

Google Earth Engine: Evaluación de los alcances y limitaciones para la resolución de problemas agrícolas en La Finca de Santa Lucía, Barva, Heredia







Google Earth Engine: Evaluation of the scopes and limitations for the resolve of agricultural problems in Santa Lucía, Barva, Heredia

Google Earth Engine: avaliação do alcance e limitações para resolução de problemas agrícolas na Fazenda Santa Lucía, Barva, Heredia

Sara Bastos Gutiérrez¹
Universidad Nacional, Costa Rica

Diana Paniagua Jiménez²
Universidad Nacional, Costa Rica

Alejandro Vargas Martínez³
Universidad Nacional, Costa Rica

-
- 1 Máster en Sistemas de Información Geográfica y Teledetección. Escuela de Topografía, Catastro y Geodesia. Universidad Nacional de Costa Rica.
 sara.bastos.gutierrez@una.cr,  <https://orcid.org/0000-0002-9167-7396>.
 - 2 Máster en Gestión de Innovación Tecnológica. Escuela de Topografía, Catastro y Geodesia. Universidad Nacional de Costa Rica.  diana.paniagua.jimenez@una.cr,  <https://orcid.org/0000-0003-2834-5310>
 - 3 Máster en Estadística Aplicada. Escuela de Ciencias Agrarias. Universidad Nacional de Costa Rica.
 alejandro.vargas.martinez@una.cr,  <https://orcid.org/0000-0001-8039-8984>.



Resumen:

Google Earth Engine (GEE) se presenta como una herramienta innovadora para la gestión agrícola, mediante el análisis geoespacial de imágenes satelitales. En este estudio, se utilizó GEE para clasificar suelos en la Finca Experimental Santa Lucía (FESL), empleando imágenes de la colección Sentinel-2 del programa Copérnico con una resolución de 20×20 m durante todo el año 2022. Se consideraron cuatro clases de entrenamiento (pastos, bosques, café e infraestructura) con el clasificador Random Forest. Adicionalmente, se realizó una matriz de confusión píxel por píxel para el proceso de entrenamiento y validación del modelo. Los resultados muestran una precisión global del 96 % en el entrenamiento y del 61 % en la validación, destacando la eficiencia del modelo en la distinción de clases, aunque con áreas potenciales de mejora en la diferenciación entre café y bosque.

Palabras clave: Agricultura, Google Earth Engine, imágenes satelitales, Sentinel 2, uso de suelo



Abstract:

Google Earth Engine (GEE) is presented as an innovative tool for agricultural management through the geospatial analysis of satellite imagery. In this study, GEE was used to classify soils at the Santa Lucía Experimental Farm (FESL), employing Sentinel-2 imagery from the Copernicus program with a resolution of 20×20 m throughout the year 2022. Four training classes were considered (pastures, forests, coffee, and infrastructure) using the Random Forest classifier. Additionally, a pixel-by-pixel confusion matrix was generated for both the training and validation processes. The results show an overall accuracy of 96% for the training set and 61% for the validation set, highlighting the model's efficiency in class distinction, although with potential for improvement in distinguishing between coffee and forest classes.

Keywords: Agriculture, Google Earth Engine, land use, satellite images, Sentinel 2



Resumo:

O Google Earth Engine (GEE) é apresentado como uma ferramenta inovadora para a gestão agrícola, por meio da análise geoespacial de imagens de satélite. Neste estudo, o GEE foi utilizado para classificar solos na Fazenda Experimental Santa Lucía (FESL), utilizando imagens da coleção Sentinel-2 do programa Copernicus com resolução de 20×20 m ao longo de 2022. Quatro classes de treinamento (pastagem, floresta, café e infraestrutura) foram consideradas com o classificador Random Forest. Além disso, uma matriz de confusão pixel a pixel foi gerada para o processo de treinamento e validação do modelo. Os resultados mostram uma precisão geral de 96% no treinamento e 61% na validação, destacando a eficiência do modelo na distinção de classes, embora com potenciais áreas de melhoria na diferenciação entre café e floresta.

Palavras-chave: Agricultura, Google Earth Engine, imagens de satélite, Sentinel-2, uso do solo

INTRODUCCIÓN

La agricultura es fundamental en la sociedad actual, ya que desempeña un papel crucial en la seguridad alimentaria, el desarrollo económico y la sostenibilidad ambiental. Este sector no solo asegura el suministro constante de alimentos necesarios para una población en crecimiento, sino que, también impulsa la economía de varias regiones, especialmente en países en desarrollo. Según la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), la agricultura es esencial para erradicar el hambre y la pobreza, promoviendo un desarrollo rural inclusivo y sostenible (FAO, 2019).

En Costa Rica, la fuerza laboral en el sector agropecuario representa el 13,0 % del total, mostrando una tendencia al crecimiento con un aumento del 7,2 % respecto a 2016 (Mora, 2018). Sin embargo, el año 2020 presentó una contracción en la actividad agropecuaria debido a una disminución en la demanda externa de productos como piña, flores, follajes, sandía y melón. Contrariamente, los productos destinados al mercado local, como hortalizas, raíces y tubérculos experimentaron un crecimiento, gracias a condiciones climáticas estables y una mayor demanda interna, además de una recuperación en la producción de banano, lo cual mitigó el impacto de la pandemia en este sector (Mora, 2020).

La calidad y cantidad exigida en los productos agrícolas ha impulsado la implementación de estrategias significativas, como el uso intensivo de agroquímicos y fertilizantes para proteger la producción, a menudo con un alto impacto ambiental. En este contexto, la toma de decisiones, en la agricultura, desde la etapa inicial hasta el mercado final, se basa en variables que proporcionan información de calidad, permitiendo procesos agrícolas más eficientes y con menor impacto en el ambiente.

