

**UNIVERSIDAD NACIONAL DE CAJAMARCA**

**ESCUELA DE POSGRADO**



**UNIDAD DE POSGRADO DE LA FACULTAD DE CIENCIAS  
PECUARIAS**

**PROGRAMA DE DOCTORADO EN CIENCIAS**

**TESIS:**

**MODELAMIENTO DEL COMPORTAMIENTO PRODUCTIVO DE  
PASTURAS RYE GRASS – TRÉBOL UTILIZANDO IMÁGENES  
MULTIESPECTRALES, CAJAMARCA**

Para optar el Grado Académico de

**DOCTOR EN CIENCIAS**

**MENCIÓN: PRODUCCIÓN ANIMAL**

Presentada por:

**M.Cs. FELIPE BALTAZAR GUTIÉRREZ ARCE**

Asesor:

**Dr. LUIS ASUNCIÓN VALLEJOS FERNÁNDEZ**


Cajamarca, Perú

2025

### CONSTANCIA DE INFORME DE ORIGINALIDAD

1. Investigador:  
Felipe Baltazar Gutiérrez Arce  
DNI: 42235034  
Escuela Profesional/Unidad de Posgrado de la Facultad de Ingeniería en Ciencias Pecuarias. Programa de Doctorado en Ciencias. Mención: Producción Animal
2. Asesor: Ph.D. Luis Asunción Vallejos Fernández
3. Grado académico o título profesional  
 Bachiller       Título profesional       Segunda especialidad  
 Maestro       Doctor
4. Tipo de Investigación:  
 Tesis       Trabajo de investigación       Trabajo de suficiencia profesional  
 Trabajo académico
5. Título de Trabajo de Investigación:  
  
Modelamiento del comportamiento productivo de pasturas rye grass - trébol utilizando imágenes multiespectrales, Cajamarca
6. Fecha de evaluación: 28 / 03 /2025
7. Software antiplagio:       TURNITIN       URKUND (OURIGINAL) (\*)
8. Porcentaje de Informe de Similitud: 22 %
9. Código Documento: 3117:443576735
10. Resultado de la Evaluación de Similitud:  
 APROBADO       PARA LEVANTAMIENTO DE OBSERVACIONES O DESAPROBADO

Fecha Emisión: 28 / 03 /2025

<i>Firma y/o Sello Emisor Constancia</i>
 <hr/> <i>Ph.D. Luis Asunción Vallejos Fernández</i> DNI: 26673237

\* En caso se realizó la evaluación hasta setiembre de 2023

COPYRIGHT © 2025 by  
**FELIPE BALTAZAR GUTIÉRREZ ARCE**  
Todos los derechos reservados



**Universidad Nacional de Cajamarca**  
LICENCIADA CON RESOLUCIÓN DE CONSEJO DIRECTIVO N° 080-2018-SUNEDU/CD  
**Escuela de Posgrado**  
CAJAMARCA - PERU



**PROGRAMA DE DOCTORADO EN CIENCIAS**

**ACTA DE SUSTENTACIÓN DE TESIS**


**MENCIÓN: PRODUCCIÓN ANIMAL**

Siendo las 4:10pm horas, del día 27 de marzo del año dos mil veinticinco, reunidos en el Auditorio de la Escuela de Posgrado de la Universidad Nacional de Cajamarca, el Jurado Evaluador presidido por el Dr. JORGE PIEDRA FLORES, Dr. JOSÉ FERNANDO CORONADO LEÓN, Dr. GILBERTO FERNÁNDEZ IDROGO y en calidad de Asesor, el Ph.D. LUIS ASUNCIÓN VALLEJOS FERNÁNDEZ. Actuando de conformidad con el Reglamento Interno de la Escuela de Posgrado y el Reglamento del Programa de Doctorado de la Escuela de Posgrado de la Universidad Nacional de Cajamarca, se inició la SUSTENTACIÓN de la tesis titulada: **MODELAMIENTO DEL COMPORTAMIENTO PRODUCTIVO DE PASTURAS RYE GRASS -TRÉBOL UTILIZANDO IMÁGENES MULTIESPECTRALES, CAJAMARCA**; presentado por el Magister en Ciencias Agrarias – Opción Ciencias Animales FELIPE BALTAZAR GUTIÉRREZ ARCE.


Realizada la exposición de la Tesis y absueltas las preguntas formuladas por el Jurado Evaluador, y luego de la deliberación, se acordó aprobar con la calificación de Distinta (18) la mencionada Tesis; en tal virtud, el Magister en Ciencias Agrarias – Opción Ciencias Animales FELIPE BALTAZAR GUTIÉRREZ ARCE, está apto para recibir en ceremonia especial el Diploma que lo acredita como **DOCTOR EN CIENCIAS**, de la Unidad de Posgrado de la Facultad de Ingeniería en Ciencias Pecuarias, Mención **PRODUCCIÓN ANIMAL**.

Siendo las 5:17pm horas del mismo día, se dio por concluido el acto.

  
.....  
Ph.D. Luis Asunción Vallejos Fernández  
Asesor

  
.....  
Dr. Jorge Piedra Flores  
Presidente-Jurado Evaluador

  
.....  
Dr. José Fernando Coronado León  
Jurado Evaluador

  
.....  
Dr. Gilberto Fernández Idrogo  
Jurado Evaluador

## **DEDICATORIA**

A mi esposa, Zulema, por enseñarme a ser fuerte y por ser siempre esa fuente de fortaleza para cumplir mis objetivos, entre ellos, este doctorado.

A mis pequeños hijos, Pedrito y Francisquito, que siempre me miran como un modelo a seguir y con su manera de amarme, me enseñan el amor perfecto.

A mis padres, Walter y Gloria, fundamentales en mi vida. Porque con su compañía y consejos son la luz que guían mi existencia.

A mis hermanos, Walter, Ángel, Humberto, Marita y Carlitos, esos ángeles que Dios pone en tu camino, que con su sola presencia elevan tu ánimo y cambian positivamente la manera de ver el mundo.

## **AGRADECIMIENTOS**

En primer lugar, a Dios, porque cada día de mi vida, me muestra que aún en las adversidades, está conmigo, no me suelta. Y me da la fortaleza para conseguir los objetivos propuestos, siempre y cuando estén a la par de su plan para mí.

A mi asesor, el Dr. Luis A. Vallejos Fernández, al que considero mi AMIGO. Por haberme acompañado a lo largo de todo este camino con actitud paternal.

A mis jurados, Dr. Piedra, Dr. Fernández, Dr. Coronado, por mostrarse no sólo como personas sabientes y con mucha expertiz en el área, sino por enseñarme que el camino del saber también de debe llevar a la sencillez y la mejora de nuestra humanidad. Gracias por su amistad.

A mis docentes y a mi Facultad de Ing. en Ciencias Pecuarias. Por su dedicación y por haberme dado las primeras herramientas para dedicarme a esta profesión, a la que amo mucho.

A la Universidad Nacional de Cajamarca, por abrirme las puertas al mundo del saber y especialmente por haber financiado el proyecto de investigación macro, que dio lugar a mi tesis de investigación.

A los administrativos, tanto de mi Facultad como de la Escuela de Posgrado, por su gran capacidad de servicio y porque siempre te atienden con una sonrisa que influye positivamente en el día a día, especialmente, a la sra. María, secretaria de la Unidad.

## ÍNDICE

<b>DEDICATORIA</b> .....	v
<b>AGRADECIMIENTOS</b> .....	vi
<b>LISTA DE TABLAS</b> .....	ix
<b>LISTA DE FIGURAS</b> .....	x
<b>LISTA DE ANEXOS</b> .....	xi
<b>LISTA DE ABREVIATURAS</b> .....	xii
<b>RESUMEN</b> .....	xiii
<b>ABSTRACT</b> .....	xiv
<b>CAPÍTULO I</b> .....	1
<b>INTRODUCCIÓN</b> .....	1
<b>CAPÍTULO II</b> .....	4
<b>MARCO TEÓRICO</b> .....	4
2.1. Antecedentes de la investigación .....	4
2.2 Bases teóricas .....	12
2.2.1. La región Cajamarca como zona forrajera .....	12
2.2.2 Imágenes multiespectrales y biomasa forrajera .....	13
2.2.3 Imágenes Multiespectrales y sus fundamentos .....	14
2.2.4 Técnicas para la estadística espacial .....	16
<b>CAPÍTULO III</b> .....	18
<b>PLANTEAMIENTO DE LAS HIPÓTESIS Y VARIABLES</b> .....	18
<b>CAPÍTULO IV</b> .....	19
<b>MARCO METODOLÓGICO</b> .....	19
4.1. Ubicación del área de estudio .....	19
4.2. Universo, muestra y unidad de análisis .....	20
4.3. Técnicas de investigación .....	20
4.4. Análisis de datos y desarrollo del modelo .....	22
<b>CAPÍTULO V</b> .....	27
<b>RESULTADOS Y DISCUSIÓN</b> .....	27
5.1. Resultados .....	27
5.2. Discusión .....	33
<b>CAPÍTULO VI</b> .....	35
<b>CONCLUSIONES Y SUGERENCIAS</b> .....	35

6.1. Conclusiones .....	35
6.2. Sugerencias .....	35
<b>CAPÍTULO VII .....</b>	<b>36</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>36</b>
<b>CAPÍTULO VIII .....</b>	<b>34</b>
<b>ANEXOS .....</b>	<b>43</b>



## LISTA DE TABLAS

<b>Tabla 1:</b> Técnicas e Instrumentos que se emplearan en el experimento. ....	21
<b>Tabla 2:</b> Índices espectrales extraídos del Micasense Rededge MX .....	23
<b>Tabla 3:</b> Estadísticas descriptivas de biomasa fresca y seca .....	27
<b>Tabla 4:</b> Test de normalidad para las variables estudiadas .....	28
<b>Tabla 5:</b> Indicadores estadísticos del análisis de correlación realizado entre la biomasa fresca y seca medida y las bandas espectrales ( $p \leq 0.05$ ) .....	29
<b>Tabla 6:</b> Evaluación de los efectos de predicción de los diferentes modelos en la predicción de la biomasa .....	32

## LISTA DE FIGURAS

<b>Figura 1:</b> Mapa de ubicación del proyecto (Fundo Huayrapongo) .....	20
<b>Figura 2:</b> Coeficientes de correlación entre la biomasa fresca y seca medida y los predictores. r: coeficiente de correlación de Pearson; significativo al 5 % de probabilidad .....	29
<b>Figura 3:</b> Proceso de construcción del mapa a través de las imágenes multiespectrales, en el Fundo Huayrapongo .....	31

## LISTA DE ANEXOS

<b>Anexo 1:</b> Características detalladas del UAV, la cámara y el plan de vuelo utilizados ...	43
<b>Anexo 2:</b> Análisis de correlación entre las variables de estudio .....	44
<b>Anexo 3:</b> Puntos de calibración (80) para la determinación de constantes del valor de las bandas .....	45
<b>Anexo 4:</b> Imágenes tomadas con la cámara multiespectral, desde el dron .....	47

## LISTA DE ABREVIATURAS

- CART:** Classification and Regression Trees.
- DEM:** Modelo digital de elevación.
- ExG:** Índice de exceso de verde.
- ExR:** Índice de exceso de rojo.
- GEE:** Google Earth Engine.
- GEE:** Imágenes ópticas de teledetección.
- GNDVI:** Índice de vegetación de diferencia normalizada verde.
- GPS:** Sistema de Posicionamiento Global.
- GSD:** distancia de superficie del terreno.
- IME:** Imágenes multiespectrales.
- IVC:** Índice de vegetación de clorofila.
- IVD:** Índice de vegetación diferencial.
- MAE:** Error absolute medio.
- ML:** Machine Learning.
- NDI:** Índice de Diferencia Normalizada.
- NDRE:** Índice de vegetación de diferencia normalizada de borde rojo.
- NDVI:** Índice de vegetación de diferencia normalizada.
- NDWI:** Índice de agua de diferencia normalizada.
- NT:** Número de árboles.
- OSAVI:** Índice de vegetación ajustado al suelo optimizado.
- RF:** Random Forest.
- RMSE:** Error cuadrático medio.
- RPAS:** Sistema de Aeronaves Pilotadas Remotamente.
- SAVI:** Índice de vegetación ajustado al suelo.
- SLR:** Regresión Lineal Simple.
- SRI:** Óxido de hierro rojo/azul de proporción simple.
- UAV:** Vehículos aéreos no tripulados.
- XGBoost:** Gradient Boosting.

## RESUMEN

El desconocimiento de tecnologías como las IME y de modelos Machine Learning, aplicables a la actividad pecuaria, ha permitido el deterioro paulatino de los recursos involucrados en esta actividad, especialmente del recurso forrajero y del recurso suelo. Por ello, esta investigación titulada Modelamiento del comportamiento productivo de pasturas rye grass – trébol utilizando imágenes multiespectrales, en Cajamarca, tuvo como objetivo modelar el comportamiento productivo del forraje, probando los modelos matemáticos CART, XGBoost y RF a través de la aplicación de la tecnología de imágenes multiespectrales (IME) utilizando drones. Se realizó en el fundo Huayrapongo, de la UNC, en potreros divididos de una hectárea, para tomar registro de la biomasa de la pastura, tanto en suelo como por aire, a través del uso del dron y de una cámara multiespectral. Se utilizaron 13 bandas espectrales y las precisiones se calcularon a partir de 100 modelos. Antes de cada repetición, el conjunto de datos se barajó aleatoriamente y se generaron nuevos pliegues para aumentar la robustez de los modelos. Para evaluar la precisión del ajuste de los modelos, se utilizaron el coeficiente de determinación ( $R^2$ ), el error cuadrático medio (RMSE) y el error absoluto medio (MAE). Se encontraron correlaciones positivas entre la Biomasa seca con la banda espectral roja y el índice de vegetación ajustado al suelo (SRI) ( $r = 0.16 - 0.39$ ); y de la Biomasa fresca con la banda espectral del infrarrojo cercano (NIR), el NDVI y los índices espectrales ExG ( $r = 0.10 - 0.33$ ). El modelo Random Forest (RF) superó a los demás (CART y XGBoost) y brindan predicciones espacialmente más consistentes ( $R^2 = 0.95$ , para Base Fresca;  $R^2 = 0.98$ , para Base Seca). Estos resultados sugieren que el procesamiento de datos de UAV con algoritmos de aprendizaje automático en la nube, ofrece una alternativa accesible para predecir la biomasa forrajera a un nivel detallado con resultados satisfactorios, bajo las condiciones de nuestra región (Cajamarca).

