

Escuela Superior Técnica EST SENCICO

MONOGRAFÍA

**USO DE INFORMACIÓN SATELITAL EN
AGRICULTURA DE PRECISIÓN:
FUNDAMENTOS TEÓRICOS Y APLICACIÓN
PRÁCTICA EN LA CLASIFICACIÓN DE
CUERPOS AGRÍCOLAS**

Para acceder al título de Profesional Técnico en Geomática

Huatay Salcedo, Luis Elías
Sede Central
Geomática
S71809640

19 de noviembre de 2025

Dedico este trabajo a mi esposa, por su notable paciencia y confianza en mí. A mis padres porque su ausencia me dio la oportunidad de descubrir mi propia fortaleza. A mis profesores, pero en especial al Ing. Kevin Sanchez por su guía e inspiración brindada durante mi camino académico.

ÍNDICE

1 INTRODUCCIÓN	1
1.1 Problematización	2
1.2 Justificación	2
1.3 Objetivos generales y específicos	3
1.4 Alcances y limitaciones	3
2 CAPÍTULO I: FUNDAMENTOS TEÓRICOS	4
2.1 Fisiología vegetal y espectro electromagnético	4
2.2 Espectroscopía	5
2.2.1 Antecedentes históricos	5
2.2.2 Principios básicos	5
2.2.3 Técnicas espectroscópicas	6
2.3 Sensoramiento y teledetección satelital	7
2.3.1 Sensores: fundamentos y tipos	7
2.3.2 Teledetección satelital	8
2.4 Estructura de datos	8
2.5 Resolución	9
2.5.1 Resolución espacial	9
2.5.2 Resolución temporal	10
2.5.3 Resolución espectral	10
2.5.4 Resolución radiométrica	10
2.6 Satélites de teledetección	12
2.6.1 Sentinel-2	12
2.6.2 Landsat 8	13
2.7 Índices espectrales aplicados a agricultura	13
2.7.1 Fundamento matemático	14
2.7.2 Índices espectrales más utilizados en agricultura	16
3 CAPÍTULO II APLICACIÓN PRÁCTICA: CLASIFICACIÓN DE COBERTURA AGRÍCOLA	17
3.1 Flujo de trabajo	17
3.2 Materiales	18
3.3 Metodología	19
3.4 Procedimiento	20
3.4.1 Etiquetas de entrenamiento	20
3.4.2 Generación del dataset de entrenamiento	22
3.4.3 Generación del modelo de clasificación supervisada	23
3.4.4 Aplicación del modelo de clasificación supervisada	23
3.5 Resultados	24
3.6 Validación del modelo de clasificación supervisada	25
3.6.1 Matriz de confusión	26
3.6.2 Precisión General del modelo (overall accuracy)	26
3.6.3 Métricas de área calculada	27
3.7 Análisis y discusión de resultados	28
3.7.1 Clasificación en áreas cubiertas por nubes	28
3.7.2 Clasificación de áreas con vegetación natural	29
3.7.3 Cuerpos de agua	30
3.7.4 Comparación con otros productos satelitales	30
3.8 Conclusiones	32
3.9 Recomendaciones	33

1. INTRODUCCIÓN

Para tratar del uso de información satelital en la agricultura de precisión, es importante entender primero el contexto histórico y tecnológico que ha llevado a su desarrollo. La agricultura ha sido una actividad fundamental para la humanidad desde hace aproximadamente 12,000 años, cuando la mayoría de las poblaciones humanas pasó de la recolección a diversas formas de cultivo. (Wells & Stock, 2020) Hasta ese momento, las sociedades humanas dependían de la caza y la recolección para su subsistencia. La transición a la agricultura permitió el desarrollo de asentamientos permanentes y el crecimiento de las civilizaciones. Desde entonces, la agricultura ha evolucionado significativamente, pasando de técnicas rudimentarias a métodos más sofisticados y eficientes. La Revolución Agrícola, que tuvo lugar entre los siglos XVIII y XIX, marcó un punto de inflexión en la historia de esta práctica. Durante este período, se introdujeron nuevas tecnologías y prácticas agrícolas, como la rotación de cultivos, el uso de fertilizantes, la mecanización de las labores agrícolas o la sintetización de productos químicos basados en nitrógeno. Estos avances permitieron un aumento significativo en la productividad agrícola y sentaron las bases para la agricultura moderna.

Sin embargo, el desarrollo de mejores prácticas agrícolas que puedan optimizar el uso de recursos, mejorar el rendimiento y mitigar los impactos negativos de factores como las enfermedades, plagas, entre otros, ha sido un desafío constante para los agricultores a lo largo de dicho tiempo. El poder identificar estos problemas de manera temprana y precisa, además de convertirse en una ventaja competitiva en el mercado actual, se ha vuelto una necesidad imperante para garantizar la seguridad alimentaria pero llega a ser una labor compleja y más, cuando se trata de grandes extensiones de terreno del orden de miles de hectáreas. En este contexto, la agricultura de precisión ha surgido como una solución innovadora, creativa y altamente tecnológica para abordar estos desafíos. La agricultura de precisión, según (Vullaganti et al., 2025), es una práctica que utiliza tecnologías avanzadas y enfoques basados en datos para ayudar en la toma de decisiones y optimizar la producción de cultivos. Dentro de estas tecnologías, la teledetección satelital ha ganado una importancia significativa debido a su capacidad para proporcionar información detallada y en tiempo real sobre las condiciones del suelo, el estado de los cultivos y otros factores relevantes para la gestión agrícola.

La teledetección satelital es el producto de la convergencia de muchas disciplinas científicas y tecnológicas que incluyen a la física, la ingeniería, la informática y las ciencias de la tierra. Desde Newton y su estudio sobre la luz, óptica, el espectro electromagnético y gravitación universal hasta la invención de los primeros satélites artificiales en la década de 1950, la humanidad ha avanzado significativamente en su capacidad para observar y comprender el planeta desde el espacio. La teledetección satelital implica la adquisición y análisis de datos obtenidos por sensores montados en satélites que orbitan la Tierra. Estos sensores pueden captar información en diferentes longitudes de onda del espectro electromagnético, lo que permite obtener una visión integral de las condiciones terrestres. Además, la conjunción entre la teledetección satelital, sistemas de información geográfica (SIG) y el impulso que ofrece la inteligencia artificial (IA) ha abierto nuevas posibilidades para el análisis y la interpretación de datos geoespaciales, facilitando la toma de decisiones informadas en la agricultura de precisión.

En el presente documento, se explorará el uso de información satelital en la agricultura de precisión, destacando sus beneficios, limitaciones y aplicaciones prácticas así como los fundamentos teóricos que sustentan su uso. Se analizarán diferentes tipos de sensores, las técnicas de procesamiento de datos y los índices vegetativos más relevantes para la agricultura. Además de la generación de un modelo de clasificación de cobertura agrícola que pueda ser útil a nivel nacional. Finalmente, se discutirán las perspectivas futuras de esta tecnología y su potencial para transformar la agricultura en los próximos años.

1.1. Problemática

La agricultura enfrenta numerosos desafíos en la actualidad, incluyendo el cambio climático, la degradación del suelo, la escasez de agua y la necesidad de aumentar la producción para satisfacer la demanda creciente de alimentos. Estos factores afectan directamente el estado de salud de los cultivos y, por ende, la productividad agrícola. La identificación temprana y precisa de problemas en los cultivos es crucial para implementar medidas correctivas que minimicen las pérdidas y optimicen el uso de recursos. Sin embargo, la monitorización tradicional de los cultivos a través de inspecciones visuales y muestreos de campo puede ser limitada en términos de cobertura espacial y temporal, especialmente en grandes extensiones agrícolas. El uso de información satelital en la agricultura de precisión ofrece una alternativa prometedora para superar las limitaciones de los métodos tradicionales, no obstante, es necesario comprender con claridad los fundamentos, alcances y limitaciones de estas tecnologías para aplicarlas adecuadamente en la clasificación de cultivos.

- ¿Cuáles son los fundamentos teóricos que sustentan el uso de información satelital en la agricultura de precisión?
- ¿Qué tipos de sensores satelitales y técnicas de procesamiento de datos son más adecuados para la monitorización de cultivos?
- ¿Cuáles son los productos derivados de la teledetección satelital más relevantes para la agricultura de precisión?
- ¿Cómo se puede aplicar la información satelital para la clasificación de cobertura agrícola en una región específica?
- ¿Cuáles son las limitaciones y desafíos asociados con el uso de información satelital en la agricultura de precisión?
- ¿Cuáles son las perspectivas futuras para la integración de tecnologías emergentes, como la inteligencia artificial, en la agricultura de precisión basada en información satelital?

1.2. Justificación

La relevancia de este estudio radica en la necesidad de optimizar la producción agrícola mediante el uso de tecnologías avanzadas que permitan una gestión más eficiente y sostenible de los recursos.

- **Relevancia científica:** La teledetección satelital es una disciplina en constante evolución que integra nuevas tecnologías, metodologías basadas en datos, enfoques interdisciplinarios y avances en sensores, plataformas satelitales, algoritmos de procesamiento y análisis de estos datos y modelos predictivos. Comprender los fundamentos teóricos que sustentan el uso de información satelital en la agricultura de precisión es esencial para aprovechar al máximo su potencial y contribuir al avance del conocimiento en este campo.
- **Relevancia práctica:** La aplicación de información satelital en la agricultura de precisión tiene un impacto directo en la gestión agrícola. Permite a los agricultores tomar decisiones informadas sobre el manejo de cultivos, la aplicación de insumos y la planificación de actividades agrícolas. Esto puede traducirse en una mayor eficiencia en el uso de recursos, reducción de costos y mejora del rendimiento agrícola.
- **Relevancia en la toma de decisiones:** La información satelital proporciona datos en tiempo real que pueden ser utilizados para una mejor política gubernamental en lo que respecta a la toma de decisiones estratégicas en la agricultura. Esto es especialmente relevante en un contexto de cambio climático y variabilidad climática, donde la capacidad de adaptarse rápidamente a las condiciones cambiantes es crucial para la seguridad alimentaria.

1.3. Objetivos generales y específicos

El objetivo general de este estudio es entender y aplicar los fundamentos teóricos del uso de información satelital en la agricultura de precisión para la generación de un modelo de clasificación supervisada de cobertura agrícola en una región específica y luego su uso en distintas regiones. Con ello se pretende contribuir al desarrollo de prácticas agrícolas más eficientes y sostenibles. De este modo se pueden desglosar algunos objetivos específicos que incluyen:

- Explicar los fundamentos teóricos del uso de información satelital en la agricultura de precisión.
- Conocer los productos derivados de la teledetección satelital más relevantes para la agricultura de precisión.
- Conocer las herramientas tecnológicas disponibles para la clasificación supervisada de cobertura agrícola.
- Desarrollar una aplicación práctica de clasificación de cobertura agrícola haciendo uso de alguna de estas herramientas.
- Identificar las limitaciones y desafíos asociados con el uso de información satelital en la agricultura de precisión.
- Explorar las perspectivas futuras para la integración de tecnologías emergentes, como la inteligencia artificial, en la agricultura de precisión basada en información satelital.

1.4. Alcances y limitaciones

Dentro de este estudio, se establecen ciertos alcances y limitaciones que definen el contexto y la aplicabilidad de los resultados obtenidos.

Alcances:

- El estudio se centrará en la generación de un modelo de clasificación que pueda ser útil a nivel nacional.
- Se utilizarán imágenes satelitales de libre acceso, como lo son las imágenes Sentinel-2, Landsat 8, MODIS y AlphaEarth Foundations.
- Se emplearán herramientas tecnológicas disponibles para la clasificación supervisada de cobertura agrícola.

Limitaciones:

- La disponibilidad y calidad de las imágenes satelitales pueden verse afectadas por condiciones climáticas y otros factores.
- La clasificación de cobertura agrícola puede verse limitada por la resolución espacial y temporal de los datos satelitales.
- La implementación de tecnologías emergentes, como la inteligencia artificial, puede requerir recursos computacionales significativos y acceso a datos de alta calidad.

2. CAPÍTULO I: FUNDAMENTOS TEÓRICOS

Para entender cómo la información satelital puede ser utilizada en la agricultura de precisión, es fundamental revisar los conceptos teóricos que sustentan esta tecnología y su aplicación en el ámbito agrícola. Es importante conocer la problemática que resuelve, de donde y cómo provienen los datos que toma y la vía por el cual se procesan para obtener información útil.

2.1. Fisiología vegetal y espectro electromagnético

La fisiología vegetal es la rama de la biología que estudia las funciones y procesos vitales de las plantas. Comprender cómo las plantas interactúan con la luz y otros factores ambientales es crucial para el uso efectivo de la información satelital en la agricultura de precisión. Las plantas por medio del proceso de la fotosíntesis convierten la energía lumínica en energía química, utilizando dióxido de carbono y agua para producir glucosa y oxígeno. La clorofila, el pigmento verde en las hojas, es responsable de absorber la luz, según (Bruning et al., 2020), las plantas absorben principalmente la luz en las regiones azul (400-500 nm) y roja (600-700 nm) del espectro electromagnético, mientras que reflejan la luz en la región verde (500-600 nm), lo que explica su color característico. Además, las plantas reflejan fuertemente la luz en el rango del infrarrojo cercano (700-1300 nm).

Rango	Factores	Interacción	Detalles
Visible (VIS) 350 – 700 nm	Pigmentos fotosintéticos: clorofila, carotenoides	absorción > reflexión	Picos de absorción de clorofila en 420, 470 y 660 nm. La fuerte absorción en azul y rojo genera un pico de reflectancia en el verde (550 nm), responsable del color característico de las plantas. Los carotenoides absorben entre 440–480 nm.
Infrarrojo cercano (NIR) 700 – 1100 nm	Estructura celular interna: paredes celulares, espacios de aire, contenido de agua y grado de desarrollo foliar.	absorción < reflexión	La transición de baja reflectancia en el rojo a alta en el NIR se denomina “borde rojo” (red edge), que ocurre entre 680–720 nm. Existen bandas de absorción por agua en 970 y 1200 nm.
Infrarrojo de onda corta (SWIR) 1100 – 2500 nm	Contenido de agua y bioquímica de la hoja: proteínas, celulosa, lignina, almidón.	Absorción dominada por el agua.	Bandas de absorción de agua en 1450 y 1940 nm. En vegetación seca, al reducirse el agua, se hacen visibles las absorciones de lignina, celulosa y almidón.

Cuadro 1: Interacción de la luz con la vegetación en diferentes rangos del espectro electromagnético.
Adaptado de (Bruning et al., 2020).

