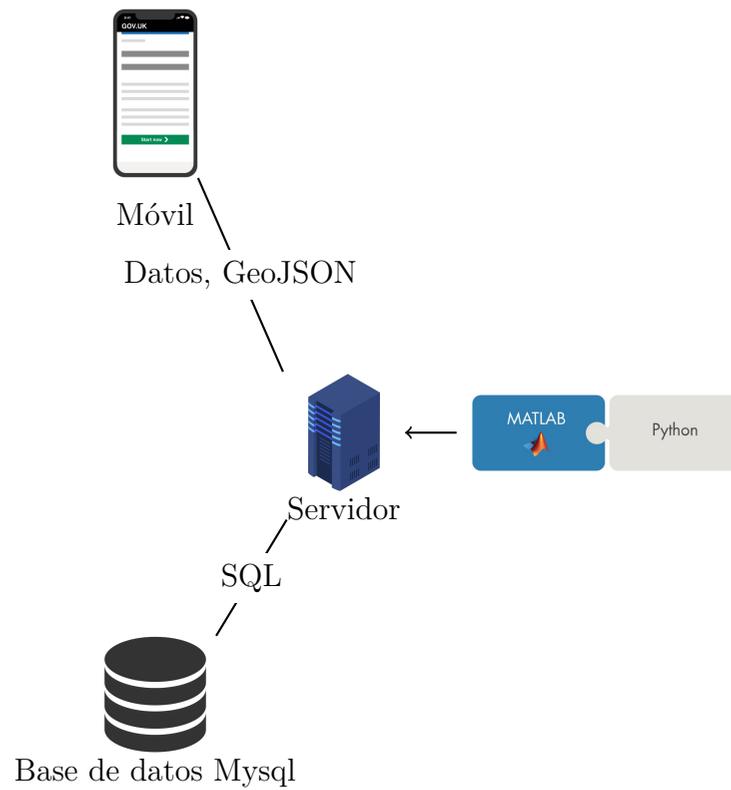


Figura 3.25: Arquitectura de software para la recolección manual de muestras de rendimiento de cultivos. Fuente: propia

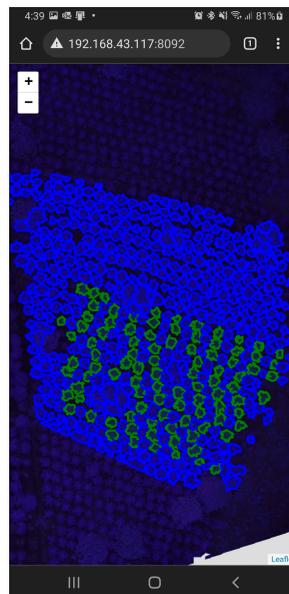


Figura 3.26: Interfaz de servicio web para muestreo manual, en verde las plantas con información completa, en azul las plantas por completar. Fuente: propia

Para el paso 2, el objetivo es medir la altura y el diámetro de la planta con la ayuda de una cinta métrica. Luego de tener la información en centímetros de este paso, se procede a ingresar los datos dentro del aplicativo basado en la interfaz de la Figura 3.27

The image shows a mobile application interface titled "Data Plant" with a close button (x). It contains three input fields: "Height(cm)", "Diameter(cm)", and "Status". Below these fields, there are two highlighted sections: a green-bordered box labeled "Header" and an orange-bordered box labeled "Footer".

Figura 3.27: Ingreso Altura y Diámetro. Fuente: propia

En el costado inferior, la Figura 3.27 expone dos secciones denominadas *header* y *footer* donde es posible registrar los datos de los pasos 3, 4 y 5 del flujo de la Figura 3.24. Estos registros son almacenados con la ayuda de la interfaz de la Figura 3.28, que solicita realizar una muestra de las ramas ingresadas en el campo inmediatamente superior.

Las muestras de nodos y granos están determinados por un tamaño de muestra estadístico calculado para cada campo con un error del 8% y un nivel de confianza del 80%. La desviación estándar se fijó en 0,5 [68]. Finalmente, si todos los datos están completos ya es posible guardar toda la información de la planta. Es importante aclarar que dichos datos irán directamente a la planta seleccionada y estarán almacenados en la base de datos relacional Mysql.

The image shows a section of the application interface titled "Header". It contains three input fields: "Branches" (with the value "34"), "Nodes", and "Beans". A red warning message is displayed: "You Must count the nodes of 22 Branches".

Figura 3.28: Muestreo de Ramas, Nodos y Granos. Fuente: propia

3.2.2. Ejecución

Teniendo en cuenta la complejidad de la tarea de obtener datos. Dicha tarea fue ejecutada el 15, 16, 17 de Diciembre del año 2021 y 6, 19 de Marzo del 2022. Este proceso contó con la participación de 5 a 8 personas voluntarias cada día.



Figura 3.29: Recolección 1



Figura 3.30: Recolección 2



Figura 3.31: Recolección 3

La recolección de muestras tuvo lugar en un lapso de tiempo de 5 días, permitiendo

conseguir información del rendimiento manual de 373 árboles de forma aleatoria en 4 lotes de diferentes edades. Los retos a los que el equipo de trabajo se enfrentó se encuentran descritos en el anexo 4.

3.3. Resultados y discusión

Con el cálculo de los descriptores físicos y espectrales, es importante estudiar el comportamiento de estos descriptores respecto al rendimiento manual. El análisis de correlación de *Pearson* permitió verificar el comportamiento de los descriptores [69]. El cálculo de los coeficientes de correlación y demás tareas del análisis fue realizado con la ayuda de paquetes para Python tales como Pandas, scikit-learn, Matplotlib, entre otras [70, 71, 72].