**Palabras clave:** Imagen multiespectral, RPAS, forraje, CART, XGBoost, Random Forest.

## ABSTRACT

The lack of knowledge of technologies such as IME and Machine Learning models applicable to livestock farming has led to the gradual deterioration of the resources involved in this activity, especially forage and soil resources. Therefore, this research, entitled "Modeling the Productive Behavior of Ryegrass-Clover Pastures Using Multispectral Imagery in Cajamarca," aimed to model forage production behavior by testing the mathematical models CART, XGBoost, and RF through the application of multispectral imaging (IME) using drones. The study was carried out at the Huayrapongo farm of the UNC (National University of Castilla-La Mancha), in divided one-hectare pastures, to record pasture biomass, both on the ground and in the air, using a drone and a multispectral camera. Thirteen spectral bands were used, and the accuracies were calculated from 100 models. Before each replicate, the dataset was randomly shuffled and new folds were generated to increase the robustness of the models. To assess the accuracy of model fit, the coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE) were used. Positive correlations were found between dry biomass with the red spectral band and the soil-adjusted vegetation index (SRI) ( $r = 0.16-0.39$ ); and between fresh biomass with the near-infrared (NIR) spectral band, NDVI, and the ExG spectral indices ( $r = 0.10-0.33$ ). The Random Forest (RF) model outperformed the others (CART and XGBoost) and provided more consistent spatial predictions ( $R^2 = 0.95$ , for Fresh Base;  $R^2 = 0.98$ , for Dry Base). These results suggest that processing UAV data with cloud-based machine learning algorithms offers an affordable alternative for predicting forage biomass at a detailed level, with satisfactory results under the conditions of our region (Cajamarca).

**Keywords:** Multispectral imaging, RPAS, forage, CART, XGBoost, Random Forest.

# CAPÍTULO I

## INTRODUCCIÓN

### 1.1. PLANTEAMIENTO DEL PROBLEMA

#### 1.1.1. Contextualización

Los pastizales se encuentran entre los ecosistemas más grandes del mundo y representan el 30 % de la biomasa terrestre del planeta. Desempeñan un papel vital en el apoyo a los sistemas alimentarios humanos al proporcionar alimento para los animales de pastoreo, especialmente en áreas no aptas para cultivos (Milazzo et al., 2023). Obtener conocimientos sobre las características de los pastizales y los pastos es esencial para mejorar nuestra comprensión de este sistema de producción y para gestionar estos ecosistemas para proporcionar servicios ecosistémicos.

La sierra norte del Perú cuenta con grandes extensiones de pastos naturales y cultivados, que constituyen la base de la alimentación del ganado lechero. El limitado uso de tecnología de teledetección ha limitado el manejo sostenible de los recursos forrajeros, restringiendo el monitoreo y avance de la actividad ganadera a procedimientos tradicionales. El desconocimiento en el uso de herramientas tecnológicas, como la visión artificial mediante imágenes multiespectrales, entre otras metodologías, limita el alcance que podríamos tener para el control de producción y calidad del recurso forrajero en Cajamarca, sobre todo considerando que limita conocer muchos indicadores productivos y nutricionales, por lo que el uso de esta tecnología puede influir en el ahorro de tiempo, además de supervisar mayor área de pasturas e incrementar la precisión de control.

Complementariamente, estas tecnologías suelen trabajar con modelos matemáticos de Inteligencia Artificial, basados en Machine Learning, que permitan la traducción de las imágenes y/o valores registrados. Existen muchos de estos algoritmos de aprendizaje automático, entre ellos, CART (Classification and Regresión Trees), XGBoost (Gradient Boosting) y RF (Random Forest), que definitivamente, son el complemento ideal para las tecnologías IME y que, en la

actual coyuntura, también son desconocidas por los principales interesados, que son los productores y la academia.

Consecuentemente, este es un problema importante si consideramos que en muchas locaciones de nuestro valle, es notoria la degradación de nuestros suelos, y si a ello sumamos la estacionalidad de la producción forrajera, el mal manejo del recurso pastoril (principalmente debido al uso inadecuado en consideración a su estado fenológico), el inadecuado uso del recurso animal y bajos niveles nutricionales de las pasturas, da como resultado una baja disponibilidad en cantidad y calidad de las pasturas, y como consecuencia, se aprecian bajos rendimientos del sector ganadero de pequeños y medianos productores lecheros del valle interandino de Cajamarca.

### **1.1.2. Formulación**

¿Bajo la utilización de los modelos matemáticos *CART* - Classification and Regression Trees, *XGBoost* – Gradient Boosting y *RF* – Random Forest, cuál será más eficiente en el modelamiento productivo de las pasturas rye grass – trébol, a través de la aplicación de la tecnología de imágenes multiespectrales (IME) utilizando drones, en Cajamarca?

### **1.2. JUSTIFICACIÓN E IMPORTANCIA**

La evaluación de pasturas, generalmente, se sustenta en la observación, medición e interpretación de distintas características como el estado fenológico, composición florística, valor nutritivo, biomasa, cobertura vegetal, estrés hídrico, entre otros, que permiten estimar la disponibilidad de forraje, calidad nutritiva y recuperación de las mismas después de haber sido pastoreadas. Sin embargo, estas actividades en la producción forrajera no se realizan de manera continua por el alto costo de su implementación y complejidad de los métodos; por ello, usar la tecnología de imágenes multiespectrales utilizando drones, permitiría superar estos problemas, dado que ayudaría al estudio de dichos parámetros, en mayores áreas, a menores costos y menos complejidad.



La implementación de sensores en vehículos aéreos no tripulados y el uso de machine learning (algoritmos de aprendizaje automático como el CART, XG Boost, RF), con soporte en nube, se maneja cada vez más como una alternativa eficiente y sostenible para la ganadería de precisión, permitiendo un procesamiento eficiente de imágenes sin servidores físicos, lo que minimiza los costos de inversión y puede utilizarse para un monitoreo más extenso en condiciones de gran altitud y permitiría un uso más eficiente de recursos, como tiempo, mano de obra y dinero, para los principales interesados, que esencialmente serían la academia y los productores.

El presente trabajo se justifica porque se empleó y aplicó las imágenes multiespectrales para el análisis de los parámetros productivos de nuestras pasturas raigrás “ecotipo cajamarquino” – trébol blanco, propias de la región, que sustentan la economía de los productores.

### 1.3. OBJETIVOS DE LA INVESTIGACIÓN

#### **General:**

- Modelar el comportamiento productivo de las pasturas rye grass – trébol, probando tres modelos matemáticos (*CART* - Classification and Regression Trees, *XGBoost* – Gradient Boosting y *RF* – Random Forest), a través de la aplicación de la tecnología de imágenes multiespectrales (IME) utilizando drones, en Cajamarca.

#### **Específicos:**

- Hallar posibles correlaciones entre la base fresca y seca de la biomasa (*in situ*), con las distintas bandas espectrales utilizados en el estudio con imágenes multiespectrales.
- Determinar el mejor modelo de predicción a utilizar (*CART* - Classification and Regression Trees, *XGBoost* – Gradient Boosting y *RF* – Random Forest) para el estudio de biomasa en sistemas pastoriles.

## CAPÍTULO II

### MARCO TEÓRICO

#### 2.1. Antecedentes de la investigación

Ganerød et al. (2023) utilizó imágenes ópticas de teledetección [GEE] de alta resolución para determinar algunas características de la superficie del suelo. Esta caracterización, bajo este método, se realiza, generalmente, de forma manual. Esta tarea, aunque sencilla, requiere mucho tiempo y una mano de obra extensa y especializada. Contribuimos a esta línea de investigación proponiendo un enfoque automatizado que utiliza computación en la nube, aprendizaje profundo, redes neuronales y un modelo U-Net aplicado en Google Collaboratory (Colab). Específicamente, probamos este método en un sitio en el suroeste de Noruega utilizando un conjunto de variables explicativas generadas a partir de un modelo digital de elevación (DEM) de 10 m de resolución y, para comparar, datos de Landsat 8 basados en la nube. En general, nuestro modelo automatizado de identificación, utilizando imágenes ópticas de teledetección [GEE], revela resultados muy prometedores, dentro de sus limitaciones.

Zhou et al. (2023), generó información para ofrecer apoyo a la toma de decisiones de los departamentos de gestión agrícola y de varios productores agrícolas. Trabajaron bajo la premisa que bajo la afirmación de que la materia orgánica del suelo (MOS) es un indicador crítico de los niveles de nutrientes del suelo, y el mapeo preciso de su distribución espacial a través de la teledetección es esencial para la regulación del suelo, la fertilización precisa y la gestión y protección científicas. Se determinaron nueve parámetros adecuados para el mapeo y modelado de la MOS, con la utilización de la reflectancia espectral, transformaciones matemáticas de reflectancia e índices de vegetación. Luego, utilizando imágenes multiespectrales basadas en vehículos aéreos no tripulados (UAV) con resolución a nivel de centímetros, se utilizó un algoritmo de aprendizaje automático de bosque aleatorio para construir el modelo de inversión de la MOS y el mapeo de la MOS en el área de estudio. Los resultados mostraron que el algoritmo de bosque aleatorio tuvo el

mejor desempeño para estimar el contenido de materia orgánica del suelo ( $R^2 = 0.91$ , RMSE = 0.95, MBE = 0.49 y RPIQ = 3.25) en comparación con otros algoritmos de aprendizaje automático. Los hallazgos indicaron una correlación negativa entre el contenido de materia orgánica del suelo y la altitud. Se concluyó que el modelado y el mapeo de SOM de alta precisión a escala de campo en el noreste de China eran factibles mediante el uso de imágenes de teledetección de UAV (satelitales) de alta resolución a nivel centimétrico.

Morais et al. (2021), sobre la base de 26 artículos recientes, realizaron una revisión de la literatura sobre el tema para identificar prácticas comunes, a saber, la relación entre el rendimiento de la estimación y el método de ML utilizado, las fuentes de datos y la escala (local/regional). Para identificar la relación entre las características de los estudios y la precisión de la estimación, se utiliza el análisis descriptivo y correlacional. A pesar de un aumento en el número de artículos y ejemplos de aplicación, no hay evidencia de que el rendimiento de estimación de los algoritmos haya mejorado con el tiempo. En todos los enfoques utilizados por los autores de los trabajos aquí considerados, el número de muestras de campo, la fuente de datos de RS y la composición de especies de los sistemas de pastizales son las variables más relevantes para explicar la precisión de la estimación. Esta precisión aumenta con el número de muestras de campo hasta que se estabiliza, lo que sugiere la existencia de un nivel óptimo para los esfuerzos de monitoreo. La precisión también aumenta con la proximidad del sensor al campo, es decir, en promedio, la precisión es mayor usando espectroscopia de campo que usando datos satelitales. No hay evidencia de que algún método de ML en particular sea más adecuado para este problema. La literatura también muestra limitaciones significativas en cuanto a las aplicaciones de los algoritmos de ML. Por ejemplo, un número limitado de artículos validaron los modelos, lo que pone en duda el potencial de los modelos para la aplicación generalizada. A pesar de estas limitaciones, y teniendo en cuenta los avances verificados, esperamos que, en un futuro próximo, los métodos de ML combinados con datos RS/proximales continúen mejorando y sean útiles para la gestión de las explotaciones.

Théau et al. (2021) observaron que los pastizales se encuentran entre los ecosistemas más extendidos de la Tierra y entre los más degradados. Su caracterización y seguimiento se basan generalmente en mediciones de campo, que son incompletas espacial y temporalmente. La reciente llegada de los vehículos aéreos no tripulados (UAV) proporciona datos con resoluciones espaciales y temporales sin precedentes. Este estudio tiene como objetivo probar y comparar tres enfoques basados en imágenes multispectrales adquiridas por UAV para estimar la biomasa de forraje o la cobertura vegetal en pastizales. El sitio de estudio está compuesto por 30 parcelas de pastoreo (25 × 50 m), 5 parcelas de suelo desnudo (25 × 50) y 6 parcelas de control (5 × 5 m) en un campo de 14 ha mantenido a varios niveles de biomasa mediante rotaciones de pastoreo y recorte durante una temporada de crecimiento completa. Se realizaron un total de 14 vuelos. Se utilizó un primer enfoque basado en la estructura a partir del movimiento para generar un modelo de estimación de biomasa basado en volumetría ( $R^2$  de 0.93 y 0.94 para la biomasa fresca [FM] y la biomasa seca [MS], respectivamente). Este enfoque no es muy sensible a niveles bajos de vegetación, pero es preciso para una estimación de FM superior a 0.5 kg/m<sup>2</sup> (0.1 kg MS/m<sup>2</sup>). Se seleccionó el Índice de Vegetación de Diferencia Normalizada Verde (GNDVI) para desarrollar dos enfoques adicionales. Uno se basa en un modelo de predicción de biomasa de regresión ( $R^2$  de 0.80 y 0.66 para FM y MS, respectivamente) y conduce a una estimación precisa a niveles de FM inferiores a 3 kg/m<sup>2</sup> (0.6 kg MS/m<sup>2</sup>). El otro enfoque se basa en una clasificación de la cubierta vegetal a partir de la agrupación de los valores del GNDVI en cuatro clases. Este enfoque es más cualitativo que los otros, pero más robusto y generalizable. Estos tres enfoques son relativamente sencillos de utilizar y aplicables en un contexto operativo. También son complementarios y se pueden adaptar a aplicaciones específicas en la caracterización de pastizales.