Con ello podemos entender que una forma de poder monitorear el estado de las plantas es a través del análisis de la luz que reflejan en diferentes longitudes de onda. Considerando que una planta saludable cumplirá con el proceso de fotosíntesis de manera óptima, mientras que una planta estresada o enferma mostrará alteraciones en su capacidad para absorber y reflejar la luz, podremos utilizar esta información para evaluar la salud y el vigor de los cultivos. Esta metodología es la base para muchas aplicaciones de teledetección en la agricultura de precisión. Teniendo en cuenta que las plantas reflejan fuertemente en el infrarrojo cercano (NIR) además de ser un indicador de contenido de agua y vigor fotosintético, los sensores satelitales que capturan esta banda son especialmente útiles para monitorear la vegetación.

De acuerdo a ello, a continuación se muestran las figuras comparativas:

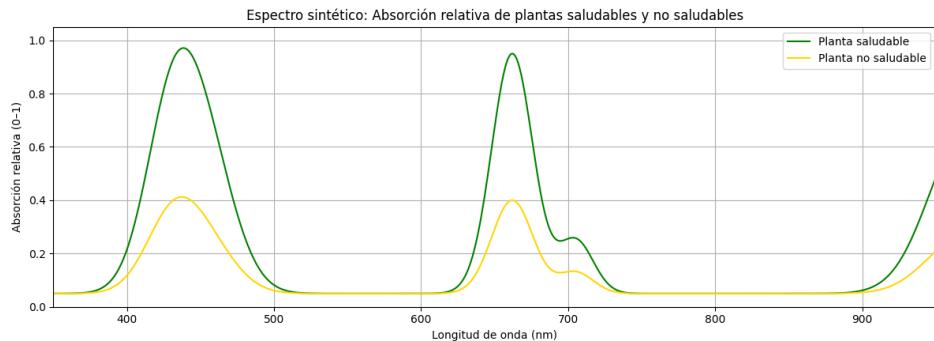


Figura 1: Curva típica de reflectanciapectral de la vegetación, mostrando la absorción en azul y rojo, el pico de reflectancia en verde y el aumento significativo en el infrarrojo cercano (NIR). Datos generales e ilustrativos. Fuente: (Bruning et al., 2020).

Una vez entendemos lo que observamos cuando vemos una planta, es importante conocer cómo aprovechar esta información, cómo tomamos los datos, y cómo los procesamos para obtener información útil.

2.2. Espectroscopía

Visto ya la interacción de la luz con la vegetación, es importante conocer el cómo podríamos medir esa luz reflejada para poder obtener los datos sobre los cuales analizaremos el estado de los cultivos. Esta es una cuestión que se aborda desde la espectroscopía, ciencia que tiene como objetivo estudiar la interacción entre la radiación electromagnética y la materia. (Xia et al., 2023) cataloga a la espectroscopía como una técnica establecida y eficaz utilizada ampliamente en las últimas décadas debido a sus excelentes resultados en la detección de la calidad agrícola. Técnicas como la espectroscopía visible y de infrarrojo cercano (Vis-NIR) y las imágenes hiperespectrales (HSI) son herramientas maduras y eficientes para la caracterización física y química de productos agrícolas sin destruirlos.

2.2.1. Antecedentes históricos

El desarrollo de la espectroscopía se remonta a varios siglos atrás. En el siglo XVII, Isaac Newton fue uno de los primeros en estudiar la dispersión de la luz a través de un prisma, lo que llevó al descubrimiento del espectro visible. Posteriormente, en el siglo XIX, científicos como Joseph von Fraunhofer y Gustav Kirchhoff realizaron importantes contribuciones al campo al analizar las líneas espectrales emitidas por diferentes elementos químicos. El avance de la espectroscopía continuó con el desarrollo de técnicas como la espectroscopía infrarroja y ultravioleta, que permitieron estudiar una gama más amplia del espectro electromagnético. En el siglo XX, la invención de instrumentos más sofisticados, como los especlómetros y los detectores electrónicos, facilitó mediciones más precisas y rápidas, ampliando aún más las aplicaciones de la espectroscopía en diversas disciplinas científicas e industriales.

2.2.2. Principios básicos

La espectroscopía se basa en la interacción de la radiación electromagnética con la materia. Cuando la luz incide sobre un objeto, parte de ella es absorbida, reflejada o transmitida. La cantidad de luz absorbida y reflejada varía según la composición química y las propiedades físicas del material. Al analizar el espectro de luz reflejada por un objeto, es posible obtener información sobre su composición y estado. Se utilizan diferentes longitudes de onda para investigar características específicas de los materiales, lo que permite identificar compuestos químicos, determinar concentraciones y evaluar propiedades físicas como la textura y la estructura, entre otras.

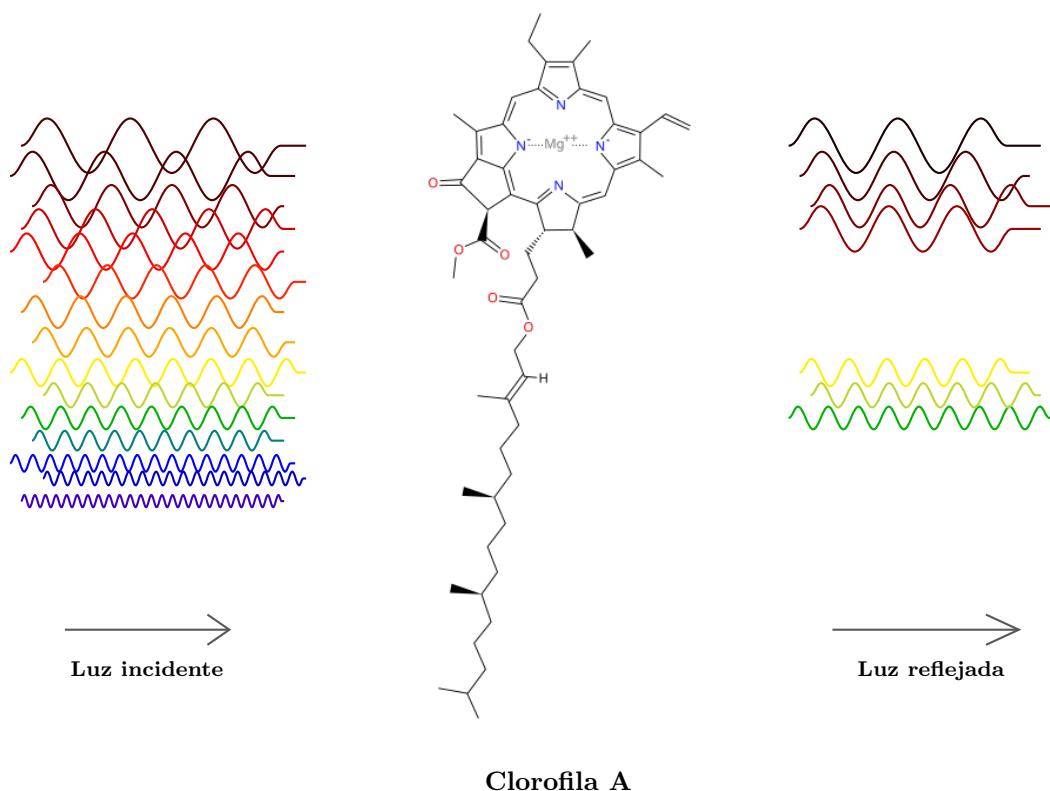


Figura 2: Interacción de la luz con una molécula de clorofila-A donde se observa cómo absorbe específicas longitudes de ondas y refleja otras. Fuente: elaboración propia.

2.2.3. Técnicas espectroscópicas

Existen diversas técnicas espectroscópicas, cada una con sus propias aplicaciones y ventajas. Algunas de las técnicas más comunes incluyen:

- **Espectroscopía de absorción:** Mide la cantidad de luz absorbida por un material en función de la longitud de onda. Es útil para identificar compuestos químicos y determinar concentraciones.
- **Espectroscopía de emisión:** Analiza la luz emitida por un material cuando es excitado por una fuente externa. Se utiliza para estudiar la composición elemental y las propiedades electrónicas.
- **Espectroscopía de reflectancia:** Examina la luz reflejada por una superficie. Es especialmente relevante en aplicaciones agrícolas, ya que permite evaluar la salud y el vigor de las plantas mediante el análisis de la luz reflejada en diferentes longitudes de onda.

Entendido estos conceptos, se puede comprender que al analizar la luz que refleja determinado cultivo en diferentes longitudes de onda, podemos obtener información valiosa sobre su estado de salud, contenido de agua y otros parámetros importantes para la agricultura de precisión, sin embargo, en la práctica no basta con analizar una sola planta, sino que se requiere analizar grandes extensiones de terreno, para lo cual es necesario poder resolver la limitación espacial que tienen los sensores espectroscópicos tradicionales, para lo cual se recurre a la teledetección satelital.

2.3. Sensoramiento y teledetección satelital

El análisis de la luz reflejada por las plantas a través de técnicas espectroscópicas es una solución efectiva para evaluar la salud y el vigor de determinado cultivo, entendido esto podríamos preguntarnos cómo obtener los datos de esta luz reflejada, si bien, los ojos humanos pueden percibir la luz visible, no son capaces de detectar otras longitudes de onda como el infrarrojo cercano (NIR), que es crucial para evaluar lo que se espera de un cultivo. Es por esta necesidad que el avance tecnológico ha permitido el desarrollo de los sensores ópticos, dispositivos diseñados para medir la radiación electromagnética en diferentes rangos del espectro, incluyendo los que no son visibles para el ojo humano. Según (Berger et al., 2022), el análisis de 96 estudios de investigación en lo que respecta a la teledetección, reveló que la combinación de sensores que cubren los dominios Visible (VIS), Infrarrojo Cercano (NIR) e Infrarrojo Térmico (TIR) fue la categoría más grande y comúnmente utilizada.

2.3.1. Sensores: fundamentos y tipos

Un sensor es un dispositivo que detecta y mide una propiedad física o química específica y la convierte en una señal que puede ser interpretada. En el contexto de los sensores utilizados en la teledetección se tienen pasivos o activos.

- **Sensores pasivos:** Estos sensores detectan la radiación electromagnética que es emitida o reflejada por los objetos en la superficie terrestre. No emiten energía propia, sino que dependen de fuentes externas de energía, como la luz solar. Un ejemplo común de sensor pasivo es una cámara fotográfica o un espectrómetro que mide la luz reflejada por algún material.
- **Sensores activos:** Estos sensores emiten su propia energía y pueden medir radiación reflejada, tiempo de captura y otros parámetros. Un ejemplo típico de sensor activo es el radar, que emite ondas de radio y mide el tiempo que tarda en regresar la señal reflejada desde la superficie terrestre.

Dentro del ámbito del sensoramiento dedicado al monitoreo agrícola, los sensores pasivos son comúnmente utilizados, ya que permiten capturar la luz reflejada por las plantas en diferentes longitudes de onda, proporcionando información valiosa sobre su estado como se explicó anteriormente.

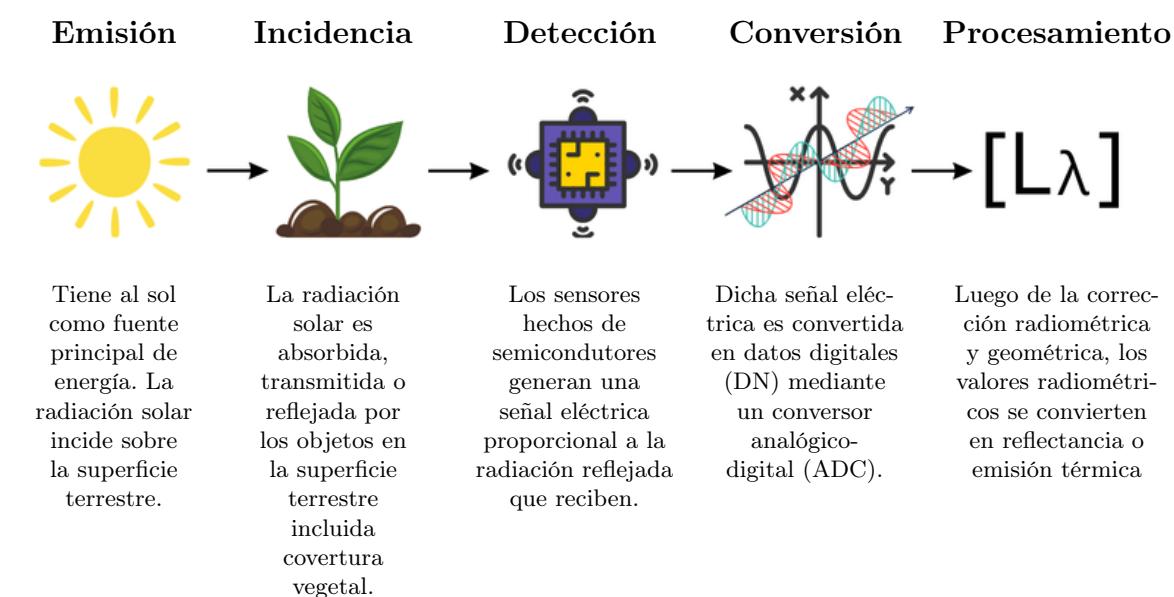


Figura 3: Flujo de trabajo general para la adquisición y procesamiento de datos mediante sensores pasivos en la detección del espectro electromagnético. Fuente: Elaboración propia.

2.3.2. Teledetección satelital

La necesidad de monitorear grandes extensiones de terreno, como campos agrícolas, ha llevado al desarrollo de la teledetección, que es la adquisición de información sobre un objeto o área sin estar en contacto directo con él. Originalmente, según (Ram et al., 2024), la imagen hiperespectral (HSI) se utilizaba predominantemente con plataformas orbitales (satélites) y suborbitales (aviones). El procesamiento de estos datos se centraba principalmente en correcciones atmosféricas y ortorectificación. Sin embargo, en los últimos veinte años, ha habido un aumento en el uso de variantes portátiles y de mano de estos sensores. Esta tendencia indica un cambio positivo desde las aplicaciones geoespaciales hacia las aplicaciones en el campo (in-field). Estos nuevos sensores, montados en vehículos terrestres (UGV), vehículos aéreos no tripulados (UAV) o usados manualmente, ofrecen una alta resolución espacial y temporal.

La teledetección utiliza una variedad de plataformas para llevar los sensores a diferentes altitudes y obtener datos desde diversas perspectivas. Estas plataformas incluyen satélites, aviones, drones y vehículos terrestres. Cada plataforma tiene sus propias ventajas y limitaciones en términos de cobertura, resolución y costos operativos. Con esta metodología de toma de datos, los resultados obtenidos por sensores simples se vuelven mucho más útiles al generar ya no arreglos de datos de una sola medición, sino que se obtienen estructuras matriciales donde cada elemento de la matriz representa un píxel, generando así una imagen con información espacial, temporal, espectral y radiométrica.