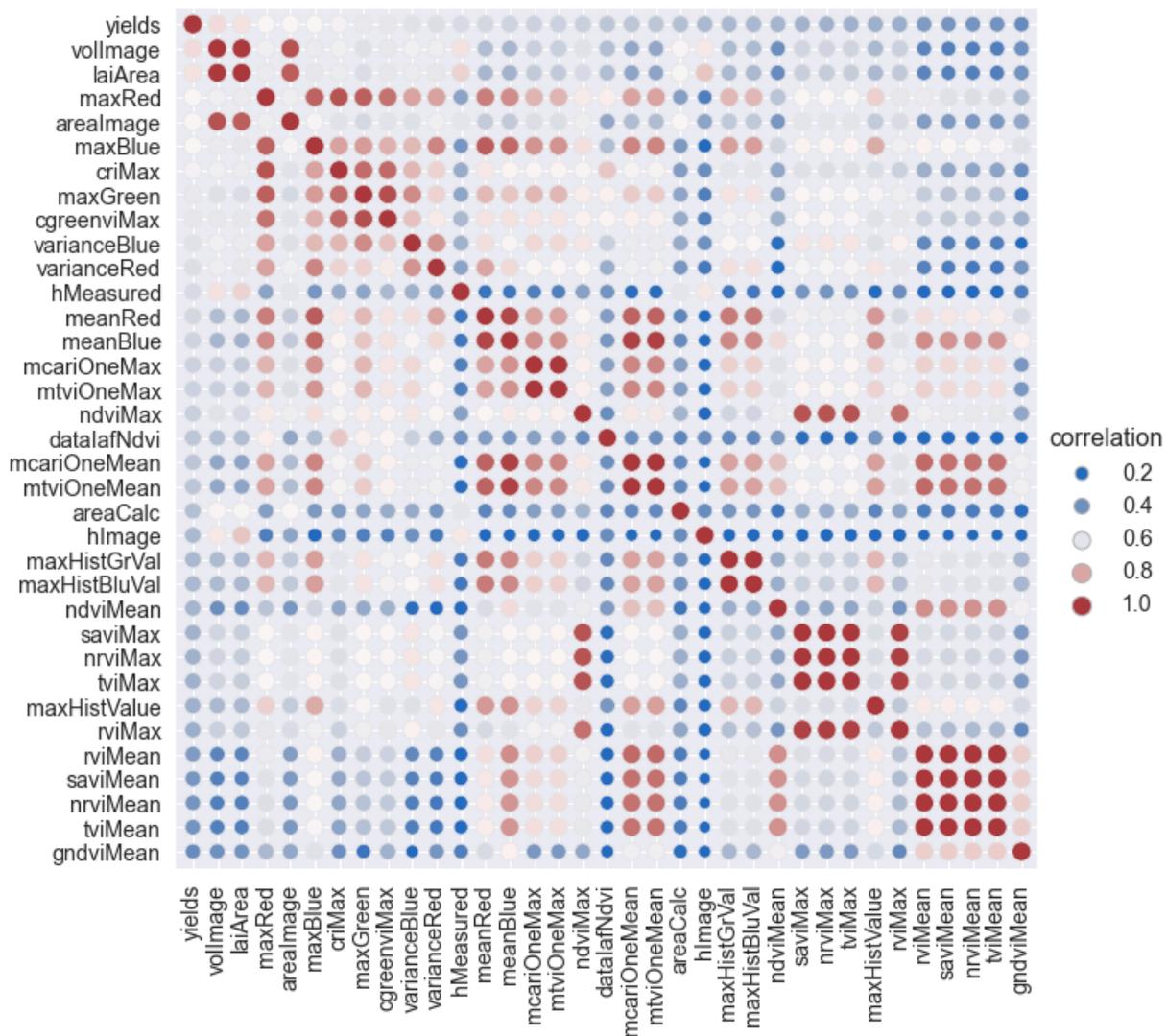


Figura 3.32: Matriz de correlación, variables correlacionadas con el rendimiento ordenadas de mayor a menor. Fuente: propia

Los resultados de la Figura 3.32 muestran las variables más correlacionadas con el rendimiento del cultivo.

Tabla 3.2: Correlación de variables con Rendimiento(Yield)

Variable	Descripción	Correlación
volImage	Volumen de la Imagen	77 %
laiArea	Índice de área foliar 3.5	77 %
NDVImean	NDVI medio	49 %
NDVImin	NDVI mínimo	-39 %
NDVImax	NDVI máximo	55 %

Los valores de volumen como volImage calculado a partir del área de contorno y altura como ecuación (3.3) obtiene una correlación de 77 %, así como el área lateral que se calculó con (3.5) obtuvo 76 % un valor esperado según [20] quien en su trabajo demuestra las relaciones entre características físicas como LAI que está fuertemente relacionada con el área lateral y el volumen [27]. En cuanto a los índices de vegetación, esta investigación encontró diferentes comportamientos. El NDVImean tuvo una correlación del 49 %, el NDVImin y el NDVImax -39 % y 55 %, respectivamente. Demostrando la importancia delNDVIque, según [56], representa el verdor y vigor de la planta [73].

El gráfico también muestra una fuerte relación entre la altura medida (hMeasured) y la altura calculada (hImage) con el DSM ya que el índice de correlación fue del 75 %, lo que valida el proceso de cálculo de la altura del árbol con base en la información de la imagen.

Los índices de vegetación RVI, SAVI, NRVI y TVI tienen una correlación cercana al 45 %, por lo que es fundamental tenerlos en cuenta en el análisis del rendimiento de los cultivos. Este trabajo midió todas estas variables en la etapa de floración de los cultivos, que representa el estado de las plantas antes del llenado de grano. Los demás índices con sus valores medio, máximo y mínimo tuvieron valores ligeramente superiores pero inferiores a 0,5, probablemente debido a su baja relación fisiológica con el rendimiento del cultivo. Al incluir valores máximos y mínimos en el análisis para el estudio de los índices de vegetación, fue posible encontrar que elNDVItenía valores de correlación más altos en el promedio de los máximos.

De acuerdo con estas correlaciones, los predictores se prepararon en marcos de datos para ejecutar los modelos de regresión. Algunas variables involucradas fueron volumen, NDVImax, maxRed, maxBlue y criMax, entre otras. La Figura 3.33 muestra un gráfico de dispersión con volImage, maxRed y maxBlue que muestra una relación lineal entre todas estas variables y el rendimiento(yields). El volumen se correlaciona en un 77 % con el rendimiento del cultivo y tiene menos dispersión en comparación a las otras dos variables. Los maxRed y maxBlue frente a yields presentan gráficas similares pero con menor concentración en sus valores altos.

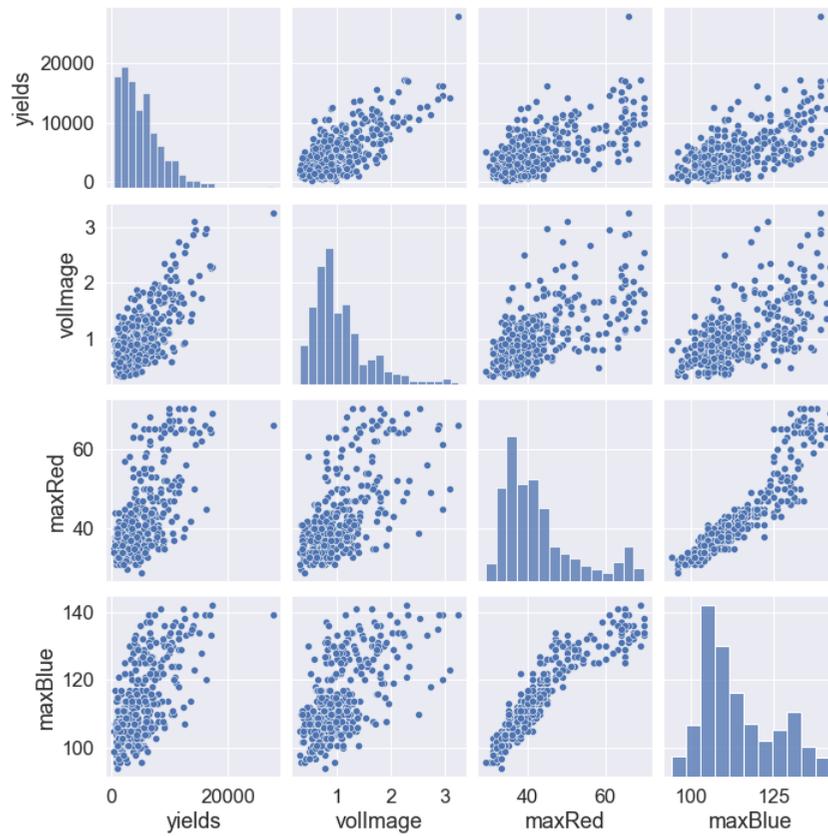


Figura 3.33: Gráfico Dispersión. Fuente: propia

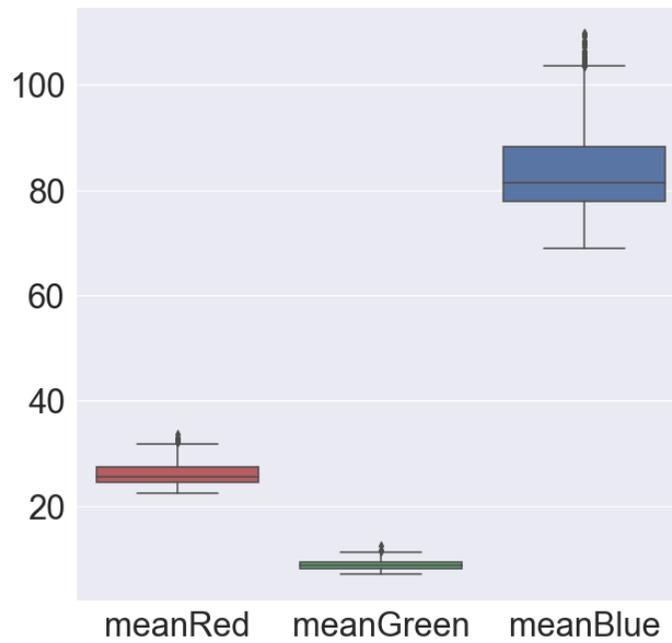


Figura 3.34: Media de las bandas roja, azul y verde. Fuente: propia

La Figura 3.34 muestra el comportamiento de las bandas espectrales presentado por Rosas *et al.*(2021) [56], con respecto a la reflectancia espectral en plantas. Este resultado valida la corrección radiométrica y muestra que la reflectancia de la cámara está de acuerdo con lo esperado ya que para las plantas la reflectancia NIR es la de menor absorción y mayor reflectividad [56].