En la actualidad, la agricultura enfrenta desafíos significativos, entre ellos el cambio climático, la escasez de recursos y la creciente demanda de alimentos. El cambio climático afecta la productividad agrícola y la disponibilidad de agua, mientras que la creciente población mundial incrementa la presión sobre los recursos naturales (Intergovernmental Panel on Climate Change [IPCC], 2019). Además, la agricultura debe adaptarse a la escasez de recursos, como el agua y la tierra arable, y, a la necesidad de producir más alimentos con menos recursos (Godfray et al.,

2010). Estos desafíos requieren soluciones innovadoras que permitan a los agricultores tomar decisiones más informadas y eficientes.

La gestión de actividades agrícolas se realiza manualmente, complementada por herramientas como sistemas de posicionamiento global y muestreos manuales o de laboratorio para evaluar el estado fitosanitario de los cultivos. Aunque existen *softwares* avanzados para la gestión agrícola, su alto costo y requerimientos técnicos limitan su accesibilidad.

Google Earth Engine (GEE) emerge como una solución innovadora para la gestión agrícola, mediante el uso de imágenes satelitales. Esta plataforma gratuita en la nube permite el almacenamiento y procesamiento de grandes bases de datos, ofreciendo capacidades de análisis geoespacial sin necesidad de *hardware* costoso. Con un vasto repositorio de imágenes satelitales y una API accesible, GEE facilita la toma de decisiones informadas en la agricultura, abordando problemas globales como la deforestación, sequías y cambios climáticos (Gorelick et al., 2017; Shelestov et al., 2017). Su capacidad para manejar datos masivos ha sido aprovechada en diversas aplicaciones agrícolas, incluyendo la planificación espacial y el monitoreo de cultivos (Yang, et al, 2024). Además, GEE ha sido fundamental en la integración de tecnologías emergentes como la inteligencia artificial y el aprendizaje automático, contribuyendo así a una agricultura más eficiente y sostenible, especialmente en contextos de cambio climático y escasez de recursos (Lottering, et al, 2025). Su uso, además, facilita procesos de análisis geoespacial en contextos de investigación aplicada y toma de decisiones, tanto en instituciones gubernamentales como en unidades académicas con recursos computacionales limitados (Tan., et al, 2024).

En Costa Rica, varias entidades como el Ministerio de Ambiente y Energía (MINAE) y el Laboratorio PRIAS del Centro Nacional de Alta Tecnología (CENAT) ya están explorando el uso de GEE para mejorar sus procesos. El Sistema Nacional de Monitoreo de Cobertura y Uso de la Tierra y Ecosistemas (SIMOCUTE) utiliza GEE para el inventario forestal y monitoreo ambiental (MINAE, 2020). Además, el Laboratorio PRIAS realiza capacitaciones en el uso de la plataforma y estudios en agricultura que permiten usar las experiencias de otras entidades como colaboración para el desarrollo de investigaciones. Debido a las características de GEE, que incluyen procesamiento en la nube, capacidad computacional en paralelo y un amplio catálogo de datos, se presenta como una herramienta

poderosa para afrontar los problemas de Big Data y ampliar los horizontes de la ciencia de datos y la percepción remota (Gorelick et al., 2017; Perilla & Mas, 2019).

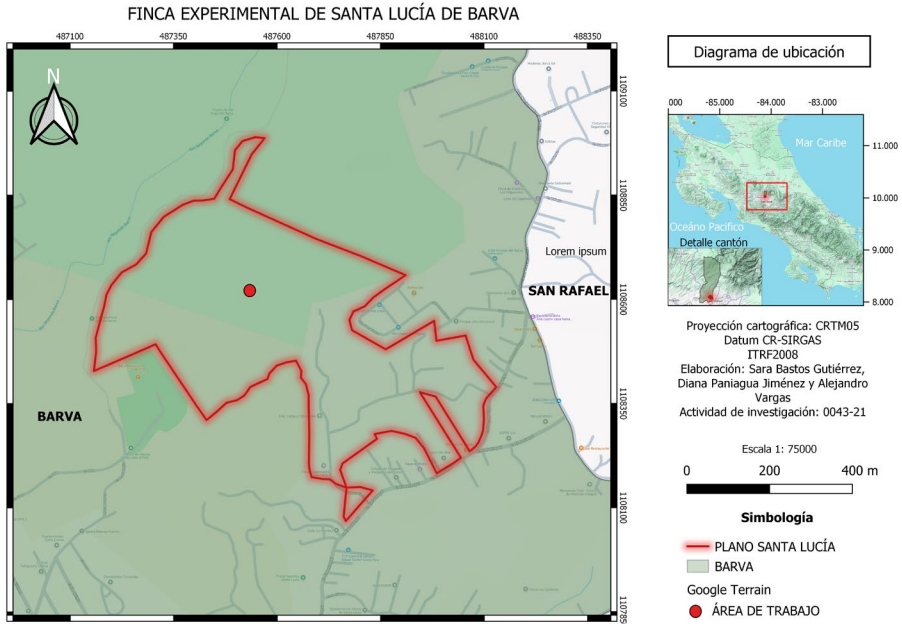
GEE permite a cualquier usuario con conocimientos básicos de programación y acceso a internet realizar investigaciones utilizando grandes cantidades de datos, fomentando una comunidad científica más equitativa y democrática. En el sector agrícola, GEE facilita la generación de mapas de clasificación de vegetación, monitoreo de uso de suelo, estimación de rendimientos de cultivos, mapeo de áreas de cultivo, evaluación de vulnerabilidad a plagas y enfermedades y evaluaciones de idoneidad, entre otras aplicaciones cruciales para mantener la productividad y seguridad alimentaria.

El propósito de esta investigación fue evaluar los alcances y las limitaciones de la herramienta GEE para la gestión de actividades agrícolas, y determinar cómo esta plataforma puede contribuir a mejorar la eficiencia y sostenibilidad de la agricultura con el uso de datos satelitales y análisis geospaciales. Por tanto, el proyecto de investigación generó un mapa clasificación de la Finca Experimental de Santa Lucía de 20x20 m de resolución, producto de las imágenes Sentinel-2, en el cual se utilizó un procedimiento de clasificación supervisada y validación basado en la plataforma de GEE.