A continuación se muestra el proceso de captura de datos mediante teledetección satelital:

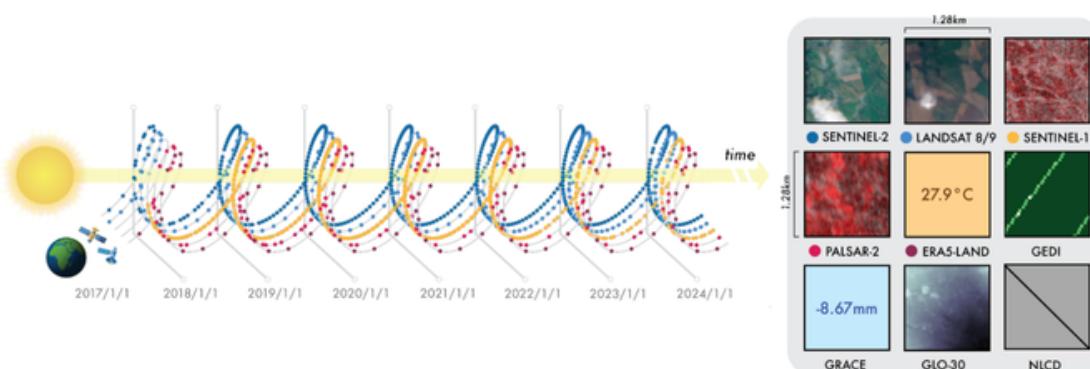


Figura 4: Dinámica del movimiento del planeta al rededor del Sol y los satélites artificiales de teledetección (izquierda) muestra tanto la fuente de emisión como la superficie de reflectancia que finalmente es capturada por los sensores de distintos tipos, este es el proceso por el cual se consiguen los diferentes productos derivados (derecha). Fuente: (Brown et al., 2025).

2.4. Estructura de datos

El objeto de medición de los sensores enfocados a teledetección y que están montados en satélites es la radiación electromagnética que es reflejada por la superficie terrestre, la cual es capturada en específicas longitudes de onda. A fin de poder manipular computacionalmente estos datos, la conversión de señal eléctrica a datos digitales (DN) mediante un conversor analógico-digital (ADC) genera un valor numérico en base 2 que representa la intensidad de la radiación detectada en una banda espectral específica. La capacidad de un sensor para discriminar entre niveles de intensidad se denomina resolución radiométrica, la cual está determinada por el número de bits utilizados en el ADC. Por ejemplo, un sensor con una resolución radiométrica de 8 bits puede representar 256 niveles diferentes de intensidad $2^8 = 256$, mientras que un sensor con una resolución de 16 bits puede representar 65,536 niveles $2^{16} = 65,536$. Una mayor resolución radiométrica permite detectar diferencias más sutiles en la intensidad de la radiación, lo que puede ser crucial para aplicaciones que requieren un análisis detallado de la reflectancia o emisión térmica. (Aburaed et al., 2023) menciona que la mayoría de los sensores satelitales comerciales tienen una resolución radiométrica de 8 bits, aunque algunos sensores avanzados pueden ofrecer resoluciones de hasta 12 o 16 bits.

Estos valores digitales (DN) son organizados en una estructura matricial donde cada elemento de la matriz representa el nivel de intensidad de radiación detectada en una ubicación espacial específica, esta forma de organizar los datos es particularmente útil porque puede guardar información espacial y más importante, puede ser representada visualmente como una imagen, donde cada píxel de la imagen corresponde a un elemento de la matriz. Esto facilita la interpretación y análisis de los datos, permitiendo identificar patrones espaciales y variaciones en la reflectancia o emisión térmica a lo largo de la superficie terrestre.

A continuación se muestra el proceso de recolección de datos y su estructura matricial:

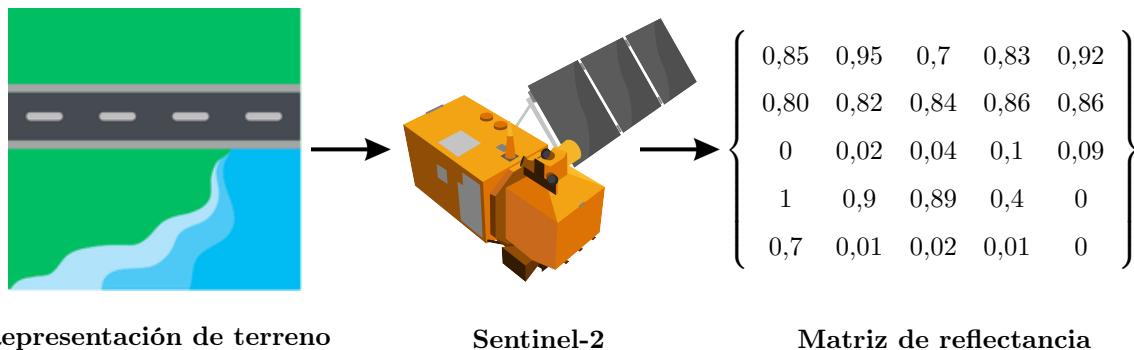


Figura 5: Proceso de captura de datos mediante teledetección satelital y su estructura matricial como producto. Nótese como la matriz de datos representa en valores de 0 a 1 el grado de energía reflejada por la superficie vegetal, esto gracias a los fundamentos de la fisiología vegetal y la espectroscopía. Fuente: elaboración propia.

2.5. Resolución

Los fundamentos teóricos de toma de datos mediante teledetección satelital se basan en la captura de energía electromagnética reflejada, como si de una cámara fotográfica se tratase, en ese sentido, existen ciertas propiedades que se comparten, entre ellas la resolución. La resolución se refiere a la capacidad de conseguir detalle en una toma de datos, especialmente en lo que respecta a sensores de alta precisión montados en plataformas satelitales, esta resolución se da en el espacio de captura, en el tiempo, la intensidad de radiación o en el espectro electromagnético. Para entender mejor estos conceptos, a continuación se describen los tipos de resolución que son relevantes en el contexto de la teledetección satelital enfocada al monitoreo agrícola:

2.5.1. Resolución espacial

Según (Kupidura & Lesisz, 2022), la resolución espacial en una imagen satelital se describe como un factor determinante para entender la calidad de visibilidad de los objetos en la superficie terrestre. Esto se relaciona profundamente con el tamaño de pixel que se presenta como una medida directa de resolución espacial.

La resolución espacial se refiere al tamaño del área en la superficie terrestre que es representada por un solo píxel en una imagen satelital. Una resolución espacial más alta significa que cada píxel cubre un área más pequeña, lo que permite distinguir detalles más finos en la imagen mientras que una resolución espacial más baja significa que cada píxel cubre un área más grande, lo que puede resultar en una pérdida de detalle. Sin embargo, una resolución espacial baja no necesariamente implica la pérdida de información, en todo ámbito esto depende de la aplicación específica y los objetivos del análisis. Por ejemplo, si se quisiera monitorear cambios meteorológicos a gran escala, una resolución espacial baja podría ser suficiente e inclusive, conveniente debido a la necesidad de cubrir grandes áreas, mientras que para aplicaciones agrícolas que requieren un análisis detallado de cultivos específicos, una resolución espacial alta sería más adecuada.

2.5.2. Resolución temporal

En líneas generales, la resolución temporal en una imagen satelital se refiere a la frecuencia teórica con la que un satélite pasa sobre el mismo lugar. Por ejemplo, Sentinel-2 tiene un tiempo de revisita de 5 días en el área de estudio mientras que Landsat 8 tiene un tiempo de revisita de 16 días. Según (Gao et al., 2024), la resolución temporal es crucial para monitorear cambios dinámicos en la superficie terrestre concluyendo que para aplicaciones agrícolas una resolución temporal no mayor a 10 días es adecuada. La muestra de datos con alta resolución temporal permite capturar cambios rápidos en la superficie terrestre, como el crecimiento de cultivos, la detección de plagas o enfermedades, y la respuesta a eventos climáticos. Esto es especialmente importante en la agricultura de precisión donde las condiciones pueden cambiar rápidamente y requieren una respuesta oportuna. Algunos conceptos importantes sobre la resolución temporal incluyen:

- **Tiempo de revisita del satélite:** Es el intervalo de tiempo entre dos pasadas consecutivas del satélite sobre la misma área. Un tiempo de revisita más corto permite obtener datos más frecuentes.
- **Frecuencia de observaciones utilizables:** No todas las observaciones son útiles debido a factores como la cobertura de nubes. La frecuencia de observaciones utilizables suele ser menor que el tiempo de revisita debido a estas limitaciones.

2.5.3. Resolución espectral

Refiere a la capacidad de un sensor para discriminar entre diferentes longitudes de onda del espectro electromagnético dispuestas en distintas bandas espectrales. Una mayor resolución espectral permite capturar información bioquímica más detallada sobre la vegetación y otros materiales en la superficie terrestre. La resolución espectral dentro de una imagen satelital puede presentarse como “stack” de imágenes (bandas) que representan diferentes rangos del espectro electromagnético, como el visible, el infrarrojo cercano (NIR) y el infrarrojo de onda corta (SWIR). Cada banda captura información específica sobre cómo los objetos en la superficie terrestre interactúan con la luz en esas longitudes de onda particulares. Al combinar diferentes bandas espectrales, es posible obtener una visión más completa o un enfoque específico al momento de analizar la información. Por ejemplo, una imagen en color verdadero es la combinación de las bandas roja, verde y azul (RGB), mientras que una imagen en falso color puede combinar bandas NIR, roja y verde para resaltar la vegetación.

Según (Sani et al., 2022) la resolución espectral en imágenes satelitales está fuertemente relacionada con la resolución espacio-temporal, concluye que productos de satélites que ofrecen valores altos de esta tienen a tener una resolución espectral limitada y viceversa. A continuación se hace un comparativo entre los productos de imagen satelital con diferentes resoluciones espetrales:

2.5.4. Resolución radiométrica

Podemos encontrar en («What is radiometric resolution? | U.S. Geological Survey», s.f.) que la resolución radiométrica refiere a la capacidad de un sensor para discriminar cambios muy finos entre niveles de intensidad de radiación. Es decir, entendiendo como las bandas espectrales capturan la energía reflejada en diferentes longitudes de onda, la resolución radiométrica determina cuántos niveles distintos de intensidad pueden ser registrados dentro de cada banda. Esta capacidad está directamente relacionada con el número de bits utilizados en el conversor analógico-digital (ADC) del sensor. Se suele decir que mientras mayor sea la resolución radiométrica de un sensor mayor es los matices de gris que puede representarse en una imagen, lo que permite detectar diferencias más sutiles en la reflectancia o emisión térmica de los objetos en la superficie terrestre. Al momento de representar esto en una imagen satelital se considera el color negro para un nivel nulo de radiación y el color blanco para el nivel máximo, creando así, con todos los niveles intermedios, una escala de grises que representa los diferentes niveles de intensidad.

Satélite	Resolución espacio-temporal	Resolución espectral (bandas)
Sentinel-2	Alta resolución espacial (10-20 m) y temporal (5 días)	13 bandas espectrales, incluyendo visible, NIR y SWIR.
Landsat 8	Resolución espacial moderada (30 m) y temporal (16 días)	11 bandas espectrales, incluyendo visible, NIR y SWIR.
PlanetScope	Alta resolución espacial (3-5 m) y muy alta resolución temporal (diaria)	4 bandas espectrales (RGB y NIR).

Cuadro 2: Comparativo entre productos de imagen satelital con diferentes resoluciones espectrales y su relación con las resoluciones espacial y temporal. Fuente: elaboración propia basada en datos de agencias espaciales y proveedores comerciales.

La resolución radiométrica se mide en bits, y un mayor número de bits permite una representación más precisa de los niveles de intensidad. Por ejemplo, un sensor con una resolución de 8 bits puede representar 256 niveles de gris, mientras que uno de 12 bits puede representar 4096 niveles. Esto es crucial para aplicaciones que requieren un alto grado de detalle, como la detección de cambios sutiles en la vegetación o calidad del aire. A continuación, se presentan algunos datos sobre la resolución radiométrica de diferentes satélites y la representación matricial de dos arreglos con diferentes resoluciones radiométricas.

Satélite	Resolución radiométrica (bits)
Landsat 5 TM	8 bits (256 niveles de gris)
Sentinel-2	12 bits (4096 niveles de gris)
NOAA AVHRR	10 bits (1024 niveles de gris)

Cuadro 3: Comparativo entre productos de imagen satelital con diferentes resoluciones radiométricas. Fuente: elaboración propia basada en datos de agencias espaciales y proveedores comerciales.

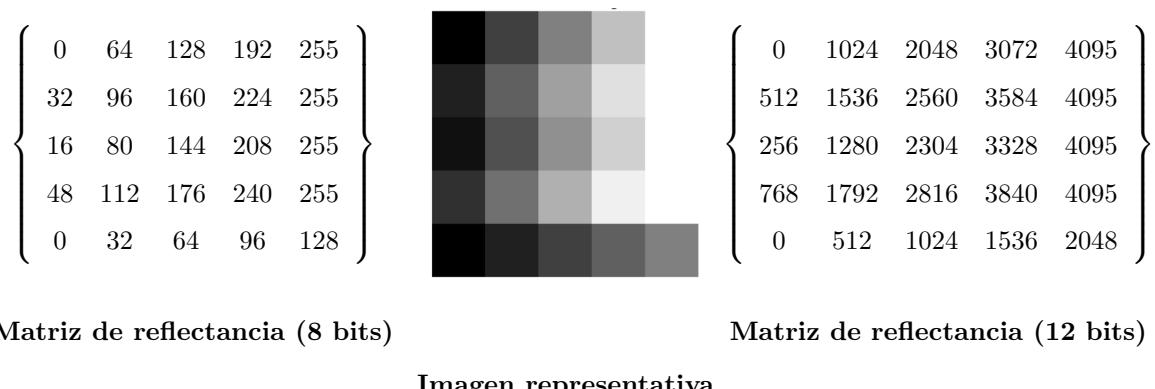


Figura 6: Comparativo entre dos matrices de reflectancia con diferentes resoluciones radiométricas (8 bits y 12 bits) y su representación visual en escala de grises (caso general). Nótese la diferencia en la cantidad de niveles de gris representados en el arreglo matricial. Fuente: elaboración propia.