Este capítulo expuso el proceso completo para la obtención de las diferentes descriptores mediante la cámara Survey3W disponible para esta investigación. Al finalizar todo el proceso, es posible afirmar que el método propuesto es robusto, confiable y suficiente para efectuar la segmentación de plantas.

En resumen, como resultado del proceso de segmentación fue posible obtener descriptores espectrales entre los que se encontraron variables estadísticas que corresponden a diferentes cálculos realizados en las bandas espectrales disponibles tales como la media, la varianza, máximos y mínimos, y valores de histograma de cada una de las bandas espectrales de la imagen. Así como también fue posible calcular algunos índices de vegetación calculados a partir de las bandas disponibles RGN resumidos en la Tabla 3.1. Y variables físicas que representan mediciones sobre las características fisiológicas de las plantas, en este caso contamos con la altura dada en metros y el área dada en m^2 .

Capítulo 4

Modelo de Predicción de rendimiento

Este capítulo presenta el proceso realizado para definir, entrenar y validar los modelos de regresión para la predicción del rendimiento de los cultivos de café. Este estudio está orientado hacia la predicción temprana del rendimiento del cultivo de café solo con datos de imágenes aéreas obtenidos utilizando herramientas de bajo costo. Las imágenes multispectrales obtenidas de los UAV permiten un análisis focalizado de alta precisión, lo que para esta investigación condujo a la individualización precisa y automática de las plantas, generando descriptores físicos y espectrales con potencial para predecir los rendimientos de los cultivos [46, 47, 48].

Con los datos de las plantas de café a partir de las imágenes multispectrales y los valores del rendimiento manual, el paso siguiente es la implementación de los modelos de predicción del rendimiento. Posteriormente, múltiples modelos de regresión son propuestos con el fin de establecer la viabilidad y la capacidad de los modelos para realizar una predicción efectiva del rendimiento. Finalmente se ejecuta una validación total del proceso en un cultivo que permita establecer la factibilidad para el uso del modelo en un cultivo de café para un ciclo productivo controlado.

4.1. Modelos de regresión

Los modelos de predicción pueden definirse como una representación de la relación entre dos o más variables de manera formal [74]. Para este trabajo, los modelos más ajustados a la cantidad de datos disponibles son los de regresión. Todos los modelos de predicción de este tipo tienen sus fundamentos en la teoría del álgebra lineal [75]. Existen diferentes tipos de modelos de regresión. Uno de ellos es el modelo de regresión lineal, que dependiendo de la cantidad de entradas es catalogado como modelo de regresión simple o múltiple. Este tipo de modelos son descritos por ecuaciones como la 4.1 que involucra un vector de entradas (X) y un vector de coeficientes que son calculados mediante la ejecución de un proceso de mínimos cuadrados para aproximar el modelo lineal a los datos de entrenamiento. Por lo general la operación matemática de la ecuación 4.1 pretende

obtener los valores estimados de la variable y que se suele llamar variable objetivo.

$$y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_m * X_m + \epsilon \tag{4.1}$$

Todos los modelos de regresión necesitan evaluarse mediante una métrica que permita conocer la calidad del proceso de predicción que están realizando. Existen distintas métricas que están disponibles para este propósito, entre ellas se encuentran el R^2 o coeficiente de determinación que mide qué tan bueno es un algoritmo para adecuarse a una variable. También se define como la porción de la variable independiente que puede ser predicha por las variables independientes. Se calcula mediante la ecuación 4.2

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})P_i - \bar{P}}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2 \tag{4.2}$$

$$RMSE = \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \right) \tag{4.3}$$

Donde \bar{O} y \bar{P} expresan las medias de los valores observados y predichos, respectivamente; y n es el número de muestras;

Además, para realizar un entrenamiento dinámico de los modelos es posible aplicar técnicas como la validación cruzada como la presentada en la Figura 4.1. Esta metodología divide el conjunto de datos en grupos de datos más pequeños, haciendo validaciones cruzadas entre todos los grupos resultando en un valor de R^2 para cada iteración. El resultado final del proceso corresponde a la media de todos los valores [76]

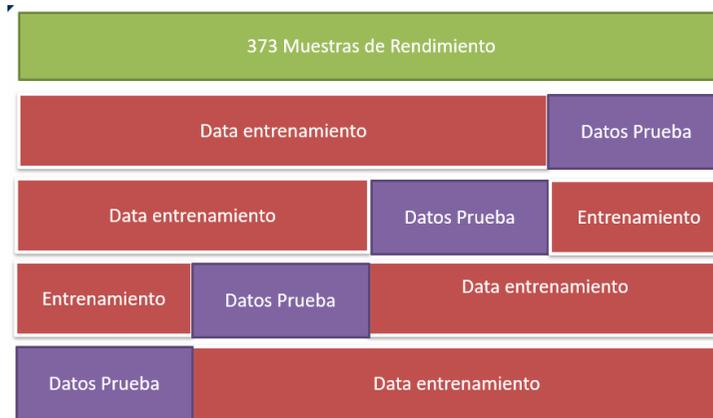


Figura 4.1: Proceso de validación cruzada con las 373 muestras de datos manuales Fuente [76]

La Tabla 4.1 presenta las variables obtenidas a través del método de selección de características que obtuvieron los mejores resultados del coeficiente de determinación R^2 . En este proceso, algunas variables correlacionadas, como criMax con una correlación del 63 %, disminuyeron el comportamiento del coeficiente R^2 , y variables con una correlación

Tabla 4.1: Variables para modelo de regresión múltiple

Variable	Resumen
vollImage	Volumen obtenido a partir de la altura y el área
maxRed and maxBlue	Valor máximo de las bandas roja y azul del árbol
varianceBlue	Valor de varianza de la banda azul (en este caso, banda NIR)
laiArea	Índice de área foliar de la ecuación 3.6
saviMax	Máximo del índice de vegetación ajustado por suelo
tviMax	Máximo del Índice de Vegetación Triangular
mtviOneMean	Media del índice de vegetación triangular modificado
ndviMax and ndviMean	Índice de vegetación de diferencia normalizado media y máximo

del 45 % como saviMax, tviMax y ndviMean aumentó el valor de R^2 para la regresión lineal.

La Figura 4.2 muestra algunos modelos entrenados con el volumen para una representación 2D. Este gráfico muestra el comportamiento de los algoritmos de árboles de decisión en comparación con RandomForest, mostrando sus características no lineales. Además de mostrar que SVR, Linear Regression y SDG tienden a modelar los datos de manera similar.

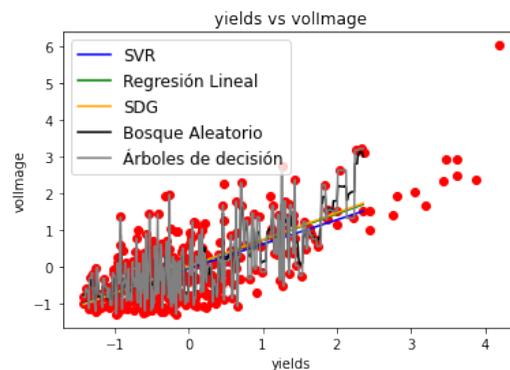


Figura 4.2: Predicción de modelos entrenados con el volumen. Fuente: propia

La Figura 4.3 permite evidenciar los rangos de predicción que tienen los modelos de regresión más relevantes, con esta gráfica es posible determinar la dispersión que tienen los rendimientos predichos respecto al rendimiento original.