ÁREA DE ESTUDIO

La Finca Experimental Santa Lucía (FESL) se encuentra ubicada entre una latitud de 10° 1' 0" N y una longitud de 84° 06' 59" O a una altitud de 1200 m s. n. m., provincia de Heredia, cantón de Barva y en el distrito de Santa Lucía. La zona de estudio corresponde a un área de 34 hectáreas (Figura 1), y es administrada por la Escuela de Ciencias Agrarias de la Universidad Nacional.

Figura 1: Mapa de ubicación geográfica del área de interés



Fuente: Elaboración propia.

La Finca Experimental Santa Lucía tiene un 65 % de área ocupada por los apartos y respastos para actividades pecuarias, el 15 % por el cultivo de café y el restante 20 % por actividades relacionadas a cultivos, obras grises, entre otras.

La topografía varía de plana a escarpada, de superficiales a profundos, moderadamente fértiles, bien estructurados y mucha pedregosidad. La zona cuenta con una precipitación anual de 2403 mm, una humedad relativa de 80 %, temperatura media anual de 20,30 °C, con 15,40 °C de promedio entre las mínimas y 25,20 °C de promedio entre las máximas.

MARCO TEÓRICO-CONCEPTUAL

1. Google Earth Engine

Google Earth Engine (GEE) es una plataforma desarrollada por Google que permite realizar procesamientos geospaciales a gran

escala, utilizando bases de datos con millones de estos. Uno de los principales intereses de esta plataforma es reducir el tiempo invertido en el preprocesamiento y facilitar los análisis realizados con información geoespacial (Solórzano & Perilla, 2022).

Es muy importante considerar que la plataforma presenta 4 principales elementos para el manejo de los datos:

- La capacidad en el procesamiento de la información (servidores con capacidad de hasta 10 000 CPU) (Perilla & Mas, 2020).
- El acervo de los datos (imágenes de sensores Landsat, Sentinel, MODIS, entre otros). Estas bases de datos se actualizan a medida que se toman nuevas imágenes (cerca de 6000 nuevas escenas diarias), creando así un enorme catálogo de datos geoespaciales. (Perilla & Mas, 2020).
- La API (Application Program Interface), la cual consiste en una serie de comandos y funciones preestablecidas, escritas en lenguaje JAVA, que permite una programación sencilla. Además, permite una interface con Python el cuál puede usar complementariamente bibliotecas de Python, para así, hacer procesamientos más complejos o funcionalidades que la API de JavaScript no permite (Solórzano & Perilla, 2022).
- Y, por último, el usuario puede, a través de código de trabajo (scripts), llamar a los datos, procesar y visualizarlos, de manera virtual, con los servidores de Google, teniendo así sus resultados e información en la nube, a partir de code Editor. (Perilla & Mas, 2020).

Adicionalmente, presenta características y ventajas para el análisis de la información geoespacial ya que contiene un catálogo de datos con imágenes con datos históricos (<https://earthengine.google.com/datasets/>), desde el clima, agricultura, imágenes con datos geofísicos, entre otros; otra de las ventajas es que se pueden tener imágenes preprocesadas, análisis radiométrico, detección de cambios y análisis de series de tiempo.

Estudios recientes han mostrado cómo el uso de GEE, combinado con algoritmos como Random Forest, permite identificar patrones temporales en cultivos, mejorar la precisión de mapas de uso del suelo y reducir

significativamente el tiempo de procesamiento de imágenes satelitales (Gorelick et al., 2017; Yang et al., 2024). Gracias a estas ventajas, GEE se ha convertido en una herramienta clave en la democratización del análisis geoespacial, brindando acceso a investigadores, Gobiernos y productores agrícolas.

2. Sentinel-2

Sentinel-2 es una misión de observación terrestre diseñada específicamente para ofrecer datos e imágenes, desarrollada por Agencia Espacial Europea (ESA) dentro del programa Copérnico (<https://www.esa.int>), Sentinel-2 es una constelación formada por satélites gemelos; Sentinel-2A y Sentinel-2B (<https://sentinel.esa.int/web/sentinel/missions/Sentinel-2>), los satélites Sentinel-2 llevan a bordo instrumentos de imágenes multiespectrales (MSI) con capacidad de registrar 13 bandas espectrales en las longitudes de onda visible, cercanas a infrarrojos (NIR) e infrarrojos de onda corta (SWIR), con cuatro bandas a los 10m, seis bandas a 20m y tres bandas a los 60m de resolución espacial (Phiri et al., 2020) (Wang et al., 2017) (Chanev., et al, 2025). El objetivo principal de la misión es proporcionar datos de satélites de alta resolución para monitoreo de cobertura/uso, cambio climático y monitoreo de desastres, además de complementar los otros programas de satélites globales como Landsat y SPOT (Satellite Pour l'Observation de la Terre) (Phiri et al., 2020).

El uso de Sentinel-2 ha permitido el monitoreo de cobertura y el uso de suelo como una de las aplicaciones esenciales de la misión (Bruzzone et al., 2017), adicionalmente, Copérnico da los servicios de cobertura terrestre de alta resolución espacial. Sentinel-2 es la primera misión óptica de observación de la Tierra de su tipo, proporciona información clave sobre el estado de la vegetación y se ha demostrado que el satélite discrimina entre los diferentes cultivos, como, por ejemplo, se logró diferenciar girasoles y maíz en una zona cerca de Toulouse, Francia (<https://www.esa.int>).

Sentinel-2 se ha convertido en una herramienta importante para monitorear las actividades agrícolas, esto lo demuestran diversos estudios que se ha centrado en el desarrollo de productos globales para apoyar actividades agrícolas (Bontemps et al., 2015). Las aplicaciones específicas de Sentinel-2 en agricultura, incluyen monitoreo de la producción de cultivos, mapeo de tipos de cultivos, monitoreo de la agricultura de riego,

evaluación de contenido de nitrógeno y la salud de los cultivos (Phiri et al., 2020). Este satélite también ha sido utilizado en combinación con modelos de aprendizaje automático en plataformas como GEE para predecir rendimientos y evaluar el estado fisiológico de los cultivos. Por ejemplo, en un estudio reciente se logró identificar la salud de los cultivos de invierno mediante análisis multitemporal de imágenes Sentinel-2 Deep Resolution 3.0 (Chanev et al., 2025).