2.6. Satélites de teledetección

Los sensores de teledetección montados en satélites son herramientas que revolucionaron la forma en que monitoreamos y gestionamos los recursos naturales. Estos están equipados en satélites avanzados que juntos pueden capturar datos en múltiples longitudes de onda del espectro electromagnético proporcionando información valiosa sobre la salud de los cultivos, la calidad del suelo y otros parámetros ambientales. En (Wulder et al., 2022) se menciona una revisión de 50 años de teledetección satelital, donde se concluye que los satélites Landsat han sido los más utilizados en estudios científicos, seguidos por los satélites Sentinel-2 y MODIS. A continuación se describen algunos de los satélites más relevantes en el monitoreo agrícola:

Satélite	Org.	Inicio	Sensor	Bandas	R. R. (bits)	R. E. (metros)	R. T. (días)
Landsat 1	NASA USGS	1972	MSS	4	6	80	18
Landsat 5	NASA USGS	1984	TM	7	8	30 - 120	16
Landsat 8	NASA USGS	2013	OLI + TIRS	11	12	15 - 30	16
Sentinel 2A	ESA	2015	MSI	13	12	10 - 60	5
Sentinel 1A	ESA	2014	SAR	1	16	5 - 40	6 - 12
MODIS	NASA	1999	Multiespectral	36	12	250 - 1000	1-2
NOAA AVHRR	NOAA	1978	Radiometro AVHRR	6	10	1100	0.5
PlanetScope	Planet	2014	RGB+NIR	4	12	3	1
WorldView 3	Maxar	2014	MS + Pan + SWIR	29	11	0.31 (pan), 1.24 (MS), 3.7 (SWIR)	<1
GeoEye 1	Maxar	2008	MS + Pan	5	11	0.41 (pan), 1.65 (MS)	2 - 3

Cuadro 4: Comparativo de satélites relevantes para observación de la Tierra y agricultura de precisión, con sus resoluciones radiométricas (R.R.), espaciales (R.E.) y temporales (R.T.). Fuente: elaboración propia basada en datos de agencias espaciales y proveedores comerciales.

Es importante considerar un apartado para los satélites más utilizados en la agricultura de precisión, los cuales son los satélites Sentinel-2 y Landsat 8, estos son de libre acceso, es decir, sus datos pueden ser descargados gratuitamente desde las plataformas oficiales de las agencias espaciales que los operan (ESA y USGS respectivamente).

2.6.1. Sentinel-2

Sentinel-2 es una misión de la Agencia Espacial Europea (ESA) que forma parte del programa Copernicus. Está compuesta por dos satélites idénticos, Sentinel-2A y Sentinel-2B, lanzados en 2015 y 2017 respectivamente. Estos satélites están equipados con el sensor Multispectral Instrument (MSI) que captura datos en 13 bandas espectrales que van desde el visible hasta el infrarrojo de onda corta (SWIR). La resolución espacial varía entre 10, 20 y 60 metros dependiendo de la bandapectral, mientras que la resolución temporal es de aproximadamente 5 días en el ecuador gracias a la combinación de ambos satélites. Sentinel-2 es ampliamente utilizado en aplicaciones agrícolas debido a su capacidad para monitorear la salud de los cultivos, evaluar la cobertura vegetal y detectar cambios en el uso del suelo.

2.6.2. Landsat 8

Landsat 8 es parte del programa Landsat, una colaboración entre la NASA y el Servicio Geológico de los Estados Unidos (USGS). Fue lanzado en 2013 y está equipado con dos sensores principales: el Operational Land Imager (OLI) y el Thermal Infrared Sensor (TIRS). Landsat 8 captura datos en 11 bandas espectrales que incluyen el visible, el infrarrojo cercano (NIR), el infrarrojo de onda corta (SWIR) y el infrarrojo térmico. La resolución espacial es de 30 metros para la mayoría de las bandas, con una resolución de 15 metros para la banda pancromática y 100 metros para las bandas térmicas. La resolución temporal es de 16 días. Landsat 8 es utilizado en una variedad de aplicaciones agrícolas, incluyendo la monitorización del crecimiento de los cultivos, la gestión del agua y la evaluación del impacto ambiental.

2.7. Índices espectrales aplicados a agricultura

Los índices espectrales son herramientas matemáticas que combinan diferentes bandas espectrales para resaltar características específicas de la vegetación y otros materiales en la superficie terrestre. (Radočaj et al., 2023) menciona cómo los índices espectrales proporcionan información cuantitativa sobre la salud y el crecimiento de los cultivos, por lo tanto, desempeñan un papel crucial al ofrecer una evaluación sencilla y fiable del estado y la salud de estos. Entre los principales aspectos de la fisiología vegetal que los índices espectrales pueden ayudar a evaluar se incluyen:

- **Contenido de clorofila:** La clorofila es el pigmento responsable de la fotosíntesis y su concentración puede indicar la salud y el vigor de las plantas.
- **Área foliar:** El área foliar es un indicador del tamaño y la densidad de la vegetación, lo que puede influir en la capacidad de las plantas para captar luz y realizar la fotosíntesis.
- **Estrés hídrico:** Los índices espectrales pueden detectar cambios en la reflectancia causados por la falta de agua, lo que es crucial para la gestión del riego.
- **Biomasa:** La biomasa es una medida de la cantidad total de materia vegetal y puede ser estimada utilizando índices espectrales.
- **Actividad fotosintética:** Algunos índices pueden proporcionar información sobre la eficiencia fotosintética de las plantas, lo que es importante para evaluar su productividad.

Conocer estos aspectos es fundamental para la toma de decisiones responsables, permitiendo a los agricultores optimizar el uso de recursos como agua, fertilizantes y pesticidas, mejorar la productividad de los cultivos y reducir el impacto ambiental. Se describen a continuación algunos de los beneficios clave de utilizar índices espectrales en la agricultura de precisión:

- **Optimizar la tasa de prescripción:** Permite la aplicación viable de insumos agrícolas en función del área que los necesite, evitando el desperdicio y reduciendo costos.
- **Identificar zonas con problemas:** Ayudan a detectar áreas dentro de un campo que tienen un rendimiento inferior o están experimentando estrés, ya sea por deficiencia de agua o de nutrientes.
- **Predicción del rendimiento de los cultivos:** Se utilizan con frecuencia en la predicción del rendimiento y la biomasa de diversos cultivos.
- **Monitoreo del crecimiento de los cultivos:** Permiten un seguimiento continuo del desarrollo de los cultivos a lo largo del tiempo, facilitando la identificación de patrones y tendencias.
- **Detección y manejo de plagas y enfermedades:** Ayudan a identificar áreas afectadas por plagas o enfermedades, permitiendo una intervención temprana y localizada.
- **Evaluación del impacto ambiental:** Facilitan la evaluación del impacto de las prácticas agrícolas en el medio ambiente, como la erosión del suelo y la contaminación del agua.

2.7.1. Fundamento matemático

Los índices espectrales se calculan utilizando fórmulas matemáticas que combinan las reflectancias medidas en diferentes bandas espectrales. Estas fórmulas están diseñadas para resaltar ciertas características de la vegetación o del suelo, y pueden variar en complejidad desde simples relaciones entre dos bandas hasta combinaciones más elaboradas que involucran múltiples bandas y coeficientes de ponderación que dependen de características específicas del sensor, del objeto de estudio, de una región geográfica específica o de una banda del espectro electromagnético en específico, sin embargo, todas estas se basan en operaciones matemáticas que manipulan los valores entregados por los sensores de teledetección, es decir, manipulan elementos de una matriz de datos.

Para entender cómo se calculan de forma general, analizaremos uno de los más utilizados, el Índice Diferencial de Vegetación Normalizada (NDVI, por sus siglas en inglés), que, matricialmente se define como:

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

Donde:

- *NIR* es la reflectancia en la banda del infrarrojo cercano.
- *RED* es la reflectancia en la banda roja del espectro visible.

Matemáticamente podemos interpretar esta relación como la diferencial entre la reflectancia en el infrarrojo cercano y la reflectancia en la banda roja, normalizada por la suma de reflectancias. A continuación se muestra un ejemplo del cálculo de este índice utilizando matrices de reflectancia:

Consideremos la captura de datos de la siguiente área de cultivo (A), se tiene que esta área de tiene una extensión de 3x3 kilómetros cuadrados y se han capturado la radiancia reflejada por el satélite (B) que es un satélite multiespectral (representación del satélite Landsat 8), por lo que se ha obtenido un estack de imágenes (C) que representan las diferentes bandas espectrales, entre ellas, la banda roja (RED) y la banda del infrarrojo cercano (NIR).

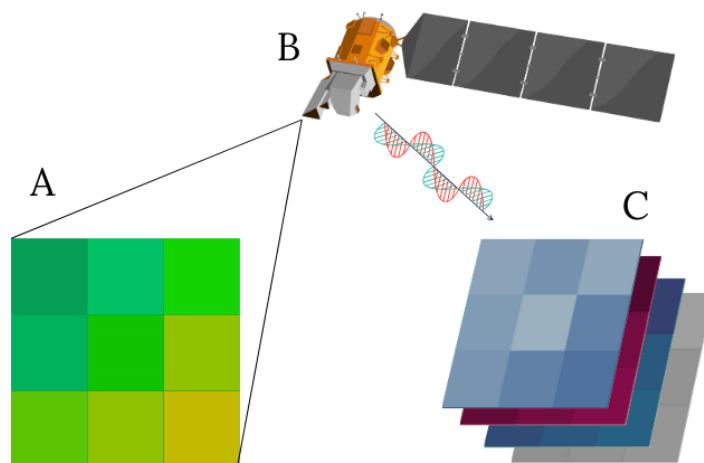


Figura 7: Ejemplo de área de cultivo (A), satélite multiespectral (B) y la captura de imagen satelital, stack de imágenes multiespectrales (bandas) (C). Fuente: elaboración propia.

Posteriormente se extraen las bandas espectrales de interés, en este caso, la banda roja (RED) y la banda del infrarrojo cercano (NIR), las cuales son representadas como matrices de reflectancia, es

importante reconocer cómo se están representando los niveles de reflectancia para el color rojo y el infrarrojo cercano, recordemos, según la fisiología vegetal, que la vegetación sana absorbe fuertemente la luz roja para la fotosíntesis y refleja fuertemente la luz del infrarrojo cercano, esto se traduce en valores mas bajos para la banda roja y valores más altos para la banda del infrarrojo cercano si es que la vegetación es en efecto, sana. sin embargo si la vegetación está estresada o enferma, como son los casos de los cuadrantes del extremo inferior derecho, los valores de reflectancia en la banda roja serán más altos y los valores en la banda del infrarrojo cercano serán más bajos. Esto es particularmente útil a la hora de calcular la diferencia entre ambas bandas, ya que si la vegetación está sana, la diferencia será mayor y si está estresada o enferma, la diferencia será menor o incluso negativa.

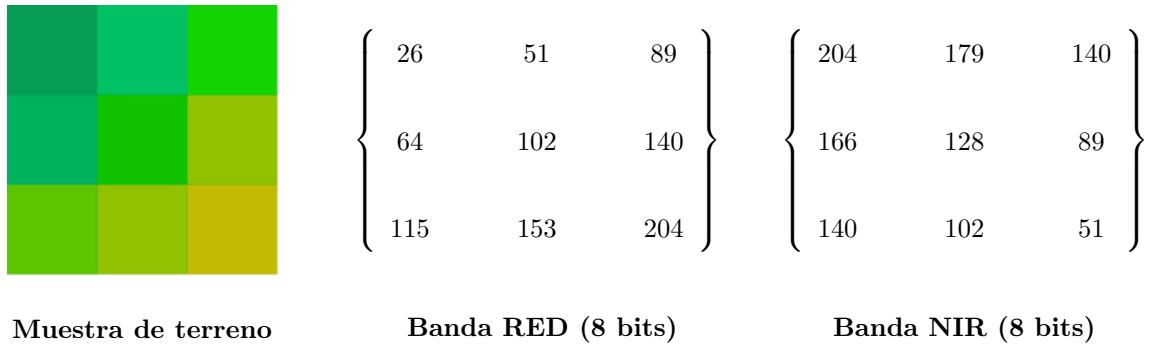


Figura 8: Representación de la banda roja (RED) y la banda del infrarrojo cercano (NIR) como matrices de reflectancia extraídas del stack de imágenes multiespectrales. Nótese cómo los valores de reflectancia varían en función de la salud de la vegetación. Fuente: elaboración propia.

Calculando la expresión para el NDVI, tenemos:

Calculando la suma:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Matricialmente no se define la división, por lo que se tiene:

$$NIR + RED = \begin{pmatrix} 230 & 230 & 229 \\ 230 & 230 & 229 \\ 255 & 255 & 255 \end{pmatrix}$$

Luego, calculando la inversa de la suma:

$$NDVI = (NIR - RED) \cdot (NIR + RED)^{-1}$$

$$(NIR+RED)^{-1} = \begin{pmatrix} 0,0043 & 0,0043 & 0,0044 \\ 0,0043 & 0,0043 & 0,0044 \\ 0,0039 & 0,0039 & 0,0039 \end{pmatrix}$$

Calculando la diferencia:

Finalmente, calculando el NDVI:

$$NIR - RED = \begin{pmatrix} 178 & 128 & 51 \\ 102 & 26 & -51 \\ 25 & -51 & -153 \end{pmatrix}$$

$$NDVI = \begin{pmatrix} 0,76 & 0,56 & 0,22 \\ 0,44 & 0,11 & -0,22 \\ 0,10 & -0,22 & -0,60 \end{pmatrix}$$

De aquí se puede concluir que el índice NDVI varía entre -1 y 1 dada la razón entre la diferencia y la suma de las reflectancias, entendiendo que la diferencia siempre será menor o igual a la suma, mientras esta última sirve para la normalización, es decir, para mantener el índice dentro de un rango específico.

Finalmente es posible representar visualmente el índice NDVI utilizando una escala de colores que va

desde el rojo (valores bajos o negativos) hasta el verde (valores altos), pasando por tonos intermedios como el amarillo y el naranja. Esta representación facilita la interpretación de los datos, permitiendo identificar rápidamente áreas con vegetación sana (valores altos de NDVI) y áreas con vegetación estresada o sin vegetación (valores bajos o negativos de NDVI).



Muestra de terreno

Matriz NDVI

Representación Visual

Figura 9: Comparativo entre la muestra de terreno, la matriz NDVI calculada y su representación visual utilizando una escala de colores. Nótese cómo las áreas con vegetación sana se representan en tonos verdes, mientras que las áreas con vegetación estresada o sin vegetación se representan en tonos rojos. Fuente: elaboración propia.

2.7.2. Índices espectrales más utilizados en agricultura

A continuación se describen algunos de los índices espectrales más comunes utilizados en la agricultura de precisión, junto con sus fórmulas matemáticas y aplicaciones específicas: Fuente: (Radočaj et al., 2023; Xue & Su, 2017)

- **Índice de Vegetación de Diferencia Simple (DVI):**

$$DVI = NIR - RED$$

Medición directa de la cantidad de vegetación, aunque poco robusto frente a condiciones variables de suelo y atmósfera. Se usa como base para índices normalizados.

- **Índice de Vegetación Mejorado (EVI):**

$$EVI = G \frac{(NIR - RED)}{(NIR + C_1 \cdot RED - C_2 \cdot BLUE + L)}$$

Donde:

G : Factor de ganancia

C_1 y C_2 : Coeficientes de corrección atmosférica

L : Factor de ajuste para el suelo.

Mejora la sensibilidad en áreas con alta biomasa y reduce la influencia del suelo y la atmósfera.