Bennett et al. (2020) proponen la utilización de imágenes satelitales de alta resolución para el estudio de ecosistemas tanto terrestres como acuáticos. Esto, a razón de que los métodos *in situ* suelen ser caros e inconsistentes en términos de tiempo y espacio. Ahora, si bien las imágenes satelitales de alta resolución también pueden ser costosas de adquirir y estar sujetas a condiciones ambientales que

ocultan las características del objetivo, y las imágenes de alta resolución recopiladas de sistemas de aeronaves pilotadas a distancia (RPAS o drones) podrían parecer una alternativa económica; el procesamiento de imágenes de drones para su análisis requiere mucho tiempo y es complejo. Este estudio presenta el primer flujo de trabajo semiautomático para el procesamiento de imágenes de drones con Google Earth Engine (GEE) y software libre y de código abierto. Con este flujo de trabajo, procesó 230 imágenes de drones de Heron Reef, Australia, y clasificamos ecosistemas terrestres con el clasificador Random Forest. Nuestra clasificación logró una precisión general del 86 % y mapeó la cobertura con una precisión del 92 %. Los métodos presentados permiten el procesamiento eficiente de imágenes de drones de cualquier entorno y pueden ser útiles cuando se procesan imágenes de drones para calibrar y validar imágenes satelitales.

Batistoti et al. (2019), bajo la premisa que el territorio brasileño contiene aproximadamente 160 millones de hectáreas de pastos, vieron necesario desarrollar técnicas para automatizar su gestión y aumentar su producción. Esta investigación tiene dos objetivos: En primer lugar, estimar la altura de la cubierta mediante fotogrametría de vehículos aéreos no tripulados (UAV); en segundo lugar, proponer una ecuación para la estimación de la biomasa de los pastos de sabana brasileña (Cerrado) a partir de la altura del dosel del UAV. Se evaluaron cuatro unidades experimentales de *Panicum maximum* cv. BRS Tamani. Se realizó simultáneamente muestreo de masa de forraje, mediciones de altura y recolección de imágenes de UAV. Los vehículos aéreos no tripulados volaron a una altura de 50 m y las imágenes se generaron con una distancia media de muestra del suelo (GSD) de aproximadamente 1.55 cm. La altura del dosel forrajero estimada por los UAV se calculó como la diferencia entre el modelo digital de superficie (DSM) y el modelo digital del terreno (DTM). La  $R^2$  entre la altura de la regla y la altura del UAV fue de 0.80; entre la biomasa ( $\text{kg ha}^{-1}$  GB: biomasa verde) y altura de la regla, 0.81; y entre la biomasa ( $\text{kg ha}^{-1}$  GB) y la altura del UAV, 0.74. La fotogrametría UAV demostró ser una técnica potencial para estimar la altura y la biomasa en los pastizales de *Panicum maximum* cv. BRS Tamani ubicados en el bioma de sabana brasileña (Cerrado) en peligro de extinción.

Näsi et al. (2018), investigaron bajo la premisa que la estimación oportuna de la biomasa y el contenido de nitrógeno de los cultivos es un paso crucial en diversas tareas de la agricultura de precisión, por ejemplo, en la optimización de la fertilización. La teledetección mediante drones y aviones ofrece una herramienta factible para llevar a cabo esta tarea. El objetivo fue desarrollar y evaluar una metodología para la estimación de biomasa y nitrógeno de cultivos, integrando características espectrales y 3D que puedan ser extraídas utilizando cámaras aerotransportadas miniaturizadas multispectrales, hiperspectrales y en color (RGB). Se utilizó el Random Forest (RF) como estimador, y además se utilizó la Regresión Lineal Simple (SLR) para validar la consistencia de los resultados del RF. El método se evaluó con conjuntos de datos empíricos capturados de un campo de cebada y un sitio de ensayo de ensilado de pasto utilizando una cámara hiperspectral basada en el interferómetro de Fabry-Pérot (FPI) y una cámara RGB regular a bordo de un dron y una aeronave. Las mediciones agrícolas de referencia incluyeron el rendimiento fresco (AF), el rendimiento de materia seca (DMY) y la cantidad de nitrógeno. En la estimación DMY de la cebada, el coeficiente de correlación de Pearson (PCC) y el error cuadrático medio normalizado (RMSE %) fueron, en el mejor de los casos, del 0.95 % y del 33.2 %, respectivamente; y en la estimación del DMY de gramíneas, los mejores resultados fueron 0.79 % y 1.9 %, respectivamente. En las estimaciones de la cantidad de nitrógeno de la cebada, el PCC y el RMSE % fueron, en el mejor de los casos, de 0.97 % y 21.6 %, respectivamente. En la estimación de la biomasa, los mejores resultados se obtuvieron al integrar las características hiperspectrales y 3D, pero la integración de imágenes RGB y características 3D también proporcionó resultados casi igual de buenos. En la estimación del contenido de nitrógeno, la cámara hiperspectral dio los mejores resultados. Concluimos que la integración de características espectrales y 3D de alta resolución espacial y la calibración radiométrica eran necesarias para optimizar la precisión.

Sendin et al. (2018), evaluaron las imágenes multispectrales para determinar su potencial para clasificar granos enteros de maíz blanco mediante el análisis multivariante (análisis discriminante de PLS), se analizaron las regiones UV, visible y

NIR (375 – 970 nm), y obtuvieron un buen rendimiento al distinguir entre maíz sano y materiales indeseables, con  $R^2$  varían entre 0.35 y 0.99 y las longitudes de onda que indicaron la presencia de licopeno y antocianina fueron 505, 525, 570 y 590 nm, y la presencia de grasa es predicha por 890, 940nm y 970nm y está asociadas con el agua.

Tang et al. (2018), realizaron un estudio de la aplicación de las imágenes multiespectrales para predecir el contenido de azúcares en manzanas, en las cuales utilizaron cuatro longitudes de onda optimas (461 nm, 469 nm, 947 nm y 1049 nm) y emplearon como modelo matemático la regresión lineal múltiple paso a paso. El modelo de regresión lineal múltiple predijo el contenido de azúcar con un coeficiente de determinación ( $R^2$ ) de 0.8861 y un error cuadrático medio (RMSE) de 0.8738° Brix.

Por otro lado, el uso en el control de pastizales también ha sido estudiado por estas tecnologías como lo muestran Magiera et al. (2017), quienes modelaron un mapa de rendimiento de pastizales y evaluaron si los problemas de saturación pueden corregirse por la composición de especies de plantas mediante la generación de los modelos, debido a que los índices de vegetación tienden a saturarse para una biomasa alta, pero al incorporar un mapa de la composición de especies como variable predictiva para el rendimiento de pastizales permite la conservación de características ecológicas por lo que lo convierte en una herramienta valiosa para determinar el uso adecuado de la tierra, sobre todo de los pastizales.

También se ha empleado la espectroscopia de imágenes para evaluar la fracción de vegetación muerta en pastizales mixtos en Nueva Zelanda, para ello, Pullanagari et al. (2017) usaron una alta resolución de pixel de 1 m y resolución espectral de 3.5 a 5.6 nm en el rango de 380 – 2500 nm. Para recuperar la fracción de vegetación muerta de los espectros; se emplearon los índices de vegetación de banda estrecha, regresión de mínimos cuadrados parciales basados en espectro completo (PLS) y selección de características basada en regresión PLS; con el modelo de PLS basado en la selección de características exhibió un mejor rendimiento en términos de precisión de predicción ( $R^2CV = 0.73$ ,  $RMSECV = 6.05$ ,  $RPDCV = 2.25$ ). Los resultados fueron consistentes con los datos de validación y también tuvieron un buen

rendimiento en los datos de prueba externa ( $R^2 = 0.62$ , RMSE = 8.06, RPD = 2.06), indicando que los datos de espectroscopia de imágenes podrían ser una herramienta útil para mapear la fracción de vegetación muerta con precisión.

Ramoelo et al. (2015) vieron que el nitrógeno (N) de las hojas y la biomasa aérea son algunos de los factores clave que limitan la producción agrícola y el funcionamiento de los ecosistemas. El N foliar y la biomasa pueden utilizarse como indicadores de la calidad y cantidad de los pastizales. Los métodos convencionales para evaluar estos parámetros de vegetación a nivel de paisaje son lentos y tediosos. La teledetección proporciona una vista aérea del paisaje, lo que brinda la oportunidad de evaluar estos parámetros de vegetación en áreas de pastoreo más amplias. El objetivo de este estudio es monitorear el N foliar y la biomasa aérea como indicador de la calidad y cantidad de los pastizales utilizando imágenes satelitales WorldView-2 y la técnica de bosques aleatorios en la parte noreste de Sudáfrica. Se realizaron series de trabajo de campo para recolectar muestras de N foliar y biomasa en marzo de 2013, abril o mayo de 2012 (final de la estación húmeda) y julio de 2012 (estación seca). Se calcularon varios índices de vegetación convencionales y basados en el borde rojo. Los resultados generales indican que los índices aleatorios de bosque y vegetación explicaron más del 89 % de las concentraciones de N en las hojas para el pasto y los árboles, y menos del 89 % para todos los años de evaluación. Los índices de vegetación basados en el borde rojo estuvieron entre las variables importantes para predecir el N de las hojas. Para la biomasa, el modelo de bosque aleatorio explicó más del 84 % de la variación de la biomasa en todos los años, y se encontró que las bandas visibles, incluidos los índices de vegetación basados en el borde rojo, eran importantes. El estudio demostró que la hoja N puede ser monitoreada usando alta resolución espacial con la capacidad de banda de borde roja, y es importante para la evaluación y monitoreo de los pastizales.

Barrachina et al. (2015), aplicaron un análisis de regresión múltiple para comparar los datos de Landsat-5TM con el cálculo de biomasa y los índices de vegetación y humedad en pastizales y generar un modelo predictivo, los modelos para biomasa vegetal aérea arrojan mejores resultados a mediados y finales del verano antes de



las cosechas, con valores de  $R^2$  de 0.76 – 0.74, RMSE de 95g – 128g.m<sup>-2</sup> y un porcentaje de error (PE) de 27 a 36 %; consideran que las bandas del infrarrojo medio son mejores descriptores de la biomasa. Así mismo, evaluaron la capacidad de los datos obtenidos por teledetección con alta resolución espacial (Imágenes SPOT) para estimar la biomasa de los pastizales en crecimiento, donde calcularon el Índice de vegetación de diferencia normalizada (NDVI), el Índice de área foliar (LAI) y la fracción de Cubierta vegetal (fCOVER); el análisis de la relación entre las variables derivadas de los datos obtenidos por teledetección y la biomasa observada en el campo muestra que el LAI supera el NDVI y fCOVER para estimar la biomasa ( $R^2$  de 0.68 contra 0.30 y 0.50, respectivamente).

En Colombia, Meneses et al. (2015), en su investigación “Uso de drones para el análisis de imágenes multiespectrales en agricultura de precisión”, tuvo como objetivo la implementación de herramientas tecnológicas que permitan mejorar la planificación de las actividades agrícolas, predecir daños y tomar decisiones adecuadas ante situaciones que afectan el desarrollo de los cultivos de papa (*Solanum tuberosum* variedad Diacol Capiro) en Cundinamarca. Para la selección del cultivo y el área objeto de estudio se realizó un análisis de los sistemas productivos de mayor importancia en la región (Cundinamarca), así como de los impactos en los rendimientos y productividad de tales cultivos en los últimos años, dados por la presencia de enfermedades y plagas, deficiencias nutricionales en los suelos o deficiencias y excesos hídricos, en un contexto de cambio y variabilidad climática. Para la obtención de mejores resultados, los vuelos se realizaron de las 11:30 hasta las 13:00 horas, ya que es el momento del día con mayor actividad fotosintética debido a la mayor radiación solar incidente. El uso de las imágenes multiespectrales sobre estos cultivos permitió identificar que, en algunos puntos, el suelo se encontraba afectado por la aplicación de agroquímicos, lo que permitió considerar una posible sobrecarga química del producto final (tubérculo). También se encontró exceso de agua en cada uno de los surcos de esta área del cultivo, suelo con evidente saturación hídrica y plantas que a pesar de haber emergido al mismo tiempo que las demás, presentaban un crecimiento tardío de dos a tres entrenudos, así como fragmentos de suelo desnudo y mínima cobertura de follaje, en

comparación a otras áreas del cultivo. Finalmente se observó que la zona cercana a un canal de desagüe, generó un área de exceso hídrico por encharcamiento, lo que afectó negativamente el desarrollo y calidad del cultivo de papa.