3. Teledetección aplicada a la agricultura

La teledetección se define como el estudio y medida de las características de una serie de objetos (superficie terrestre), sin que exista contacto físico. Para ello se miden las perturbaciones que el objeto provoca en su entorno, principalmente las de tipo electromagnético (Olaya, 2020). Las aplicaciones terrestres de la teledetección incluyen la cartografía, detección de cambios, monitoreo, modelado y otras observaciones de usos de la tierra, derivado de la actividad humana (Khorram et al., 2016). Una de las aplicaciones más utilizadas de la teledetección es la creación de mapas de uso y cobertura de suelo. Estos mapas se crean tomando una imagen y categorizando las características naturales y artificiales en representativos grupos o clases. Además, la utilización de sensores remotos puede ser el único mecanismo viable para determinar cambios en la vegetación ya sea a nivel subregional, regional y global (Khorram et al., 2016). La teledetección ha revolucionado la forma en que se gestiona la agricultura moderna. Permite la observación remota de cultivos, la detección temprana de enfermedades, el seguimiento de la humedad del suelo y el control del estrés hídrico. Estas capacidades son fundamentales para la agricultura de precisión y la gestión sostenible de los recursos (Lottering et al., 2025).

Por otra parte, la radiación electromagnética que reflejan las plantas contiene información sobre su composición biofísica y estado fisiológico, y se puede medir con sensores remotos. En la agricultura, en situaciones donde los cultivos interactúan con cualquier aspecto dado de su entorno (variaciones climáticas estacionales, eventos meteorológicos extremos, plagas, propiedades del suelo, etc.) o a medida que los cultivos crecen y pasan por diferentes etapas fenológicas, las interacciones entre las plantas y la reflectancia de la luz se traducen en cambios o patrones de señales

de las plantas que pueden ser interpretadas utilizando datos satelitales (Segarra et al., 2020).

El uso combinado de sensores remotos con modelos predictivos ha mejorado sustancialmente la planificación del riego en cultivos de zonas tropicales y subtropicales. Investigaciones recientes han mostrado cómo los sensores de humedad, acoplados a las imágenes de satélite, pueden identificar zonas críticas dentro de un cultivo y optimizar el uso del agua (Yang et al., 2024). Asimismo, el análisis bitemporal de índices espectrales ha permitido cuantificar pérdidas vegetales con alta precisión, como lo evidencian estudios realizados en ecosistemas agrícolas del norte de Europa (Rynkiewicz., et al, 2025).

4. Aprendizaje automático en la clasificación de imágenes

Actualmente dentro del mundo de Machine Learning (Aprendizaje Automático) subcampo de la Inteligencia Artificial (IA) existen diversas técnicas, que hoy en día se utilizan para la clasificación supervisada, utilizando etiquetas de clases previamente establecidas para entrenar al modelo y luego este aprende a reconocer y clasificar automáticamente las clases (Hernández et al., 2023).

Entre las técnicas más utilizadas está Random Forest (RF) (Bosques Aleatorios) por su habilidad de generar resultados de clasificación precisos, por medio del ensamblaje de múltiples árboles de decisión para llegar a una predicción con la menor varianza (Mahapatra, 2014). Dentro de sus características se menciona que es un algoritmo versátil, es estable frente al ruido y valores atípicos, menos propenso al sobreajuste, permite el uso de bases de datos con alta dimensionalidad, rápida velocidad de entrenamiento, entre otros (Xi, 2022; Alonso y Hoyos, 2025). Lee et al. (2024) reporta que el 90 % de técnicas de AI disponibles en GEE son de aprendizaje automático, aplicadas a la clasificación de imágenes para mapeo de cultivos y que RF es el algoritmo más utilizado.

MARCO METODOLÓGICO

En la siguiente sección, se detalla cada etapa del estudio, proporcionando una descripción de los procedimientos utilizados, los datos analizados y las herramientas aplicadas, con el fin de ofrecer una evaluación objetiva y fundamentada sobre los alcances y limitaciones de GEE.

Adquisición y procesamiento de imágenes

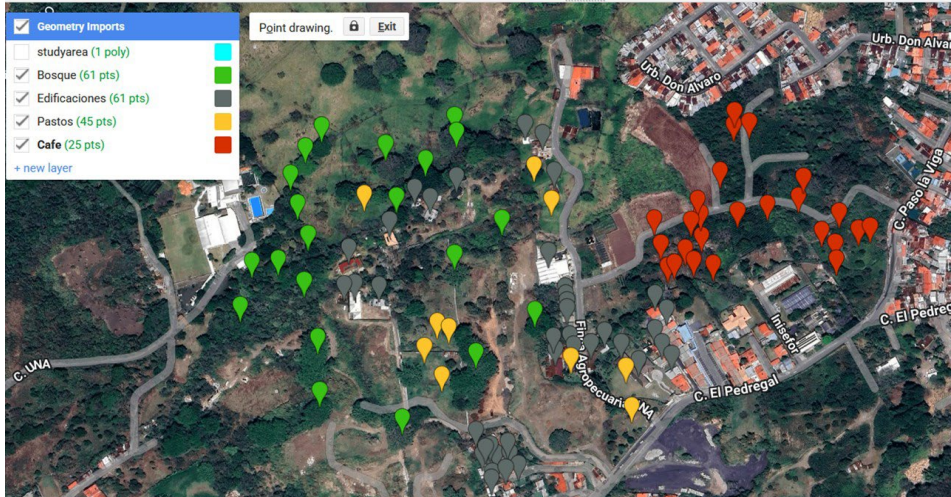
Se utilizaron las imágenes del satélite Sentinel-2 (Multiespectrales) y se descargaron del portal de acceso de datos de la Agencia Espacial Europea (ESA). Se trabajó con las 13 bandas espectrales (resolución espacial de 10, 20 y 60 metros) y la colección de imágenes utilizadas comprendió el año 2022 completo. Para el procesamiento primero se filtró por la propiedad de la imagen “Porcentaje de nublado” de hasta un 20 % de nubosidad y segundo, se obtuvo de todas las colecciones una imagen formada con el valor de la mediana del pixel para cada una de las bandas.