- **Índice de Agua Normalizado (NDWI):**

$$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)}$$

Detección del contenido de agua en la vegetación y monitoreo del estrés hídrico.

- **Índice de Clorofila Verde (GCI):**

$$GCI = \frac{NIR}{GREEN} - 1$$

Estimación del contenido de clorofila en las hojas, útil para evaluar la salud nutricional y el estado nitrogenado de las plantas.

- **Índice de Vegetación Ajustado al Suelo (SAVI):**

$$SAVI = \frac{(NIR - RED) \cdot (1 + L)}{(NIR + RED + L)}$$

Donde: L : Factor de influencia del suelo.

Mejora la precisión en áreas con baja cobertura vegetal y alta influencia del suelo.

- **Índice de Diferencia Normalizada del Red Edge (NDRE):**

$$NDRE = \frac{(NIR - RED_{edge})}{(NIR + RED_{edge})}$$

Sensible al contenido de clorofila y vigor en etapas avanzadas del cultivo, reduce la saturación del NDVI en coberturas densas.

3. CAPÍTULO II APLICACIÓN PRÁCTICA: CLASIFICACIÓN DE COBERTURA AGRÍCOLA

La clasificación de cobertura agrícola es una tarea fundamental en la agricultura de precisión, esto porque permite identificar y mapear las áreas cultivadas y no cultivadas en determinada región, incluso poder diferenciarlas de aquellas donde solo hay vegetación natural. Esta información es crucial para la toma de decisiones informadas en la gestión agrícola, optimización del uso de recursos y monitoreo del crecimiento de los cultivos. En este capítulo, se presenta una aplicación práctica de generación de un modelo de clasificación supervisada de cobertura agrícola utilizando unos de los datasets de información espacial más recientes y de libre acceso como lo es AlphaEarth Foundations.

3.1. Flujo de trabajo

A continuación se muestra el flujo de trabajo seguido para la generación del modelo de clasificación supervisada de cobertura agrícola:



Figura 10: Diagrama de flujo que describe el proceso de generación y uso del modelo de clasificación supervisada de cobertura agrícola.

3.2. Materiales

Las herramientas que se utilizarán para la generación y uso del modelo de clasificación supervisada de cobertura agrícola representan un poderoso conjunto de recursos para el análisis geoespacial que combina tecnología, datos, algoritmos avanzados e inteligencia artificial. Estas herramientas permiten a los usuarios procesar, analizar e interpretar grandes volúmenes de datos geoespaciales de manera eficiente y efectiva.

- **AlphaEarth Foundations:** Creado por Google DeepMind, AlphaEarth Foundations es un modelo de campos de embeddings (*embedding field model* EFM) que combina datos geoespaciales de muchas diferentes fuentes y utiliza un algoritmo de aprendizaje automático para crear representaciones vectoriales (embeddings) de ubicaciones geográficas en todo el mundo. Estas representaciones vectoriales capturan características espaciales y temporales de las ubicaciones generando así una especie de "huella digital" para cada lugar en la Tierra. Estas representaciones vectoriales pueden ser utilizadas para una variedad de aplicaciones, incluyendo la clasificación de imágenes satelitales, la detección de cambios en el uso del suelo, la predicción de fenómenos naturales y la mejora de modelos climáticos. De manera simplificada, AlphaEarth Foundations es un dataset que proporciona información geoespacial rica y detallada en sus "bandas" que son 64 en total, las cuales por sí solas no tienen un significado específico, ya que son representaciones matemáticas de características geoespaciales. Sin embargo, cuando se combinan y analizan en conjunto, estas bandas pueden proporcionar información valiosa sobre la superficie terrestre y sus cambios a lo largo del tiempo. Brown et al., 2025
- **Google Earth Engine (GEE):** Propiedad de Google, es una plataforma en la nube que permite a los usuarios analizar y visualizar grandes conjuntos de datos geoespaciales. Proporciona acceso a una vasta colección de imágenes satelitales y datos geoespaciales, junto con herramientas de procesamiento y análisis avanzadas. GEE es ampliamente utilizado en la investigación ambiental, la gestión de recursos naturales, la agricultura de precisión y otros campos relacionados con el análisis geoespacial. La plataforma permite a los usuarios realizar análisis complejos sin necesidad de descargar grandes volúmenes de datos, lo que facilita el trabajo con conjuntos de datos masivos. Pizarro et al., 2022. Para el caso de estudio, esta plataforma se utiliza para la generación del modelo, pruebas y uso del mismo.
- **QGIS (Quantum Geographic Information System):** Es un sistema de información geográfica (SIG) de código abierto que permite a los usuarios crear, editar, visualizar, analizar y publicar información geoespacial. QGIS es una herramienta poderosa y versátil que es ampliamente utilizada en diversas disciplinas, incluyendo la agricultura de precisión, la gestión ambiental, la planificación urbana y la investigación científica. QGIS ofrece una amplia gama de funcionalidades, incluyendo la capacidad de trabajar con múltiples formatos de datos geoespaciales, realizar análisis espaciales avanzados, crear mapas personalizados y automatizar tareas mediante scripts. La plataforma es altamente extensible a través de complementos (plugins) que permiten a los usuarios agregar funcionalidades específicas según sus necesidades. Cubillas et al., 2024. Este software se utilizará para la generación de las etiquetas de entrenamiento que representan junto al dataset de AlphaEarth Foundations, los datos de entrada para la generación del modelo de clasificación supervisada de cobertura agrícola.
- **Random Forest:** Es un algoritmo de aprendizaje automático supervisado que se utiliza para tareas de clasificación y regresión. Fue desarrollado por Leo Breiman y Adele Cutler en la década de 1990 y se basa en la idea de combinar múltiples árboles de decisión para mejorar la precisión y robustez del modelo. En lugar de construir un solo árbol de decisión, Random Forest crea un conjunto (o "bosque") de árboles de decisión independientes, cada uno entrenado en una muestra aleatoria del conjunto de datos original. Durante el proceso de entrenamiento, cada árbol toma decisiones basadas en diferentes subconjuntos de características y datos, lo que ayuda a reducir el sobreajuste y mejora la capacidad del modelo para generalizar a nuevos datos. Para la generación del modelo de clasificación supervisada de cobertura agrícola, se utiliza la implementación de Random Forest disponible en Google Earth Engine. Castillo et al., 2022

3.3. Metodología

La metodología seguida para la clasificación supervisada de cobertura agrícola utilizando AlphaEarth Foundations y Google Earth Engine (GEE) se detalla a continuación, enfocando el área geográfica en el departamento de Lambayeque, Perú:

- **Tipo de análisis:** Aplicado, cuantitativo, descriptivo y explicativo.
- **Unidad de análisis:** Pixel satelital (10 m x 10 m).
- **Periodo de estudio:** Enero 2020 - Diciembre 2024.
- **Área de estudio:** Modelo de clasificación de cobertura agrícola válido para el departamento de Lambayeque, Perú.
- **Justificación del área de estudio:** Lambayeque es una región costera con una importante actividad agrícola, caracterizada por una cobertura agrícola bien delimitada y homogénea, lo que facilita la generación y validación del modelo. Además, la diferenciación entre áreas agrícolas y vegetación natural es más clara en comparación con otras regiones del país como la Selva o la Sierra, donde la diversidad de cultivos y la presencia de vegetación natural pueden complicar el proceso de clasificación.
- **Fuentes de datos:** AlphaEarth Foundations y Google Earth Engine.
- **Software:** Google Earth Engine y QGIS.
- **Complementarios:** Google Maps, límites administrativos del Perú (INEI).

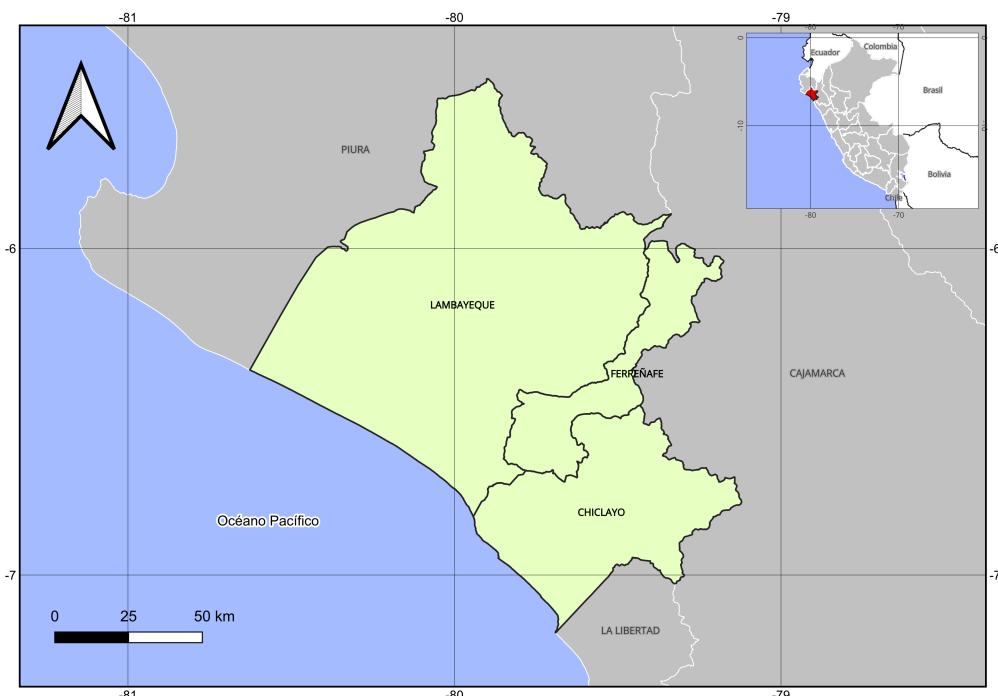


Figura 11: Mapa de ubicación del departamento de Lambayeque, Perú, área de estudio para la clasificación de cobertura agrícola.

3.4. Procedimiento

De acuerdo al flujo de trabajo, la generación del modelo de clasificación supervisada de cobertura agrícola y posterior uso del mismo contempla la secuencia de pasos que inician con la generación de etiquetas de entrenamiento. AlphaEarth Foundations brinda 64 bandas que representan características geoespaciales, las cuales se toman como entrada para la generación de un dataset nuevo que describe las características de estas bandas para cada polígono de las etiquetas de entrenamiento, lo que posteriormente se utiliza para entrenar el modelo de clasificación.

3.4.1. Etiquetas de entrenamiento

Para la generación de etiquetas de entrenamiento, se generará un recurso vectorial en QGIS que contenga polígonos que, a modo de muestra, representen los diferentes tipos de cobertura agrícola presentes en el departamento de Lambayeque, así como áreas sin cobertura agrícola, como zonas urbanas, cuerpos de agua y vegetación natural. Estos polígonos deben ser lo más representativos y variados posibles para asegurar que el modelo pueda aprender a diferenciarlas, lo cual es importante dada la naturaleza estadística del algoritmo.

- **Área de Interés (AOI):** Para la delimitación del área de estudio, se utilizará un recurso vectorial con los límites departamentales de Lambayeque, el cual puede ser descargado desde el portal del INEI «INEI», s.f. mediante la capa de "Departamentos." actualizado al año 2023. Este recurso será útil tanto para la delimitación del área de entrenamiento del modelo como para la posterior aplicación del mismo, además, servirá para delimitar los polígonos de las etiquetas de entrenamiento. La elección del departamento de Lambayeque se justifica por ser una región con una importante actividad agrícola, productos como caña de azúcar, paltos, arándanos, limones, etc, son cultivados en grandes extensiones y representan uno de los motores económicos de la región por su gran volumen de exportación.
- **Imagen de referencia:** Se utilizarán imágenes compuestas de Sentinel-2. Es importante que la generación de etiquetas de entrenamiento se realice con imágenes de la misma fecha de adquisición que las imágenes que se utilizarán para la clasificación, ya que las características espectrales de los cultivos pueden variar dependiendo de la etapa fenológica en la que se encuentren.
 - Fecha de adquisición: 01-01-2024 - 02-01-2024.
 - Fuente: Sentinel-2.
 - Resolución espacial: 10 m.
 - Bandas: B4 (Red), B3 (Green), B2 (Blue).
 - Área de interés: Departamento de Lambayeque, Perú.
 - Plataforma: Google Earth Engine.

A continuación se muestra el código en JavaScript utilizado en Google Earth Engine para la generación de la imagen de referencia:

```

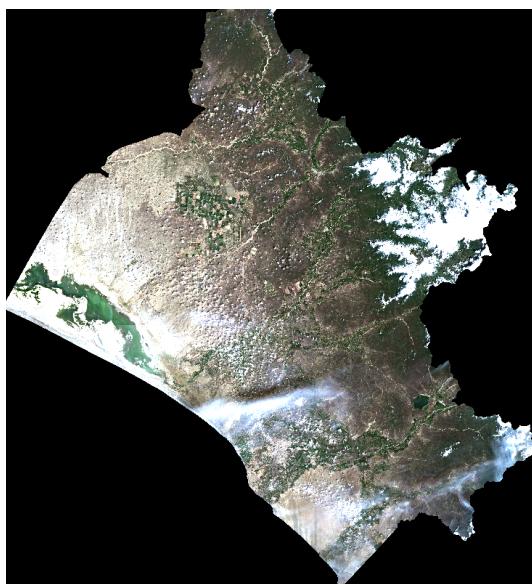
1  var sentinel = ee.ImageCollection("COPERNICUS/S2_SR_HARMONIZED")
2      .filterDate('2024-01-01', '2024-01-02')
3      .filterBounds(aoiLambayeque);
4
5  var visParamsSentinel = {
6      bands: ['B4', 'B3', 'B2'],
7      min: 0,
8      max: 3000,
9      gamma: 1.2
10 };

```

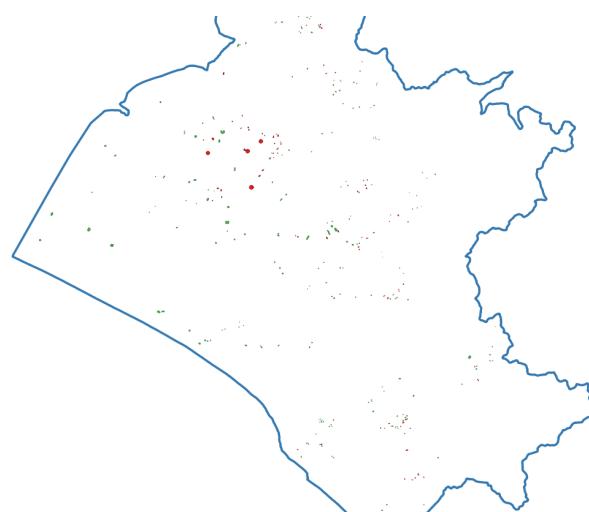
Figura 12: Código en JavaScript para la generación de la imagen de referencia en GEE. De la línea 1 a 4, se carga el dataset de Sentinel-2 y se filtra por fecha y área de interés, y de la línea 6 a 11, se definen los parámetros de visualización para la imagen.