## **2.2. Bases teóricas**

### **2.2.1. La región Cajamarca como zona forrajera**

Zarria (2015) define a las pasturas como los recursos forrajeros que crece naturalmente sin la intervención del hombre, más abundante y también el más económico para la producción de carne y leche. Además, la importancia de estudiar a los pastos nativos o naturalizados es debido a la invasión de especies de pastos exóticos, lo cual no se está teniendo en consideración las especies nativas a pesar de que muchos agricultores manifiestan que los pastos nativos son más importantes para mejorar la producción de leche y carne (Vásquez, 2012), considerando que se debe de considerar estas variedades de especies naturales de mayor importancia para el ganadero debido a que sirven para la producción ganadera (Oliva et al., 2015). Por otro lado, las pasturas nativas y naturalizadas son más resistentes al frío, sequías y no son muy exigentes en la fertilidad pues son plantas perennes (Ramirez, 2009), en las cuales se encuentran ubicadas las pasturas raigrás trébol.

La región de Cajamarca, ubicada en la cuenca norte del Perú, es una de las principales cuencas lecheras del país, representando el 17.5 % de la producción nacional (INEI, 2012). Esta producción depende principalmente, de una combinación de pasturas cultivadas y naturales (Carrasco, 2019). Típicamente, estos pastizales cultivados incluyen 17.3 mil hectáreas de raigrás y trébol, junto con 6.1 mil hectáreas de alfalfa en la Región Cajamarca (INEI, 2012).

El reygrass que encontramos en Cajamarca en su mayoría es el “ecotipo cajamarquino”. Una de sus características es la tolerancia a la sequía, las heladas, plagas, enfermedades y al pisoteo del ganado; bondad que ha adquirido en un periodo de 70 años desde su introducción, al igual que el Rye grass Ingles (*Lolium perenne*), Rye grass italiano (*Lolium multiflorum*) y Trébol Blanco (*Trifolium repens*). Esta especie es una gramínea bianual, pero pasa a comportarse como perenne,

debido al periodo de descanso que supera los 50 días en promedio, lo cual posibilita la maduración de las semillas, y estas al caer al suelo permitiendo la “Resiembra Natural”. El Ecotipo cajamarquino logra establecerse rápidamente y se asocia muy bien con el *Trifolium repens*, el cual está naturalizado en toda la cuenca ganadera de Cajamarca. Además, las pasturas de raigrás “ecotipo cajamarquino” - trébol blanco, representa un alto potencial alimenticio para los vacunos, siendo de alta palatabilidad y ofrece una alta proporción de nutrientes.

### **2.2.2. Imágenes multiespectrales y biomasa forrajera**

La biomasa es la cantidad de materia vegetal que se encuentra en un determinado espacio y tiempo, se expresa en unidades de peso de materia seca por superficie ( $\text{g}/\text{m}^2$ ,  $\text{kg}/\text{ha}$ , etc.), normalmente se mide la parte aérea por la dificultad de acceder a la materia subterránea (rizomas, bulbos, raíces, etc.) y porque es la sección de la planta que tiene interés de uso pastoral. Se puede distinguir la biomasa verde de la seca (necromasa), se considera un parámetro importante para evaluar el interés pastoral de las distintas comunidades vegetales y la potencialidad de su aprovechamiento por el ganado (Gómez, 2008).

Ahora, el monitoreo de la distribución de la biomasa de cultivos forrajeros a lo largo de su ciclo de crecimiento ayuda en el manejo sustentable de los pastizales, alineándose con la demanda de alimentos y la disponibilidad de forrajes, y apoya prácticas de pastoreo efectivas (Théau et al. 2021). La biomasa, definida como la producción de energía primaria por las plantas dentro de un ecosistema, se puede evaluar a través de varias técnicas directas o indirectas (Villalobos-Villalobos et al., 2023). Para su estudio, los enfoques directos, como el método de cosecha, tienden a requerir más tiempo que los métodos indirectos como el uso de microsensores multiespectrales a bordo de UAV, que se emplean ampliamente como herramientas flexibles para el monitoreo de la biomasa (De Rosa et al., 2021).

### **2.2.3. Imágenes Multiespectrales y sus fundamentos**

El análisis de imágenes hiperespectrales (HSI) se basa en la interacción entre la materia y la energía electromagnética, es decir, los átomos y moléculas pueden pasar a un estado activado durante un corto periodo de tiempo – relajación, regresando luego a su estado fundamental. Este análisis integra espectroscopia e imagen suministrando información espectral y física de una muestra, estas imágenes se obtienen de manera remota o local a través de sensores que hacen uso de cientos de bandas espectrales contiguas y discretas (Mundaca, 2017). En cambio, las imágenes multiespectrales están formadas por relativamente pocas bandas (normalmente entre 3 y 20) y son bandas no necesariamente contiguas unas a otras, y podemos obtener los valores de intensidad en las longitudes de onda discretas en las que el sistema capte radiación. Con la creciente disponibilidad de conjuntos de datos de alta resolución temporal y densos espacialmente, tanto de plataformas satelitales como de vehículos aéreos no tripulados, los datos resultantes pueden volverse demasiado densos y engorrosos para los sistemas informáticos tradicionales y los métodos de análisis clásicos. Como resultado, el machine learning (ML) y la computación en la nube surgen como soluciones efectivas, que permiten la exploración de valores de parámetros en una amplia gama de posibilidades, guiada por datos de entrenamiento, para identificar soluciones que maximicen métricas de rendimiento específicas (Morais et al., 2021).

Sin embargo, hay que conocer el concepto de “espectroscopía” para entender mejor estas definiciones. Esta palabra deriva de la raíz latina “*spectrum*” que significa apariencia, imagen y de la palabra griega “*skopia*” que significa ver (Reeves, 2000). La espectroscopia es un conjunto de técnicas de análisis que se fundamentan en la interacción de la radiación electromagnética con la muestra, consiste en la medición e interpretación de fenómenos de absorción de la luz que entra en contacto con la muestra, es decir, la espectroscopia involucra transferencia de energía entre la luz y la muestra. También Armenta et al. (2007), mencionan que el principio fundamental de la espectroscopia es la radiación, y al interactuar con la materia sufre una serie de modificaciones observables y medibles que dependen de sus características fisicoquímicas y estructurales.

Con respecto al espectro electromagnético, se ha afirmado que está formado por microondas, luz infrarroja, luz visible, luz ultravioleta, rayos X y rayos gamma (Oliveira y Santos, 2016), de los cuales el rango de la luz visible se encuentra entre 400 y 780nm y el rango del infrarrojo cercano esta entre 780 y 2500 nm (Bosoon y Renfu, 2015), estas son las regiones de mayor interés analítico (Feng et al., 2013). Para utilizar la espectroscopia y fotogrametría se han implementado tecnologías como las imágenes multiespectrales, por las cuales es posible realizar evaluaciones de componentes químicos atributos físicos de múltiples productos que se pueden encontrar en el tiempo y espacio, siendo una metodología no destructiva y práctica. Finalmente, también debemos hablar del “Índice de vegetación”. Al respecto, Verdín et al. (2003), mencionan que el índice de vegetación de diferencia normalizada (NDVI), es usado para estimar la cantidad, calidad y desarrollo de la vegetación con base en la medición, por medio de sensores remotos instalados comúnmente desde una plataforma espacial, de la intensidad de la radiación de ciertas bandas del espectro electromagnético que la vegetación emite o refleja. Así, el índice de NDVI cuantifica, a través de la reflectancia, áreas de alto vigor (densidad vegetal) que poseen una mayor reflectancia en el infrarrojo cercano y una menor reflectancia en el rojo, debido a que poseen un mayor índice NDVI. Esser et al. (2002) explica que el principio básico del NDVI se basa en el hecho de que las capas esponjosas de las hojas reflejan mucha luz en el infrarrojo cercano, en marcado contraste con la mayoría de los objetos no vegetales. Cuando la planta se deshidrata o estresa, la capa esponjosa colapsa y las hojas reflejan menos luz infrarroja, pero la misma cantidad en el rango visible, esta combinación matemática de las dos señales puede ayudar a diferenciar las plantas de lo no vegetal o plantas saludables de plantas enfermas. Es decir, los sensores multiespectrales capturan la reflectancia de la vegetación como imágenes. Los diferentes tipos de vegetación muestran comportamientos espectrales únicos; por ejemplo, la radiación infrarroja cercana (760-900 nm) es fuertemente reflejada por las estructuras celulares de las hojas, mientras que la radiación visible roja (630-690 nm) es absorbida por la clorofila (Goebel e Iwaszczuk., 2023).

#### ***2.2.4. Técnicas para la estadística espacial***

##### *Classification and Regression Trees – Árboles de clasificación y regresión (CART)*

CART, es un algoritmo no paramétrico utilizado para la clasificación o análisis de regresión, con la capacidad de suprimir el ruido de los datos, utilizando un método de regresión no paramétrica que agrega un conjunto de árboles de decisión en una partición binomial (Jin et al., 2016). Además, los árboles de regresión reemplazan los datos faltantes y gestionan los datos anormales, la estructura jerárquica de la clasificación también permite interacciones del modelo entre las variables predictoras (Schuler et al., 2010). Sin embargo, no es un modelo estable en el sentido de que pequeños cambios en el espacio de entrada pueden generar un árbol completamente diferente. Por esta razón, CART se utiliza como un aprendizaje base para construir modelos más complejos como RF y XGBoost (Jain et al., 2020).

##### *Random Forest (RF)*

Breiman (2001) afirma que el Random Forest construye múltiples árboles de decisión que se muestrean de forma independiente durante el entrenamiento, lo que mejora la clasificación mediante los resultados de las votaciones en comparación con un modelo de árbol de decisión único. El algoritmo no hace suposiciones sobre la distribución de los datos, puede manejar puntuaciones y variables continuas simultáneamente y es muy eficiente en la minería, para el análisis de datos no lineales y capacidades de generalización (Gholizadeh et al., 2018).

##### *Gradient Boosting (XGBoost)*

XGBoost es una mejora del algoritmo de potenciación de gradiente y se ha utilizado ampliamente en la clasificación y el análisis de regresión (Mayr et al., 2014.), con una precisión generalmente buena. Los árboles de decisión en XGBoost se entrenan secuencialmente con ajustes realizados a partir del error del árbol anterior, mientras que en RF se construyen en paralelo e independientemente (Jain et al., 2020). Además, selecciona subconjuntos aleatorios para ajustar predictores

individuales de forma iterativa, con el fin de obtener la función de pérdida minimizada e introduce el procedimiento de potenciación de gradiente estocástico, que puede reducir el riesgo de sobreajuste y mejorar la generalización de modelos con regularización.

## CAPÍTULO III

### PLANTEAMIENTO DE LAS HIPÓTESIS Y VARIABLES

#### 3.1. Hipótesis de la investigación

**Ho:** Ninguno de los tres modelos matemáticos probados (*CART* - Classification and Regression Trees, *XGBoost* – Gradient Boosting y *RF* – Random Forest), permitirán modelar el comportamiento productivo de las pasturas rye grass – trébol, registrados con la aplicación de tecnología de imágenes multiespectrales (IME) utilizando drones, en Cajamarca.

**Ha:** El modelo RF supera a los demás modelos (*CART* y *XGBoost*) para el modelamiento del comportamiento productivo de las pasturas rye grass – trébol, registrados con la aplicación de tecnología de imágenes multiespectrales (IME) utilizando drones, en Cajamarca, debido a que brinda predicciones espacialmente más consistentes.



## CAPÍTULO IV

### MARCO METODOLÓGICO

#### 4.1. Ubicación del área de estudio

El presente trabajo se llevó a cabo en el Centro de Investigación y Producción Pecuaria del C.I.P.P. Huayrapongo de la Facultad de Ingeniería en Ciencias Pecuarias (FICP) de la Universidad Nacional de Cajamarca, en el distrito de Baños del Inca, situado al Noroeste de Cajamarca a una altitud de 2525 msnm, con una temperatura promedio anual de 14.17 °C (figura 1).

Las características meteorológicas son las siguientes:

Altitud : 2643 m.s.n.m

Latitud sur : 7°10' 36"

Latitud oeste : 78°28' 07"

Temperaturas promedio / año : 13 a 20°C

Humedad Relativa : 68%

Clima frío y seco, la temporada de lluvias es de diciembre a marzo.

(Fuente: SENAMHI – Cajamarca, 2020).