Clasificación de las imágenes

Para la clasificación supervisada se establecieron cuatro clases de entrenamiento: cultivo de café (clase 1), pastos (clase 2), bosques (clase 3) e infraestructura (clase 4), como se puede ver en la Figura 2. Luego se hizo la base de datos utilizando la colección de imágenes como base, para ello se estableció una zona buffer alrededor del área de estudio (FESL) y se seleccionaron puntos de muestreos para cada clase. La base de datos se dividió en dos, una base de datos de entrenamiento y una base de datos de evaluación, con un porcentaje de 75 % y 25 % respectivamente. Se utilizó el algoritmo de clasificación supervisado de bosques de decisión o *Random Forest (RF)* como clasificador principal, debido a su rendimiento competitivo en tareas de clasificación multiclase, su capacidad para manejar datos de alta dimensionalidad y su bajo riesgo de sobreajuste. Este algoritmo construye múltiples árboles de decisión y emite una predicción final, por votación mayoritaria entre ellos, lo que lo hace especialmente robusto frente a ruido en los datos y eficaz para diferenciar clases en entornos espectralmente complejos, como los asociados al uso del suelo en imágenes multiespectrales (Tan., et al, 2024; Kumar & Singh, 2023).

Las variables utilizadas para generar la clasificación fueron las que se obtuvieron cuando se tomó la mediana de cada banda, para un total de 23 variables.

Figura 2: Captura de pantalla de las variables de geometría tipo punto en GEE



Fuente: Elaboración propia.

Evaluación del modelo

Para la evaluación del desempeño del modelo, tanto para los datos de entrenamiento como para los datos de la validación, se hizo uso de una matriz de confusión con las predicciones realizadas por el modelo con los valores reales, permitiendo determinar cuántas veces el modelo acierta o se equivoca en la clasificación de cada clase. La estructura de la matriz de confusión tiene en filas las clases reales y en sus columnas las clases predichas por el modelo y derivan las siguientes métricas:

- Verdadero positivo (TP): casos en los que el modelo predice correctamente la clase
- Verdadero negativo (TN): casos en los que el modelo predice correctamente la clase negativo
- Falso positivo (FP): casos en los que el modelo predice la clase positiva siendo esta negativa
- Falso negativo (FN): Casos en los que el modelo predice la clase negativa, pero en realidad es positiva

Se utilizaron dos parámetros cuantitativos, el primero fue la exactitud promedio como el cociente entre la suma de las predicciones correctas (diagonal principal de la matriz de confusión) y el total de observaciones evaluadas, expresado como porcentaje, utilizando la siguiente fórmula.

$$Exactitud = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k (TP_i + FP_i + FN_i + TN_i)}$$

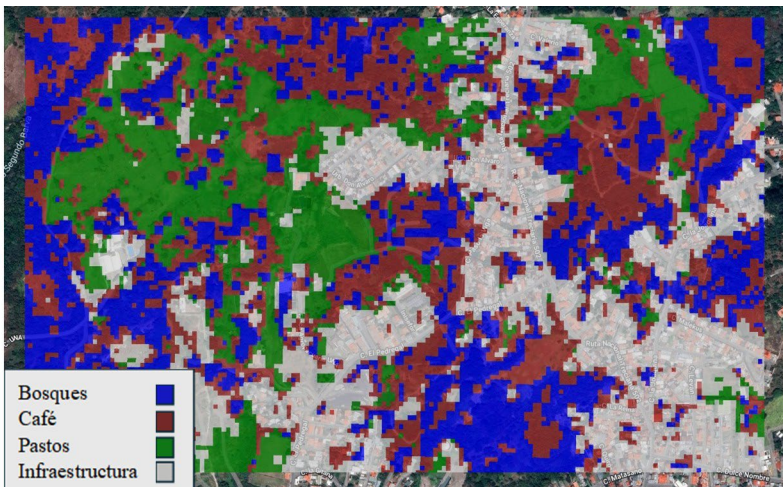
Y el segundo fue el error global, que representa la tasa de error del clasificador y se obtuvo como

$$Error\ global = 1 - Exactitud$$

RESULTADOS

La clasificación obtenida se presenta detalladamente en la Figura 3, donde se identifican las siguientes clases: bosque (azul), pastos (verde), café (marrón) e infraestructura (gris). De acuerdo con los resultados, las clases pastos e infraestructura muestran un traslape mínimo, lo que indica una clara diferenciación entre estas categorías en el análisis de imágenes, mientras que, bosque y café tienden a confundirse; esto implica la necesidad de mejorar los algoritmos de clasificación o emplear técnicas adicionales para aumentar la precisión en la distinción entre bosque y café.

Figura 3: Mapa de clasificación de clases utilizando el algoritmo



Fuente: Elaboración propia.

En relación con las medidas de desempeño, la precisión global alcanzada fue del 96 %, mientras que el error global para el conjunto de entrenamiento se mantuvo en un 4 %, como se detalla en la Tabla 1. Estos resultados demuestran un alto nivel de exactitud en la clasificación de las diferentes clases.

La Tabla 1 presenta la matriz de confusión correspondiente al proceso de entrenamiento, donde se pueden observar las siguientes asignaciones:

- **Clase café:** 70 casos correctamente clasificados, con 1 caso clasificado erróneamente como bosque y 1 como pasto.
- **Clase bosque:** 64 casos correctamente clasificados, con 6 casos clasificados erróneamente como café.
- **Clase infraestructura:** 54 casos correctamente clasificados, con 1 caso clasificado erróneamente como bosque.
- **Clase pasto:** 48 casos correctamente clasificados, sin errores de clasificación.