■ **Generación de etiquetas:** Con la imagen de referencia descargada y en QGIS, se crearán polígonos que representen las diferentes clases de cobertura agrícola y no agrícola. Cada polígono debe ser etiquetado con la clase correspondiente, se utilizan 0 para áreas sin cobertura agrícola, 1 para área de cultivos. Es recomendable crear múltiples polígonos para cada clase para capturar la variabilidad dentro de cada categoría. La cantidad de polígonos serán de 150 por cada clase.

A continuación, se muestra un ejemplo de las etiquetas de entrenamiento generadas en QGIS para el departamento de Lambayeque:



(a) Imagen de referencia (Sentinel-2)



(b) Etiquetado de polígonos en QGIS



(c) Muestra de polígonos realizados

Figura 13: Muestra del etiquetado de polígonos en QGIS para la generación del dataset de entrenamiento en el departamento de Lambayeque, Perú. Nótese que en (a) se muestra la imagen de referencia utilizada, en (b) el proceso de etiquetado de polígonos y en (c) una muestra de los polígonos realizados.

- **Exportación de etiquetas:** Una vez hechos y revisados los polígonos en QGIS, se exportan como un archivo shapefile (.shp), formato compatible con Google Earth Engine. Este archivo contendrá las coordenadas y las etiquetas de clase para cada polígono, que serán utilizadas en el siguiente paso para extraer las características de AlphaEarth Foundations.

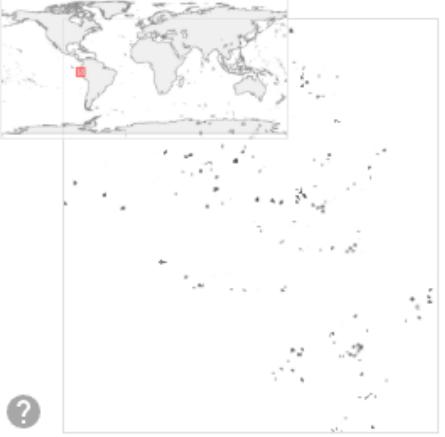


Table ID 
 projects/test-hsluis4326/assets/etiquetas

Date
 Start date: NA 
 End date: NA 

File Size 67.08KB

Number of Features 314

Last modified 2025-10-06 19:44:44 UTC

DESCRIPTION	FEATURES	PROPERTIES
Feature Index	clase (Integer)	system:index (String)
0	0	
1	0	
2	0	
3	0	
4	0	
5	0	
6	0	
7	0	
8	0	
9	0	

*Limited to the first 10 features.

Recurso shapefile subido en GEE

Figura 14: Subida del archivo shapefile con las etiquetas de entrenamiento a Google Earth Engine.

3.4.2. Generación del dataset de entrenamiento

Una vez subidos los recursos necesarios para la generación del modelo de clasificación supervisada de cobertura agrícola, se procede a la generación del recurso necesario que describe las características de las 64 bandas de AlphaEarth Foundations para cada polígono de las etiquetas de entrenamiento, este recurso alojado en el ambiente de desarrollo de Google Earth Engine, es un *FeatureCollection* que contiene las características de las mencionadas 64 bandas.

```

1  var aoiLambayeque =
2    ee.FeatureCollection("projects/test-hsluis4326/assets/aoiLambayeque"),
3    etiquetas =
4      ee.FeatureCollection("projects/test-hsluis4326/assets/etiquetas");

```

Figura 15: Código en JavaScript para la importación de los recursos necesarios en GEE. La ubicación de los recursos varían dependiendo del usuario y proyecto GEE.

A continuación, se muestra el código en JavaScript utilizado en Google Earth Engine para la generación del dataset de entrenamiento, nótese que se utiliza el método `sampleRegions()` para extraer las características para cada polígono de las etiquetas de entrenamiento.

```

1 var dataset = ee.ImageCollection("GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL")
2   .filterDate('2024-01-01', '2024-01-02')
3   .filterBounds(aoiLambayeque);
4
5 var imageAlpha = dataset.mosaic().clip(aoiLambayeque);
6
7 var samples = imageAlpha.sampleRegions({
8   collection: etiquetas,
9   properties: ['clase'],
10  scale: 20,
11});
```

Figura 16: Código en JavaScript para la generación del dataset de entrenamiento en GEE. De la línea 1 a 3, se carga el dataset de AlphaEarth Foundations, de la línea 5 a 7, se genera una imagen mosaico y se recorta al área de interés, y de la línea 9 a 13, se genera el dataset de entrenamiento utilizando las etiquetas de entrenamiento. Obs: se utiliza una escala de 20 metros para la extracción de las características por su alta carga computacional.

3.4.3. Generación del modelo de clasificación supervisada

Con el dataset de entrenamiento generado, se procede a la creación del modelo de clasificación supervisada utilizando el algoritmo de Random Forest disponible en Google Earth Engine. A continuación, se muestra el código en JavaScript utilizado para la generación del modelo:

```

1 var classifier = ee.Classifier.smileRandomForest(200).train({
2   features: samples,
3   classProperty: 'clase',
4   inputProperties: imageAlpha.bandNames()
5});
```

Figura 17: Código en JavaScript para la generación del modelo de clasificación supervisada en GEE. En este caso, se utiliza 200 árboles de decisión para la creación del modelo. Nótese, además, que se especifica la propiedad de clase y las bandas de entrada.

3.4.4. Aplicación del modelo de clasificación supervisada

El modelo de clasificación supervisada generada se aplica sobre el mismo dataset de AlphaEarth Foundations utilizado para la generación del modelo, pero esta vez, para todo el departamento de Lambayeque. Una vez procesada la imagen, se visualizan los resultados obtenidos en un mapa de clasificación, donde cada píxel resulta en una máscara que indica la clase asignada por el modelo, 0 para áreas sin cobertura agrícola y 1 para áreas con cobertura agrícola.

```

1 var classified = imageAlpha.classify(classifier);
2
3 Map.addLayer(classified, {min:0, max:1, palette:['red','green']},
4   'Clasificación');
```

Figura 18: Código en JavaScript para la aplicación del modelo de clasificación supervisada en GEE. La generación de la imagen clasificada se realiza mediante el método `classify()` y la visualización de la misma en el lienzo mediante `Map.addLayer()` con una paleta de colores tal que las áreas sin cobertura agrícola se muestran en rojo y las áreas con cobertura agrícola en verde.

3.5. Resultados

El producto obtenido tras la aplicación del modelo de clasificación supervisada de cobertura agrícola es una capa ráster que muestra la clasificación de cada píxel en el área de estudio, como se explicó en los fundamentos teóricos, la salida es la representación de una máscara binaria (matriz) donde cada píxel es clasificado en una de las dos categorías: áreas con cobertura agrícola (1) y áreas sin cobertura agrícola (0). A continuación, se muestra el mapa de clasificación obtenido para el departamento de Lambayeque sobre la imagen de referencia, entrenada con etiquetas y generada con AlphaEarth Foundations:

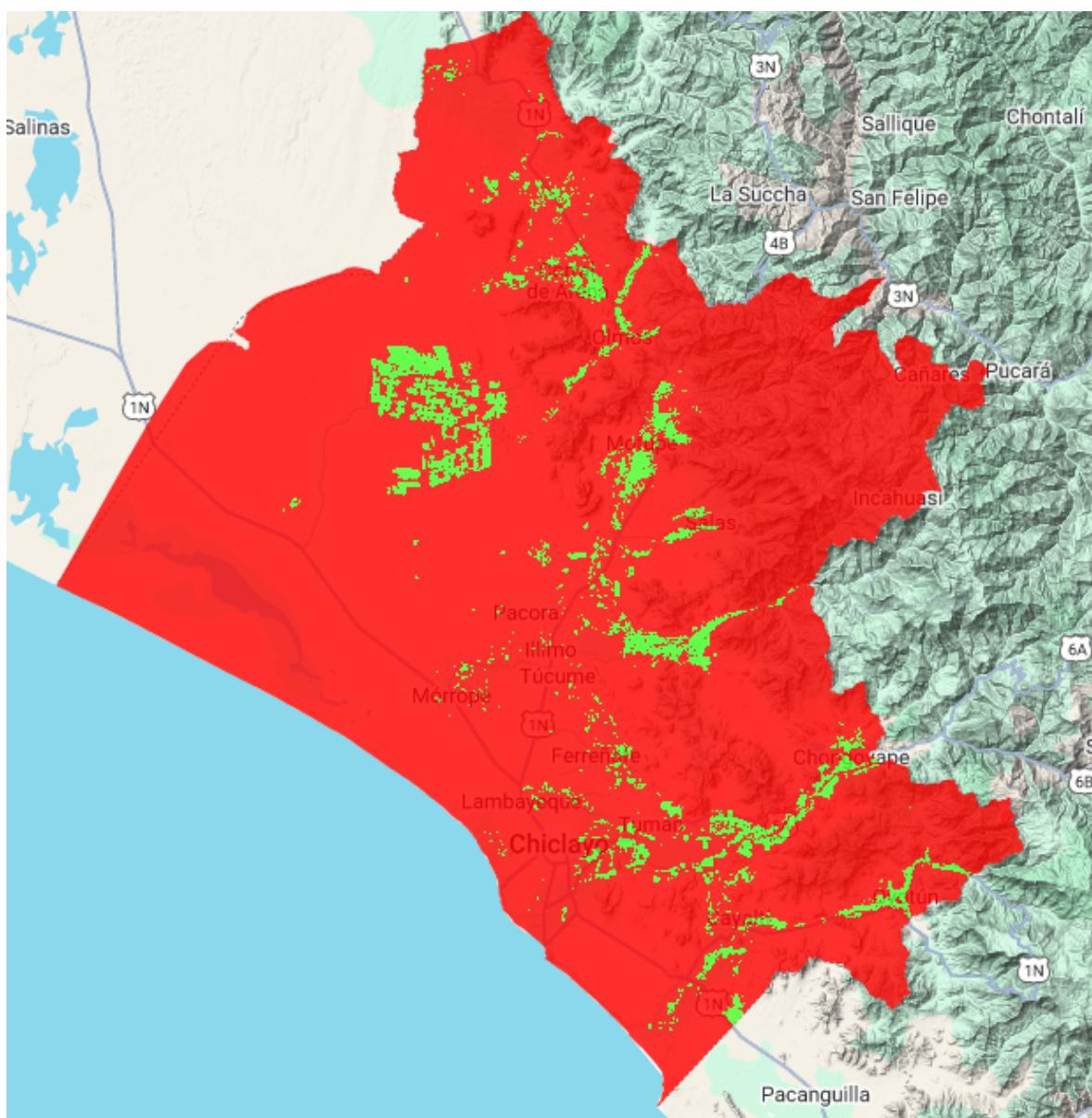


Figura 19: Mapa de clasificación de cobertura agrícola obtenido para el departamento de Lambayeque, Perú, utilizando el modelo de clasificación supervisada generado con AlphaEarth Foundations y Google Earth Engine. En verde se muestran las áreas con cobertura agrícola y en rojo las áreas sin cobertura agrícola.

3.6. Validación del modelo de clasificación supervisada

Para la validación de la clasificación de cobertura agrícola mediante el modelo de Random Forest, se utilizan diversas métricas que permiten evaluar la precisión y confiabilidad del modelo aprovechando la naturaleza binaria de la clasificación. Los métodos de validación surgen a partir de la segmentación del dataset de entrenamiento en dos subconjuntos: uno para el entrenamiento del modelo (70 % de los datos) y otro para la validación (30 % de los datos). Esta división asegura que el modelo se evalúe con datos que no ha visto durante el entrenamiento, proporcionando una medida más realista de su rendimiento, posteriormente con el modelo ya validado y con métricas satisfactorias, se reentrena el modelo con el 100 % de los datos para su uso en producción. Se puede observar a continuación el código que genera la segmentación de los datos de entrenamiento y que servirán para la validación del mismo. Finalmente, se entrena el modelo utilizando solo las muestras de entrenamiento y se imprimen en la consola las cantidades de muestras totales, de entrenamiento y de validación.

```

1  var withRandom = samples.randomColumn('random');
2  var split = 0.7;
3  var trainingSamples = withRandom.filter(ee.Filter.lt('random', split));
4  var validationSamples = withRandom.filter(ee.Filter.gte('random', split));
5
6  var classifier = ee.Classifier.smileRandomForest(200).train({
7    features: trainingSamples,
8    classProperty: 'clase',
9    inputProperties: imageAlpha.bandNames()
10   });
11
12 print('Total de muestras:', samples.size());
13 print('Muestras de entrenamiento:', trainingSamples.size());
14 print('Muestras de validación:', validationSamples.size());
15

```

Figura 20: Código en JavaScript para la segmentación del dataset de entrenamiento y generación del modelo de clasificación supervisada en GEE. Se utiliza el método `randomColumn()` para asignar un valor aleatorio a cada muestra, luego se filtran las muestras en dos subconjuntos: uno para entrenamiento (70 %) y otro para validación (30 %).

Considérese que los valores reales corresponden a una escala de 20 metros por píxel lo que reduce la carga computacional, valores más bajos en escala generan más muestras y mejor precisión espacial, pero a costa de mayor tiempo de procesamiento. La salida de la consola de Google Earth Engine tras la ejecución del código anterior es la siguiente:

```

1      Total de muestras:
2      51321
3      Muestras de entrenamiento:
4      35821
5      Muestras de validación:
6      15500

```

Figura 21: Salida de la consola de Google Earth Engine tras la ejecución del código para la segmentación del dataset de entrenamiento y generación del modelo de clasificación supervisada, los valores corresponden a la cantidad de píxeles muestreados (51,321), de los cuales 35,821 se utilizan para el entrenamiento del modelo y 15,500 para la validación del mismo.

3.6.1. Matriz de confusión

La matriz de confusión es una herramienta fundamental para evaluar el rendimiento de un modelo de clasificación. En este caso, se utiliza para comparar las predicciones del modelo con las etiquetas reales de un conjunto de datos de validación independiente. A continuación se muestra el proceso de obtención de la matriz de confusión, resultado e interpretación:

```

1 var validation = validationSamples.classify(classifier);
2
3 var confusionMatrix = validation.errorMatrix('clase', 'classification');
4 print('Matriz de Confusión:', confusionMatrix);

```

Figura 22: Código en JavaScript para la obtención de la matriz de confusión en GEE. Se utiliza el método `errorMatrix()` para comparar las etiquetas reales con las predicciones del modelo.

Con la comparación entre las etiquetas reales y las predicciones del modelo, se obtiene la siguiente matriz de confusión:

	Predicción: Cultivo	Predicción: No Cultivo
Clase: Cultivo	9481	4
Clase: No Cultivo	14	6001

Cuadro 5: Matriz de confusión obtenida para la validación del modelo de clasificación supervisada de cobertura agrícola.