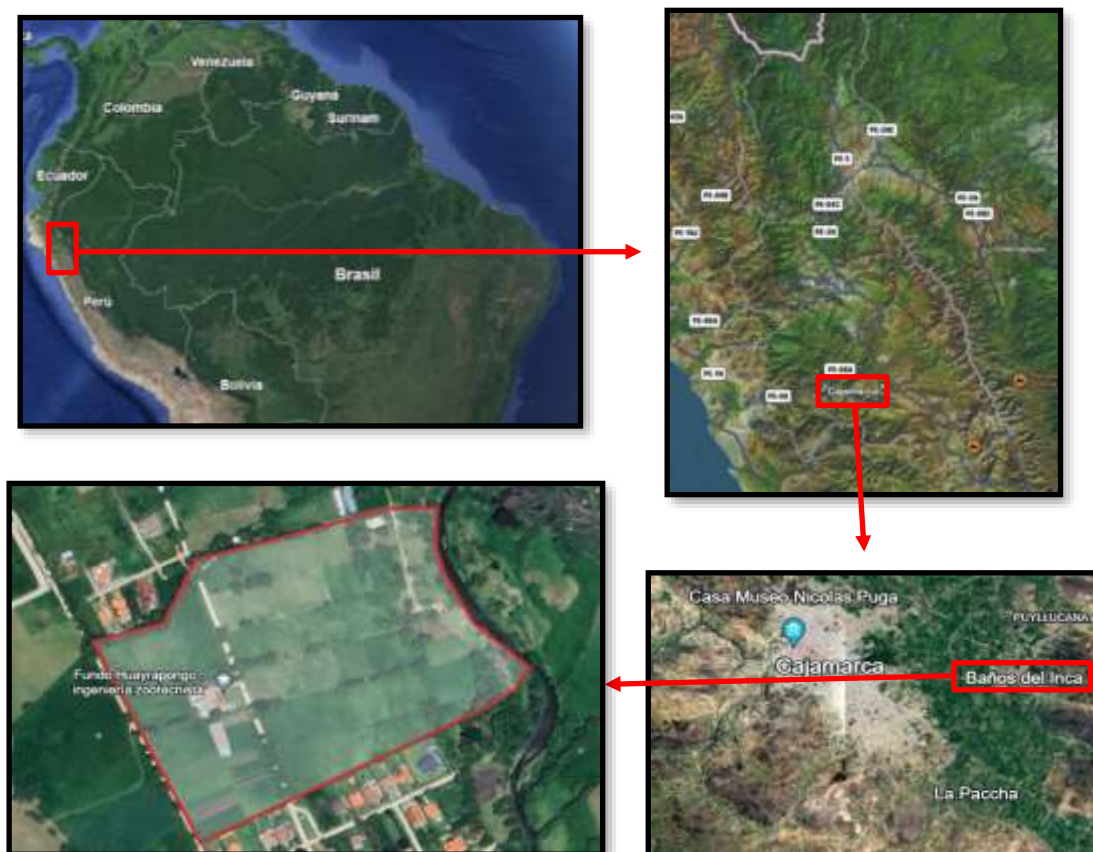


Figura 1: Mapa de ubicación del proyecto (Fundo Huayrapongo).

#### 4.2. Universo, muestra y unidad de análisis

**Universo:** Área total de pradera del fundo Huayrapongo.

**Muestra:** Área total de pradera del fundo Huayrapongo.

**Unidad de análisis:** Fracción de área, del predio.

#### 4.3. Técnicas de investigación

Las técnicas e instrumentos que se emplearán para obtener los datos y el procesamiento de estos se muestran en la tabla 1.

Tabla 1.

*Técnicas e Instrumentos que se emplearan en el experimento.*

<b>TÉCNICAS</b>	<b>INSTRUMENTOS</b>
Observación y medición en campo y en laboratorio.	* Fichas de recolección de datos. * Registros de rendimiento de biomasa. * Registros de control de materia seca.
Toma de imágenes multiespectrales.	* Dron. * Cámara de imágenes multiespectrales. * Software Spectronon 2.62 (RESONON Inc. – USA).
Análisis y modelación	* Software R-project

Para muestreo de las pasturas *in situ*: se trabajaron en potreros de una hectárea (metodología transectos y cuadrantes), debidamente georreferenciados. Se recolectaron un total de 12 muestras de biomasa, en la segunda semana de noviembre de 2021 con un cuadrante de 0,5 x 0,5, para fines de poder correlacionar con la unidad área de estudio de la cámara multiespectral (pixel = < 7 cm<sup>2</sup>).

Para el muestreo de las pasturas con drones se utilizó una cámara multiespectral MicaSense RedEdge MX (MicaSense, Inc., Seattle, WA, EE. UU.), a bordo de un DJI Matrice 600 (DJI, Shenzhen, China) para tomar fotos multiespectrales de 16 bits, con 5 bandas espectrales (azul (475 ± 20 nm), verde (560 ± 20 nm), rojo (668 ± 10 nm), NIR (840 ± 40 nm) y RE (717 ± 10 nm) a una resolución de 3,2 megapíxeles (2064 × 1544 píxeles). Las características detalladas del UAV, la cámara y el plan de vuelo utilizados se muestran en el Anexo 1. Además, el plan de vuelo se llevó a cabo alrededor del mediodía, hora local, a una altitud entre 100 y 150 metros sobre el suelo. Las fotografías se capturaron cada 2 segundos, asegurando un solapamiento del 75 % tanto de adelante hacia atrás como de lado a lado. Finalmente, estas imágenes se almacenaron en formato TIFF de 16 bits.

El procesamiento fotogramétrico se realizó con el software Pix4D Pro Mapper (Prilly, Suiza). Las diferencias relativas entre los parámetros internos iniciales y optimizados

fueron mínimas (0.48 %), lo que indica que los parámetros iniciales eran confiables para construir el ortomosaico.

Recolectamos cuatro puntos de control terrestre por paddock, utilizando un GPS GNSS D-RTK 2-DJI (Horizontal: 1 cm + 1 ppm (RMS); Vertical: 2 cm + 1 ppm (RMS)) y los incorporamos al flujo de trabajo de procesamiento para mejorar la precisión topográfica de la nube de puntos y las bandas de reflectancia del ortomosaico, logrando una distancia de superficie del terreno (GSD) final de 15.42 cm. Basado en la nube de puntos, se generó un modelo de superficie digital (DSM) a la misma resolución que el ortomosaico y se exportó en formato TIFF. Se vigiló, en suelo, el transecto de recorrido del dron, para controlar datos de interpretación dudosa (altura por encima del promedio que no respondan al patrón de conducta de la pastura, que puede ser un arbusto, una vaca, etc.).

***Medidas de control experimental:*** Se determinó y manejó condiciones ambientales y del manejo del dron. Por ejemplo:

- Se utilizó, para el análisis de información, los datos provenientes de días soleados, para no afectar las lecturas de la cámara multiespectral.
- Se buscó que los días de muestreo sean días con baja intensidad de corriente de viento.
- La altura del vuelo del dron, fue aproximadamente, de 100 m.
- Se realizó diez vuelos de dron, por parcela.

#### **4.4. Análisis de datos y desarrollo del modelo**

##### ***Extracción de variable***

Para desarrollar los modelos espaciales de biomasa, utilizamos 13 índices espectrales comúnmente empleados en el análisis de la vegetación, como se muestra en la Tabla 2. Estos índices abarcan índices de vegetación, suelo y agua. Se utilizó un cuadrado rectangular de 0,5 x 0,5 m para cada muestra, y los valores de reflectancia de cada píxel se convirtieron en réplicas de observación para contrastar con el peso de la biomasa. Los índices de vegetación se calcularon utilizando varias

combinaciones de valores de reflectancia y se compilaron como predictores junto con las bandas espectrales puras.

Tabla 2.

*Índices espectrales extraídos del Micasense Rededge MX.*

BANDAS	LONGITUD DE ONDA (NM)
Índice de vegetación de diferencia normalizada (NDVI) (Rouse et al., 1974).	$\frac{(NIR - RED)}{(NIR + RED)}$
Índice de agua de diferencia normalizada (NDWI) (McFeeters, 1996).	$NDWI = \frac{(GREEN - NIR)}{(GREEN + NIR)}$
Índice de vegetación ajustado al suelo (SAVI) (Qi et al., 1994).	$L = 0,6$ $\left( \frac{(NIR - RED)}{(NIR + RED + 1)} \right) (1 + L)$
Índice de vegetación de diferencia normalizada verde (GNDVI) (Gitelson et al., 1996).	$\frac{(NIR - GREEN)}{(NIR + GREEN)}$
Índice de vegetación diferencial (IVD) (Richardson y Everitt, 1992).	$(NIR - RED)$
Índice de vegetación ajustado al suelo optimizado (OSAVI) (Rondeaux et al., 1996).	$(1 + 0.16) \left( \frac{(NIR - RED)}{(NIR + RED + 0.16)} \right)$
Índice de exceso de verde (ExG) (Woebbecke et al., 1995).	$2 \times GREEN - RED - BLUE$
Índice de exceso de rojo (ExR) (Hindman y Meyer, 1998).	$2 \times RED - GREEN$
ExG-ExR (Meyer y Neto, 2008).	$ExG - ExR$
Índice de Diferencia Normalizada (NDI) (Bannari et al., 1995).	$\frac{(GREEN - RE)}{(GREEN + RE)}$
Índice de vegetación de diferencia normalizada de borde	$\frac{(NIR - RED)}{(NIR + RED)}$

rojo (NDRE) (Gitelson y Merzlyak, 1994).

Índice de vegetación de clorofila (IVC) (Vincini et al., 2008).

$$NIR \times \left( \frac{RED}{GREEN^2} \right)$$

Óxido de hierro rojo/azul de proporción simple (SRI) (Hewson et al., 2001).

$$RED/BLUE$$

---

Longitudes de onda centrales multiespectrales de Micasense Rededge MX: B, G, R, RE y NIR: 474, 560, 668, 717 y 840 nm.

### ***Análisis espacial***

Para modelar biomasa fresca y seca, utilizamos técnicas de regresión de aprendizaje automático basadas en lógica, incluidos métodos de árboles de decisión como CART, así como Gradient Boosting (XGBoost) y Random Forest (RF).

Las precisiones se calcularon a partir de 100 modelos. Antes de cada repetición, el conjunto de datos se barajó aleatoriamente y se generaron nuevos pliegues para aumentar la robustez de los modelos. Se deben elegir dos parámetros antes de ejecutar el bootstrap: el tamaño de la muestra y el número de repeticiones. En nuestro caso, el tamaño de la muestra fue el mismo que el del conjunto de datos original (Phinzi et al., 2021) y aplicamos 100 repeticiones.

La implementación de XGboost y GEE requiere seis parámetros. Para ello, se examinaron el número de árboles (NT) de 50 a 500 en intervalos de 50 árboles. Como resultado, se generaron y evaluaron un total de 50 combinaciones (NT = 10, LR = 5) para cada k-fold (k – vez). Para el ajuste del algoritmo RF, la plataforma GEE requiere seis parámetros. En este estudio, se examinaron el número de árboles (NT) de 50 a 500 en intervalos de 50 árboles, las variables por división (VPS) de 1 a 5 en intervalos de 1 variable y la fracción de bolsa (BF) de 0.1 a 1 en intervalos de fracción de 0.1 para identificar los parámetros óptimos de RF. Como resultado, se generó un total de 500 combinaciones (NT = 10, VPS = 5 y BF = 10) y se aplicaron iterativamente para cada k-fold. Luego, se seleccionaron los valores óptimos de NT y SR para

XGboost y NT, VPS y BF para RF en función de las métricas de regresión y se validaron en las muestras que no se incluyeron en la muestra de bootstrap.

Para evaluar el desempeño de los modelos desarrollados, realizamos evaluaciones de precisión. El coeficiente de determinación ( $R^2$ ), el error cuadrático medio (RMSE) y el error absoluto medio (MAE) se utilizaron para comparar la precisión de diferentes modelos (Chai y Draxler, 2014). Más específicamente, el  $R^2$  se utilizó para medir la variación entre los parámetros del suelo medidos y predichos; el RMSE se utilizó para evaluar la magnitud del error entre las mediciones y los parámetros del suelo predichos y el MAE se utilizó para medir la magnitud promedio de los errores en un conjunto de predicciones, sin considerar su dirección. MAE y RMSE expresan el error de predicción promedio en unidades de la variable de interés. Cuanto más cerca esté  $R^2$  de 1, y cuanto más cerca estén RMSE y MAE de 0, mejor será el ajuste del modelo.

En las siguientes ecuaciones, podemos ver las fórmulas utilizadas para determinar el coeficiente de determinación, el error absoluto medio y el error cuadrático medio (fórmula (1), (2) y (3), respectivamente).

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \quad (3)$$

donde  $n$  es el número de muestras (parcela individual) en el conjunto de datos,  $y_i$  es la propiedad del suelo medida;  $i$ ,  $y_i$  es la propiedad del suelo prevista en función de las imágenes del UAV y  $\hat{y}_i$  indica el promedio de la propiedad del suelo medida.

Evaluamos la importancia relativa de los predictores en los modelos de mejor desempeño. Las variables más influyentes se utilizaron para dividir los nodos de decisión en la raíz, mientras que las menos importantes se utilizaron en niveles más

profundos (Ishwaran y Kogalur, 2010). Este análisis proporcionó una contribución porcentual normalizada para cada predictor, cuantificando su impacto en la precisión de nuestro modelo optimizado de predicción de parámetros del suelo.

Finalmente, para cuantificar la incertidumbre en los mapas predichos, calculamos el coeficiente de variación entre los mapas k-fold generados para el mejor modelo seleccionado para cada elemento. Este análisis resalta las áreas con variaciones significativas en los valores.



## CAPÍTULO V

### RESULTADOS Y DISCUSIÓN

#### 5.1. RESULTADOS

Las estadísticas descriptivas de la biomasa fresca y seca se muestran en la Tabla 3. La biomasa fresca varió de 52.64 a 3985.40 g/m<sup>2</sup>. Los valores de biomasa seca variaron de 34.74 a 392.58 g/m<sup>2</sup>. La desviación estándar y el coeficiente de variación para la biomasa fresca y seca indican una alta dispersión.

Tabla 3.

*Estadísticas descriptivas de biomasa fresca y seca.*

Variable	Mínimo (g/m <sup>2</sup> )	Máximo (g/m <sup>2</sup> )	Media	Mediana	Desvío estándar	Coficiente Variabilidad (%)
Biomasa fresca	52.64	3.985.40	966.5	840.32	816.11	84.44
Biomasa seca	34.74	392.58	99.2	89.34	53.08	53.51

Se destaca la alta variabilidad de los datos, que muestra la tabla 3. Sin embargo, al realizar el test de normalidad (Kolmogorov – Smirnov y Shapiro-Wilk), los resultados mostraron que los datos cumplen la normalidad. Esto es debido a que los datos para este análisis fueron posteriores a la corrida de los algoritmos, que automáticamente, corrigieron esta variabilidad (tabla 4).