Tabla 1: Matriz de confusión para el proceso de entrenamiento

Clases	Café	Bosque	Pasto	Infraestructura
Café	70	1	1	0
Bosque	6	64	0	0
Infraestructura	0	1	54	0
Pasto	0	0	0	48

Fuente: Elaboración propia.

En cuanto a la base de datos de validación, la precisión global alcanzada fue del 61 %, mientras que el error global se situó en un 39 %. Estos resultados reflejan un desafío significativo en la precisión del modelo cuando se evalúa con datos independientes de validación. En particular, las clases café y bosque mostraron las menores precisiones, ambas con un 50 %, lo que indica una considerable confusión entre ellas. En contraste, la clase pasto destacó por su alta precisión, logrando un 93 %, lo que sugiere una diferenciación y correcta identificación de esta categoría por parte del modelo. Estos resultados se detallan en la Tabla 2.

La Tabla 2 presenta la matriz de confusión correspondiente al proceso de validación, con las siguientes asignaciones y precisiones por clase:

- **Clase café:** 15 casos correctamente clasificados, con 8 casos clasificados erróneamente como bosque, 1 como pasto y 1 como Infraestructura, resultando en una precisión del 50 %.
- **Clase bosque:** 18 casos correctamente clasificados, con 14 casos clasificados erróneamente como café y 4 como pasto, resultando en una precisión del 50 %.
- **Clase infraestructura:** 11 casos correctamente clasificados, con 3 casos clasificados erróneamente como café, 2 como bosque y 2 como pasto, resultando en una precisión del 61 %.
- **Clase pasto:** 13 casos correctamente clasificados, sin errores de clasificación, resultando en una precisión del 93 %.

Tabla 2: Matriz de confusión para el proceso de validación

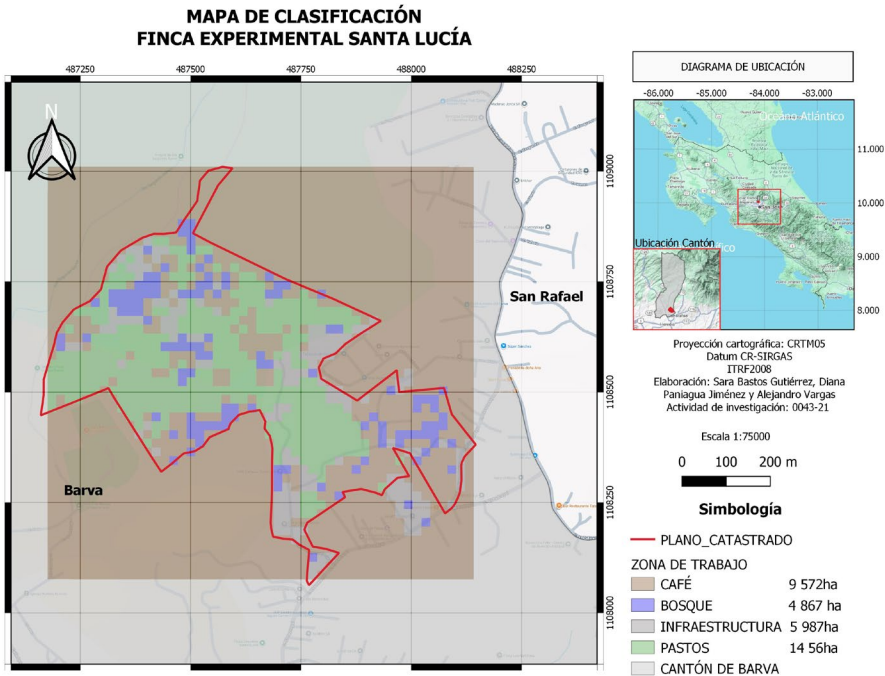
Clases	Café	Bosque	Pasto	Infraestructura	Precisión clase
Café	15	8	1	1	50 %
Bosque	14	18	4	0	50 %
Infraestructura	3	2	11	2	61 %
Pasto	1	0	0	13	93 %

Fuente: Elaboración propia.

En la Figura 4 se muestra el mapa de clasificación de las distintas clases, basado en el plano de la FESL, lo cual permitió definir, con precisión, las áreas de la finca en estudio. Este análisis facilitó contabilizar cada clase y la determinación de la superficie correspondiente en hectáreas a café, bosque, infraestructura y pastos.

En detalle, la clase de café abarca un total de 9 572 hectáreas, lo que representa una porción significativa del área total. La clase de bosque ocupa 4 867 hectáreas, destacando la importancia de estas áreas forestales dentro de la finca. La infraestructura, que incluye caminos, edificaciones y otras construcciones, se extiende a lo largo de 5 987 hectáreas. Por último, la clase de pastos, utilizada principalmente para la ganadería y otras actividades agrícolas, comprende 14 56 hectáreas.

Figura 4: Mapa de clasificación por clases



Fuente: Elaboración propia.

DISCUSIÓN DE RESULTADOS

Para mejorar la precisión en la colecta de muestras, se recomienda el uso de puntos de referencia marcados físicamente en el campo o área de estudio. Estos puntos deben ser representativos de las diversas clases de interés, tales como cultivos, pastos, bosques e infraestructuras. Esta estrategia asegura que la base de datos de entrenamiento sea lo más representativa y diversa posible.

Se recomienda segmentar la colección de imágenes en períodos más cortos o definidos, especialmente en zonas donde el uso del suelo presenta una dinámica temporal significativa. Esta estrategia permite detectar, con mayor precisión, los cambios estacionales o transitorios en la cobertura terrestre, lo que favorece una clasificación más detallada y actualizada. Asimismo, la incorporación de un análisis multitemporal enriquecería el

diagnóstico territorial al capturar la variabilidad fenológica de los cultivos y otras transiciones sutiles del paisaje, aumentando así la sensibilidad del modelo frente a transformaciones graduales o episódicas (Khan., et al, 2024).

Las limitaciones de resolución en las imágenes del satélite Sentinel-2 pueden complicar la separación de puntos de muestra en áreas con superposición de diferentes clases dentro de una misma imagen. Esta superposición puede comprometer la precisión de la selección de muestras representativas para cada clase y, por ende, influir negativamente en la exactitud general de la clasificación realizada con algoritmos como Random Forest.