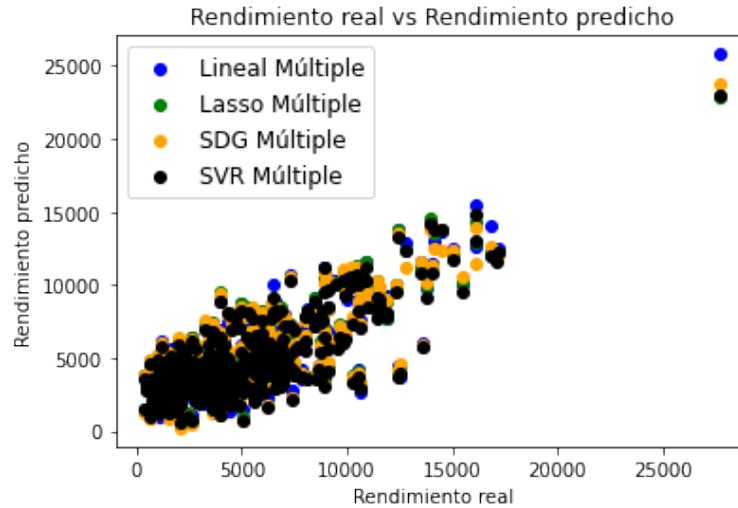


Figura 4.3: Comparación del rendimiento original con el predicho en diferentes modelos de predicción. Fuente: propia

Los modelos que estuvieron involucrados en el proceso, con sus respectivos resultados pueden ser encontrados en la Tabla 4.2.

Tabla 4.2: Resultados de los modelos de regresión [1]

Modelo	Tipo	R^2	RMSE
Regresión Lineal	Simple	0.48	2648
	Múltiple	0.576	2361
Regresión Lasso	Simple	0.48	2648
	Múltiple	0.55	2442
Regresión PLS	Simple	0.48	2648
	Múltiple	0.544	2462
Regresión SDG	Simple	0.1	3474
	Múltiple	0.51	2546
SVR	Simple	0.47	2660
	Múltiple	0.53	2477
Regresión Bosque Aleatorio	Simple	0.22	3226
	Múltiple	0.50	2526
Regresión Árbol de decisión	Simple	-0.05	3747
	Múltiple	0.12	3333

Los resultados de la Tabla 4.2 muestran que para este proceso, el método que mejores resultados obtuvo es el método de regresión lineal con 57.6 % y un RMSE de 2361, seguido de Lasso con un R^2 de 55 % y con RMSE de 2442. El modelo SVR con kernel lineal obtuvo un 53 %. El modelo SVR con kernel lineal obtiene un 53 %. En este modelo, cuando se cambió el kernel a RBF, el resultado disminuyó al 46 % [77]. Como los modelos anteriores tienen una base lineal, ajustan mejor los datos, pero la diferencia entre Random Forest aplicado al volumen con un 22 % y multivariable con un 52 % es un fenómeno interesante. En este escenario, el modelo menos adecuado para los datos fue Decision Tree, como se puede ver en la Figura 4.2, el modelo se ve afectado por la dispersión de los datos. El modelo Random Forest se ve menos afectado por la dispersión y su rendimiento se acerca al SVR, siendo superior a este modelo con kernel tipo RBF con un 48 %. Similar a [37] en este trabajo, la altura y el área no son relevantes por separado hasta que se conviertan en volumen. El LAI se comporta de manera similar al volumen cuando se hace una regresión simple, demostrando la importancia de este índice definido en múltiples trabajos anteriores [37, 27].

El proceso de aplicación del modelo de predicción se basa en la biblioteca scikit-learn utilizando modelos de regresión de aprendizaje supervisado. Las entradas a los modelos fueron rendimiento manual y predictores. Se probó un modelo de variable simple con volumen por ser el de mayor correlación con el rendimiento del cultivo y modelos múltiples con las variables posteriores para comparar los modelos de regresión simple y múltiple. Se realizaron múltiples iteraciones usando diferentes configuraciones con los modelos de regresión disponibles, por ejemplo, para SVR, el kernel se cambió de RBF a lineal. Para SDG, la función de pérdida se cambió entre huber, squared_error o epsilon_insensitive, mostrando mejores resultados con la configuración squared_error. Para los demás modelos se mantuvo la configuración por defecto. En general se probó escalando todos los datos usando la función StandardScaler. Los modelos fueron evaluados mediante validación cruzada con un puntaje R^2 que nos permite determinar el porcentaje de ajuste del modelo a los datos. Los predictores y los modelos con la puntuación R^2 más alta se muestran en la Tabla 4.2. El papel del volumen en los modelos de regresión puede explicarse por la relación entre la arquitectura del árbol y la edad [21]. Además, se puede complementar con un indicador de estado fisiológico como el NDVI [37].

Teniendo en cuenta los resultados anteriores es posible afirmar que el mejor modelo es el de regresión lineal para la predicción del rendimiento de un cultivo de café.

4.2. Validación

El proceso de validación de todo el proyecto consistió en ejecutar todos los pasos del experimento en un ciclo de cultivo controlado. Este proceso se ejecutó en dos cultivos que tuvieron las mismas condiciones que los usados para el proceso de entrenamiento de los modelos de regresión.

4.2.1. Cultivos de prueba

Los cultivos involucrados en la validación de la predicción del rendimiento tienen un control de su ciclo de producción que está determinado por variables como sus fertilizantes, manejo de plagas, variedad, edad entre otros. Esta validación estuvo centrada en un cultivo involucrado dentro del muestreo usado para el entrenamiento y uno completamente independiente pero bajo condiciones de manejo y ambientales similares.

Para realizar la validación fue necesario contar con el apoyo de las personas expertas encargadas de los cultivos, inicialmente el objetivo fue consultar los manejos de los cultivos y consolidar la información de la Tabla 4.3

Elemento	La Sultana	San Antonio
Edad	3 Años	5 Años
Variedad	Castillo Tambo	Castillo Tambo
Cantidad	3000	4000
Potencial	210 Arrobas	300Arrobas
Año Anterior	50 Arrobas	80 Arrobas
Dosis Fertilizante Marzo	25-04-24 70Gr/planta	100Gr/Planta
Dosis Fertilizante Septiembre	17-06-18-02 70Gr/Planta	100Gr/Planta
Abono Orgánico	1Kg/planta Mayo	0
Limpiezas	3 Por Año	2 por Año

Tabla 4.3: Información de Manejo de los Cultivos * 1 Arroba = 11.34Kg

De la información anterior es importante resaltar para el cultivo de La Sultana el potencial de producción a la que se espera llegar en algún momento del desarrollo del cultivo, pero donde es importante resaltar que el año anterior obtuvo un rendimiento de 50 arrobas posiblemente debido a las condiciones climáticas del año 2020. De igual forma el cultivo de la finca también presenta una producción potencial de 300 arrobas para 4000 plantas teniendo una edad de 4 años. Toda esta información solo es útil para realizar una comparación de las condiciones de los cultivos con su rendimiento final.

4.2.2. Validación de la segmentación

Las pruebas del proceso de segmentación de la sección 3.1.2.3 inicia con el post procesamiento de las imágenes generando los ortofotos de los cultivos mediante la herramienta WebODM que permite la combinación de las imágenes capturadas bajo las mismas condiciones para el proceso de entrenamiento. Las figuras 4.4 y 4.5 muestran la delimitación de los cultivos a procesar para el proceso de validación.

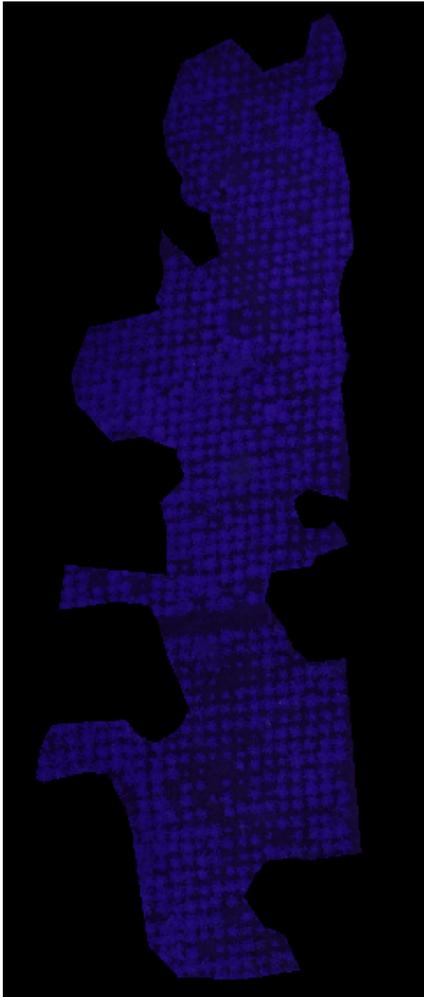


Figura 4.4: RGB Imagen original finca La Sultana. Fuente: propia



Figura 4.5: RGB Imagen Original finca San Antonio. Fuente: propia

Con las áreas ya delimitadas es posible pasar a ejecutar los procesos de marcado de Foreground y Background que son requeridos para ejecutar el proceso de segmentación con el algoritmo Watershed descrito en la sección 3.1.2.3 . Al ejecutar el proceso antes planteado fue posible obtener plantas individuales que era el objetivo inicial del proceso de segmentación. Las figuras 4.6 y 4.7 muestran el resultado del proceso concluyendo que fue posible ejecutar el proceso planteado sin mayores ajustes a los valores inicialmente definidos para los cultivos de entrenamiento.