Tabla 4.

*Test de normalidad para las variables estudiadas.*

Prueba de normalidad					
Variable	n	Kolmogorov-Smirnov		Shapiro-Wilk	
		Estadístico	Sig.	Estadístico	Sig.
Dry_Matter	12	0.037	0.56	0.027	0.055
red	12	0.047	0.68	0.035	0.058
SRI	12	0.014	0.57	0.014	0.058
Fresh_Matter	12	0.048	0.55	0.011	0.059
NIR	12	0.025	0.57	0.039	0.062
NDVI	12	0.016	0.67	0.044	0.061
ExG	12	0.028	0.54	0.024	0.064

La Figura 2 y la tabla 5 ilustran los coeficientes de correlación de Pearson ( $r$ ) entre la biomasa fresca y seca y los predictores derivados de las bandas espectrales y los índices espectrales obtenidos de las imágenes de Micasense, calculados utilizando la biblioteca *corrplot* (Wei y Simko, 2017), en el entorno *R* (R Core Team, 2021).

En general, se identificaron varios grupos de correlación, que muestran que la biomasa seca está correlacionada positivamente con la banda espectral roja y el índice de vegetación ajustado al suelo (SRI) ( $r=0.16-0.39$ ). Por el contrario, la biomasa fresca exhibió correlaciones positivas con la banda espectral del infrarrojo cercano (NIR), el NDVI y los índices espectrales ExG ( $r=0.10-0.33$ ).

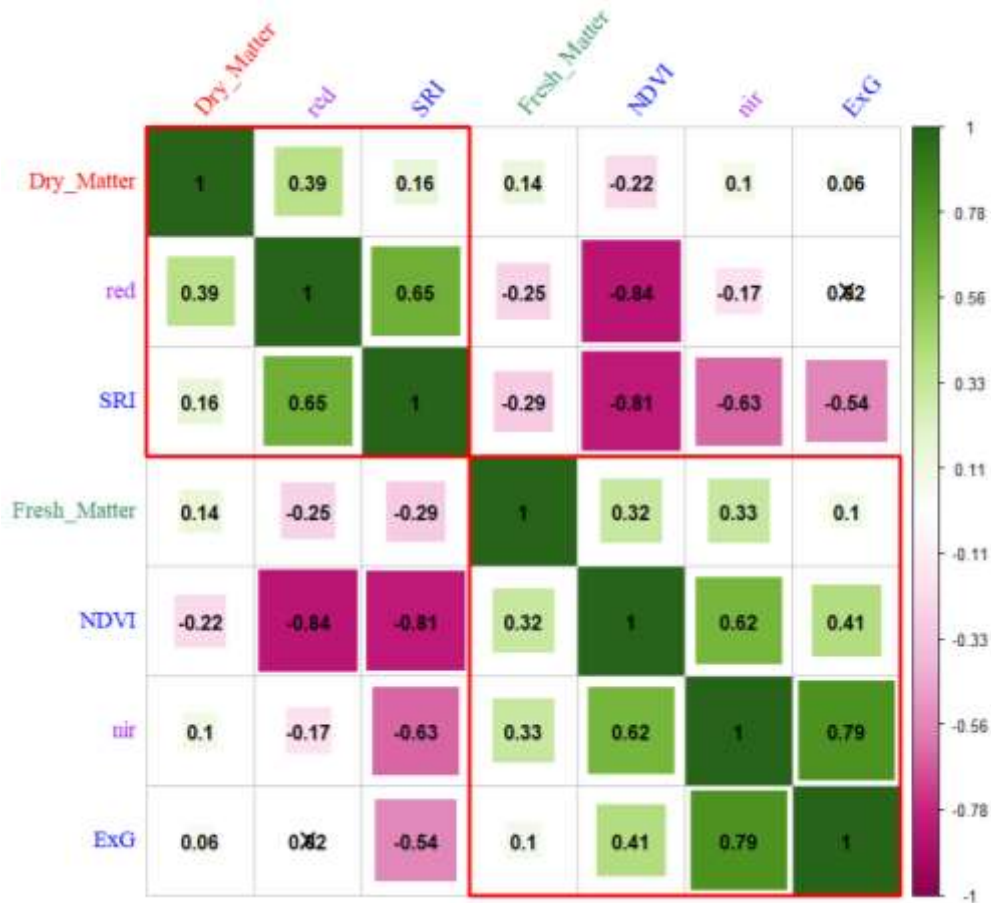


Figura 2. Coeficientes de correlación entre la biomasa fresca y seca medida y los predictores. r: coeficiente de correlación de Pearson; significativo al 5 % de probabilidad.

Tabla 5.

*Indicadores estadísticos del análisis de correlación realizado entre la biomasa fresca y seca medida y las bandas espectrales ( $p \leq 0.05$ ).*

Variable (1)	Variable (2)	Coef. Correlación(r)	Probabilidad (p-valor)
Dry_Matter	Dry_Matter	1.00	$\leq 0.0001$
Dry_Matter	red	0.39	0.02487
Dry_Matter	SRI	0.16	0.03654
Fresh_Matter	Fresh_Matter	1.00	$\leq 0.0001$
Fresh_Matter	NIR	0.33	0.00455
Fresh_Matter	NDVI	0.32	0.00127
Fresh_Matter	ExG	0.10	0.04568

Con respecto a las imágenes obtenidas a través de las imágenes multiespectrales (figura 3), estas pudieron ser traducidas por las constantes calculadas en base a la calibración realizada, previo al inicio del experimento. La calibración requirió la utilización de 80 puntos de muestreo (Anexo 2).

Las imágenes pudieron ayudar a construir y delimitar el mapa del área de estudio (Fundo Huayrapongo). Como se puede observar en la figura 3, primero se obtiene el perímetro de ubicación (1), para, posteriormente, pasar a la detección de bandas oscuras y claras (2) (3) (4), luego se detectan infraestructura específica (5) y finalmente, se obtiene las bandas de colores (NDVI, verdes, etc.) (6).

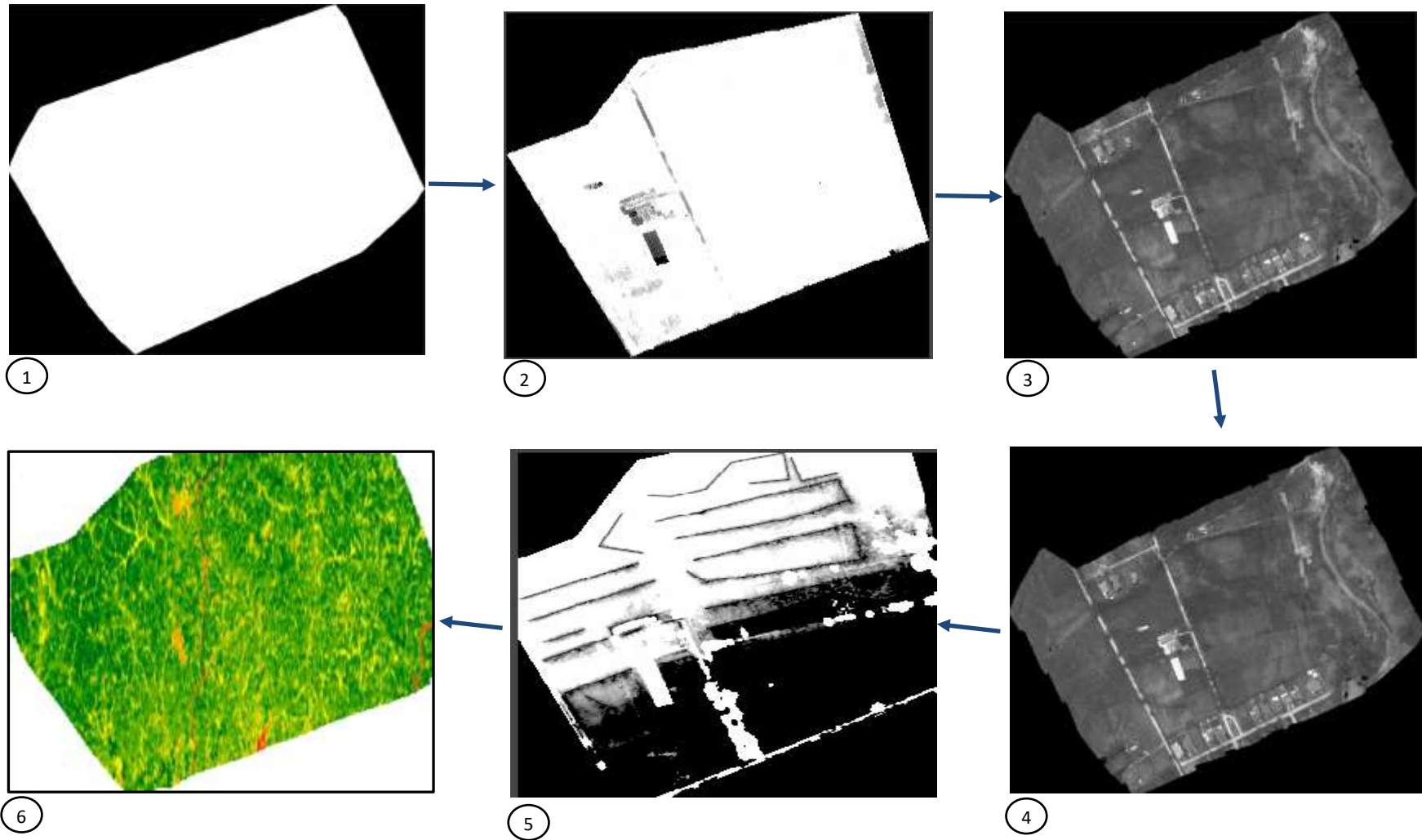


Figura 3. Proceso de construcción del mapa a través de las imágenes multiespectrales, en el Fundo Huayrapongo.

Con respecto a la evaluación del modelado, la tabla 6 muestra la evaluación de los modelos de regresión con las métricas de precisión más altas. Para biomasa fresca, el modelo más eficaz es Random Forest. Para biomasa seca, el algoritmo con mejor rendimiento es Random Forest. XGBoost muestra una precisión moderada, mientras que CART demuestra el rendimiento menos eficaz para las variables evaluadas.

Tabla 6.

*Evaluación de los efectos de predicción de los diferentes modelos en la predicción de la biomasa.*

Variable	Algoritmo	Hiperparámetro	Valores óptimos	R cuadrado	RMSE	MAE
<b>Biomasa fresca</b>	CART	-		-1.156	11.894	8.729
	RF	Número árboles	350	0,9556	1.6531	0,5054
		Variables por división	4			
		Fracción bolsa	1			
	XGBoost	Número árboles	500	0,5652	5.3452	3.333
		Tasa de aprendizaje	0,08			
<b>Biomasa seca</b>	CART	-		0,3387	0,4139	0,1873
	RF	Número árboles	300	0,9844	0,0610	0,0124
		Variables por división	3			
		Fracción bolsa	0.9			
	XGBoost	Número árboles	500	0,5592	0,1687	0,3450
		Tasa de aprendizaje	0,08			

## 5.2. DISCUSIÓN

Nuestros hallazgos muestran que las imágenes multiespectrales basadas en vehículos aéreos no tripulados pueden predecir eficazmente la biomasa. Esto se logra utilizando la plataforma GEE (Google Earth Engine) de acceso abierto a través de modelos de regresión con técnicas de aprendizaje automático como Random Forests, CART y XGBoost, sin la necesidad de comprar o descargar software (Zhou et al., 2023). Si bien la preparación de datos con GEE requiere mucho tiempo (Ganerød et al., 2023), las fases de capacitación y construcción de modelos son más rápidas que los métodos de computación tradicionales para imágenes basadas en vehículos aéreos no tripulados (Bennett et al., 2020; Magiera et al., 2017).

La generación de modelos de biomasa utilizando Machine Learning (ML) es una forma factible de predecir la biomasa, en este caso encontramos un  $R^2 = 0.95$  para biomasa fresca y  $R^2 = 0.98$  para biomasa seca utilizando Random Forest (Tabla 4), comparado con investigaciones previas (Batistoti et al., 2019; Tang et al. 2018; Barrachina et al., 2015; Ramoelo et al., 2015), nuestros resultados son similares, por lo que se afirma que el uso de imágenes multiespectrales combinadas con índices espectrales como predictores mejora la correlación con la biomasa, especialmente utilizando el Random Forest.

Esta mejora se debe principalmente a las características típicas de absorción de la vegetación en las regiones espectrales VIS y NIR (Pullanagari et al., 2017; Barrachina et al., 2015; Meneses et al., 2015). La productividad de la vegetación está fuertemente asociada con la forma en que la radiación solar interactúa con el dosel de la planta. Esta idea proviene de Monteith (1972), quien sugirió que la productividad de los cultivos anuales libres de estrés está relacionada linealmente con la radiación fotosintéticamente activa (PAR) absorbida por la vegetación. En consecuencia, se desarrollaron índices espectrales de vegetación (IV) como el índice de vegetación de diferencia normalizada (NDVI) para su aplicación en tierras de cultivo y pastizales (Sendin et al, 2018). Théau et al. (2021) emplearon un modelo de predicción de biomasa de regresión para un sistema basado en pasturas utilizando el Índice de Vegetación de Diferencia Normalizada Verde (GNDVI), logrando un  $R^2$  de 0.80 para materia fresca y 0.66 para materia seca. Esto indica

estimaciones precisas cuando los niveles de materia fresca están por debajo de los 3 kg/m<sup>2</sup>.