Además de las limitaciones derivadas de la resolución espacial, es importante considerar que el rendimiento del modelo depende de una base de datos de entrenamiento equilibrada. La presencia de clases desbalanceadas o poco representadas puede inducir sesgos en el modelo de clasificación (Khan., et al, 2024). Asimismo, Random Forest puede mostrar una disminución en el rendimiento cuando se aplican a áreas con alta heterogeneidad interna o coberturas espectralmente similares (Mroczek & Wójcik, 2023).

Por otro lado, los resultados obtenidos subrayan la eficiencia del modelo de clasificación, especialmente en la distinción entre las clases, con una tasa de errores mínima. Sin embargo, se observó una mayor confusión entre las clases café y bosque, lo que sugiere una similitud en las características utilizadas para la clasificación y señala áreas potenciales para la mejora del modelo. Estos resultados reafirman la robustez del método aplicado, aunque también destacan la importancia de continuar optimizando los parámetros y las características para reducir aún más los errores de clasificación.

Una forma de mejorar la diferenciación entre clases vegetales con reflectancias similares, como el café y el bosque, es incorporar índices espectrales como el NDVI. Este índice ha demostrado ser eficaz para capturar variaciones en el vigor de la vegetación y puede ser útil como variable adicional en la clasificación supervisada (EOS Data Analytics, 2025).

La variabilidad en las precisiones por clase resalta la necesidad de mejorar las técnicas de clasificación y selección de características, especialmente para mejorar la diferenciación entre las clases de café y bosque. Algunas categorías podrían beneficiarse de ajustes específicos en

el modelo, lo que contribuiría a reducir la tasa de error global y aumentar la precisión de la clasificación.

Además, se sugiere considerar otros métodos de clasificación supervisada disponibles en la plataforma GEE que podrían evaluarse en investigaciones futuras, tales como Support Vector Machines (SVM) o k-Nearest Neighbors (k-NN), los cuales han demostrado resultados prometedores en diversos estudios recientes sobre clasificación de cobertura del suelo (Tan., et al, 2024; Khan., et al, 2024; Mroczek & Wójcik, 2023; Lefulebe & Musakwa, 2023; Kumar & Singh, 2023).

La incorporación de índices espectrales, como el NDVI, puede mejorar la diferenciación entre clases de cobertura vegetal, proporcionando una herramienta adicional para mejorar la precisión de la clasificación (EOS Data Analytics, 2025).

Por último, los datos generados en esta investigación, en particular los mapas de clasificación y la distribución detallada de las áreas en hectáreas, constituyen una herramienta esencial para la gestión y planificación eficiente del uso del suelo. Esta información facilita la toma de decisiones informadas en torno a la conservación de áreas forestales, la expansión racional de cultivos y la organización de infraestructura productiva. Los productos cartográficos obtenidos pueden ser utilizados como insumos estratégicos para la formulación de planes agrícolas, la delimitación de zonas prioritarias para la conservación y la evaluación del uso actual del territorio en fincas con características similares. Además, su aplicación puede extenderse a procesos de capacitación de productores y técnicos, así como al diseño de estrategias de manejo sostenible del paisaje (Lefulebe & Musakwa, 2023).

REFERENCIAS

- Alonso, J. C. y Hoyos, C. C. (2025). Una introducción a los modelos de Clasificación empleando R. Cali: Editorial Universidad Icesi. DOI: <https://doi.org/10.18046/EUI/bda.h.5>.
- Bontemps, S.; Arias, M.; Cara, C.; Dedieu, G.; Guzzonato, E.; Hagolle, O.; Inglada, J.; Matton, N.; Morin, D.; Popescu, R. J. R. S. (2015). Building a data set over 12 globally distributed sites to support the development of agriculture monitoring applications with Sentinel-2. *Remote Sensing*, 7, 16062-16090.
- Bruzzone, L.; Bovolo, F.; Paris, C.; Solano-Correa, Y.T.; Zanetti, M.; Fernández-Prieto. (2017) D. *Analysis of multitemporal Sentinel-2 images in the framework of the ESA Scientific Exploitation of Operational Missions*. In Proceedings of the 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp), Brugge, Belgium, 27-29; pp. 1-4.
- Chaney, M., Kamenova, I., Dimitrov, P., & Filchev, L. (2025). Evaluation of Sentinel-2 Deep Resolution 3.0 Data for Winter Crop Identification and Organic Barley Yield Prediction. *Remote Sensing*, 17(6), 957. <https://doi.org/10.3390/rs17060957>.
- EOS Data Analytics. (2025). *Índices de vegetación en la agricultura digital*. Recuperado de <https://eos.com/es/blog/indices-de-vegetacion/>.
- FAO. (2019). *The State of Food and Agriculture 2019: Moving forward on food loss and waste reduction*. FAO. Recuperado de: <https://www.fao.org/publications/home/fao-flagship-publications/the-state-of-food-and-agriculture/2019/en>.
- Godfray, H. C. J., Beddington, J., Crute, I., Haddad, L., Lawrence, D., Muir, J., Pretty, J., Robinson, S., Thomas, S., Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. *Science*, 327(5967), 812-818.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27. <https://doi.org/10.1016/j.rse.2017.06.031>.