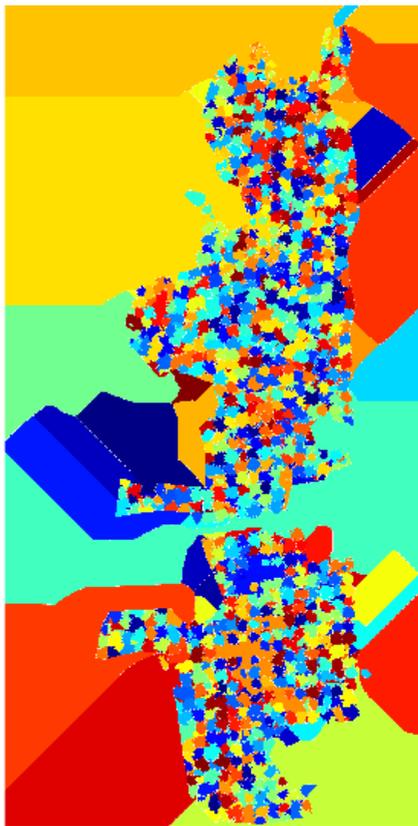


Figura 4.6: Plantas individualizadas de la finca La Sultana

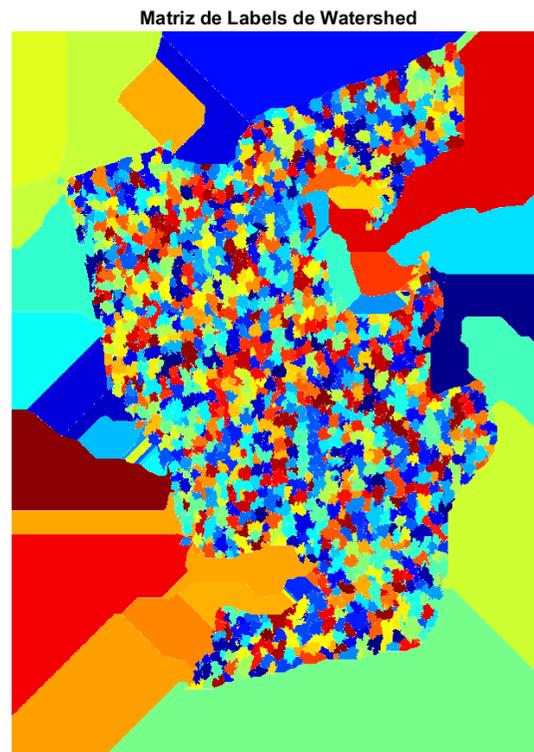


Figura 4.7: plantas individuales de la finca San Antonio

4.2.3. Validación del modelo de regresión

El proceso de generación de datos concluyó con la consolidación de 716 muestras válidas para el lote de La Sultana y de 1172 para la finca San Antonio. Con las plantas ya procesadas es requerido diseñar un método que permita entrenar el modelo de regresión y ejecutar la predicción con las plantas de validación para ambos lotes conservando los parámetros iniciales de predicción. Como los resultados de la Tabla 4.2 determinan que el modelo con mejor desempeño es el de regresión lineal, los pasos siguientes se ejecutarán con este mismo modelo. El proceso diseñado para la validación es el mostrado en la Figura 4.8.

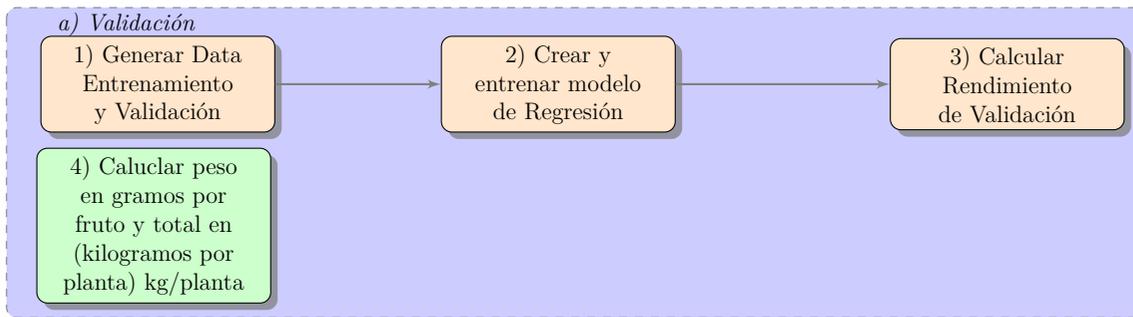


Figura 4.8: Proceso de validación de modelo de predicción donde se general los datos, se entrenan los modelos para finalmente llegar a una medida de kg/planta. Fuente: propia

4.2.3.1. Generación de datos de entrenamiento y validación

El proceso de la generación de datos consiste en la consolidación de los objetos de datos que definen los predictores como las variables físicas y espectrales y el rendimiento por planta manual en la sección de entrenamiento. De igual forma se crea una segunda lista de datos de validación pero sin los valores del rendimiento manual pues estos no existen para la data de validación.

4.2.3.2. Creación y entrenamiento del modelo

Este punto es posible ejecutarlo mediante la creación del modelo de regresión. En este punto, como el modelo es capaz de funcionar correctamente con la data no escalada es posible ejecutar la función "fit" para entrenar el modelo con las variables de entrenamiento con lo que este quedará ya definido teniendo en cuenta los predictores de entrada establecidos en la sección 3.3.

4.2.3.3. Calcular el rendimiento de validación

Con el modelo ya entrenado es posible utilizar las variables de validación definidas en un arreglo de las mismas dimensiones que en el proceso de entrenamiento para evitar conflictos. Este proceso generó los valores del rendimiento resultado de la aplicación del modelo de regresión lineal múltiple que es posible utilizar para el cálculo de la carga de granos de café por planta.

4.2.3.4. Calcular peso en gramos por fruto y total kg/Ha

Para lograr establecer una medida aproximada del peso de cada fruto fue necesario medir repetitivamente una cantidad apreciable de granos de café pergamino seco para luego promediarlos.

4.2.4. Resultados

Teniendo en cuenta el proceso propuesto y las variables obtenidas mediante la ejecución de cada uno de los pasos se tienen los siguientes resultados:

La Tabla 4.4 expone el peso del café pergamino seco. Con este proceso es posible tener una idea de la equivalencia en kilogramos del número de granos obtenido de la predicción del rendimiento.

Tabla 4.4: Medidas de granos de café:

Cantidad(Granos)	Peso(Kg)
50	0.012
100	0.023
100	0.022
50	0.012
100	0.023
100	0.023
100	0.023
100	0.024
100	0.023
100	0.023
300	0.071
Total Granos: 1200	Peso:0.279

En la Tabla 4.5 se presentan los resultados del proceso de validación del proceso de predicción del rendimiento.