Dado que los modelos de aprendizaje automático no necesitan asumir normalidad (Hengl y Macmillan, 2019), no se aplicaron transformaciones de datos para el modelo utilizado en este estudio. Muestra de ello son los resultados del test de normalidad realizado (tabla 4). El RF se considera una técnica que mitiga la incertidumbre asociada con CART porque RF toma el promedio de un conjunto de datos. RF ha mostrado niveles de precisión similares en la estimación de biomasa en comparación con otros métodos utilizados en agricultura (Näsi et al., 2018). Morais et al. (2021), examinaron cómo se aplica el aprendizaje automático para estimar la biomasa en pastizales, destacando que RF fue el método más utilizado.

Al respecto, la variabilidad obtenida en nuestro muestreo, es considerable. Se sabe que la investigación en pasturas siempre muestra altos valores en sus estadígrafos de dispersión, dado que la pastura es un componente eminentemente plástico, cuyos cambios inherentes a su producción (calidad o cantidad), pueden darse en períodos muy cortos (minutos). Pero, el RF mostró su fortaleza en la corrección de esta alta variabilidad, dado que, como ya se dijo, el k-fold y su capacidad de aprendizaje continuo, permitió la el ajuste de estos valores, incluso de su distribución.

En este estudio, nuestros resultados pueden verse limitados por la rutina de rotación de pastoreo de cada potrero, lo que nos impide obtener predicciones de la carga ganadera durante un período prolongado. Pueden surgir otras limitaciones debido a la complejidad inherente del manejo de los potreros, incluidos los planes de fertilización y estiércol. Sin embargo, esta limitación podría mitigarse incorporando un monitoreo continuo vinculado a datos satelitales, aumentando así la resolución temporal de las observaciones.



## CAPÍTULO VI

### CONCLUSIONES Y SUGERENCIAS

#### 6.1. Conclusiones

- Se encontró correlaciones positivas entre Biomasa seca y Biomasa Fresca con las distintas bandas espectrales:
  - ✓ Biomasa seca con la banda espectral roja y el índice de vegetación ajustado al suelo (SRI) ( $r = 0.16 - 0.39$ ).
  - ✓ Biomasa fresca con la banda espectral del infrarrojo cercano (NIR), el NDVI y los índices espectrales ExG ( $r = 0.10 - 0.33$ ).
- Al comparar tres técnicas de aprendizaje automático (CART, RF y XGBoost), descubrimos que los modelos RF superan a los demás y brindan predicciones espacialmente más consistentes ( $R^2 = 0.95$ , para Base Fresca;  $R^2 = 0.98$ , para Base Seca).

Se demostró, entonces, que el procesamiento de datos de UAV con algoritmos de aprendizaje automático en la nube, utilizando plataformas como Google Earth Engine, ofrece una alternativa accesible a los servidores físicos para predecir la biomasa forrajera a un nivel detallado con resultados satisfactorios, bajo las condiciones de nuestra región (Cajamarca).

#### 6.2. Sugerencias

- Si bien la regresión de bosque aleatorio (RF) se reconoce como una metodología adecuada para tamaños de muestra pequeños, y la validación cruzada ayuda a evaluar el sobreajuste del modelo, aumentar la cantidad de muestras aún sería valioso para futuras investigaciones. A pesar de la carga adicional del muestreo y las mediciones *in situ*, un conjunto de datos más grande mejoraría la solidez del modelo.

## CAPÍTULO VII

### REFERENCIAS BIBLIOGRÁFICAS

- Armenta, S., Garrigues, S., de la Guardia, M. (2007). Determination of edible oil parameters by near infrared spectrometry. *Anal. Chim. Acta* 596(2), 330-337.
- Bannari, A., Morin, D., Bonn, F., Huete, A.R. (1995). A review of vegetation indices. *Remote Sens Rev.* 1995;13(1–2):95–120.
- Barrachina, M., Cristóbal, J., & Tulla, A. F. (2015). Estimación de biomasa aérea en montañas y pasturas a través de sensores remotos. *International Journal of Applied Earth Observation and Geoinformation*, 38, 184–192. <https://doi.org/10.1016/j.jag.2014.12.002>
- Batistoti, J., Marcato, J., Itavo, L., Matsubara, E., Gomes, E., Oliveira, B. (2019). Estimación de biomasa y altura del dosel en la Sabana Brasileña utilizando fotogrametría por UAV. *Remote Sens.* 2019;11(20):1–12.
- Bennett, M.K., Younes, N., Joyce, K. (2020). Automatización del procesamiento de imágenes de drones para mapear sustratos de arrecifes de coral utilizando Google Earth Engine. *Drones.* 2020;4(3):1–13.
- Bosooin, P., & Renfu, L. (2015). *Hyperspectral Imaging Technology in Food and Agriculture*. Springer.
- Breiman, L. (2001). Random Forests. *Mach Learn.* 2001;45(1):5–32.
- Carrasco Chilón, W. (2019). Determinación del estado actual de la composición florística del piso forrajero en la campiña de Cajamarca [Internet] [Tesis de maestría]. [Cajamarca]: Universidad Nacional de Cajamarca; 2019. Disponible en: <http://hdl.handle.net/20.500.14074/3270>
- Chai, T., Draxler, R.R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geosci Model Dev.* 2014;7(3):1247–50.
- De Rosa, D., Basso, B., Fasiolo, M., Friedl, J., Fulkerson, B., Grace, P.R. (2021). Predicting pasture biomass using a statistical model and machine learning

algorithm implemented with remotely sensed imagery. *Comput Electron Agric.* 2021; 180 (October, 2020): 105880.

- Esser, A., Ortega, R., Santibáñez, O. (2002). Nuevas Tecnologías para Mejorar la Eficiencia Productiva en Viñas. *Revista Agronomía y forestal*. Vol. 1 (15). Centro de Agricultura de Precisión, Universidad de Chile, p. 4-9. Disponible en [https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiB2g-Sh\\_uKAXrp5UCHbfsGz8QFnoECBoQAQ&url=https%3A%2F%2Fagronomia.uc.cl%2Fcomponent%2Fcom\\_sobipro%2FItemid%2C232%2Ffid%2C218.884%2Fsid%2C87%2Ftask%2Cdownload.file%2F&usg=AOvVaw0ylhUOErlvSdpu4XQzsBaV&opi=89978449](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiB2g-Sh_uKAXrp5UCHbfsGz8QFnoECBoQAQ&url=https%3A%2F%2Fagronomia.uc.cl%2Fcomponent%2Fcom_sobipro%2FItemid%2C232%2Ffid%2C218.884%2Fsid%2C87%2Ftask%2Cdownload.file%2F&usg=AOvVaw0ylhUOErlvSdpu4XQzsBaV&opi=89978449)
- Feng, Y. Z., Elmasry, G., Sun, D. W., Scannell, A. G. M., Walsh, D., & Morcy, N. (2013). Near-infrared hyperspectral imaging and partial least squares regression for rapid and reagentless determination of Enterobacteriaceae on chicken fillets. *Food Chemistry*, 138(2–3), 1829–1836. <https://doi.org/10.1016/j.foodchem.2012.11.040>
- Ganerød, A.J., Bakkestuen, V., Calovi, M., Fredin, O., Rød, J.K. (2023). Where are the outcrops? Automatic delineation of bedrock from sediments using Deep-Learning techniques. *Appl Comput Geosci.* 2023;18(May).
- Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L. (2018). Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens Environ.* 2018;218(September 2017):89–103.
- Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sens Environ.* 1996;58(3):289–98.
- Gitelson, A.A., Merzlyak, M.N. (1994). Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *J Plant Physiol.* 1994;143(3):286–92.

- Goebel, M., Iwaszczuk, D. (2023). Spectral Analysis of Images of Plants Under Stress Using a Close-Range Camera. *Int Arch Photogramm Remote Sens Spat Inf Sci - ISPRS Arch.* 2023;48(1/W3-2023):63–9.
- Gómez, D. (2008). Métodos para el estudio de los pastos, su caracterización ecológica y valoración.
- Hengl, T., Macmillan, R.A. (2019). Predictive Soil Mapping with R. *Predictive soil mapping with R.* 2019. 227–273 p.
- Hewson, R.D., Cudahy, T.J., Huntington, J.F. (2001). Geologic and alteration mapping at Mt fitton, South Australia, using ASTER satellite-borne data. *Int Geosci Remote Sens Symp IGARSS.* 2001;2(C):724–6.
- Hindman, T., Meyer, G.E. (1998). Machine Vision Detection Parameters for Plant Species Identification. *Syst Eng.* 1998;(3543):327–35.
- INEI. (2012). IV Censo Nacional Agropecuario. Result Definitivo IV Censo Nac Agropecu. 2012; 62.
- Ishwaran, H., Kogalur, U.B. (2010). Consistency of Random Survival Forests. *Stat Probab Lett.* 2010;80(1):1056–64.
- Jain, P., Coogan, S.C.P., Subramanian, S.G., Crowley, M., Taylor, S., Flannigan, M.D. (2020). A review of machine learning applications in wildfire science and management. *Environ Rev.* 2020;28(4):478–505.
- Jin, X., Du, J., Liu, H., Wang, Z., Song, K. (2016). Remote estimation of soil organic matter content in the Sanjiang Plain, Northeast China: The optimal band algorithm versus the GRA-ANN model. *Agric For Meteorol.* 2016;218–219:250–60.
- Magiera, A., Feilhauer, H., Waldhardt, R., Wiesmair, M., & Otte, A. (2017). Modelling biomass of mountainous grasslands by including a species composition map. *Ecological Indicators*, 78, 8–18.  
<https://doi.org/10.1016/j.ecolind.2017.02.039>
- Mayr, A., Binder, H., Gefeller, O., Schmid, M. (2014). The Evolution of Boosting Algorithms. *Methods Inf Med.* enero de 2014;53(06):419–27.

- McFeeters, S.K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int J Remote Sens.* 1996;17(7):1425–32.
- Meneses, V. A. B., Téllez, J. M., Velásquez, D. F. A. (2015). Uso de drones para el análisis de imágenes multiespectrales en agricultura de precisión. @ limentech, *Ciencia y Tecnología Alimentaria*, 13(1).
- Meyer, G.E., Neto, J.C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Comput Electron Agric.* 2008;63(2):282–93.
- Milazzo, F., Francksen, R.M., Zavattaro, L., Abdalla, M., Hejduk, S., Enri, S.R. (2023). The role of grassland for erosion and flood mitigation in Europe: A meta-analysis. *Agric Ecosyst Environ.* 2023;348(August 2022).
- Monteith, J.L. (1972). Solar Radiation and Productivity in Tropical Ecosystems. *J Appl Ecol.* 1972;9(3):747.
- Morais, T.G., Teixeira, R.F.M., Figueiredo, M., Domingos, T. (2021). The use of machine learning methods to estimate aboveground biomass of grasslands: A review. *Ecol Indic.* 2021; 130 (October, 2020): 108081.
- Mundaca, G. A. (2017). Análisis de la calidad del grano de cacao mediante imágenes hiperespectrales usando técnicas de visión artificial.
- Näsi, R., Viljanen, N., Kaivosoja, J., Alhonoja, K., Hakala, T., Markelin, L. (2018). Estimating biomass and nitrogen amount of barley and grass using UAV and aircraft based spectral and photogrammetric 3D features. *Remote Sens.* 2018;10(7):1–32.
- Oliva, M., Rojas, D., Morales, A., Oliva, C., & A. Oliva, M. (2015). Nutritional content, digestibility and performance of native grasses biomass that dominate livestock Molinopampa, Pomacochas and Leymebamba basins, Amazonas, Peru. *Scientia Agropecuaria*, 6(3), 211–215. <https://doi.org/10.17268/sci.agropecu.2015.03.07>
- Oliveira, E., & Santos, H. M. (2016). An over view on sensing materials depending on the electromagnetic spectra region applied. *Dyes and Pigments*, 135, 3–25. <https://doi.org/10.1016/j.dyepig.2016.07.012>

- Phinzi, K., Abriha, D., Szabó, S. (2021). Classification efficacy using k-fold cross-validation and bootstrapping resampling techniques on the example of mapping complex gully systems. *Remote Sens.* 2021;13(15).
- Pullanagari, R. R., Kereszturi, G., & Yule, I. J. (2017). Mapping of macro and micro nutrients of mixed pastures using air born eAisa FENIX hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117, 1–10. <https://doi.org/10.1016/j.isprsjprs.2016.03.010>
- Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sens Environ.* 1994;48(2):119–26.
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria; 2021.
- Ramirez, R. (2009). Forrajes nativos. Una alternativa sustentable en la alimentación de rumiantes. *Ciencia UANL*, XII(1), 4–5.
- Ramoelo, A., Cho, M.A., Mathieu, R., Madonsela, S., van de Kerchove, R., Kaszta, Z. (2015). Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *Int J Appl Earth Obs Geoinformation*. 2015;43:43–54.
- Reeves, J. (2000). Use of near infrared reflectance spectroscopy. In: D’Mello, J. (ed). *Farm Animal Metabolism and Nutrition*. CABI. Wallingford, UK. pp: 185-207.
- Richardson, A.J., Everitt, J.H. (1992). Using spectral vegetation indices to estimate rangeland productivity. *Geocarto Int.* 1992;7(1):63–9.
- Rondeaux, G., Steven, M., Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sens Environ.* 1996;55(2):95–107.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. En: *Proceedings of Third Earth Resources Technology Satellite Symposium*. Washington, DC: Remote Sensingcenter, Texas A&M hivemity, Colfegp Station, Texas; 1974. p. 309.