- Hernández, H. J. I., Martínez, F. A., Torres, R. A. A., Orozco, S. A. P., Torres, L. R., Cienfuegos, M. F., Cerón, P. J. L. (2023). *Desarrollo e implementación de clasificadores de imágenes mediante el empleo de algoritmos de aprendizaje por transferencia utilizando Labview, Python y Google Teachable Machine*. Doi: [10.59920/JBSW8482](https://doi.org/10.59920/JBSW8482).
- IPCC. (2019). *Special Report on Climate Change and Land*. Intergovernmental Panel on Climate Change. Recuperado de: <https://www.ipcc.ch/srccl/>.
- Khan, M. A., Khan, S., & Khan, M. A. (2024). Comparative study of multiple algorithms classification for land use and land cover change detection in Mardan. *Heliyon*, 10(5), e15849. <https://doi.org/10.1016/j.envc.2024.101069>.
- Khorram. S, van der Wiele. C, Koch. F, Nelson. S, Potts. M. (2016). *Principles of Applied Remote Sensing*. Springer. DOI [10.1007/978-3-319-22560-9](https://doi.org/10.1007/978-3-319-22560-9).
- Kumar, R., & Singh, A. (2023). Accurate classification of land use and land cover using a boundary-specific two-level learning approach based on SVM and kNN. *Journal of Environmental Management*, 328, 116949. <https://doi.org/10.1016/j.jenvman.2023.116949>.
- Lee, J., Kim, K., Lee, K. 2024. Multi-Sensor Image Classification Using the Random Forest Algorithm in Google Earth Engine with KOMP-SAT-3/5 and CAS500-1 Images. *Remote Sens*. 16, 4622. <https://doi.org/10.3390/rs16244622> .
- Lefulebe, L. M., & Musakwa, W. (2023). Classification of urban land use and land cover with k-nearest neighbour classifier in the city of Cape Town, South Africa – Cape Flats case study. *ISPRS Archives*, XLVIII-1/W2-2023, 967-974. <https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-967-2023>.
- Lottering, R., Peerbhay, K., & Adelabu, S. (2025). Remote Sensing Applications in Agricultural, Earth and Environmental Sciences. *Applied Sciences*, 15(8), 4537. <https://doi.org/10.3390/app15084537>.
- Mahapatra, D. (2014). Analyzing Training Information From Random Forests for Improved Image Segmentation. *IEEE Transactions on Image Processing*, 23(4), 1504-1512. doi:10.1109/tip.2014.2305073
- MI-NAE. (2020). *Sistema Nacional de Monitoreo de Cobertura y Uso de la Tierra y Ecosistemas*. Recuperado de: www.simocute.gov.cr.

- MINAE. (2020). *Diagnóstico de mapeo sobre cobertura y uso de la tierra y ecosistemas: Sistema Nacional de Monitoreo de Cobertura y Uso de la Tierra y Ecosistemas (SIMOCUTE)*. Ministerio de Ambiente y Energía. Recuperado de <http://www.simocute.go.cr>
- Mora Ramírez, S. (2018). *Desempeño Sector Agro*. Secretaría Ejecutiva de Planificación Sectorial Agropecuaria. http://www.sepsa.go.cr/docs/2018-007_Desempenno_SectorAgro_2017.pdf.
- Mora Ramírez, S. (2020, agosto). *Indicadores Macroeconómicos 2016-2020*. Secretaría Ejecutiva de Planificación Sectorial Agropecuaria. http://www.sepsa.go.cr/docs/2020-012-Indicadores_Macroeconomicos_2016-2020.pdf.
- Mroczek, P., & Wójcik, M. (2023). Support vector machine algorithm for mapping land cover types in Senegal. *Land*, 12(3), 24. <https://doi.org/10.3390/earth5030024>.
- Olaya, V. (2020) *Sistemas de Información Geográfica*. ISBN: 978-1-71677-766-0
- Perilla, A. & Mas, F. (2020). Google Earth Engine (GEE): una poderosa herramienta que vincula el potencial de los datos masivos y la eficacia del procesamiento en la nube. *Investigaciones Geográficas Instituto de Geografía*. UNAM SSN: 2448-7279 DOI: [dx.doi.org/10.14350/rig.59929](https://doi.org/10.14350/rig.59929). NOTA TÉCNICA Núm. 101 Abril 2020 E59929. www.investigacionesgeograficas.unam.mx.
- Phiri, D, Simwanda, M, Salekin, S, Nyirenda, V, Muruyama, Y, Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing*. 12(14), 2291. <https://doi.org/10.3390/rs12142291>.
- Rynkiewicz, A., Hościło, A., Aune-Lundberg, L., Nilsen, A. B., & Lewandowska, A. (2025). Detection and Quantification of Vegetation Losses with Sentinel-2 Images Using Bi-Temporal Analysis of Spectral Indices and Transferable Random Forest Model. *Remote Sensing*, 17(6), 979. <https://doi.org/10.3390/rs17060979>.
- Segarra, J., Buchailot, M. L., Araus, J. L., & Kefauver, S. C. (2020). *Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications*. *Agronomy*, 10(5), 641. <https://doi.org/10.3390/agronomy10050641>

- Shelestov, A., Lavreniuk, M., Kussul, N., & Skakun, S. (2017). Exploring Google Earth Engine platform for Big Data processing: classification of multi-temporal satellite imagery for crop mapping. *Frontiers in Earth Science*, 5, 17.
- Solórzano, J. V., & Perilla, G. A. (2022). *Cómo usar Google Earth Engine y no fallar en el intento*. Universidad Nacional Autónoma de México, Centro de Investigaciones en Geografía Ambiental. <https://doi.org/10.22201/ciga.9786073066969e.2022>
- Tan, Y.-C., Duarte, L., & Teodoro, A. C. (2024). Comparative study of random forest and support vector machine for land cover classification and post-wildfire change detection. *Land*, 13(11), 1878. <https://doi.org/10.3390/land13111878>.
- Wang, Q., Blackburn, G. A., Onojeghuo, A. O., Dash, J., Zhou, L., Zhang, Y., & Atkinson, P. M. (2017). *Fusion of Landsat 8 OLI and Sentinel-2 MSI data*. IEEE Transactions on Geoscience and Remote Sensing. <https://doi.org/10.1109/TGRS.2017.2683444>
- Xi, E. 2022 Image Classification and Recognition Based on Deep Learning and Random Forest Algorithm, *Wireless Communications and Mobile Computing*. 2013181, 9 pages, 2022. <https://doi.org/10.1155/2022/2013181>.
- Yang, Y., Wang, Y., & Li, X. (2024). Generating high-precision farmland irrigation pattern maps using Google Earth Engine. *Agricultural Water Management*, 287, 109302. <https://doi.org/10.1016/j.agwat.2025.109302>.