Tabla 4.5: Resumen proceso de detección

Lote	Plantas Detectadas	Rendimiento Detectado promedio	Rendimiento por Planta
La Sultana	716	5380	1.25 kg/planta
San Antonio	1172	21245	4.9 kg/planta

Teniendo en cuenta los resultados de la Tabla 4.5 es posible inferir que el cultivo separado de la finca San Antonio tiene unas características detectadas que hacen que el rendimiento por planta esté cuatro veces por encima de la finca La Sultana. Este

comportamiento podría estar determinado por la edad y el manejo que reciben cada uno de los cultivos. Es importante notar que el rendimiento por planta de 4.9Kg está más cerca del promedio que establece [78].

El resultado obtenido anteriormente es prometedor para la investigación pues determina que la cantidad obtenida está dentro de lo esperado para los cultivos de café.

Capítulo 5

Conclusiones y trabajos futuros

Este capítulo describe las conclusiones obtenidas durante el desarrollo de la presente investigación así como algunos trabajos futuros propuestos. Este trabajo analiza la viabilidad de predecir el rendimiento de un cultivo de café a través de imágenes aéreas multiespectrales. El trabajo comienza con la discriminación de condiciones para la captura óptima de fotografías en la etapa de floración. Posteriormente, este trabajo describió el procesamiento de imágenes para la segmentación e individualización de las plantas donde el algoritmo de cuencas hidrográficas obtuvo los resultados esperados

5.1. Conclusiones

- El desarrollo del trabajo encontró una fuerte correlación entre variables físicas como el volumen y variables espectrales como el NDVI máximo y el rendimiento del cultivo con el 75 % y 55 % de correlación respectivamente permitiendo conseguir un índice de determinación R^2 del 57 % que demuestra la factibilidad de la predicción del rendimiento con los modelos de regresión analizados. La segmentación de las plantas ayuda a evitar la influencia del suelo y otras plantas en el cálculo de los índices de vegetación.
- Es posible segmentar cultivos de café mediante el algoritmo de cuenca hidrográfica *watershed* de forma confiable permitiendo la obtención de información de forma ágil y automática en comparación con trabajos como Barbosa *et al.* [37]. Este proceso permite la segmentación de las plantas que es integrado en una aplicación web que permitió tener agilidad y precisión en el etiquetado de los datos manuales del rendimiento.
- El porcentaje de precisión del modelo no es concluyente. La precisión se puede mejorar al ampliar la base de datos incluyendo plantas de distintas edades, aumentando la cantidad de datos y también realizando muestras manuales de cosechas acumuladas en diferentes años.
- El conocer el rendimiento del cultivo tiene un alto impacto en las actividades agrícolas. Este trabajo demostró que es factible la predicción del rendimiento de forma

temprana mediante el uso de imágenes aéreas y herramientas de bajo costo utilizando solamente información espectral.

- La segmentación de las plantas mediante el procesamiento de imágenes permite la clasificación y el etiquetado automatizados de muestras de rendimiento manual para el entrenamiento del modelo de regresión. Considerando los resultados de la tabla 4.2, es posible afirmar que el mejor modelo es el modelo de regresión lineal para predecir el rendimiento de un cultivo de café logrando un coeficiente de determinación R^2 del 57% y un RMSE de 2361 granos de café.
- El uso de imágenes multiespectrales que involucran la banda NIR tienen el potencial de permitir la predicción del rendimiento. Sin embargo, se requiere extender la investigación sobre otras bandas para verificar su correlación con el rendimiento.

5.2. Trabajos Futuros

Esta investigación estuvo enfrentada a múltiples retos técnicos y limitaciones que podrían ser mejoradas y superadas en investigaciones posteriores. Las investigaciones venideras en este tema fundamental podrían realizarse en estos trabajos futuros:

- Refinar y reproducir todo el proceso de segmentación para que sea posible ejecutar una validación teniendo en cuenta imágenes hiperespectrales para ampliar la cantidad de bandas disponibles en el análisis. Esto podría permitir el análisis con índices de vegetación relevantes encontrados en la literatura como el NDWI que permite establecer un análisis del estado hídrico de las plantas.
- La exactitud en las medidas como el área y el volumen podrían mejorarse con el uso de puntos de control en tierra mediante el uso de un GPS de precisión que permita el marcado de estos puntos dentro de las imágenes y así mejorar la precisión de la georreferenciación del ortofoto y el DSM. Así como el uso de un GPS a bordo de mejor calidad que evite los problemas en la marcación de altura que afectaron esta investigación.
- Una de las limitaciones de este trabajo es la cantidad de datos disponibles para el uso de los diferentes modelos de predicción. Es deseable obtener más datos de rendimiento manual en lotes de café con más características, como la edad o la variedad.
- Los modelos de predicción están limitados por la cantidad de datos de entrenamiento disponibles. Este trabajo involucró modelos de regresión simples que podrían mejorar su comportamiento con más muestras del rendimiento. Esto también permitiría el uso de modelos más avanzados como las redes neuronales o aprendizaje profundo. Este proceso puede ser complementado con los datos ya existentes y con información de cosechas acumuladas en distintas condiciones climáticas a través de los años.

- La interfaz web de obtención de muestras se puede adaptar para la recolección de otro tipo de datos a partir de las plantas, como las muestras de clorofila o el marcado de fotografías para el conteo manual mejor adaptado que el utilizado en la presente investigación. También es posible implementar otro tipo de aplicaciones geolocalizadas en los cultivos como sensores en tiempo real o mejoras al geolocalizar la logística del manejo de los cultivos como la fertilización o la cosecha dentro del ortofoto. También permitiría realizar un seguimiento del personal en un proceso de gestión de las diferentes tareas del cultivo.

Bibliografía

- [1] Scikii-learn. Supervised learning. [Online]. Available: https://scikit-learn.org/stable/supervised_learning.html
- [2] O. L. Ocampo-López and L. M. Alvarez-Herrera, “Tendencia de la producción y el consumo del café en colombia,” Apuntes del CENES, vol. 36, no. 64, pp. 139–165, 2017. [Online]. Available: <http://www.scielo.org.co/pdf/cenes/v36n64/0120-3053-cenes-36-64-00139.pdf>
- [3] F. Cafeteros, “Producción de café de colombia cerróel 2019 en 14,8 millones de sacos,” 2020. [Online]. Available: <https://federaciondefeteros.org/wp/listado-noticias/produccion-de-cafe-de-colombia-cerro-el-2019-en-148-millones-de-sacos/>
- [4] —, “Producción de café de colombia cae 15 % en octubre,” 2021. [Online]. Available: <https://federaciondefeteros.org/wp/listado-noticias/produccion-de-cafe-de-colombia-cae-15-en-octubre/>
- [5] J. Arcila, F. Farfan, A. Moreno, L. F. Salazar, and E. Hincapié, “Sistemas de producción de café en colombia,” 2007. [Online]. Available: <https://biblioteca.cenicafe.org/bitstream/10778/720/1/Sistemas%20producci%C3%B3n%20caf%C3%A9%20Colombia.pdf>
- [6] J. M. Miranda, R. A. Reinato, and A. B. d. Silva, “Modelo matemático para previsão da produtividade do cafeeiro,” Revista Brasileira de Engenharia Agrícola e Ambiental, vol. 18, no. 4, pp. 353–361, 2014.
- [7] J. Rendón, J. Arcila, and E. Montoya, “Estimación de la producción de café con base en los registros de floración,” 2008. [Online]. Available: [https://www.cenicafe.org/es/publications/arc059\(03\)238-259.pdf](https://www.cenicafe.org/es/publications/arc059(03)238-259.pdf)
- [8] E. Montoya-Restrepo, Modelo para simular la producción potencial del cultivo del café en Colombia, ser. Boletín técnico. FNC-Cenicafé, 2009. [Online]. Available: <https://books.google.com.co/books?id=jOwBaAEACAAJ>
- [9] T. W. Idol and A. H. Youkhana, “A rapid visual estimation of fruits per lateral to predict coffee yield in hawaii,” Agroforestry Systems, vol. 94, no. 1, pp. 81–93, 2020.