- Schuler, U., Herrmann, L., Ingwersen, J., Erbe, P., Stahr, K. (2010). Comparing mapping approaches at subcatchment scale in northern Thailand with emphasis on the Maximum Likelihood approach. *Catena*. 2010;81(2):137–71.
- Sendin, K., Manley, M., & Williams, P. J. (2018). Classification of white maize effects with multispectral imaging. *Food Chemistry*, 243(June 2017), 311–318. <https://doi.org/10.1016/j.foodchem.2017.09.133>
- Tang, C., He, H., Li, E., & Li, H. (2018). Multispectral imaging for predicting sugar content of ‘Fuji’ apples. *Optics & Laser Technology*, 106, 280-285.
- Théau, J., Lauzier-Hudon, É., Aubé, L., Devillers, N. (2021). Estimation of forage biomass and vegetation cover in grasslands using UAV imagery. *PLoS ONE*. 2021;16(1 January):1–18.
- Thenkabail, P.S., Smith, R.B., De Pauw, E. (1995). Wiegand and Richardson, † International Center for Agricultural Research in the Dry Areas 1990, natural vegetation (Friedl et al., 1994), and in (ICARDA). *Environ*. 1995;71(99):158–82
- Vásquez, E. H. (2012). Caracterización de tres especies de plantas (*Rumex Crispus*, *Azolla Anabaena*, *Beta Vulgaris*) con potencial forrajero para consumo en alimentación animal.
- Villalobos-Villalobos, L.A., Wing Ching-Jones, R. (2023). Forage Biomass Estimated with a Pre-Calibrated Equation of a Rising Platometer in Pastures Grown in Tropical Conditions. *Grasses*. 2023;2(2):127–41.
- Vincini, M., Frazzi, E., D’Alessio, P. (2008). A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precis Agric*. 2008;9(5):303–19.
- Wei, T., Simko, V. (2017). *Corrplot: Visualization of a Correlation Matrix (Version 0.84)*. 2017. p. 18.
- Woebbecke, D.M., Meyer, G.E., Von Bargen, K., Mortensen, D.A. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Trans Am Soc Agric Eng*. 1995;38(1):259–69.

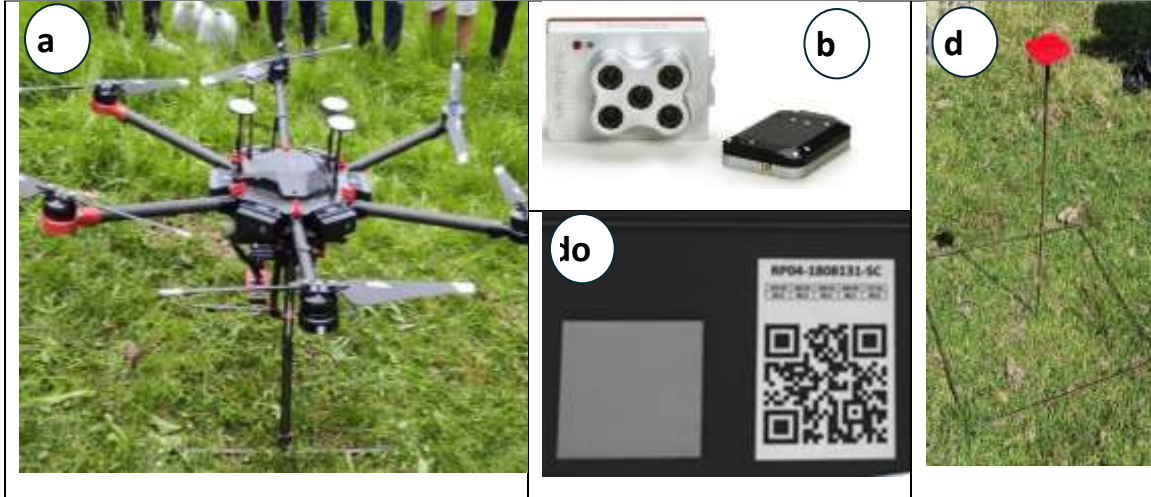
- Zarria, M. R. (2015). Inventario y estrategias de mejora de los pastizales de los sistemas de producción de alpacas en la Sierra Central. Universidad Nacional Agraria la Molina.
- Zhou, J., Xu, Y., Gu, X., Chen, T., Sun, Q., Zhang, S. (2023). High-Precision Mapping of Soil Organic Matter Based on UAV Imagery Using Machine Learning Algorithms. Drones. 2023;7(5).



## CAPÍTULO VIII

### ANEXOS

Anexo 1. Características detalladas del UAV, la cámara y el plan de vuelo utilizados.



Anexo 2. Análisis de correlación entre las variables de estudio.

Resumen

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.998408606
Coefficiente de determinación R <sup>2</sup>	0.996819744
R <sup>2</sup> ajustado	0.995627148
Error típico	0.051846459
Observaciones	12

ANÁLISIS DE VARIANZA

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	3	6.740354224	2.246784741	835.8402249	2.51418E-10
Residuos	8	0.021504442	0.002688055		
Total	11	6.761858667			

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 95.0%</i>	<i>Superior 95.0%</i>
Intercepción	-5.282687201	5.884062904	-0.89779584	0.395509807	-18.85136059	8.285986189	-18.85136059	8.285986189
Variable X 1	0.772869364	0.100473901	7.692240011	0.0345245	0.541176132	1.004562596	0.541176132	1.004562596
Variable X 2	-0.01047166	0.008930509	-1.172571473	0.0248792	-0.03106545	0.01012213	-0.03106545	0.01012213
Variable X 3	0.068318342	0.061454993	1.111680902	0.0365478	-0.073397126	0.210033811	-0.073397126	0.210033811

Anexo 3. Puntos de calibración (80) para la determinación de constantes del valor de las bandas.

<i>Fundo</i>	<i>Cod. Punto</i>	<i>Punto</i>	<i>X</i>	<i>Y</i>
<i>Fundo 1</i>	<i>Inicio</i>	<i>PM1</i>	779509.44	9206058.07
<i>Fundo 1</i>	<i>W2</i>	<i>PM2</i>	779480.82	9206065.79
<i>Fundo 1</i>	<i>W29</i>	<i>PM3</i>	779471.22	9206104.68
<i>Fundo 1</i>	<i>W3</i>	<i>PM4</i>	779460.45	9206086.38
<i>Fundo 1</i>	<i>W4</i>	<i>PM5</i>	779441.00	9206082.00
<i>Fundo 1</i>	<i>W7</i>	<i>PM6</i>	779390.53	9206070.18
<i>Fundo 1</i>	<i>W6</i>	<i>PM7</i>	779417.33	9206016.79
<i>Fundo 1</i>	<i>W9</i>	<i>PM8</i>	779539.09	9206096.23
<i>Fundo 1</i>	<i>W10</i>	<i>PM9</i>	779540.44	9206132.36
<i>Fundo 1</i>	<i>W11</i>	<i>PM10</i>	779591.85	9206153.59
<i>Fundo 1</i>	<i>W13</i>	<i>PM11</i>	779661.24	9206194.46
<i>Fundo 1</i>	<i>PM12</i>	<i>PM12</i>	779572.00	9206040.00
<i>Fundo 1</i>	<i>PM13</i>	<i>PM13</i>	779630.00	9206056.00
<i>Fundo 1</i>	<i>PM14</i>	<i>PM14</i>	779641.00	9205993.00
<i>Fundo 1</i>	<i>PM15</i>	<i>PM15</i>	779579.00	9206000.00
<i>Fundo 1</i>	<i>PM16</i>	<i>PM16</i>	779626.00	9205945.00
<i>Fundo 1</i>	<i>PM17</i>	<i>PM17</i>	779689.00	9205978.00
<i>Fundo 1</i>	<i>PM18</i>	<i>PM18</i>	779745.00	9205985.00
<i>Fundo 1</i>	<i>PM19</i>	<i>PM19</i>	779706.00	9206071.00
<i>Fundo 1</i>	<i>PM20</i>	<i>PM20</i>	779675.00	9206100.00
<i>Fundo 1</i>	<i>PM21</i>	<i>PM21</i>	779685.00	9206151.00
<i>Fundo 1</i>	<i>PM22</i>	<i>PM22</i>	779632.00	9206169.00
<i>Fundo 1</i>	<i>PM23</i>	<i>PM23</i>	779539.00	9206098.00
<i>Fundo 1</i>	<i>PM24</i>	<i>PM24</i>	779518.44	9206062.07
<i>Fundo 1</i>	<i>PM25</i>	<i>PM25</i>	779489.82	9206069.79
<i>Fundo 1</i>	<i>PM26</i>	<i>PM26</i>	779480.22	9206108.68
<i>Fundo 1</i>	<i>PM27</i>	<i>PM27</i>	779469.45	9206090.38
<i>Fundo 1</i>	<i>PM28</i>	<i>PM28</i>	779450.00	9206086.00
<i>Fundo 1</i>	<i>PM29</i>	<i>PM29</i>	779399.53	9206074.18
<i>Fundo 1</i>	<i>PM30</i>	<i>PM30</i>	779426.33	9206020.79
<i>Fundo 1</i>	<i>PM31</i>	<i>PM31</i>	779548.09	9206100.23
<i>Fundo 1</i>	<i>PM32</i>	<i>PM32</i>	779549.44	9206136.36
<i>Fundo 1</i>	<i>PM33</i>	<i>PM33</i>	779600.85	9206157.59
<i>Fundo 1</i>	<i>PM34</i>	<i>PM34</i>	779670.24	9206198.46
<i>Fundo 1</i>	<i>PM35</i>	<i>PM35</i>	779581.00	9206044.00
<i>Fundo 1</i>	<i>PM36</i>	<i>PM36</i>	779639.00	9206060.00
<i>Fundo 1</i>	<i>PM37</i>	<i>PM37</i>	779650.00	9205997.00
<i>Fundo 1</i>	<i>PM38</i>	<i>PM38</i>	779588.00	9206004.00
<i>Fundo 1</i>	<i>PM39</i>	<i>PM39</i>	779635.00	9205949.00
<i>Fundo 1</i>	<i>PM40</i>	<i>PM40</i>	779698.00	9205982.00
<i>Fundo 1</i>	<i>PM41</i>	<i>PM41</i>	779754.00	9205989.00
<i>Fundo 1</i>	<i>PM42</i>	<i>PM42</i>	779715.00	9206075.00
<i>Fundo 1</i>	<i>PM43</i>	<i>PM43</i>	779511.44	9206104.00
<i>Fundo 1</i>	<i>PM44</i>	<i>PM44</i>	779482.82	9206155.00

Fundo 1	PM45	PM45	779473.22	9206173.00
Fundo 1	PM46	PM46	779462.45	9206102.00
Fundo 1	PM47	PM47	779443.00	9206066.07
Fundo 1	PM48	PM48	779392.53	9206073.79
Fundo 1	PM49	PM49	779419.33	9206112.68
Fundo 1	PM50	PM50	779541.09	9206094.38
Fundo 1	PM51	PM51	779542.44	9206090.00
Fundo 1	PM52	PM52	779593.85	9206078.18
Fundo 1	PM53	PM53	779663.24	9206024.79
Fundo 1	PM54	PM54	779574.00	9206104.23
Fundo 1	PM55	PM55	779632.00	9206140.36
Fundo 1	PM56	PM56	779643.00	9206161.59
Fundo 1	PM57	PM57	779581.00	9206202.46
Fundo 1	PM58	PM58	779628.00	9206048.00
Fundo 1	PM59	PM59	779691.00	9206064.00
Fundo 1	PM60	PM60	779747.00	9206001.00
Fundo 1	PM61	PM61	779708.00	9206008.00
Fundo 1	PM62	PM62	779504.44	9205953.00
Fundo 1	PM63	PM63	779431.33	9205986.00
Fundo 1	PM64	PM64	779553.09	9205993.00
Fundo 1	PM65	PM65	779554.44	9206079.00
Fundo 1	PM66	PM66	779605.85	9206108.00
Fundo 1	PM67	PM67	779675.24	9206159.00
Fundo 1	PM68	PM68	779586.00	9206177.00
Fundo 1	PM69	PM69	779644.00	9206106.00
Fundo 1	PM70	PM70	779655.00	9206070.07
Fundo 1	PM71	PM71	779593.00	9206077.79
Fundo 1	PM72	PM72	779640.00	9206116.68
Fundo 1	PM73	PM73	779703.00	9206098.38
Fundo 1	PM74	PM74	779759.00	9206094.00
Fundo 1	PM75	PM75	779720.00	9206082.18
Fundo 1	PM76	PM76	779642.00	9206028.79
Fundo 1	PM77	PM77	779653.00	9206108.23
Fundo 1	PM78	PM78	779591.00	9206144.36
Fundo 1	PM79	PM79	779638.00	9206165.59
Fundo 1	PM80	PM80	779701.00	9206206.46
Fundo 1	PM81	PM81	779757.00	9206052.00
Fundo 1	PM82	PM82	779718.00	9206068.00
Fundo 1	PM83	PM83	779514.44	9206005.00
Fundo 1	PM84	PM84	779485.82	9206012.00
Fundo 1	PM85	PM85	779476.22	9205957.00
Fundo 1	PM86	PM86	779465.45	9205990.00
Fundo 1	PM87	PM87	779446.00	9205997.00
Fundo 1	PM88	PM88	779395.53	9206083.00
Fundo 1	PM89	PM89	779422.33	9206112.00
Fundo 1	PM90	PM90	779544.09	9206163.00

Anexo 4. Imágenes tomadas con la cámara multiespectral, desde el dron.