Como se observa en la tabla 5, la matriz de confusión muestra que el modelo ha clasificado correctamente 9,481 píxeles como “Cultivo” y 6,001 píxeles como “No Cultivo”. Sin embargo, también ha cometido algunos errores, clasificando incorrectamente 4 píxeles como “No Cultivo” cuando en realidad son “Cultivo” (falsos negativos) y 14 píxeles como “Cultivo” cuando en realidad son “No Cultivo” (falsos positivos), sin embargo, estos errores son mínimos en comparación con la cantidad total de muestras, lo que indica un buen rendimiento del modelo.

3.6.2. Precisión General del modelo (overall accuracy)

A partir de la matriz de confusión, se pueden calcular diversas métricas de precisión que proporcionan una evaluación cuantitativa del rendimiento del modelo. Esta métrica es especialmente aplicable en este caso ya que las etiquetas de entrenamiento están balanceadas, es decir, tal como se indicó anteriormente, se generaron 150 polígonos para cada clase (cultivo y no cultivo), lo que resulta en un conjunto de datos de entrenamiento equilibrado. El cálculo de la precisión general puede ser calculado mediante:

$$\text{Precisión General} = \frac{TP + TN}{TP + TN + FP + FN}$$

Donde:

TP (True Positives): Verdaderos positivos, clasificación correcta de “Cultivo”.

TN (True Negatives): Verdaderos negativos, clasificación correcta de “No Cultivo”.

FP (False Positives): Falsos positivos, clasificación incorrecta de “Cultivo”.

FN (False Negatives): Falsos negativos, clasificación incorrecta de “No Cultivo”.

A continuación, se muestra la obtención de la precisión del modelo y su resultado:

```

1 var overallAccuracy = confusionMatrix.accuracy();
2 print('Precisión general del modelo:', overallAccuracy);

```

Figura 23: Código en JavaScript para calcular la precisión general (overall accuracy) del modelo de clasificación supervisada en GEE a partir de la matriz de confusión.

La salida de la consola de Google Earth Engine tras la ejecución del código anterior es la siguiente:

```

1 Precisión Global (%):
2 99.88387096774194

```

Figura 24: Salida de la consola de Google Earth Engine tras la ejecución del código para calcular la precisión general del modelo de clasificación supervisada.

Como se observa en la figura 24, la precisión general del modelo es del 99.88 %, lo que indica que el modelo ha clasificado correctamente el 99.88 % de las muestras de validación. Esta alta precisión sugiere que el modelo es muy efectivo para distinguir entre áreas con y sin cobertura agrícola en el área de estudio.

3.6.3. Métricas de área calculada

Además de la precisión general, se presentan los valores calculados de área general del departamento de Lambayeque, área clasificada como cobertura agrícola y área clasificada como sin cobertura agrícola, así como el porcentaje que representa cada una de estas áreas con respecto al total del departamento. Estos valores son importantes para entender la distribución espacial de la cobertura agrícola en la región y pueden ser útiles para la toma de decisiones en la gestión agrícola y planificación territorial.

Estadística	Valor
Área total (hectáreas)	1,467,943.73
Área agrícola (hectáreas)	83,381.34
Porcentaje de cobertura agrícola (%)	5.68
Porcentaje de cobertura no agrícola (%)	94.32

Cuadro 6: Estadísticas de área calculadas para el departamento de Lambayeque.

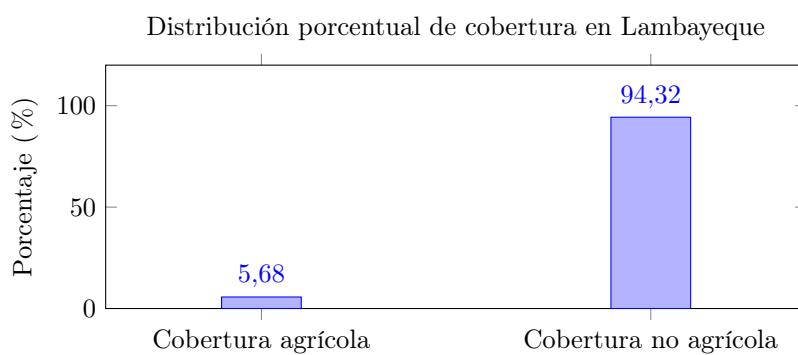


Figura 25: Gráfico de barras con la distribución porcentual de cobertura agrícola y no agrícola en Lambayeque.

3.7. Análisis y discusión de resultados

Los resultados obtenidos en la clasificación de cobertura agrícola para el departamento de Lambayeque utilizando el modelo de Random Forest entrenado con AlphaEarth Foundations y Google Earth Engine son altamente satisfactorios. La precisión general del modelo, que alcanza un 99.88 %, indica que el modelo es extremadamente efectivo para distinguir entre áreas con y sin cobertura agrícola en la región. Esta alta precisión es un indicativo claro de que las características extraídas de AlphaEarth Foundations son adecuadas para esta tarea específica de clasificación.

3.7.1. Clasificación en áreas cubiertas por nubes

Un aspecto importante a considerar en la clasificación de imágenes satelitales es la presencia de nubes, que pueden afectar la calidad de los datos y, por ende, la precisión del modelo. En este caso, se observa que el modelo ha logrado clasificar correctamente las áreas cubiertas por nubes, generalmente, un clasificador convencional etiquetaría estas áreas como “No Cultivo” debido a la falta de información visual. Sin embargo, el modelo entrenado con AlphaEarth Foundations ha demostrado ser capaz de manejar estas situaciones de manera efectiva, posiblemente debido a la riqueza de las características geoespaciales proporcionadas por las 64 bandas del dataset designando a esta área como “Cultivo” cuando corresponde. Este comportamiento es particularmente beneficioso en regiones donde la cobertura nubosa es frecuente, ya que permite mantener la integridad de la clasificación sin perder información valiosa sobre la cobertura agrícola.

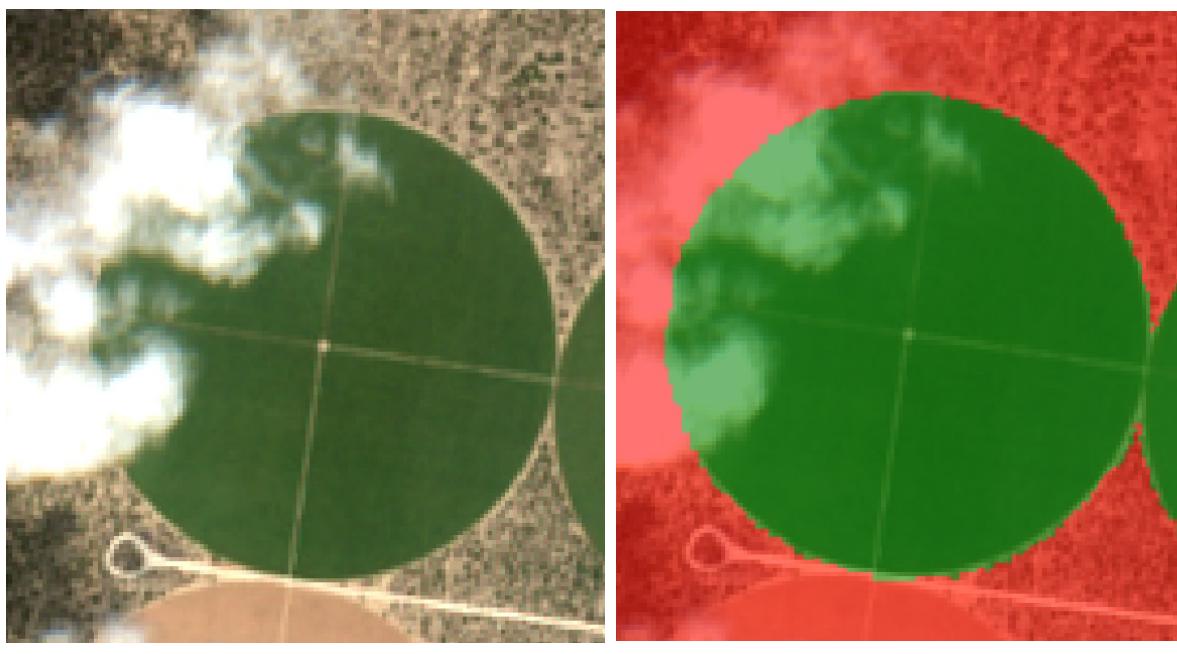


Figura 26: Comparación visual entre la imagen original (a) y la imagen clasificada (b) para el área de estudio.

El área que se presenta en la figura 26 muestra un ejemplo claro de cómo el modelo ha manejado las áreas cubiertas por nubes. En la imagen original (a), se puede observar un área con cobertura nubosa que podrían dificultar la clasificación precisa. Sin embargo, en la imagen clasificada (b), el modelo ha logrado identificar correctamente las áreas de cultivo. Es importante destacar que, si bien, esto podría ser explicado por la naturaleza temporal de las nubes, que tienden a desplazarse rápidamente, en este caso, se debe recordar que la imagen de referencia y el dataset de AlphaEarth Foundations utilizados para la clasificación corresponden a la misma fecha.



(a) Imagen original 3

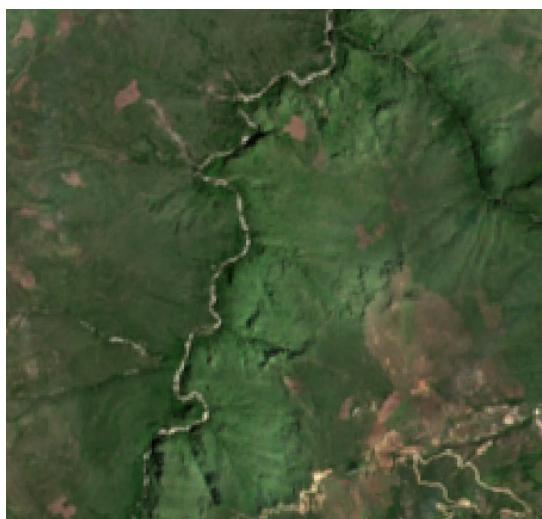


(b) Imagen clasificada 3

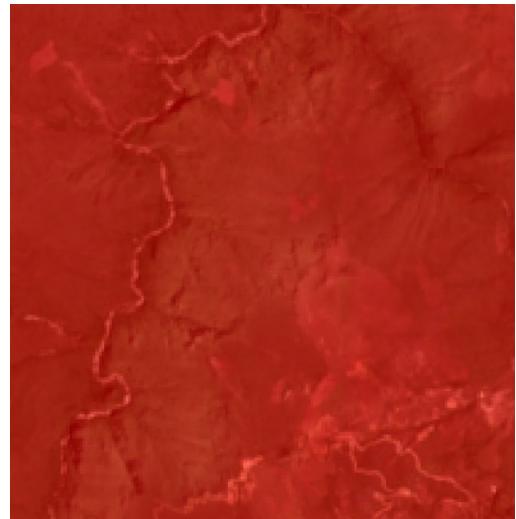
Figura 27: Tercera comparación visual entre la imagen original (a) y la imagen clasificada (b) en una región diferente del área de estudio.

3.7.2. Clasificación de áreas con vegetación natural

Dentro de los modelos o derivados de productos satelitales dedicados a la clasificación de cobertura agrícola, es común encontrar que las áreas con vegetación natural, como bosques o áreas protegidas, son clasificadas erróneamente como áreas agrícolas debido a la similitud en las características espectrales. Sin embargo, el modelo ha demostrado notable capacidad para diferenciar entre áreas agrícolas y vegetación natural. El modelo ha logrado identificar correctamente estas zonas como “No Cultivo”, evitando así la confusión comúnmente observada en otros resultados.



(a) Área con vegetación natural

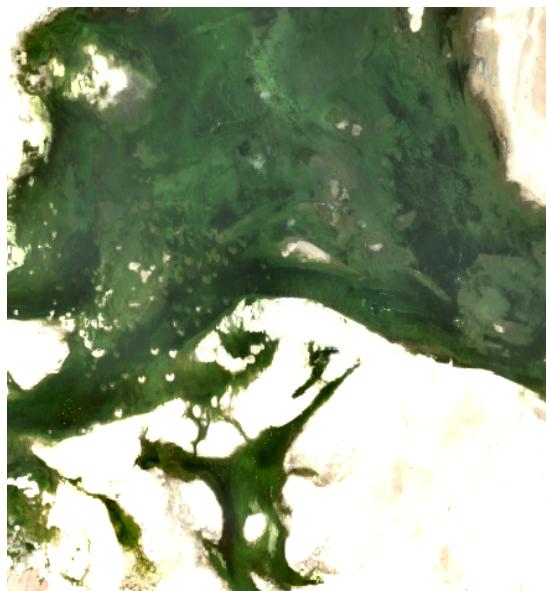


(b) Clasificación del área

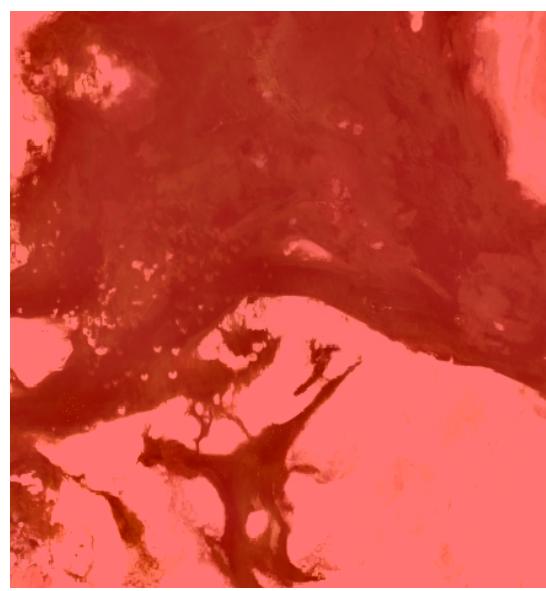
Figura 28: Comparación visual entre un área con vegetación natural (a) y su correspondiente clasificación (b), donde el modelo identifica correctamente la vegetación natural como “No Cultivo”.

3.7.3. Cuerpos de agua

Los cuerpos de agua, como ríos, lagos y embalses, presentan características espectrales muy distintas a las áreas agrícolas y vegetación natural. El modelo presenta una alta precisión al clasificar correctamente estas áreas como “No Cultivo”. La capacidad del modelo para identificar cuerpos de agua es importante, ya que estos pueden ser confundidos con áreas agrícolas en algunas circunstancias, especialmente si están rodeados de vegetación. La correcta clasificación de estos cuerpos de agua contribuye a la precisión general del modelo y a la utilidad práctica de los resultados obtenidos.