- [10] S. d. A. Silva, J. d. S. Lima, and R. de Oliveira, “Agrometeorological model estimating the productivity of two varieties of arabic coffee considering the spatial variability.” IRRIGA, vol. 16, no. 1, pp. 1–10, 2011.
- [11] V. G. C. d. Rosa, M. A. Moreira, B. F. T. Rudorff, and M. Adami, “Coffee crop yield estimate using an agrometeorological-spectral model,” Pesquisa Agropecuária Brasileira, vol. 45, no. 12, pp. 1478–1488, 2010.
- [12] C. Gonzaga Aguilar, “Aplicación de índices de vegetación derivados de imágenes satelitales landsat 7 etm+ y aster para la caracterización de la cobertura vegetal en la zona centro de la provincia de Loja, Ecuador,” Ph.D. dissertation, Universidad Nacional de La Plata, 2014.
- [13] R. E. Crippen, “Calculating the vegetation index faster,” Remote sensing of Environment, vol. 34, no. 1, pp. 71–73, 1990.
- [14] D. K. Bolton and M. A. Friedl, “Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics,” Agricultural and Forest Meteorology, vol. 173, pp. 74–84, 2013.
- [15] A. Ahmad and A. M. Samad, “Aerial mapping using high resolution digital camera and unmanned aerial vehicle for geographical information system,” in 2010 6th International Colloquium on Signal Processing & its Applications. IEEE, 2010, pp. 1–6. doi: 10.1109/CSPA.2010.5545303
- [16] B. Brisco, R. Brown, T. Hirose, H. McNairn, and K. Staenz, “Precision agriculture and the role of remote sensing: a review,” Canadian Journal of Remote Sensing, vol. 24, no. 3, pp. 315–327, 1998.
- [17] S. Liaghat and S. Balasundram, “A review: the role of remote sensing in precision agriculture.” American Journal of Agricultural and Biological Sciences, vol. 5, no. 1, pp. 50–55, 2010.
- [18] A. Vibhute and S. K. Bodhe, “Applications of image processing in agriculture: A survey,” International Journal of Computer Applications, vol. 52, no. 2, pp. 34–40, 2012. doi: 10.5120/8176-1495
- [19] V. Ramirez, “La fenología del café una herramienta para apoyar la toma de decisiones,” Centro Nacional de Investigaciones de Café (Cenicafé), Tech. Rep., 2014. [Online]. Available: <https://biblioteca.cenicafe.org/bitstream/10778/489/1/avt0441.pdf>
- [20] S. Castro-Tanzi, M. Flores, N. Wanner, T. V. Dietsch, J. Banks, N. Ureña-Retana, and M. Chandler, “Evaluation of a non-destructive sampling method and a statistical model for predicting fruit load on individual coffee (Coffea arabica) trees,” Scientia Horticulturae, vol. 167, pp. 117–126, 2014. doi: <https://doi.org/10.1016/j.scienta.2013.12.013>

- [21] C. A. U. Muñoz, R. D. M. Rivera, and C. P. F. Ramos, “Relación entre producción y características fenotípicas en *coffea arabica* l.” *Cenicafé*, vol. 68, no. 1, pp. 62–72, 2017. [Online]. Available: <https://biblioteca.cenicafe.org/handle/10778/816>
- [22] E. Montoya, J. Hernández, C. Unigarro, and C. Florez, “Estimación del área foliar en café variedad castillo a libre exposición y su relación con la producción,” 2017. [Online]. Available: <https://www.cenicafe.org/es/publications/5.Estimacion.pdf>
- [23] S. Taugourdeau, G. Le Maire, J. Avelino, J. R. Jones, L. G. Ramirez, M. J. Quesada, F. Charbonnier, F. Gómez-Delgado, J.-M. Harmand, and B. Rapidel, “Leaf area index as an indicator of ecosystem services and management practices: An application for coffee agroforestry,” *Agriculture, Ecosystems & Environment*, vol. 192, pp. 19–37, 2014.
- [24] M. C. Espindula, A. M. A. dos Passos, L. F. B. Araújo, A. L. Marcolan, F. L. Partelli, and A. R. Ramalho, “Indirect estimation of leaf area in genotypes of ‘conilon’ coffee (*coffea canephora* pierre ex a. froehner),” *Australian Journal of Crop Science*, vol. 12, no. 6, pp. 990–994, 2018.
- [25] L. M. dos Santos, B. D. de Souza Barbosa, A. V. Diotto, D. T. Maciel, and L. A. G. Xavier, “Biophysical parameters of coffee crop estimated by uav rgb images,” *Precision Agriculture*, vol. 21, no. 6, pp. 1227–1241, 2020.
- [26] J. P. R. A. D. Barbosa, G. A. Martins, R. T. Ferreira, J. P. Pennacchi, V. F. de Souza, and A. M. Soares, “Estimating the iaf of coffee from the volume occupied by leaves and plant architecture,” *Coffee Science*, vol. 7, no. 3, pp. 267–274, 2012.
- [27] J. L. Favarin, D. Dourado Neto, A. García y García, N. A. Villa Nova, and M. d. G. G. V. Favarin, “Equações para a estimativa do índice de área foliar do cafeeiro,” *Pesquisa agropecuária brasileira*, vol. 37, no. 6, pp. 769–773, 2002.
- [28] P. Ramos, F. A. Prieto, E. Montoya, and C. E. Oliveros, “Automatic fruit count on coffee branches using computer vision,” *Computers and Electronics in Agriculture*, vol. 137, pp. 9–22, 2017.
- [29] J. d. S. Lima, S. d. A. Silva, R. de Oliveira, and A. da Fonseca, “Use of kriging techniques to estimate conilon coffee productivity.” *Revista Ceres*, vol. 63, no. 1, pp. 54–61, 2016.
- [30] S. Nogueira, M. A. Moreira, and M. M. Volpato, “Relationship between coffee crop productivity and vegetation indexes derived from oli/landsat-8 sensor data with and without topographic correction,” *Engenharia Agrícola*, vol. 38, no. 3, pp. 387–394, 2018.
- [31] A. a. G. Picini, M. B. P. D. Camargo, A. A. Ortolani, L. C. Fazuoli, and P. B. Gallo, “Desenvolvimento e teste de modelos agrometeorol para a estimativa de produtividade do cafeeiro,” *Bragantia*, vol. 58, pp. 157 – 170, 00 1999. doi: 10.1590/S0006-87051999000100015

- [32] F. A. B. Arruda and M. A. Grande, “Fator de resposta da produção cafeeiro ao déficit hídrico em Campinas,” *Bragantia*, vol. 62, pp. 139 – 145, 00 2003. doi: 10.1590/S0006-87052003000100017. [Online]. Available: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0006-87052003000100017&nrm=iso
- [33] L. Eduardo, G. de Souza, C. Lamparelli, R. Augusto, P. Sergio, and E. Ribeiro, “Agrometeorological models for forecasting coffee yield,” *Agronomy Journal*, 2017.
- [34] B. T. W. Putra, P. Soni, E. Morimoto, and P. Pujiyanto, “Estimating biophysical properties of coffee (*coffea canephora*) plants with above-canopy field measurements, using cropspec®,” *International Agrophysics*, vol. 32, no. 2, 2018.
- [35] L. Kouadio, R. C. Deo, V. Byrareddy, J. F. Adamowski, and S. Mushtaq, “Artificial intelligence approach for the prediction of robusta coffee yield using soil fertility properties,” *Computers and electronics in agriculture*, vol. 155, pp. 324–338, 2018.
- [36] J. S. Sousa, J. C. L. Neves, H. E. P. Martinez, and V. H. V. Alvarez, “Relationship between coffee leaf analysis and soil chemical analysis,” *Revista Brasileira de Ciência do Solo*, vol. 42, 2018.
- [37] B. D. S. Barbosa, G. A. e Silva Ferraz, L. Costa, Y. Ampatzidis, V. Vijayakumar, and L. M. dos Santos, “Uav-based coffee yield prediction utilizing feature selection and deep learning,” *Smart Agricultural Technology*, vol. 1, p. 100010, 2021. doi: <https://doi.org/10.1016/j.atech.2021.100010>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2772375521000101>
- [38] N. T. T. Thao, D. N. Khoi, A. Denis, L. V. Viet, J. Wellens, and B. Tychon, “Early prediction of coffee yield in the central highlands of vietnam using a statistical approach and satellite remote sensing vegetation biophysical variables,” *Remote Sensing*, vol. 14, no. 13, 2022. doi: 10.3390/rs14132975. [Online]. Available: <https://www.mdpi.com/2072-4292/14/13/2975>
- [39] A. Ashapure, J. Jung, A. Chang, S. Oh, J. Yeom, M. Maeda, A. Maeda, N. Dube, J. Landivar, and S. Hague, “Developing a machine learning based cotton yield estimation framework using multi-temporal uas data,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 180–194, 2020.
- [40] A. Moghimi, C. Yang, and J. A. Anderson, “Aerial hyperspectral imagery and deep neural networks for high-throughput yield phenotyping in wheat,” *Computers and Electronics in Agriculture*, vol. 172, p. 105299, 2020.
- [41] M. Herrero-Huerta, P. Rodriguez-Gonzalvez, and K. M. Rainey, “Yield prediction by machine learning from uas-based multi-sensor data fusion in soybean,” *Plant Methods*, vol. 16, pp. 1–16, 2020.

- [42] S. S. Kunapuli, V. Rueda-Ayala, G. Benavídez-Gutiérrez, A. Córdova-Cruzatty, A. Cabrera, C. Fernández, and J. Maiguashca, “Yield prediction for precision territorial management in maize using spectral data,” in *Precision Agriculture’15*. Wageningen Academic Publishers, 2015, pp. 344–358.
- [43] E. Khosla, R. Dharavath, and R. Priya, “Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression,” *Environment, Development and Sustainability*, pp. 1–22, 2019.
- [44] F. F. Haque, A. Abdelgawad, V. P. Yanambaka, and K. Yelamarthi, “Crop yield prediction using deep neural network,” pp. 1–4, 2020. doi: 10.1109/WF-IoT48130.2020.9221298
- [45] T. van Klompenburg, A. Kassahun, and C. Catal, “Crop yield prediction using machine learning: A systematic literature review,” *Computers and Electronics in Agriculture*, vol. 177, p. 105709, 2020. doi: 10.1016/j.compag.2020.105709
- [46] L. S. Santana, G. A. e. S. Ferraz, D. B. Marin, R. d. O. Faria, M. S. Santana, G. Rossi, and E. Palchetti, “Digital terrain modelling by remotely piloted aircraft: Optimization and geometric uncertainties in precision coffee growing projects,” *Remote Sensing*, vol. 14, no. 4, 2022. doi: 10.3390/rs14040911. [Online]. Available: <https://www.mdpi.com/2072-4292/14/4/911>
- [47] B. D. S. Barbosa, G. Araújo e Silva Ferraz, L. Mendes dos Santos, L. S. Santana, D. Bedin Marin, G. Rossi, and L. Conti, “Application of rgb images obtained by uav in coffee farming,” *Remote Sensing*, vol. 13, no. 12, 2021. doi: 10.3390/rs13122397. [Online]. Available: <https://www.mdpi.com/2072-4292/13/12/2397>
- [48] R. Nogueira Martins, F. d. A. de Carvalho Pinto, D. Marçal de Queiroz, D. S. Magalhães Valente, and J. T. Fim Rosas, “A novel vegetation index for coffee ripeness monitoring using aerial imagery,” *Remote Sensing*, vol. 13, no. 2, 2021. doi: 10.3390/rs13020263. [Online]. Available: <https://www.mdpi.com/2072-4292/13/2/263>
- [49] A. Soetedjo and E. Hendriarianti, “Plant leaf detection and counting in a greenhouse during day and nighttime using a raspberry pi noir camera,” *Sensors*, vol. 21, no. 19, 2021. doi: 10.3390/s21196659. [Online]. Available: <https://www.mdpi.com/1424-8220/21/19/6659>
- [50] A. Thomas, “Aerial photography,” *University of the Western Cape*. [Online]. Available: http://www.cdioinstitute.org/papers/Day1/AERIAL%20PHOTOGRAPHY_Abraham_Thomas.pdf, 2013.
- [51] G. Canada, “Concepts of aerial photography,” [Online]. Available: <https://www.nrcan.gc.ca/maps-tools-publications/satellite-imagery-air-photos/air-photos/national-air-photos/2013>.

- [52] Ardupilot. Mission planner home¶. [Online]. Available: <https://ardupilot.org/planner/>
- [53] DJI. Phantom 3 standard. [Online]. Available: <https://www.dji.com/phantom-3-standard>
- [54] V. Technology. Litchi. [Online]. Available: <https://flylitchi.com/>
- [55] Mapir. Survey3: Multi-spectral survey cameras. [Online]. Available: <https://www.mapir.camera/pages/survey3-cameras#specs>
- [56] J. T. F. Rosas, F. d. A. de Carvalho Pinto, D. M. de Queiroz, F. M. de Melo Villar, D. S. Magalhaes Valente, and R. Nogueira Martins, “Coffee ripeness monitoring using a uav-mounted low-cost multispectral camera,” *Precision Agriculture*, pp. 1–19, 2021.
- [57] Mapir. Calibrating images in mapir camera control application. [Online]. Available: <https://www.mapir.camera/pages/calibrating-images-in-mapir-camera-control-application>
- [58] ——. Ground target package v2. [Online]. Available: <https://www.mapir.camera/en-gb/products/mapir-camera-reflectance-calibration-ground-target-package-v2>
- [59] OpenDroneMap. Webodm. [Online]. Available: <https://opendronemap.org/webodm/>
- [60] I. Dabbura. K-means clustering: Algorithm, applications, evaluation methods, and drawbacks. [Online]. Available: <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a>
- [61] MathWorks. Matlab. [Online]. Available: <https://www.mathworks.com/help/matlab/>
- [62] Oracle. Mysql. [Online]. Available: <https://www.mysql.com/>
- [63] T. C. Parreiras, G. H. E. Lense, R. S. Moreira, D. B. Santana, and R. L. Mincato, “Using unmanned aerial vehicle and machine learning algorithm to monitor leaf nitrogen in coffee,” 2020.
- [64] ayuware. Wgs84. [Online]. Available: <https://www.ayuware.es/blog/coordenadas-wgs84/>
- [65] Alireza. Area. [Online]. Available: <https://pypi.org/project/area/>
- [66] opendronemap. Tiles service. [Online]. Available: <https://community.opendronemap.org/t/using-webodm-interface-to-show-map-tiles-on-own-server/2276>
- [67] V. Agafonkin. Leafletjs. [Online]. Available: <https://leafletjs.com/>
- [68] QuestionPro. Tamaño de muestra. [Online]. Available: <https://www.questionpro.com/es/tama%C3%B1o-de-la-muestra.html>

-
- [69] R. Taylor, “Interpretation of the correlation coefficient: A basic review,” *Journal of Diagnostic Medical Sonography*, vol. 6, no. 1, pp. 35–39, 1990. doi: 10.1177/875647939000600106. [Online]. Available: <https://doi.org/10.1177/875647939000600106>
- [70] Scikit-learn. Scikit-learn. [Online]. Available: <https://scikit-learn.org/>
- [71] NumFOCUS. Pandas. [Online]. Available: <https://pandas.pydata.org/>
- [72] Matplotlib. Matplotlib. [Online]. Available: <https://matplotlib.org/>
- [73] J. Rousel, R. Haas, J. Schell, and D. Deering, “Monitoring vegetation systems in the great plains with erts,” in *Proceedings of the Third Earth Resources Technology Satellite—1 Symposium; NASA SP-351*, 1973, pp. 309–317.
- [74] G. Molina and M. Rodrigo. El modelo de regresión lineal. [Online]. Available: http://ocw.uv.es/ciencias-de-la-salud/pruebas-1/1-3/t_09nuevo.pdf
- [75] P. University. A matrix formulation of the multiple regression model. [Online]. Available: <https://online.stat.psu.edu/stat462/node/132/>
- [76] Scikit-learn. Cross validation. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html
- [77] scikit learn. Rbf svm parameters. [Online]. Available: https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html
- [78] Innovatione. El cultivo del café en colombia. [Online]. Available: <https://innovatione.eu/2022/04/30/el-cultivo-del-cafe-en-colombia/>