(a) Cuerpo de agua



(b) Clasificación del área

Figura 29: Comparación visual entre un cuerpo de agua (a) y su correspondiente clasificación (b), donde el modelo identifica correctamente el cuerpo de agua como “No Cultivo”.

3.7.4. Comparación con otros productos satelitales

Se evalúa la clasificación obtenida con el modelo de Random Forest entrenado con AlphaEarth Foundations en comparación con la clasificación de cobertura general y uso de suelo, disponible en Google Earth Engine, específicamente, el producto **ESA/WorldCover/v200**, que ofrece una clasificación global de la cobertura terrestre a una resolución de 10 metros. Este producto es ampliamente utilizado en estudios ambientales y de uso del suelo debido a su alta resolución y cobertura global. La comparación se realiza visualmente y mediante la evaluación de áreas específicas para identificar diferencias y similitudes en la clasificación de cobertura agrícola. De acuerdo con la leyenda del producto **ESA/WorldCover/v200**, la clase relevante para la cobertura agrícola es la clase 40 (Cultivo), mientras que las clases 10 (Árboles), 20 (Vegetación arbustiva), 30 (Hierba), 50 (Área húmeda), 60 (Cuerpo de agua), 70 (Nieve y hielo) y 80 (Área construida) se consideran como áreas sin cobertura agrícola.

A continuación, se muestra una comparación visual entre la clasificación propuesta, el producto ESA WorldCover, la imagen de referencia Sentinel-2 y la clasificación generada por el modelo presentado para un área seleccionada:



(a) Clasificación propuesta



(b) Producto ESA WorldCover



(c) Clasificación generada por el modelo

Figura 30: Comparación visual entre la clasificación propuesta (a) imagen Sentinel-2, el producto ESA WorldCover (b) y la clasificación generada por el modelo (c). Se puede observar que la clasificación propuesta muestra una mayor precisión en la identificación de áreas agrícolas en comparación con el producto ESA WorldCover.

3.8. Conclusiones

El presente documento logra demostrar de manera clara, detallada pero sencilla, los fundamentos teóricos y prácticos detrás de las tecnologías que permiten la teledetección con imágenes satelitales y su aplicación en la agricultura de precisión. Se ha repasado conceptos fundamentales como la radiometría, la fotogrametría, los sistemas de información geográfica (SIG), los derivados de productos satelitales, los algoritmos desarrollados para la clasificación de imágenes, así como las plataformas y herramientas disponibles para llevar a cabo estos procesos. Se ha aprovechado estos conocimientos en la aplicación práctica de un modelo de clasificación de cobertura agrícola utilizando de esta forma tecnologías de vanguardia como el dataset AlphaEarth Foundations que no tiene más unos pocos meses a la fecha de este documento y la plataforma Google Earth Engine, que ha revolucionado la forma en que se hace geomática.

En relación con los fundamentos teóricos, se ha logrado comprender cómo las imágenes satelitales capturan información de la atmósfera o superficie terrestre en diferentes longitudes de onda, y cómo esta información puede ser procesada y analizada para extraer datos útiles. Con esta información, la generación de índices espectrales como el NDVI, que es crucial para monitorear la salud de los cultivos, se ha podido entender su importancia y aplicación en la agricultura de precisión. Además, se ha explorado cómo los SIG permiten la gestión y análisis de datos geoespaciales, facilitando la toma de decisiones informadas en todo ámbito relacionado con la ciencia de datos geoespaciales.

La aplicación práctica de los conceptos teóricos ha sido elaborada mediante la creación de un modelo de clasificación binaria para identificar áreas con y sin cobertura agrícola en el departamento de Lambayeque, Perú. La elección del dataset AlphaEarth Foundations ha demostrado ser una herramienta poderosa para este propósito ya que es un recurso único en su tipo al combinar en un solo dataset características espectrales, topográficas y climáticas, lo que enriquece significativamente el análisis y mejora la precisión de los modelos de clasificación. El modelo de clasificación supervisada desarrollado ha mostrado una precisión general del 99.88 %, esto indica un rendimiento excepcional en la identificación de áreas agrícolas. La validación del modelo mediante la matriz de confusión y otras métricas confirma y valida la efectividad del enfoque adoptado. Además, el análisis detallado de los resultados ha revelado que el modelo es capaz de manejar desafíos comunes en la clasificación de imágenes satelitales, como la presencia de nubes y la diferenciación entre vegetación natural y cultivos agrícolas. La comparación con otros productos satelitales, como el ESA WorldCover, ha resaltado las ventajas del enfoque adoptado, mostrando una mayor precisión y capacidad de adaptación a las condiciones específicas del área de estudio.

Los resultados obtenidos han demostrado poder hacer frente a comparaciones entre productos satelitales ya establecidos que realizan clasificaciones de cobertura terrestre, destacando una mayor precisión y adaptabilidad a las condiciones específicas del área de estudio. El uso del dataset de comparación ESA WorldCover ha permitido validar la efectividad del modelo, destacando su capacidad para superar las limitaciones de otros productos en términos de precisión y detalle. La correcta clasificación de áreas agrícolas frente a no agrícolas bajo condiciones desafiantes como la presencia de nubes y vegetación natural, subraya la robustez del producto. Es importante mencionar, que el proceso de entrenamiento del modelo de clasificación es independiente del área de estudio, es decir, el dataset, al no ser un recurso que muestra datos de radiancia como lo son en su mayoría los productos satelitales, sino que son características vectoriales derivadas de múltiples fuentes de datos sirve como recurso de entrenamiento para cualquier área geográfica. Esto abre la posibilidad de aplicar el modelo a otras regiones con características similares, siempre y cuando se cuente con etiquetas de entrenamiento adecuadas para esas áreas que sirvan como reforzamiento y ajuste fino del modelo.

En lo que respecta a las limitaciones del estudio, es importante reconocer que la precisión del modelo depende en gran medida de la calidad y representatividad de las etiquetas de entrenamiento utilizadas. Aunque se ha logrado una alta precisión, la generación de etiquetas precisas y representativas puede ser un desafío, especialmente en áreas con alta variabilidad en el uso del suelo. Además, la resolución espacial del dataset AlphaEarth Foundations, aunque adecuada para muchos propósitos, puede no ser suficiente para aplicaciones que requieren un detalle más fino.

Las perspectivas futuras para esta línea de investigación y aplicaciones son no solo prometedoras, si no, emocionantes. La geomática y el desarrollo en sistemas de información geográfica están en constante evolución y cada día estas habilidades, conocimientos y herramientas son más demandadas en un mundo donde la gestión eficiente de los recursos naturales, organización, monitoreo, planificación, control y toma de decisiones basadas en datos geoespaciales son cruciales para enfrentar los desafíos globales. La integración de nuevas tecnologías, como la inteligencia artificial y el aprendizaje automático, con los datos geoespaciales abre nuevas fronteras en la agricultura de precisión y otras áreas relacionadas. Además, la disponibilidad creciente de datos satelitales de alta resolución y la mejora continua de las plataformas de análisis geoespacial, como Google Earth Engine y Colab, facilitarán aún más el acceso a herramientas avanzadas para profesionales y académicos en este campo.

En conclusión, este documento no solo ha proporcionado una comprensión sólida de los fundamentos teóricos y prácticos de la teledetección y su aplicación en la agricultura de precisión, sino que también ha demostrado el potencial transformador de estas tecnologías en la gestión agrícola y la toma de decisiones informadas. La combinación de estos recursos ha abierto nuevas posibilidades para el análisis geoespacial, y se espera que estas herramientas continúen evolucionando y desempeñando un papel crucial en el futuro de la geomática y la agricultura de precisión.

3.9. Recomendaciones

Basado en la experiencia adquirida durante el desarrollo de este proyecto y los resultados obtenidos, se presentan las siguientes recomendaciones para futuros trabajos y aplicaciones en el campo de la teledetección y la agricultura de precisión:

- **Generación de etiquetas de entrenamiento:** Se recomienda dedicar tiempo y recursos a la generación de etiquetas de entrenamiento precisas y representativas. La calidad de estas etiquetas es crucial para el rendimiento del modelo, por lo que se sugiere utilizar imágenes de alta resolución y, si es posible, validar las etiquetas con datos de campo o expertos locales.
- **Balance de clases:** Asegurarse de que las etiquetas de entrenamiento estén balanceadas entre las diferentes clases (cultivo y no cultivo) para evitar sesgos en el modelo. En este proyecto, se utilizó un número igual de polígonos para cada clase, lo que contribuyó a la alta precisión del modelo.
- **Concordancia temporal:** Al realizar el etiquetado para el entrenamiento del modelo, es fundamental que la imagen utilizada como referencia coincida temporalmente con el dataset que se está utilizando para la clasificación. Esto minimiza las discrepancias causadas por la variabilidad fisiológica de la vegetación y otros factores temporales.
- **Carga computacional:** Considerar que entornos como Google Earth Engine tienen limitaciones en cuanto a la carga computacional.
- **Modelo de clasificación:** Aunque el modelo de Random Forest ha demostrado ser efectivo en este caso, se recomienda explorar otros algoritmos de clasificación supervisada, como Support Vector Machines (SVM) o redes neuronales, para comparar su rendimiento y adaptabilidad a diferentes tipos de datos y áreas de estudio.
- **Validación continua:** Implementar un proceso de validación continua del modelo utilizando conjuntos de datos independientes y actualizados regularmente. Esto ayudará a asegurar que el modelo mantenga su precisión y relevancia a lo largo del tiempo.
- **Documentación y replicabilidad:** Documentar detalladamente todos los pasos del proceso, desde la generación de etiquetas hasta la validación del modelo. Esto facilita la replicabilidad y permite construir sobre el trabajo realizado.

Referencias

- Aburaed, N., Alkhattib, M. Q., Marshall, S., Zabalza, J., & Ahmad, H. A. (2023). A Review of Spatial Enhancement of Hyperspectral Remote Sensing Imaging Techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 2275-2300. <https://doi.org/10.1109/JSTARS.2023.3242048>
- Berger, K., Machwitz, M., Kycko, M., Kefauver, S. C., Wittenberghe, S. V., Gerhards, M., Verrelst, J., Atzberger, C., van der Tol, C., Damm, A., Rascher, U., Herrmann, I., & Paz, V. S. (2022). Multi-sensor spectral synergies for crop stress detection and monitoring in the optical domain: A review. *Remote Sensing of Environment*, 280, 113198. <https://doi.org/10.1016/J.RSE.2022.113198>
- Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Ruckridge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., Kohli, P., ... DeepMind, G. (2025). AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data.
- Bruning, B., Berger, B., Lewis, M., Liu, H., & Garnett, T. (2020). Approaches, applications, and future directions for hyperspectral vegetation studies: An emphasis on yield-limiting factors in wheat. *The Plant Phenome Journal*, 3, e20007. <https://doi.org/10.1002/PPJ2.20007>
- Castillo, E. B., Coronel, W. S., Paucar, D. G., Valqui, L. V., Navarro, D. S., Gonzales, J., Aldana, W., Pérez, H. V. V., & Berrocal, C. I. A. (2022). Cover and land use changes in the dry forest of Tumbes (Peru) using sentinel-2 and google earth engine data, 2. <https://doi.org/10.3390/IECF2022-13095>
- Cubillas, J. E., Torcitas, R., Mayol, J., & Bajao, P. J. (2024). Land Classification Plugin for QGIS Using Pix2Pix. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-5-2024, 33-39. <https://doi.org/10.5194/ISPRS-ANNALS-X-5-2024-33-2024>
- Gao, F., Anderson, M., & Houborg, R. (2024). Impacts of Spatial and Temporal Resolution on Remotely Sensed Corn and Soybean Emergence Detection. *Remote Sensing 2024*, Vol. 16, Page 4145, 16, 4145. <https://doi.org/10.3390/RS16224145>
- INEI. (s.f.). <https://ide.inei.gob.pe/#geo>
- Kupidura, P., & Lesisz, K. (2022). The Impact of the Type and Spatial Resolution of a Source Image on the Effectiveness of Texture Analysis. *Remote Sensing 2023*, Vol. 15, Page 170, 15, 170. <https://doi.org/10.3390/RS15010170>
- Pizarro, S. E., Pricope, N. G., Vargas-Machuca, D., Huanca, O., & Ñaupari, J. (2022). Mapping Land Cover Types for Highland Andean Ecosystems in Peru Using Google Earth Engine. *Remote Sensing*, 14, 1562. <https://doi.org/10.3390/RS14071562/S1>
- Radočaj, D., Šiljeg, A., Marinović, R., & Jurišić, M. (2023). State of Major Vegetation Indices in Precision Agriculture Studies Indexed in Web of Science: A Review. *Agriculture 2023*, Vol. 13, Page 707, 13, 707. <https://doi.org/10.3390/AGRICULTURE13030707>
- Ram, B. G., Oduor, P., Igathinathane, C., Howatt, K., & Sun, X. (2024). A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Computers and Electronics in Agriculture*, 222, 109037. <https://doi.org/10.1016/J.COMPAG.2024.109037>
- Sani, D., Mahato, S., Sirohi, P., Anand, S., Arora, G., Devshali, C. C., & Jayaraman, T. (2022). Crop Type Identification for Smallholding Farms: Analyzing Spatial, Temporal and Spectral Resolutions in Satellite Imagery. <https://arxiv.org/pdf/2205.03104>
- Vullaganti, N., Ram, B. G., & Sun, X. (2025). Precision agriculture technologies for soil site-specific nutrient management: A comprehensive review. *Artificial Intelligence in Agriculture*, 15, 147-161. <https://doi.org/10.1016/J.AIIA.2025.02.001>
- Wells, J. C., & Stock, J. T. (2020). Life History Transitions at the Origins of Agriculture: A Model for Understanding How Niche Construction Impacts Human Growth, Demography and Health. *Frontiers in Endocrinology*, 11. <https://doi.org/10.3389/fendo.2020.00325>
- What is radiometric resolution? | U.S. Geological Survey. (s.f.). <https://www.usgs.gov/faqs/what-radiometric-resolution>
- Wulder, M. A., Roy, D. P., Radeloff, V. C., Loveland, T. R., Anderson, M. C., Johnson, D. M., Healey, S., Zhu, Z., Scambos, T. A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C. J., Masek, J. G., Hermosilla, T., White, J. C., Belward, A. S., Schaaf, C., Woodcock, C. E., & L., J. (2022). Fifty years of Landsat science and impacts. *Remote Sensing of Environment*, 280. <https://doi.org/10.1016/J.RSE.2022.113195>
- Xia, Y., Liu, W., Meng, J., Hu, J., Liu, W., Kang, J., Luo, B., Zhang, H., & Tang, W. (2023). Principles, developments, and applications of spatially resolved spectroscopy in agriculture: a review. *Frontiers in Plant Science*, 14, 1324881. <https://doi.org/10.3389/FPLS.2023.1324881/FULL>
- Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, 2017, 1353691. <https://doi.org/10.1155/2017/1353691>